Journal of Multimedia

ISSN 1796-2048
Volume 8, Number 3, June 2013

Contents

Adaptive Halftone Watermarking Algorithm Based on Particle Swarm Optimization

Liuping Feng, Dongsheng Cong, Hou Shu, and Ben Liu
School of Information and Mechanical Engineering, Beijing Institute of Graphic Communication, Beijing, China
Email: lpfeng@yahoo.cn

***Abstract*—Digital watermarking has been widely used in copyright protection for digital images, but most of the watermarking algorithm can't resist print-and-scan process. Print-and-scan attack is a challenging problem for most of digital watermarks. A watermarking algorithm against print-and-scan attack based on PSO (Particle Swarm Optimization) is proposed in this paper. The watermark information is embedded in a halftone process for image screening with phase modulation method. The watermark is extracted by a template, which is optimized using the PSO algorithm. Both mean filtering and median filtering are used to remove noise from the recovered watermark image. The experimental results show the halftone watermarking algorithm optimized using PSO has better performance, it is robust under smearing attack and cropping attack, and has good resistance to print-and-scan process. It can be applied in the field of anti-counterfeit printing.**

***Index Terms*—particle swarm optimization, halftone image, digital watermark, phase modulation, print-and-scan**

I. Introduction

With the development of information technology and the popularity of the high quality image input/output devices, variety of printed matter piracy is becoming more and easier. There are many ways to protect printed matter from being counterfeited.

The traditional printing anti-counterfeiting technology, utilization of special papers and special inks, usage of not widely spread techniques and devices, are secure as long as the forger is not able to get access to the technical means or it is not profitable to reproduce a counterfeited original. However, there are many defects in the application, such as high cost, complex process, etc. And much anti-counterfeiting technology once used, the core technology and process are hard to change. Because it is difficult to renew and enhance the technology, it is easy to be illegal counterfeit and use.

Digital watermarking provides a solution for protecting the copyright of printed matter. Compared with the traditional security technology, digital watermark anti-counterfeiting printing technology will join the watermark in prepress process. It needn't change the printing material and equipment, and needn't increase the cost of printing. It is difficult to counterfeit and has strong adaptability. The application prospect and the economic benefit are vast. Digital watermark plays a more important role in the printing security.

In the existing literature, several techniques have been presented for data hiding in printed images. The methods proposed for data-embedding in hardcopy images may be grouped into two main categories. The first corresponds to robust embedding methods that are intended to survive printing and scanning but do not directly exploits characteristics of the printing process in the embedding. The second corresponds to techniques that use the particular characteristics of the printing process (e.g. halftoning) for embedding. The exploitation of this specific knowledge typically offers greater potential for embedding and is well-suited to hardcopy applications since the embedding occurs just prior to printing [1].

Until now, the various watermarking techniques are proposed and have been highly effective on academic aspect of watermarking [2]. Digital watermarking and data hiding are used in a large area of digital multimedia applications including copyright protection, copy control and authentication process.

Although there are many existing techniques in the previous work based on the gray images and color images for embedding data in continuous tone images, most are not always reliable when used directly with halftone images. The main restrictions are:

(1) In comparison with gray images, halftone images have much less bit planes where secret data can be hidden.

(2) Digital printing as well as offset printing is a physical process, which inevitably includes random components, make the original design information interference and destruction, such as dot expansion.

(3)The print matter must be scanned to obtain the image content for security verification. The image could be distorted and destructed during scanning, watermark extraction becomes more difficult.

After an image is successively printed and scanned, it is perhaps the same by naked eyes, but there is a great difference from the original image. The embedded watermark may be destroyed.

The watermark embedding in a halftone image is important in anti-counterfeiting printing and it is a very challenging task. Halftoning is typically utilized for the reproduction of images in most digital printing systems and many of the digital hardcopy watermarking techniques exploit the characteristics of the halftoning process for embedding watermarks in the printed images [3].

In recent years, several watermarking methods for halftone images have been proposed. Fu et al. [4] embedded a single watermark or multi-watermarks in the

doi:10.4304/jmm.8.3.183-190

parity domain of halftone images during halftoning, but the finite number of watermarks is not enough for copyright authentication. Fu et al. [5] also proposed the data hiding by modified stochastic error diffusion to embed a binary hidden visual pattern in two diffused halftone images However, the host image is essential in the watermark extracting procedure, and this proposal is not suitable for practical application. Pei et al. [6] proposed a technique by embedding the robust watermark into a dithered image after the pre-process of bit-interleaving and sub-image-interleaving to maintain low computational complexity of ordered dithering. This method has high visual quality with low capacity. In Hel-Or's work [7], the watermark is embedded in the printed image by using a number of different dither cells to create a threshold pattern. But, it is not robust to cropping and contains low transparency.

Several methods can improve the performance of watermarking schemes. One way is to make use of artificial intelligence techniques by considering image watermarking problem as an optimization problem [8]. However, there is a general serious problem of this kind of watermarking method, such as complicated and long training time of training algorithm.

PSO is an intelligence optimization algorithm imitating birds' clustering movement. It is widely applied to target tracking, positioning and navigation, mode identification etc. Z. J. Lee et al. proposed a hybrid watermarking algorithm based on Genetic Algorithm (GA) and Particle Swarm Optimization(PSO) in 2008 [9]. In 2011, Wang et al. proposed a blind PSO watermarking using wavelet trees quantization for finding the most suitable strength of both the robustness and the imperceptibility of watermarked image [10].

The problem of print-and-scan is critical to the future of digital watermarking. If the solution could be found, paper media including book, newspaper, magazine, could be all protected by digital watermark, which no doubt leads to a technology revolutionary in copyright protection.

An adaptive halftone watermarking algorithm resistance to print-and-scan process is proposed in this paper. The watermark information is embedded in a halftone process for image screening with phase modulation method. The watermark is extracted by a template, which is optimized using the PSO algorithm.

The rest of the paper is organized as follows. Section II describes the halftone watermarking embedding method with phase modulation. Adaptive halftone watermark extraction based on the PSO optimization is detailed in Section III. Section III also demonstrates the experimental Results of PSO optimization. To improve the quality of the recovered watermark using filtering is presented in Section IV. The algorithm performance is tested under various kind attacks to halftone watermarked image in Section V, and the experiment results are analyzed. Section VI concludes the paper by summarizing the key aspects of our scheme and point out shortcomings.

II. Halftone Watermarking Embedding Method with Phase Modulation

A. Clustered-Dot Ordered Dither Method

Halftone images contain only two tones and are generated by a procedure called halftoning from multi-tone images, which look like the original multitone images when viewed from a distance. Digital halftoning is commonly employed in newspapers, books, magazines, and computer printouts. Digital halftoning provides a mechanism for rendering continuous-tone images on devices such as printers [11].

A large number of halftoning methods have been proposed for the rendering of halftone images on bi-level printers, including order dithering, error diffusion, dot diffusion, and least square error, and the common digital halftoning methods can be divided into two categories: ordered dithering and error diffusion. The ordered dithering method can generate dots of amplitude modulation. This mean that the dots are of different sizes but the distance between them is the same. The error diffusion method generates dots of the same size but with different distances between them.

The ordered dithering technique is again divided into two parts: clustered dots and dispersed dots. Due to their stability and predictability clustered-dot halftoning is the predominant halftoning method for the xerographic and lithographic printer families [12].

A halftone image that meets the requirement that its visual appearance at normal viewing distance closely approximates $I(i,j)$, but does not carry the watermark patterns, can be obtained via a screening process that compares the image $I(i,j)$ against a periodic threshold matrix $T(i,j)$ as:

$$I^h(i,j) = \begin{cases} 1, & if \;\; I(i,j) < T(i,j) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where the convention that the values "1" and "0" correspond are adopted, respectively, to whether ink/toner is, or is not, deposited at the pixel position (i,j).

Eq. (1) shows that when the grayscale value is larger than the marginal value which has been established, this is expressed by "0". If the grayscale value is smaller than the marginal value, this is expressed by "1". Thus the halftone image can be produced by "0" and "1". In this way many brightness patterns can be defined.

In particular, clustered-dot ordered dither where the black dots grow from the centre of the threshold matrix and dispersed-dot dither where the black pixels are more or less uniformly distributed over the threshold matrix for any given grayscale value.

The clustered-dot ordered dithering method is chosen in our algorithm to create a halftone pattern, and two different threshold matrix $T0$, $T1$ for clustered-dot ordered dither are defined in Eq.(2) and Eq.(3), which screen angles are 135° and 45° respectively.

$$T0=\begin{bmatrix} 32 & 11 & 19 & 27 & 34 & 53 & 45 & 37 \\ 24 & 8 & 3 & 10 & 42 & 58 & 61 & 56 \\ 16 & 6 & 2 & 18 & 50 & 60 & 64 & 48 \\ 30 & 22 & 14 & 26 & 36 & 44 & 52 & 40 \\ 33 & 54 & 46 & 38 & 31 & 12 & 20 & 28 \\ 41 & 57 & 62 & 55 & 23 & 7 & 4 & 9 \\ 49 & 59 & 63 & 47 & 15 & 5 & 1 & 17 \\ 35 & 43 & 51 & 39 & 29 & 21 & 13 & 25 \end{bmatrix} \quad (2)$$

$$T1=\begin{bmatrix} 35 & 49 & 41 & 33 & 30 & 16 & 24 & 32 \\ 43 & 59 & 57 & 54 & 22 & 6 & 8 & 11 \\ 51 & 63 & 62 & 46 & 14 & 2 & 3 & 19 \\ 39 & 47 & 55 & 38 & 26 & 18 & 10 & 27 \\ 29 & 15 & 23 & 31 & 36 & 50 & 42 & 34 \\ 21 & 5 & 7 & 12 & 44 & 60 & 58 & 53 \\ 13 & 1 & 4 & 20 & 52 & 64 & 61 & 45 \\ 25 & 17 & 9 & 28 & 40 & 48 & 56 & 37 \end{bmatrix} \quad (3)$$

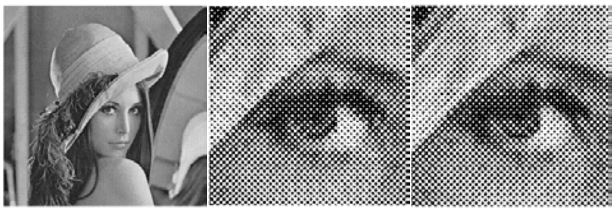
The 8×8 threshold matrix contains values from 1 to 64 that define the order that cells will be turned "on" or "white".

The original grayscale image g is converted firstly as blackness in Eq.(4), which means the percentage of black area in a 8×8 pattern.

$$g'(x,y)=64\times(1-\frac{g(x,y)}{255}) \quad (4)$$

where $g(x,y)$ is the grayscale intensity value in position (x,y) .

The 8×8 pattern I^h correspond to $g'(x,y)$ can be calculated by Eq. (1), where I is a 8×8 matrix which element is $g'(x,y)$. Halftone image is obtained by performing the iterative calculations. Lena original gray image is shown in Fig. 1 (a). The halftone images generated with the threshold matrix of *T0* and *T1* are shown in Fig.1(b) and Fig.1(c) respectively.



(a) (b) (c)

Figure 1. The halftone image with varied threshold matrix: (a) Lena original gray image, (b) the halftone image of screen angle 135°, (b) the halftone image of screen angle 45°.

B. Watermark Embedding

There are many possible ink patterns that produce printed images that appear similar to the human eye. This characteristic is exploited to embed a watermark in printed images [7]. Data hiding methods, on the other hand, carry the information in a manner that minimally disrupts the primary content included in the print-and-scan process.

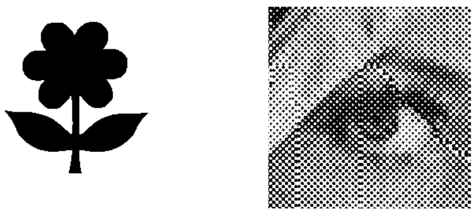
From a technical standpoint, adapting a data hiding scheme to the channel characteristics can offer significant performance improvements in terms of robustness and embedding rates. This viewpoint indicates that hardcopy data hiding schemes that aim at high embedding rates and robustness to the print-scan process must specifically adapt to the characteristics of the print-scan channel distortion. The key component of the printing process is a bit-depth reduction step called digital halftoning which produces an illusion of continuous tone (contone) by trading off amplitude resolution for spatial resolution [12].

In our algorithm, we use the different phase-modulated halftone images to embed watermark. The watermark part of the cover image is modulated the halftone-dot with a threshold matrix $T1$, and the other part of the cover image is modulated the halftone-dot with another threshold matrix $T0$.

Name the halftone images $im0$ and $im1$ modulated with $T0$ and $T1$ respectively. The binary watermark image is shown in Fig. 2(a). Enlarge it to the same size as the halftone cover image, and name it M . Generate the watermarked image im by combining images with different screen rulings described by Eq. (5).

$$im = im0\,\&\,M \,\|\, im1\&\sim M \quad (5)$$

The watermarked halftone image is shown in Fig. 2(b).



(a) (b)

Figure 2. Watermark embedded: (a) the binary watermark image, (b)the watermarked halftone image.

C. Artifact Elimination at the Boundary of the Image

Observe the watermarked image shown in Fig. 2(b), we can find there are distinct edges at the boundary of the watermark. These artifacts tremendously degrade the invisibility of the image; watermark image pre-processing is used to eliminate it.

Different methods can be used for watermark image scrambling pre-processing such as Fass Curve, Gray Code, Arnold Transform, Magic square, etc.

Due to the periodicity process of the Arnold transform, the image can be easily recovered after the permutation. Arnold transformation is applied widely in digital image scramble. In our algorithm, the Arnold transformation is employed to shuffle the pixel positions of the watermark image before embedding.

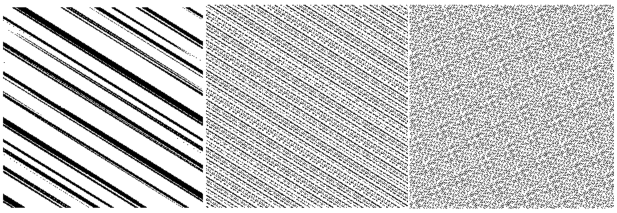
The Arnold transformation can be described by Eq. (6).

$$\begin{bmatrix} x' \\ y' \end{bmatrix}=\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}(\bmod N) \quad (6)$$

where (x, y) is a pixel of the $N \times N$ original image, and (x', y') is a pixel of the transformed image.

The special property of Arnold transform is that image comes to its original state after certain number of iterations, called "Arnold period". Because of the periodicity of Arnold transform, the original image will be recovered.

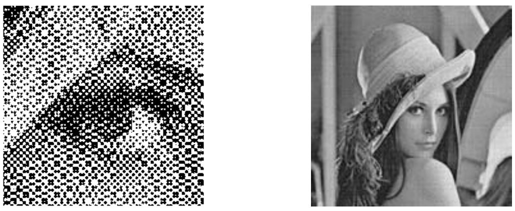
The scrambled images of the watermark image are shown in Fig.3, using Arnold transformation and the iteration time is n.



(a) (b) (c)

Figure 3. The images scrambled by Arnold transform with the different iteration time *n*: (a) *n*=3, (b) *n*=5, (c) *n*=10.

Through watermark pre-process, the watermarked image generated with the method described in Section II. B is shown in Fig. 4. The enlargement part of the watermarked halftone image is shown in Fig. 4(a), and the whole of the watermarked halftone image is shown in Fig. 4(b). There are no edges at the boundary of the watermark after pre-processing with Arnold transformation.



(a) (b)

Figure 4. The watermarked halftone image using Arnold pre-processing, (a) the enlargement part of the watermarked halftone image, (b) the whole of the watermarked halftone image.

III. Adaptive Halftone Watermark Extraction Based on the PSO Optimization

A. Watermark Extracting with a Template

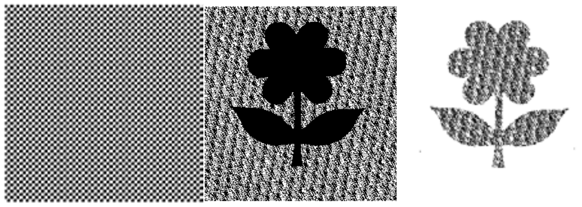
The watermark image is extracted with a template $Model$ shown in Fig. 5(a). The template $Model$ is a haltone image, which is constructed from a medium gray image, which grayscale value is 128, screening with $T0$ or $T1$.

Overlay the watermarked halftone image with the model $Model$ using Eq.(7).

$$M' = im \,\&\, Model \tag{7}$$

Because the phase of the template $Model$ is the same as $im0$ or $im1$ in Eq.(5), the watermark embedded part and the other part of the image can be distinguished by comparison the number of "1" with the threshold value for each 8×8 block of M' after overlaying, the watermark image can be extracted.

Examine and count the number of "1" for each 8×8 block of M'. If it is greater than a threshold m, the watermark bit is "1", otherwise is "0". Here we get m equal to median 30. The watermark image shown in Fig.5(b) is extracted using the template constructed with $T0$. The watermark image shown in Fig.5(c) is extracted using the template constructed with $T1$, then taken anti-bit.



(a) (b) (c)

Figure 5. The recovered watermark image: (a) the template *Model*, (b) the watermark image extracted using the template constructed with *T0*, (c) The anti-bit watermark image extracted using the template constructed with *T1*.

From Fig. 5 (b) and Fig. 5 (c) we can see, watermark information can be extracted roughly, but there is much noise in the image. The reason is the fixed template *Model* and fixed threshold value *m* are used. The average greyscale of the original cover image is changing, therefore, the watermark template *Model* and the threshold *m* must be optimized and adapted.

B. Optimization of Watermark Extraction Using PSO

PSO algorithm is proposed by J. Kennedy and R.C. Eberhart in 1995. The PSO model simulates birds swarm behavior. PSO has become an active branch of swarm intelligence during the last decade.

As a population-based technique, PSO was inspired by the emergent motion of swarms and birds behavior. Particles in PSO iteratively explore optima in a multidimensional search space by utilizing personal memories and sharing information within a specific neighborhood.

The swarm is typically modeled by particles with a position and a velocity, where each particle represents a candidate solution to the optimization problem. During the optimization procedure, particles communicate good positions to each other and adjust positions according to their history and the experience of neighboring particles. Due to the nature of individual memories and information sharing, particles can reach the best solution quickly.

PSO does not require any gradient information of the function to be optimized uses only primitive mathematical operators and is conceptually very simple. Each PSO particle moves about the cost surface with an individual velocity. It is becoming very popular due to its simplicity of implementation and ability to quickly converge to a reasonably good solution.

In PSO system, each individual (particle) represents a solution in a *D*-dimensional space. Each particle also has knowledge of its previous best experience and knows the global best experience (solution) found by the entire swarm [13].

PSO algorithm can be expressed as follows: to randomly initialize a particle swarm whose number is *m* and dimension is *D*, in which the *i-th* particle has a corresponding position and velocity, denoted as $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})^T$ and $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})^T$ $(i = 1, 2, \ldots, m)$ respectively. Each particle also has a fitness value decided by decision function. $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})^T$ stands for the best value found by the *i-th* particle and $P_g = (p_{g1}, p_{g2}, \ldots, p_{gD})^T$ denotes the global best found by the entire swarm.

During each iteration, The particle position x_{id}^k and velocity v_{id}^k are updated using current value so as to reach optimization, and the newly computed position x_{id}^{k+1} and velocity v_{id}^{k+1} are described by Eq.(8) and Eq.(9) .

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k) \tag{8}$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \tag{9}$$

where r_1 and r_2 are random number within interval (0,1), ω is the inertia coefficient (A larger inertia coefficient means stronger global search ability, while a smaller one stands for stronger local search ability), c_1 and c_2 are learning factors. This process is repeated for every dimension and for all the particles in the swarm.

Because the average grayscale value of the original cover image is different, the effect of watermark extraction is not favorable using the fixed template $Model$. However, it is difficult to determine the appropriate grayscale value of the template through the experiment. Moreover, the threshold value *m* used to determining the extracted watermark bit is also very important. Therefore, PSO algorithm is used to optimize the template $Model$ and the threshold value *m*.

Firstly, normalized correlation function NC is defined to evaluate similarity degree between the extracted watermark w' and the original watermark w in Eq.(10).

$$NC = \frac{\sum_{i=1}^{N} w(i) w'(i)}{\sqrt{\sum_{i=1}^{N} w^2(i)} \sqrt{\sum_{i=1}^{N} w'^2(i)}} \tag{10}$$

where *N* is the length of the watermark.

In order to make the watermark extraction to achieve the best effect, it must make NC maximum. PSO optimization process includes the following steps:

(1) Initialization

First, the position and velocity of particles are represented and initialized in the solution space. PSO algorithm adjusted two important parameters in the watermark extraction process: The grayscale value gv of the gray image, which constructed the halftone template $Model$, limited in 0 ~ 255; and the threshold value m, which is the number of "1" to determine watermark bit is "0" or "1" for every 8×8 block, limited in 0 ~ 64. Generate particle swarm positions (gv, m) random in this range, and initialize the velocity of the particles with a random number.

(2) Evaluation

The fitness function evaluates the goodness of position and returns a finite number. In our algorithm, for a given particle position (gv, m), calculation NC defined in Eq.(10), which is similarity degree between the extracted watermark and the original watermark in this parameter condition. NC value will be used to calculate the fitness function and it is defined in Eq.(11).

$$y = 1 - NC \tag{11}$$

(3) Updating

The position and velocity of the particles is updated according to Eq. (8) and Eq. (9), and partial extremism and global extremism is updated according to the fitness function value of the new particles.

(4) Check and iterative

The process is repeated starting at stage (2) until the stopping or the iteration number is reached, and the minimum value of *y* can be getting through the iterative optimization. The corresponding (*gv*, *m*) can be used to construct the template and extract the watermark.

C. Adaptive Mutation

However, PSO iteration is a heuristic method and the strategy of PSO algorithm is to optimize the fitness function by use of individual memories (the best particle of individual in all generations) and information sharing (the globe best particle with particle individuals). it is not able to guarantee convergence to a global optimum, but rather to a good solution or a local optimum. Sometimes, It can quickly reach the optimum, and is more likely to produce premature convergence and fall into a local optimal equilibrium state.

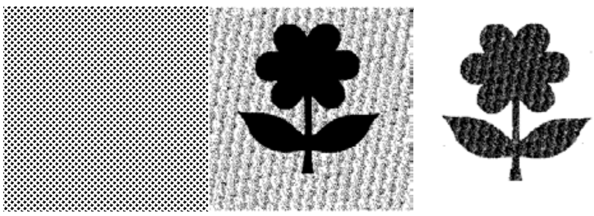
In Eq. (8), the constants c_1 and c_2 are the cognitive and social learning rates, respectively. r_1 and r_2 rand are two uniformly distributed numbers in the range (0,1). Noted that these two rates control the relative influence of the updating velocity of particle and neighborhood. In the iterative process, the particles can easily become the same state and have a rapid decline in the diversities, because particles have always had the tendencies closed to the two optimal values P_i and P_g. If the particle is not converge to the global optimal solution but a local optimal value at this time, the PSO algorithm will be difficult to jump out of local optimal values[13].

In order to make up for the shortage of PSO, to avoid a local optimum in finding the solution, referring to the mutation for genetic algorithm, the mutation operation in PSO algorithm is introduced. The mutation operation expands the swarm search space reduced in iteration, make particles jump out the optimal position searching previously, and search in the larger space. The algorithm keeps the swarm diversity, and improves the possibility to find the more optimal solution.

Therefore, a simple mutation operator is introduced in this paper. With a certain probability, that is, when r_1 and r_2 of Eq. (8) are more than 0.8, the position of the particle is initialized again after updating.

D. Experimental Results and Analysis

According the optimization algorithm described in Section III. B&C, the optimization result $(gv, m) = (173{,}16)$ is get, and the optimized template $Model'$ is shown in Fig. 6(a). The recovered watermark shown in Fig. 6(b) is extracted by the optimized template $Model'$ constructed with $T0$. If another threshold matrix $T1$ with different phase is used to construct the optimized template $Model'$, and take result anti-bit, the recovered watermark is shown in Fig.6(c).



(a) (b) (c)

Figure 6. The recovered watermark image using the template *Model'* optimized by PSO: (a) the optimized template *Model'*, (b) the recovered watermark using the optimized template constructed with *T0*, (c) The anti-bit watermark image extracted using the optimized template constructed with *T1*.

Compared with Fig. 5, the quality of the recovered watermark image is shown in Fig. 6 are improved, but there is noise. Filtering is used to remove the noise of the recovered watermark image in next section.

IV. Quality Improvement of the Recovered Watermark using Filtering

A. Mean Filtering

In image processing, the mean and median filters are simple and very effective tools for noise suppressing.

Mean filtering is usually used for suppressing Gaussian noise. The idea of mean filtering is simply to replace each pixel value in an image with the mean value of its neighbours, including itself. This has the effect of eliminating pixel values which are unrepresentative of their surroundings.

Mean filtering is usually thought of as a convolution filter. Like other convolutions, it is based around a kernel, which represents the shape and size of the neighborhood to be sampled when calculating the mean.

A 3×3 square kernel is shown in Fig. 7(a) is used in our algorithm, and the result of mean filtering on the recovered watermark of Fig.6(c) and binarization is Fig. 7(b).

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

(a) (b)

Figure 7. The result of mean filtering: (a) averaging kernel, (b) the recovered watermark after mean filtering and binarization.

B. Median Filtering

The Median Filter does somewhat the same, but, instead of taking the mean or average, it takes the median. The basic idea of the median filtering consists of simultaneous replacing every pixel of an image with the median of the pixels contained in a window around the pixel. A median filter can't be done with a convolution, and a sorting algorithm is needed. The median is gotten by sorting all the values from low to high, and then taking the value in the center. If there are two values in the center, the average of these two is taken.

Median filtering is a nonlinear operation often used in image processing to reduce high frequency noise in an Image. Median filtering is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges [14]. A median filter gives better results to remove salt and pepper noise, because it completely eliminates the noise. With an average filter, the color value of the noise particles are still used in the average calculations, when taking the median you only keep the color value of one or two healthy pixels.

Perform median filtering on the recovered watermark shown in Fig.6(c) with the 3-by-3 neighbourhood, the result of filtering and binarization is shown in Fig. 8.

Figure 8. The result of median filtering.

V. Attacks Experiment and Results analysis

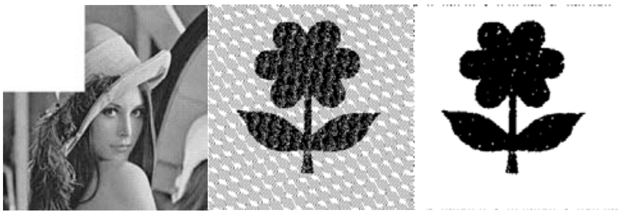
After optimization and filtering, the recovered watermark is very similar to the original image in character.

However, an efficient watermarking algorithm has to combine invisibility and robustness. The goal in halftone data hiding is to recover the embedded data from a scan of the printed image. As such, in most applications a malicious adversary is nonexistent [12]. So smearing attack and cropping attack are considered firstly, which may be encountered in print-and-scan process. Then the efficacy of the watermark extracting against print-and-scan process will be tested.

The experiment results demonstrate the algorithm is robust under smearing attack and cropping attack. The experiment results are shown in Fig. 9 and Fig. 10.

(a) (b) (c)

Figure 9. Smearing attack: (a) the attacked image, (b)the recovered watermark image (c) the result of median filtering.

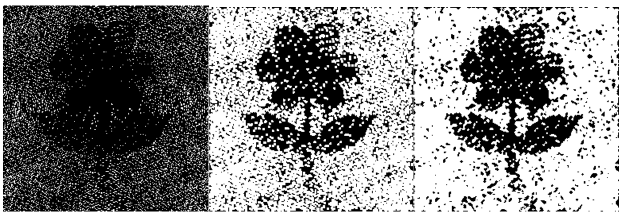


(a) (b) (c)

Figure 10. Cropping attack: (a) the attacked image, (b) the recovered watermark image, (c) the result of median filtering.

The goal of this paper is to extract the embedded watermark against print-and-scan process.

The watermarked halftone image is printed on HP LaserJet 1020 printer with resolution of 300×300 dpi, and scanned on Kodak i65 scanner with the same resolution. The recovered watermark image is shown in Fig. 11(a), and it is obscure. Both the mean filter and the median filter are used to remove noise from the recovered watermark image. The mean filtering is performed on the recovered watermark of Fig. 11(a) using averaging kernel of Fig. 7(a), and the result is shown in Fig. 11(b). Then the median filtering is performed with the 3-by-3 neighbourhood, and the result is shown in Fig. 11(c).



(a) (b) (c)

Figure 11. The recoveres watermark image after print-and-scan process: (a)the recovered watermark image, (b)the result of mean filtering, (c)the result of mean and median filtering.

In order to qualitative comparison and quantitative analysis the performance of the watermark extraction in all in all kinds of situation above, a unified method is adopted. That is, the threshold matrix *T1* is used to construct the optimized template, and recovered watermark is taken anti-bit. Comparison of *NC* values is shown in Table I.

TABLE I. *NC* VALUE COMPARISON

The method of watermark extraction	*NC* value
Not PSO optimization	0. 9276
PSO optimization	0. 9736
PSO optimization and Mean filtering	0. 9950
PSO optimization and Median filtering	0. 9830
Print-and-scan process and mean filtering	0.8643
Print-and-scan process, mean and median filtering	0.9123

VI. CONCLUSION

In this paper, a clustered-dot halftone watermark using phase modulation is proposed. In the watermark embedding, the gray image is transformed to a halftone image with the threshold matrixes of two different phases in the watermark region and the other region respectively. The watermark is extracted by overlaying a template, which phase is the same as one of them. In order to eliminate edges at the boundary of the watermark, Arnold transformation is used for watermark pre-processing. PSO algorithm is used to optimize the watermark extraction template. Experiments results demonstrate the efficacy of our algorithm; it is robust under smearing, cropping and pint-and-scan process. Because of a slight distortion and deformation in print-and-scan process, and our algorithm have not considered these factors, therefore, the efficacy of the watermark extraction is not excellent after print-and-scan process.

Print-and-scan attack is a challenging problem for most of digital watermarks. Some necessary or unnecessary parameters are adjusted by operator occasionally, and the mechanism of print-and-scan process has not been fully investigated, so that it is difficult to find the watermarking countermeasure against the attacks. Particle swarm optimization has become an active branch of swarm intelligence during the last decade. Compared with GAs, PSO is also a population-based stochastic search technique, but does not have any crossover and mutation operators [15]. In the paper, the watermarking method using PSO tries to strengthen the ability of watermark extraction in kinds of situations.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Funds project, No.61170259; Beijing Municipal Natural Science Foundation, No. KZ201010015013.

REFERENCES

[1] O. Bulan, V. Monga, G. Sharma, and B. Oztan, “Data Embedding In Hardcopy Images Via Halftone-Dot Orientation Modulation”, *in Proc. SPIE: Security, Forensics, Steganography, and Watermarking of Multimedia Contents*, vol. 6819, pp.68190C-1-12, 2008

[2] Huang-Chi Chen, Yu-Wen Chang, and Rey-Chue Hwang, “A Watermarking Technique based on the Frequency Domain”, *Journal of Multimedia*, vol.7, no.1, pp.82-89, February 2012

[3] B. Oztan, G. Sharma, “Multiplexed clustered-dot halftone watermarks using bi-directional phase modulation and detection”. *Proceeding of IEEE 17th International*

Conference on Image Processing, Hong Kong, pp.981-984, September 2010
[4] M. S. Fu and O.C. Au, “A robust public watermark for halftone images”, *Proc. Circuits and Systems*, vol. 3, pp. 639-642, 2002.
[5] M. S. Fu and O.C. Au, “A novel method to embed watermark in different halftone images: data hiding by conjugate error diffusion (DHCED) ”, *Proc. Multimedia and Expo*, vol. 1, pp.609-612, 2003.
[6] S.C. Pei, J.M. Guo, H. Lee, “Novel robust watermarking technique in dithering halftone images”, *IEEE Letter on Signal Processing*, vol. 12, pp. 333-336, 2005
[7] Hagit, Z. Hel-Or., “Copyright labeling of printed images”, *Proc. Image Processing*, vol.3, pp.702-705, 2000.
[8] P. Fakhari, E. Vahedi and C. Lucas, “Protecting Patient Privacy From Unauthorized Release of Medical Images Using a Bio-nspired Wavelet-based Watermarking Approach”. Digital Signal Processing, vol.21, pp.433-446, 2011
[9] Z. J. Lee, S. W. Lin, S. F. Su, and C. Y. Lin, “A hybrid watermarking technique applied to digital images”, *Applied Soft Computing*, pp. 798-808, 2008.
[10] Y.R. Wang, W.H. Lin, and L. Yang, “A blind PSO watermarking using wavelet trees quantization”, *International Conference on Machine Learning and Cybernetics*, pp.1612-1616, 2011
[11] Y. Ju, D. Saxena, T. Kashti, D. Kella, et al. “Shaked Modeling Large-Area Influence in Digital Halftoning for Electrophotographic Printers”, *Proc. of SPIE-IS&T Electronic Imaging, SPIE* Vol. 8292, 82920X, 2012
[12] O. Bulan, G. Sharma, and V. Monga “Orientation Modulation for Data Hiding in Clustered-Dot Halftone Prints”, *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp.2070-2084,August 2010.
[13] Benzheng WEI, Zhimin ZHAO, and Xin PENG, “Spatial Information Based Medical Image Registration using Mutual Information”, *Journal of Multimedia*, vol.6, no.3, pp.236-243, February 2011
[14] R. Reddy, M.V.N. Prasad, and D.S. Rao, “Robust Digital Watermarking of Images using Wavelets”, *International Journal of Computer and Electrical Engineering*, vol. 1, no. 2, pp. 1793-8163, June 2009
[15] Jun Tang, Xiaojuan Zhao, “A Hybrid Particle Swarm Optimization with Adaptive Local Search”, *Journal of Networks*, vol. 5, no. 4, pp.411-418, April 2010

Liu Ping Feng is an associate professor of Beijing Institute of Graphic Communication, China. Her research focuses on information security, digital rights management and digital watermark technology. She received her Ph.D. degrees in Computer Application Technology from Beijing Institute of Technology, China.

Dong Sheng Cong is a graduate student in Signal and Information Processing of Beijing Institute of Graphic Communication, China. His research work is focusing on digital watermarking technology and digital image processing.

Hou Shu is an associate professor of Beijing Institute of Graphic Communication, China. Her research interests include digital media technology and watermark technology.

Ben Liu is an associate professor of Beijing Institute of Graphic Communication, China. He specializes in network publishing technology and digital media technology.

A 3D Simplification Method based on Dual Point Sampling

Juan Cao
College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China
Email: caojuan6666@gmail.com

Yitian Zhao, Ran Song
Department of Computer Science, Aberystwyth University, UK
Email: {yyz10, res}@aber.ac.uk

Yingchun Zhang
College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China
Email: twsbh@sina.com

***Abstract*—For efficiently processing integration, registration, representation and recognition of large-scale 3D point clouds stored in computer disks or other hardware, it is important to simplify their sheer volume, discarding redundant information and meanwhile preserving the most important information as much as possible. This paper proposes an effective point cloud simplification method which is based on data points sampling. The dual sampling scheme considers both the local details and the overall shape. The local details analysis approach is based on graph-based segmentation, while for the overall shape analysis, the approach voxelizes the model and samples points in terms of the entropy, based on the shape index of vertices. Compared to other simplification methods, this approach reduces the number of vertices in a 3D model described by a point cloud and better preserves local details. We present a number of results to show that the method performs well both visually and quantitatively.**

***Index Terms*—simplification, dual sampling, segmentation, voxelization**

I. Introduction

High-resolution laser range scanners have been widely used for acquiring 3D data from a 3D surface model. Typically, a laser range scanners produces a set of structured data points with or without reflectance strength information, depicting the reflectance characteristics of the 3D objects of interest. The structured data points, usually stored as range images in computer disks, can easily be triangulated and rendered as 3D meshes which physically describes a discredited surface. Most of the range images have considerably large sizes due to the high resolution of scanners. A complete 3D model could contain millions of points and often leads to expensive processing time and intensive memory demand. This creates great challenges, such as storage, editing and transmission. Consequently, simplification is becoming necessary in many applications. It can accelerate either the display or processing of the large 3D data. So the simplified data has become a powerful alternative to the original data.

A. Previous Method

Generally, existing simplification methods can be categorized into three classes: vertex decimation [1-5], vertex clustering [6-9] and edge contraction [10-13]. Some techniques offer efficient processing but produce simplified meshes visually undesirable [1, 4, 7]. Others create more considerable approximations but take an expensive processing time and are difficult to implement [2]. Normally, most of the simplification methods firstly define a mesh operation and then apply that to a mesh, act on a small collection of its points and produce a new mesh with fewer points.

Schroeder et al [1] introduced the vertex decimation, which operates on a single vertex by deleting that vertex and re-tessellating the resulting hole, it measures the distance from the vertex to average plane by its adjacent triangles, then uses the distance to decide the order in which vertices are removed. This method is simple to implement, but generates low-quality approximated models. Wu and Kobbelt [4] used a random selection of vertices to be removed in the context of streaming to simplify large meshes. These methods are based on the removal of vertices from the mesh. Once a vertex is removed, all faces using that vertex are also removed and then the hole is retriangulized. Because of the way it creates triangles, this kind of algorithm is limited to manifold meshes [16].

In [7], a vertex clustering method is proposed to handle meshes of arbitrary topological structure. The authors assigned a weight to each vertex on the input mesh by its perceptual importance, then subdividing the mesh into a three-dimensional grid, and finally, all the vertices in a given grid cell are clustered to the position of the vertex with maximum weight. This method tends to be very fast but the visual appearance of the final mesh is not relatively very good.

doi:10.4304/jmm.8.3.191-197

Hoppe et al [11, 12] investigated the edge contraction and it has become the most common simplification operation. They used the edge collapse operator to construct a progressive mesh, and measured the distance from the proposed new triangles to a set of sample points from the original mesh to decide which edge to collapse. Gueziec [17] presented a technique for simplifying the triangulated surface by using tolerance volume which is built to forbid simplification errors exceeding a local tolerance, then applied this to the surfaces using edge collapses. Cohen et al [18-20] presented a completely different approach which is appearance based, and measures the amount of deviation caused by the operation in the screen-space representation of the mesh. Garland and Hechbert [14, 15] proposed the vertex pair-collapse operator, which can be considered the topology modifying variant of the edge-collapse operator. In practice, this method of simplification is probably making achieve the most efficient results.

Nehab et al [21] presented a stratified sampling strategy for 3D data that is a technique to generate evenly spaced samples by subdividing the sampling domain into non-overlapping parts and sampling independently from each part. Osada et al [22] proposed a method which uniformly sampled on a triangulation mesh. This sampling approach basically generates random sample points in the randomly picked triangles with equal probability per unit area.

B. Our Method

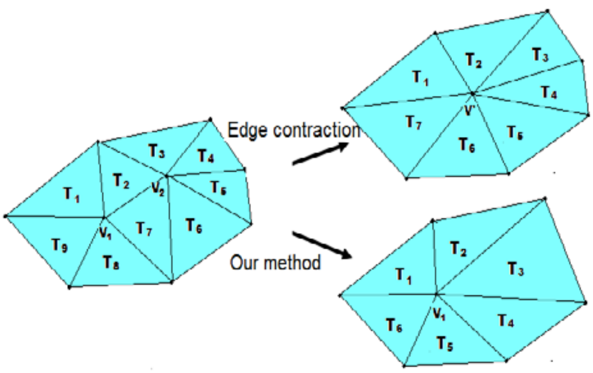


Figure 1. Simplification by using edge contraction and our algorithm

This paper proposes a new mesh simplification method based on dual point sampling. Fig. 1 shows a simplification result after using edge contraction and our method. The topology of the mesh is changed during edge contraction but not by our method. The sampling includes two steps: global sampling and local sampling. In the global points sampling, a uniform and even sampling approach will be described further in the next section and normally it captures the overall shape but sometimes misses the local details. To improve this, the local sampling is applied. In this step, the graph-based segmentation is employed. The mesh can be segmented into a number of patches in terms of their own similarity of curvature or shape index. The local sampling method will select the points from each segmented region individually. Hence, the points which are sampled could present the overall shape and some local details. The algorithm proposed in this paper consists of four phases: preprocessing, local points sampling, global points sampling and surface reconstruction.

Preprocessing: In order to sample the data accurately, the mesh smoothing technique is employed. The Laplacian smoothing method is chosen because of its simplicity and efficiency. Then based on Taubin's algorithm [23], the curvature and shape index can be estimated.

Local points sampling: After estimating the shape index of the model, employing the graph-based segmentation method to segment the surface, the deterministic sampling approach is used, which will sample the maximum absolute shape index values from each segment.

Global points sampling: This estimates the entropy of all points, uses a window operator for overall surface and samples the points that have the largest entropies.

Surface reconstruction: After local and global points sampling, this combines the sampled data by using the objective function.

II. Preprocessing

A. Data Smoothing

The acquired data of the 3D model usually contains imaging noise from various sources, such as scanning noise. It is necessary to remove the noise while preserving the underlying sampled surface, in particular its fine features. In order to produce more accurate segmentation results, the Laplacian smoothing is employed. It is a simple and efficient method to smooth the polyhedral surface and improve the appearance. The Laplacian smoothing algorithm reduces the high frequency surface information and tends to flatten the surface.

B. Curvature and Shape Index Estimation

In Taubin's work [23], principal curvatures and principal directions are obtained by computing the eigenvalues and eigenvectors of 3 by 3 symmetric matrices defined by integral formulas in closed form, and closely related to the matrix representation of the tensor of the curvature. The principal curvature can be obtained as functions of the nonzero eigenvalues of symmetric matrices.

It has been suggested that the shape index s could give a simple measurement of the local shape. The shape index can be obtained by the principal curvature

$$s = \frac{2}{\pi} arctan \frac{k_2+k_1}{k_2-k_1} \quad (1)$$

where k_2 and k_1 are the principle curvatures. Fig. 2 illustrates several examples of the mean curvature and shape index mapped on the model surface. Compared with the results of mean curvature, the shape index is recognized the flat, concave and convex regions more clearly, because it can represent the geometric information more accurately.

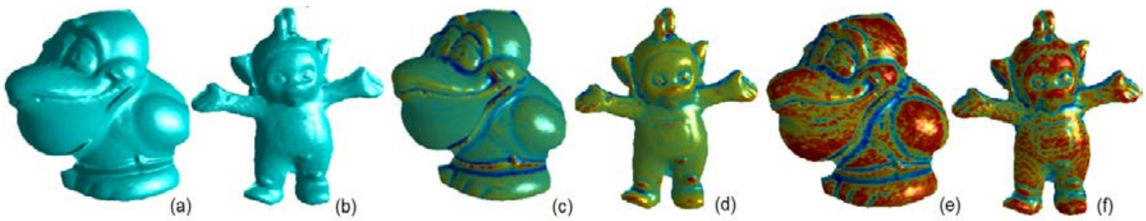


Figure 2. (a), (b): input models bird and tubby; (c), (d):curvature estimation; (e), (f): shape index estimation

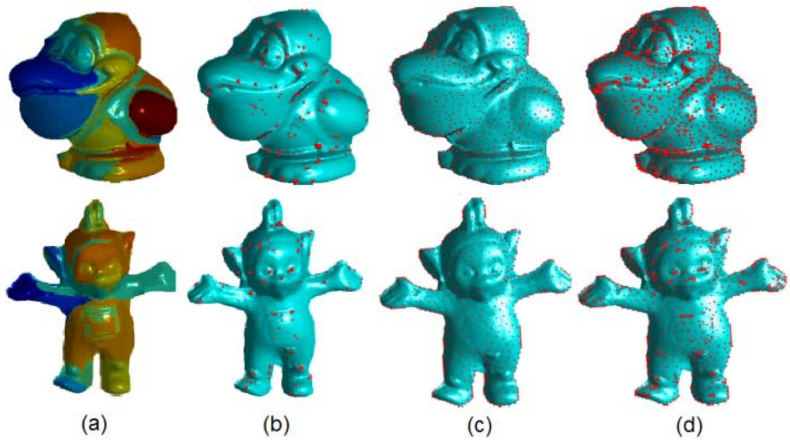


Figure 3. Step results of bird and tubby(a) Segmentation;(b) Local sampling;(c)Global sampling; (d)Points combination

C. Algorithm

Felzenszwalab [24] introduces an efficient graph-based image segmentation based on pairwise region comparison. It produces segmentations that obey the global properties of being not too coarse and not too fine by using particular region function. Due to its elegance in implementation, efficiency in computation and good results in segmentation, the graph based method [24] has been employed to segment the range images. For this method initially applied on the objective images, while we have to adapt it to the range images. The points cloud can be defined as $\mathrm{x(i,j)}$, $\mathrm{y(i,j)}$, $\mathrm{z(i,j)}$ and we estimated curvature and shape index of each point. The pairwise comparison predicate can be defined as follow in order to evaluate whether or not there is evidence for a boundary between two regions in segmentation.

$$D(R_1,R_2)=\begin{cases} true, if \;\; Dif(R_1,R_2)>MInt(R_1,R_2) \\ false, \qquad\qquad\qquad\qquad otherwise \end{cases} \tag{2}$$

where $Dif(R_1,R_2)$ is the difference of two regions R_1,R_2. $MInt(R_1,R_2)$ denotes the minimal invariant difference

$$D(R_1,R_2)=\min\,(v_{i\in R_1},v_{j\in R_2},(v_i,v_j)\in E^{W(e)}) \tag{3}$$

$$\mathrm{MInt(R1,R2)}=\min\,(\mathrm{Int}(\mathrm{R_1})+\tau(\mathrm{R_1}),\mathrm{Int}(\mathrm{R_2})+\tau(\mathrm{R_2})) \tag{4}$$

where $\mathrm{W\,(e)}=\parallel \mathrm{Int}(\upsilon_i)-\mathrm{Int}(\upsilon_j)\parallel$ is the corresponding weight of the edge which connects the υ_i and υ_j , and was estimated from either curvature or shape index. The invariants difference $\mathrm{Int(R)}=\max_{\max_{e\in \mathrm{MST(R;E)}}} \mathrm{w(e)}$ where the MST is the minimum spanning tree and E corresponding to pairs of neighboring vertices. $\tau=k/|\mathrm{R}|$ where the k are some constant parameters and $|\mathrm{R}|$ is the size of R. The segmentation obeys the properties of being neither too coarse nor too fine, according to the following definitions:

Definition 1: A segmentation S is too fine if there are some pairs of region $R1$, $R2 \in S$ where no evidence for a boundary exists between them.

Definition 2: A segmentation S is too coarse if there exists a proper refinement of S that is not too fine.

The most important reason to implement the segmentation in this method is that it can perceptually capture important groupings or regions that often reflect global aspects of the image. Figure 3(a) show the results of the graph-based segmentation.

The points of the input surface with a high information content related to the surface feature need to be detected. The data sampling can be grouped into two categories: deterministic sampling and probabilistic sampling. Because the surface has been segmented into a limited number of regions, the feature points can be selected from each region. The selection criterion is based on each segment and selects points inside them by thresholding the absolute shape index values. Figure 3(b) shows the results after point sampling. It can be found few points are selected in the flat region, whereas more are selected in the high information content related regions, such as the eyes of bird and tubby. The points that have been sampled by the local sampling algorithm sometimes focus too much on the local details and overlook the overall shape, hence, the selected points cannot represent the model shape well and might miss meaningful parts of the shape. In order to address the problem, the global sampling is applied.

$$\mathrm{H(X)}=\sum_{\mathrm{i=1}}^{\mathrm{n}}-\mathrm{P(r_i)}*\log_2(\mathrm{P(r_i)}) \tag{5}$$

where $\mathrm{P(r_i)}$ is the probability of vertex in a local voxel, the number of vertices in a local voxel allows to be controlled by a parameter, n denotes the total number of the vertices.

- The third step chooses one or more points from each sub box based on the estimated entropies. Normally, the maximum entropy is chosen.
- The minimum distance between the samples is employed to address the possibility that the generated samples are close to the boundary between two or more adjacent voxels that might be too close to each other.

As shown in Figure 3(c), we can see the sampled points are globally distributed. The points which have been sampled could represent the overall mode shape well, but miss a lot of particular features such as the face of tubby, and the ear of bird.

As we have presented above, the local and global points have been sampled. In order to achieve a better simplification result, the sampled points are combined together here. By denoting the points from local sampling as Ls and the points from global sampling as Gs, the simplification approach can be defined as follows:

$$\mathrm{SIM} = \frac{T}{10}L_s + \left(1 - \frac{T}{10}\right)G_s \tag{6}$$

where T controls the ratio of local sampling and global sampling points in the simplification step. If T is a large number, the method still can represent the global shape well, but it may lose a lot of local shape information. When the T goes down, the local shape could be represented well but the global shape may be destroyed. So an appropriate value for T needs to be found.

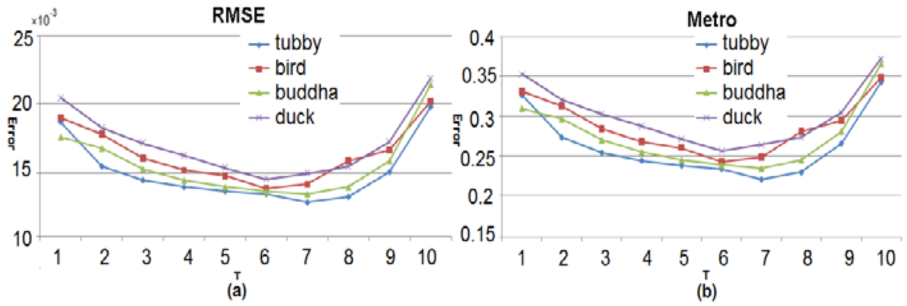


Figure.4 The evolution of the RMSE errors and Metro errors with different parameter T, tested on tubby, bird, buddha and duck. (a) RMSE; (b) Metro.

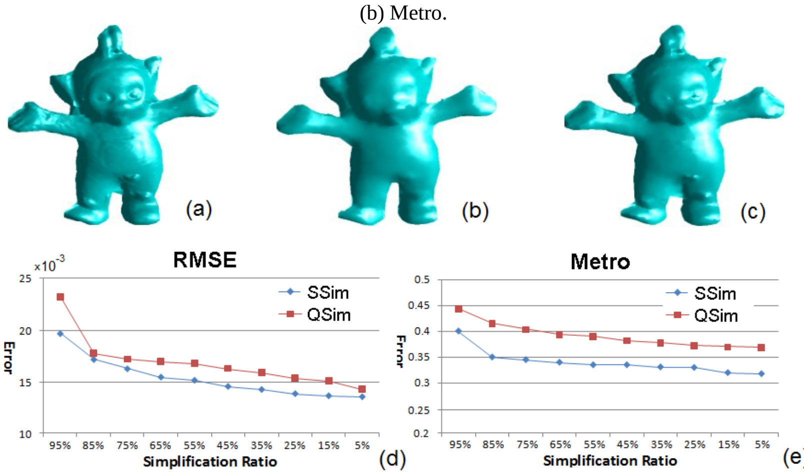


Figure 5. Simplification by using QSim and SSim in simplification rate 90% and the evolution of the RMSE errors and Metro errors with different simplification rate tested on tubby. (a) Original; (b) QSim; (c) SSim; (d) RMSE; (e) Metro.

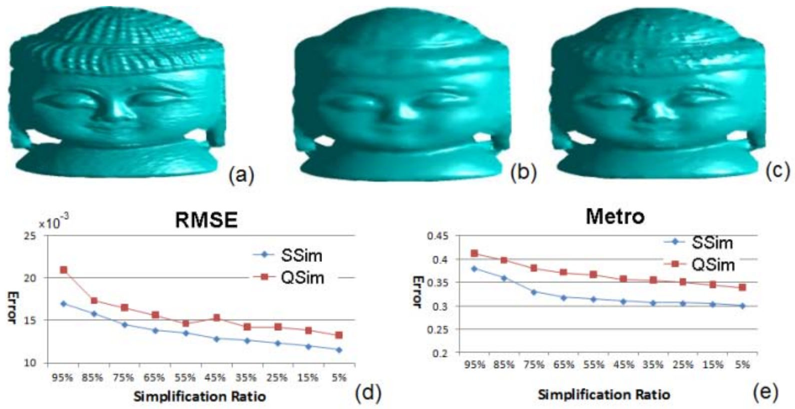


Figure 6. Simplification by using QSim and SSim in simplification rate 90% and the evolution of the RMSE errors and Metro errors with different simplification rate tested on Buddha. (a) Original; (b) QSim; (c) SSim; (d) RMSE; (e) Metro.

III. Experiments

In this paper, a publicly available range image database hosted by the signal analysis and machine perception laboratory at Ohio State University is employed. To this end, several experiments with meshes of differing complexity were performed. All models were simplified on a computer with Intel Core 2 Duo CPU E8400, 3.00GHz. Examples of results are illustrated in Figures 4-9. In order to evaluate the effect of the simplification criteria, our algorithm (SSim) has been compared with the geometric QSlim algorithm [11] which uses the best half-edge collapse. The QSlim algorithm was chosen for the high-quality of its approximations and its code is freely available. Although the QSlim algorithm has been proposed for a long time, it still achieves accurate results. The visual and geometric errors between the original mesh and simplified mesh were measured using root mean squared error (RMSE). RMSE is usually used as an efficient measurement in the mesh simplification since it measures the global average error between the original and simplified meshes. Also the mesh comparison tool named Metro [25] is applied to compare with the RMSE.

In the tool Metro, the Hausdorff distance between two polygonal meshes was measured. The error can be defined as

$$E(S,S') = \sqrt{\frac{1}{\mathrm{AREA}}\int d(p,S')^2 ds} \tag{7}$$

where S is an orientable surface and d is the distance between a point *p* of *S* and *S0*.

In order to make an efficient fusion of global sampled points and local sampled points, determining an appropriate value for the parameter T is necessary. Fig. 4 shows the errors of tubby which were measured by RMSE and Metro for different values of T. For T=0, it means there in this case all the points are selected globally and it obtains a large error. For T=10, only local points are selected at the point cloud, the global structure information is missing and it makes a messy mesh. Large errors occurred for both cases of above. When the T>0, it seems that there is a lower simplification error. It can be observed that when T=6, angel and bird meet the minimal error value of both RMSE and Metro. When T=7, tubby and duck meet the minimal error value of both RMSE and Metro. This means when T=6 and T=7, the simplification creates the most accurate results. Thus, we choose the average value T=6.5.

Fig. 5 (a), (b), (c) and Fig. 6 (a), (b), (c) show the results for the tubby and buddha models in terms of 90% simplification rate by using QSim and SSim respectively. In these two cases it can be observed that they still contain the major topology characteristics of the initial models. But it can be seen that the results of SSim are better than QSim on both models. For the model tubby, the eyepit and ears are retained better in SSim. For model buddha, the hair and eyebrow are preserved better by SSim but the hair is removed completely when by QSim.

Fig. 5 (d), (e) and Fig. 6 (d), (e) show the SSim errors of tubby and buddha, which were measured by RMSE and Metro for different simplification rate. It can be observed if we increase the simplification rates, the RMSE and Metro errors will become much larger in both QSim and SSim as we expected. Even though, the errors for SSim are much lower than that of QSim.

Fig.7 shows the results of the proposed dual sampling method and Figure 8 shows the examples of models using the sampled points at different simplification rate. Table Ⅰ shows the errors of RMSE and Metro for all models including tubby, bird, Buddha, duck and lobster after SSim and QSim respectively. It can be observed if we increase the simplification rates, the RMSE and Metro errors will become much larger in both QSim and SSim

TABLE I.
RMSE and Metro Errors measured for all models using QSlim and SSim

Model	Simpification rate	RMSE ($\times 10^{-3}$)		Metro	
		Qsim	Ssim	Qsim	Ssim
tubby	50%	14.61	12.11	0.26	0.23
	90%	17.2	13.32	0.31	0.27
	95%	19.73	15.3	0.33	0.32
bird	50%	16.81	9.45	0.35	0.21
	90%	16.22	11	0.41	0.33
	95%	21	11.63	0.5	0.39
buddha	50%	16.33	10	0.47	0.33
	90%	17.85	12.12	0.58	0.42
	95%	23.28	13.39	0.61	0.51
duck	50%	12.44	11.16	0.21	0.19
	90%	15.33	14.21	0.3	0.23
	95%	18.46	17.34	0.37	0.31
lobster	50%	21.54	16.53	0.45	0.39
	90%	25.42	18.21	0.57	0.42
	95%	28.32	21.24	0.67	0.53

as we expected. Even though, the errors for SSim are much lower than that of QSim.

Fig. 9 shows snapshot of the small topological feature preservation in terms of 90% simplification rate. As shown in Figure 6, the eye region on the bird surface can be persevered well after SSim, but the shape of eye is over smoothed by QSim. For the surface buddha the hair is removed completely by QSim, but the outline of the curls still presented well by SSim. That shows the SSim can preserve the small topological feature that is visually important.

IV. Conclusions

In this paper, we present a framework for mesh simplification by combining surface segmentation and the points sampling and the method can retain the topology and small features well. Firstly, it estimates the shape index of all vertices and then the model is segmented into patches using the graph-based segmentation method.

Secondly, it detects the local feature and samples the feature points from the model. Then the global points sampling is applied by using the surface voxelization.

Finally, it combines the points from local and global sampling. Compared with the simplification algorithms such as edge collapse, a number of experimental results have illustrated that the proposed method not only preserve the small features on the models but also retain the overall mesh topology, because the local features have been extracted. Therefore, the geometric quality is not affected.

The combination ratio from local and global sampled points plays a key role during the simplification. So we intend to design more efficient methods to select the parameters of the combination ratio in future work.

Acknowledgment

This work is supported in part by the science and technology project of Chongqing Education Commission (No. KJ120413).

References

[1] W. Schroeder, J. Zarge, and W. Lorensen, “Decimation of Triangle Meshes,” In Proceedings of *SIGGRAPH*, pp. 65-70, 1992.

[2] K. Reinhard, “Multiresolution representations for surfaces meshes based on the vertex decimation method,” *Computers & Graphics*, vol. 22, no. 1, pp.13-26, 1998.

[3] Y. Y. Tsai, “An adaptive steganographic algorithm for 3D polygonal models using vertex decimation,” *Multimedia Tools and Applications*, pp. 1-18, 2012.

[4] J. Wu and L. Kobbelt, “A stream algorithm for the decimation of massive meshes,” In: *Graphics Interface*, pp.185-192, 2003.

[5] F. Larue, M. D. Benedetto, M. Dellepiane, and R. Scopigno, “From the digitization of cultural artifacts to the Web publishing of digital 3D collections: an Automatic Pipeline for Knowledge Sharing,” *Journal of multimedia*, vol. 7, no. 2, pp. 132-144, April 2012.

[6] H. Song and H.Y. Feng, “A global clustering approach to point cloud simplification with a specified data reduction ratio,” *Computer-Aided Design*, vol. 40, no. 3, pp.281-292, 2008.

[7] J. Rossignac, and P. Borrel, “Multi-resolution 3D approximations for rendering complex scenes,” In *Modeling in Computer Graphics*, pp.455-465, 1993.

[8] Y. Q. Lei, “Geometric features of 3D face and recognition of It by PCA,” *Journal of multimedia*, vol. 6, no. 2, pp. 207-216, April 2011.

[9] T. Boubekeur, and M. Alexa, “Mesh simplification by stochastic sampling and topological clustering,” *Computers & Graphics*, vol. 33, pp. 241-249, 2009.

[10] P. Castello, M. Sbert, M. Chover, and M. Feixas, “Viewpoint-based simplification using f-divergences,” *Information Sciences*, vol. 178, pp. 2375-2388, 2008.

[11] H. Hoppe, “Progressive Meshes,” In Proceedings of *SIGGRAPH*, pp. 99-108, 1996.

[12] H. Hoppe, “View-depended refinement of progressive meshes,” In Proceedings of *SIGGRAPH*, pp.189-198, 1997.

[13] D. Mendez, “On sensor data verification for participatory sensing systems,” *Journal of network*, vol. 8, no. 3, pp. 576-587, March 2013.

[14] M. Garland, and P.S. Heckbert, “Surface simplification using quadric error metrics,” In Proceedings of *SIGGRAPH*, pp. 209-221, 1997.

[15] M. Garland, and P.S. Heckbert, “Simplifying surfaces with color and texture using quadric error metrics,” In Proceedings of *Visualization*, IEEE Computer Soc. Press, pp. 263-269, Oct. 1998.

[16] C. Gonzalez, J. Gumbau, M. Chover, F. Ramos, and R. Quiros, “User-assisted simplification method for triangle meshes preserving boundaries,” *Computer-Aided Design*, vol. 41, pp.1095-1106, 2009.

[17] A. Gueziec, “Surface Simplification with Variable Tolerance,” *Second Annual International Symposium on Medical Robotics and Computer Assisted Surgery*, pp.132-139, 1995.

[18] J. Cohen, A. Varshney, D. Manocha, G. Turk, H. Weber, P. Agarwal, F. Brooks, and W. Wright, “Simplification Envelopes,” In *Proceedings of SIGGRAPH*, ACM SIGGRAPH, pp.119-128, 1996.

[19] J. Cohen, Manocha, D. and M. Olano, “Simplifying Polygonal Models Using Successive Mappings,” In Proceedings of *Visualization*, IEEE Computer Soc. Press, pp.395-402, Oct.1997.

[20] J. Cohen, M. Olano, and D. Manocha, “Simplifying Polygonal Models Using Successive Mappings,” In Proceedings of *Visualization, IEEE Computer Soc. Press*, pp.395-402, Oct. 1997.

[21] D. Nehab and P. Shilane, “Stratified point sampling of 3d models”, In Proceedings of *the symposium on point-based graphics*, pp. 49–56, 2004.

[22] R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin,” Shape distributions”, *ACM Transactions on Graphics*, vol. 21, no.4, pp.807-832, 2002.

[23] G. Taubin, “Estimating the tensor of curvature of a surface from a polyhedral approximation,” *ICCV*, vol.4, no.2, pp. 107-114, October 1986.

[24] F. Felzenszwalab, “Efficient graph-based image segmentation,” *IJCV*, vol.59, no.2, pp.167-181, 2004.

[25] P. Cignoni, C. Rocchini and R. Scopigno, “Metro: measuring error on simplified surfaces,” *Computer Graphics forum*, vol.17, no.2, pp.167-174, 1998.

Juan Cao received the B.Sc. degree in Computer Science and Technology from Chongqing Jiaotong University, China, in

2001 and the M.Sc. degree in Computer Software and Theory from Chongqing University, China, in 2008.

She is currently a lecture in the College of Information Science and Engineering, Chongqing Jiaotong University. Her research interests include computer vision, mobile robot visual navigation and artificial intelligence.

Yitian Zhao received the B.Sc. degree in Electronic Engineering from the University of Central Lancashire, UK, in 2008. He also received M.S. on Digital Communication from Newcastle University, UK, in 2009.

He is currently a PhD student in Department of Computer Science, Aberystwyth University, UK. So for, he has published several papers in international's conference proceedings and journals. His research interests lie in 3D image processing.

Ran Song received his B.Sc degree from Shandong University, China in 2005 and his PhD degree from the University of York, UK, in 2009.

He is now a postdoctoral research associate in Department of Computer Science, Aberystwyth University, UK. So far, he has published several papers in international conference proceedings and journals. His research interests mainly include image mosaicing, Markov Random Field and 3D surface proces sing, etc.

Yingchun Zhang received the B.Sc. degree in Computer Science and Technology from Southwest University, China, in 1990.

She is currently an associate professor in the College of Information Science and Engineering, Chongqing Jiaotong University. Her research interests are computer vision, computer education.

Figure 7. Point sampling results on the tubby, bird, Buddha and duck models with simplification rate 90%. Top row: local points sampled. Bottom row: combined global and local sampled points in appropriate combination rate.

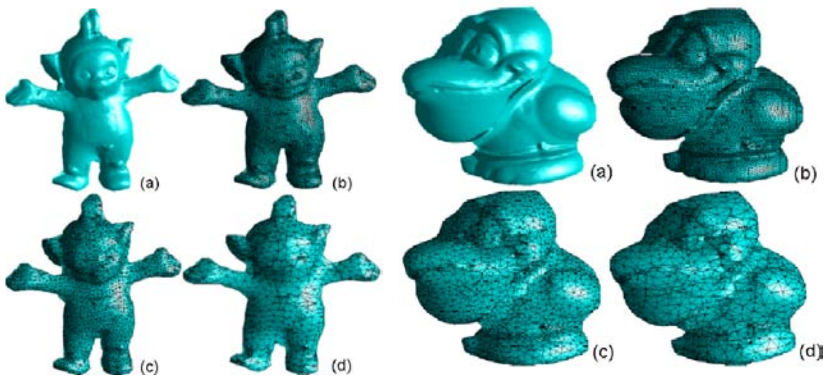


Figure 8. Examples of models simplified at decreasing simplification rate

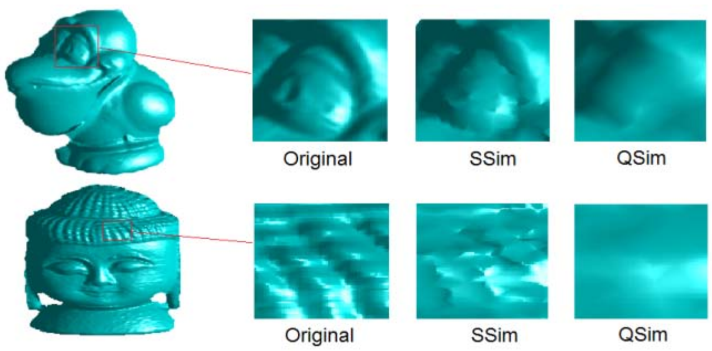


Figure 9. Small topological feature reservation on bird and buddha with 90% simplification rate

Characteristic Value Extraction of Gear Defect based on Image Processing

Cheng Pengfei and Feng Changyong
Henan Mechanical and Electrical Engineering College, Mechanical and Electrical Engineering, Xinxiang Henan, China
Email: chengpengfei1888@163.com

***Abstract*—Aiming at pitting of gear failure forms and attrition, this paper uses image processing software of MATLAB and the research image extractive technique of image character. Based on the two different image expressions of defect forms, this paper abstract defect area ratio as well as the characteristic value and abstract contrast with gray co-occurrence matrix as image characteristics. In order to verify the abstracted image characteristic for evaluating pitting and wearing effectiveness, we use the neighboring method to construct a recognition model, and analyze the collected image samples. The result is correct, but the recognition accuracy needs to be improved. At last, eccentricity ratio and circularity is abstracted as the characteristic value based on image morphological characteristics.**

***Index Terms*—gear defection, image processing, characteristic value**

I. Introduction

Digital image processing processes and evaluates images through computer with particular algorithm. At present, image processing techniques have been applied and researched in various fields with great achievement. Digital image processing can divide into: image transformation, image intensification and restoration, image segmentation, image analysis, image recognition and other technique branches. MATLAB as one kind of high-level computer language, it has a powerful data processing ability that obtains widely application in digital image processing [1].

This paper takes advantage of MATLAB image processing for gear defect detection. Gear shows different defect forms during the using process [2-3]. Corrosive pitting and attrition is the most important type. Different defect forms will create different image characteristics. We can effectively distinguish and decide the defect type through these image characteristics [4-5]. An image processing toolbox of Matlab has great functions. It has an abundant supported image format, and this software provides 15 types of image processing function which includes all the research findings of image processing methods in recent days [6]. Take the advantages of these image processing toolboxes and combine the powerful data handling capacity, it is necessary to pay attention to image format, read-write, display and other details. Instead, we can focus on algorithm research, which greatly improve the working efficiency. Moreover, when testing these algorithms, we can conveniently get the statistical data and view the graphical representation. Reference [7] uses Matlab software to make pre-processing for disadvantage images of silicon solar cell. Then we can operate picture segmentation, characteristic abstract and image recognition. At last, it is necessary to judge and test disadvantages of silicon solar cell. Reference [8] researches about digital image processing that how to realize the non-contact measurement of geometric dimensioning in straight spur gear. Reference [9-10] aims at the characteristics of gear defect image, and provides the research of these defect abstract technologies. Although there have relevant references about image characteristics of gear defect, the above researches only find out the defects by image processing technology. It cannot distinguish the gear defect formats, and the characteristic differences among the various defects. Therefore, it is necessary to find the deeper research, which will provide powerful technical support for the working condition of high-efficiency gear monitoring [11].

Gear drive is the most important transmission type of mechanical transmission [12]. The gear will operate monitoring and control during the process. For the normal machine usage, especially the emphasis machine such as gearbox, and reducer, it is always the difficulty and key technology [13]. We can take the advantage of image processing to detect gear defect. It is an effective method and still belongs to the starting. During the entire research, we combine image character of gear defect, user the mature image processing theory, and graying the selected defect image samples. Moreover, we provide relative weighted factor. Filtering and noise reduction considers about the much noise point after long-term usage of the gear. At the same time, it considers about the image feature requirement [14-15]. The key point of image filtering is to keep defect contour feature. Moreover, the gear defect image has differences in shape and location. How to extract texture feature and morphological characteristics is the key point and innovation point of this research [16].

II. Image Collecting and Processing

A. Experimental Material and Equipment

Computer and digital camera belong to hardware system. Use CCD (Charge Coupled Device)

doi:10.4304/jmm.8.3.198-205

photosensitive sensor can reduce effects from other lights. Image collection processes in camera obscura. Then the typical gear defect image will be collected. MATLAB software is an image processing instrument.

B. Graying Processing of Image

The color has three characteristics. They are lightness, tone, and saturation. Human eyes can distinguish different colors through these three characteristics. Color lightness is one kind of energy measurement. The more energy that reflected on color surface, the object will have higher lightness. However, the tone is the characteristic to distinguish different colors. It indicates the color type depends on dominating wave length. Saturation means the close degree between one color and monochromatic light of the relative tone. It is as well as the color clarity. Human eyes have different sensitivities about R, G, and B lightness. We can use lightness to be the basis of image graying, which means ignore tone and saturation of various pixels and take the lightness information only. This method will obtain a better subjective color effect. Theoretically, the same R, G, B addition can turn into white.

MTALAB image can read image data through imread command. Due to the RGB file that we obtained, it includes color information and luminance information which has great processing difficulty. Therefore, we have to project colorful information on the gray space. RGB image transforms into a gray image we need to use rgb2gray function of Matlab. [1]

$$f(i,j) = \alpha R(i,j) + \beta G(i,j) + \gamma B(i,j)) \qquad (1)$$

In this formula, R , G , B is three components of one pixel of color image. α , β , γ is strength coefficient. People have the greatest sensitive to green, the lowest sensitive to blue. Therefore, under the following formula we can operate weighted average to RGB for obtaining the reasonable gray image. In function rgb2gray, the parameter value is [0.30 0.59 0.11] [3].

We process the collected image under the above formula, 256 -level gray image will be obtained. The image after processing is in figure 1.

Figure 1. Image after graying

C. Image Smoothing

During the collection, conversion and transformation procedure, the image will be influenced by the imaging device, transmission equipment, and outside noise interference. It will infect some noise signals and influence image quality. In order to reduce this kind of effect as far as possible, we need to depress noise process. This paper provides two general filtering methods [2]: neighborhood averaging and median filtering.

Neighborhood averaging: neighborhood averaging is one kind of linear smoothing filtering. It is suitable for removing gain noise in the images. Every image point of (m,n) in the provided image $f(i,j)$ we take the neighborhood S . If we suppose S has M pixels, and the average value is the gray level of image point (m,n) after processing. The shape and size of the neighborhood S will be confirmed under image point. The ordinary obtained shapes are square, rectangle, and cross. The point (m,n) usually places in the center of S . Neighborhood averaging means average window pixel replaces the original gray value. This paper uses 3×3 pixel image $F(x,y)$. After smoothing process, we can obtain an image $G(x,y)$. And $G(x,y)$ is decided by the following formula.

$$G(x,y) = \sum_{(x,y)=L} G(m,n) / L2 \qquad (2)$$

In the formula, $x, y = 0,1,2,\cdots, N-1$, and L is a neighbor coordination collection of (x,y) . However, (x,y) is not included, and L is the coordination amount of the collection.

(2) Median filtering

Median filter is one common low-pass smoothing filter. Through median filtering we can effectively remove noise. At the same time, it can protect the image edge. The principle is sequence area pixels follow gray levels. The middle value is output pixel. [15]The expression is:

$$G(i,j) = \underset{(x,y)\in A}{Med}[G(i,j)] \qquad (3)$$

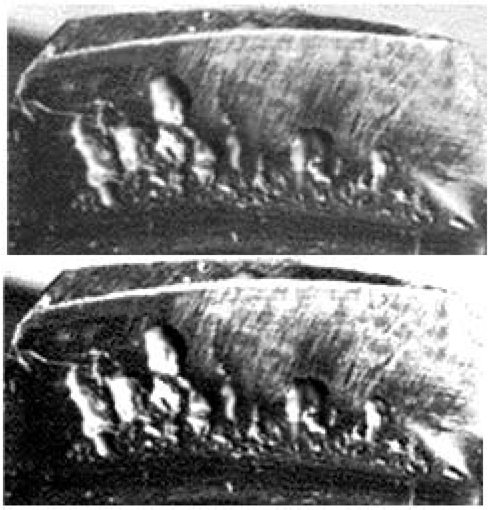


Figure 2. Images after two kinds algorithms

The median filter is easy that can better protect the edge. However, the window size and shape will influence the filtering effect. This paper compares the different windows of 3×3，5×5，7×7. The too large window will leach the lesion part. Moreover, this paper uses 5×5 rectangular window for filtering. The figure 2 is the comparison of two different filtering effects. The upper is neighborhood average and the lower one is median filtering.

Compare these two flatting methods, neighborhood averaging has the characteristic of simple and practicable, fast operating rate, and perfect denoising effect. However, the neighborhood averaging leads a certain degree of boundary dim when removing the noise and it will affect the accuracy of boundary extraction. Although, the median filtering is better than the neighborhood averaging in boundary conservation and texture detail, it is necessary to consider about the size sequence of each pixel.

Compare the above images we can find out, the image after median filtering has clear contour. Therefore, this paper selects median filtering for the smoothing process before texture analysis.

D. Image Segmentation

For the convenient analysis of gear defect detection, we need to test image edge. Matlab software provides differential operator functions for edge test [2]. The common differential operators are Roberts operator, prewitt operator, Sobel operator, canny operator, and so on. Canny operator is the best edge detection operator of trap cut. We use canny edge detection for image strengthen processing. The image after edge detection processing is figure 3.

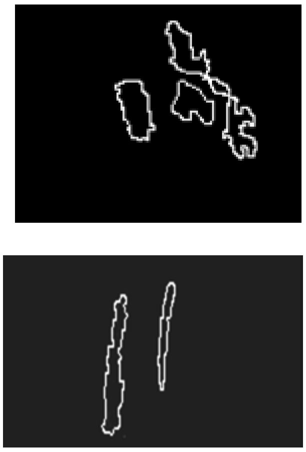


Figure 3. Image after edge detection processing

In the image area after segmentation, for pitting defect concentrates upon the gear mesh pitch line with dotted distribution. Moreover, gear attrition expresses teeth surface scarification. The area calculation principle is: binary image f (x , y) = 1, 1 stands for the object, 0 is the background, and the area is the number of f (x , y) = 1 after statistic. From the area static we can find out two types of gear defect forms have great differences in the area occupancy. We can effectively distinguish gear defect form through area ratio. Therefore, area ratio is one reference characteristic value to distinguish normal tooth surface and the defected tooth surface. The pixel area ratio is in table 1.1. .

TABLE I. TWO TYPES OF DEFECT FORMS AND NORMAL PIXEL STATISTIC

Lose efficiency	Normal pixel	Defect pixel	Total pixel	Area ratio
Attrition	51215	88261	139476	63.28%
Pitting	119740	19736	139476	14.15%

III. IMAGE CHARACTERISTIC EXTRACTION

A. Parameter Research of Texture Characteristic

Texture is one common method during image evaluation. Because the texture formed through the repeatable appearance of gray distribution on the location. Therefore, two pixels with distances in the image space will has gray relationship. They have abundant periodicity. The general three texture descriptions are [5]: statistical method, structural method, and frequency spectrum. The statical method describes the texture with gray co-occurrence matrix [3]. It is a common method to describe the texture through researching space correlated characteristic. We set S is the target area, and R is a pixel collection with special space relation. The co-occurrence matrix P can be defined as

$$P(g_1, g_2) = \frac{\#\left\{\left[(x_1, y_1), (x_2, y_2)\right] \in S \;\middle|\; f(x_1, y_1) = g_1 \,\&\, f(x_2, y_2) = g_2\right\}}{\# S}$$

In this formula, the numerator stands for the pixel number g_1 and g_2 with space relationship and gray value. The denominator stands for pixel amount. If we divide denominator S and it stands for gray co-occurrence matrix after normalization (# stands for the number), we set step length is 2, and the direction is 45 degrees. Then we can get the following texture description value [5]:

(1) Energy: $W_1 = \sum_{i=0}^{N} \sum_{j=0}^{N} [p(i, j)]^2$

The energy character is one uniformity metrical method of image gray level distribution. The larger W_1 wheel has thicker texture with more energy. On the contrary, the smaller W_1 value will have a thicker texture with less energy.

(2) Contrast: $W_2 = \sum_{t=0}^{N-1} t^2 \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} p(i, j) \right\}$

$t = |i - j|$

The contrast reflects image intelligibility. The deeper groove in this image will has larger value.

(3) Entropy: $W_3 = -\sum_{i=1}^{N} \sum_{j=1}^{N} p(i, j) \log\left[P(i, j)\right]$

Entropy value can reflect an information amount. The larger entropy value means fine texture. The smaller entropy value stands for fewer texture distribution in the image. We select various defects of obvious image samples, after three above calculation, the result is figure 4.

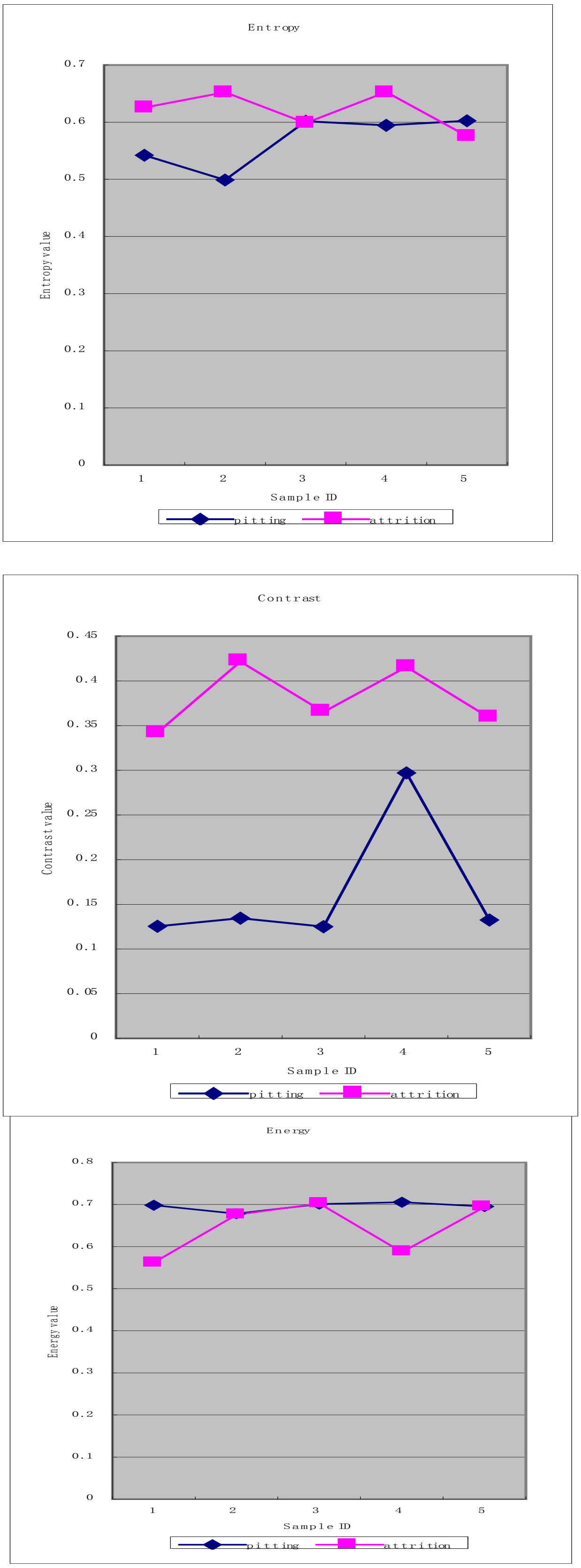


Figure 4. Characteristic value of gray co-occurrence matrix

In figure 4 we can find out the contrast value can obviously distinguish two different gear defects. Because the contrast value reflects the groove depth in the texture. The image display of the tooth's surface pitting has a deeper expression of the tooth surface attrition. Therefore, contrast can be characteristic value for judging pitting and attrition. However, the capacity and entropy has no obvious characteristic expression that we can abandon them.

B. Extraction of image morphological characteristics

Image morphological characteristic is one common feature of the image characteristic. Morphological characteristic has high stability that cannot change with image color variation and other mutative conditions.[5] Except texture features, different gear defect form has different morphological characteristics. Therefore, the characteristic can be judgment basic of gear defect form.

We can abstract morphological characteristic parameters. The most used method has contoured tracing [5]. Contour tracing means tracing the boundary through finding edge points in sequence. If we use contour tracing we need to make image binaryzation.(Image binaryzation is figure 2.3)It will random take one point near the edge, and trace it under the following rules.

(1) One pixel at one step.

(2) When step from the white area into the black area, turn left until going out the black area.

(3) When step from the black area into the white area, turn right until going out the white area.

One circulation of the above steps, the trace is image drawing object.

Figure 5. Image after binaryzation

This paper uses three parameters of morphological characteristics [5-6].

(1) Eccentricity ratio (E)

$$E = \frac{\min\{W, H\}}{W, H}$$

The long and thin object will has smaller E . If the image is a circle, eccentricity ratio will has the greatest value 1. We set direction code a_i has a component of a_{ix}, a_{iy} on x，y axis and x_o, y_o is the initiative coordination. Therefore, the width is

$$W = \max\left(\sum_{(k=1)}^{i} a_{kx} + x_o\right) - \min\left(\sum_{k=1}^{i} a_{kx} + x_o\right) \quad (8)$$

and the height is

$$W = \max\left(\sum_{(k=1)}^{i} a_{ky} + y_o\right) - \min\left(\sum_{k=1}^{i} a_{ky} + y_o\right) \quad (9)$$

(2) Circularity (C)

$$C = \frac{\mu_R}{\sigma_R} \quad (10)$$

In this formula, μ_R is the average distance from area center to the contour point. σ_R is the means square error from the main part to the contour point.

$$\mu_R = \frac{1}{K}\sum_{K=1}^{K-1}\left\|(x_k, y_k) - (\overline{x}, \overline{y})\right\|$$

$$\sigma_R = \frac{1}{K}\sum_{K=1}^{K-1}\left[\left\|(x_k, y_k) - (\overline{x}, \overline{y})\right\| - \mu_R\right]^2$$

Circularity uses for description of image shape and circle deviation degree. If the image looks like a circle, it will have higher value. If the image looks like a rectangle, it will have smaller value.

(3) Rectangularity (R)

$$R = \frac{S}{W \times H}$$

In this formula, circle $S = \sum_{i=1}^{n} a_{ix}\left(y_{i-1} + \frac{1}{2}a_{iy}\right)$, and $y_i = y_{i-1} + a_{iy}$. Rectangularity explains the image shape and the rectangle deviation degree. If the image looks like a matrix, R will have the largest value 1. The morphological character value is in table 2.

TABLE II. MORPHOLOGICAL CHARACTERISTIC VALUE OF GEAR DEFECT

Type	Eccentricity ratio	Circularity	Rectangularity
Pitting	0.698	36.219	0.689
Attrition	0.302	2.365	0.654

From the above characteristics, we can find out much more pitting areas are approachable to circle and near by the pitch line. However, attrition parts are on the whole gear surface with strip shapes. The statistical data shows these two types of defect forms have great differences in eccentricity ratio and circularity about pitting and attrition. However, the un-obvious differences of rectangularity cannot work as a character value.

IV. Classifier Establishment.

Aiming at the different objects and targets, there will have various pattern recognition theory and method. The common pattern recognition methods have the traditional mode recognition, syntactic pattern recognition, fuzzy pattern recognition, neural network pattern recognition, and so on. For the few testing samples in this paper, we select neighborhood decision under the limited samples. It selects being recognized neighborhood. If the recognition sample x has many neighborhood sample belong to type w_i, we can decide x belongs to w_i [14].

We suppose there has one c type, $w_1, w_2, w_3 ......w_c$ is the pattern recognition problems. Each type has N_i marked category samples. We can set judgment function of w_i is:

$$g_i(x) = \min\left\|x - x_i^k\right\|, k = 1, 2, ...N_i \tag{12}$$

The corner mark i of x_i^k stands for category w_i, k stands for the k th sample in N_i of category w_i. Follow the above regulation, we can write the decision rule as:

$$\text{If } g_j(x) = \min g_i(x), i = 1, 2, ..., c \tag{13}$$

Ten the decision $x \in w_j$

This decision method is the nearest neighbor method. The explanation is very simple. For the unknown sample x, we only need to compare the Euclidean distance between x and the given category samples of $N = \sum_{i=1}^{c} N_i$. Then we can decide the sample similar which is closest to it. The error ratio is:

$$p^* \le p \le p^*(2 - \frac{c}{c-1}p^*)$$

p^* is the Bayes error ratio, c is the number of categories. Therefore, it is the important method of pattern recognition.

However, the k neighborhood method has one obvious disadvantage that it only consider one category number which belongs to the neighborhood without thinking about the distance disparity. Apply fuzzy classification and here we can use the subordinating degree function to get over this defect.

The detailed method is in the following:

The k th sample $\{x_i, i = 1, 2, ...k\}$ of the under recognized sample x, we can use the following formula to calculate x degree of membership to the various categories.

$$u_j(x) = \frac{\sum_{i=1}^{k} u_j(x_i)(1/\left\|x - x_i\right\|^2)}{\sum_{i=1}^{k}(1/\left\|x - x_i\right\|^2)}, j = 1, 2, ...c$$

In this formula, c is category number, $\left\|x - x_i\right\|$ is Euclidean distance between x and x_i, $\left\|x - x_i\right\| = [\sum_{i=1}^{n}(x_i - y_i)^2]^{1/2}$. $u_j(x_i)$ is the j th degree of membership of sample x_i, j is the category.

Actually, $u_j(x_i)$ is the weighting based on different neighbor distance of under recognized category samples. The formula is in the following.

$$u_j(x_i) = \frac{1/\left\|x_i - m_j\right\|^2}{\sum_{k=1}^{c} 1/\left\|x_i - m_k\right\|^2}, i = 1, 2......n, j = 1, 2.....c$$

m_j is the j th category average value, $m_k = \frac{1}{N_k}\sum_{i=1}^{N_T} x_i$, and N_K is the k th category sample number.

In order to get over the out wild-value effect (if one point is far from each center of clustering, under the requirement of $\sum_{j=1}^{c} u_j(x_i) = 1$, it will have larger degree of membership in different categories), we uncage the normalizing condition, and the membership of various categories will havea totall of n, which means:

$$\sum_{j=1}^{c}\sum_{i=1}^{n} u_j(x_i) = n$$

$$u_j(x_i)=\frac{n(1/\|x_i-m_j\|^2)}{\sum_{k=1}^{c}\sum_{j=1}^{k}(1/\|x_j-m_k\|^2)},i=1,2...n,\ j=1,2...c$$,based on $u_j(x_i)$ we can obtain category sample x_i has degree of membership to category j , compare the different degree of membership and get the maximum as the type.

A. Structure of Recognition Model

Binary tree is the simple form of decision tree. The binary tree means each tree node has only two branches, except leaf notes. The structure makes easy principle of categorizer. Moreover, each node can select different characteristics and use various decision rules.

The defect formats have complicated conditions. We cannot obtain the correct judgment of character value through one gear defect sample. Therefore, we need to build recognition model. In order to reduce category level and refer to artificial discrimination, this paper uses textural features and morphological characteristics. The model establishment is under the defect, and the defect type. The grading judgment builds two categorizers. The first categorizer judges normal gear tooth surface and the second categorizer judges the defect type. Each categorizer uses neighborhood method of fuzzy K. the structure chart is figure 6.

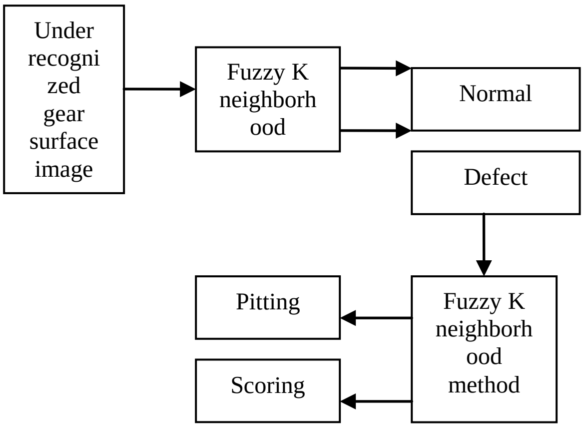


Figure 6. Structure chart of gear defect judgment

B. Categorizer Establishment

Follow the above methods and we need to recognize normal gear surface and defected gear with fuzzy K neighborhood method. We need to select each ten normal samples and ten defect sample. Based on the previous, we select texture and morphology parameter as well as contrast value of gray-level cooccurrence Matrix and feature vector of eccentricity ratio and circularity. For the balance of input parameter and avoid excessive overstate or decrease some factor functions, it is better to normalize all the input samples. The normalization processing follows the next formula [21].

$$x_j=\frac{0.8}{x_{\max}-x_{\min}}x_j+\frac{0.1x_{\max}-0.9x_{\min}}{x_{\max}-x_{\min}}$$

x_j is the j th sample character parameter, $x_{\max}x_{\min}$ is the maximum and minimum value of sample character parameter.

The mean vector of two types of samples:

$$m_z=\frac{1}{10}\sum_{i=1}^{10}x_i^z$$

$$m_b=\frac{1}{10}\sum_{i=1}^{10}x_i^b$$

Each sample x_i has degree of membership to normal tooth surface and defect tooth surface:

$$u_z(x_i)=\frac{20/\|x_i-m_z\|^2}{\sum_{k=1}^{2}\sum_{j=1}^{20}(1/\|x_j-m_k\|^2)},i=1,2......20$$

$$u_b(x_i)=\frac{20/\|x_i-m_b\|^2}{\sum_{k=1}^{2}\sum_{i=1}^{20}(1/\|x_j-m_k\|^2)},i=1,2......20$$

In order to avoid $k_1=k_2$ and k_1 is odd number, we select the nearest three samples from under recognized sample x and given 20 samples. The membership of under recognized sampel x and defect tooth surface is:

$$u_z(x)=\frac{\sum_{i=1}^{3}u_z(x_i)(1/\|x-x_i\|^2)}{\sum_{i=1}^{3}(1/\|x-x_i\|^2)}$$

$$u_b(x)=\frac{\sum_{i=1}^{3}u_b(x_i)(1/\|x-x_i\|^2)}{\sum_{i=1}^{3}(1/\|x-x_i\|^2)}$$

If $u_z(x)>u_b(x)$, the system will judges it into normal tooth surface. If $u_z(x)<u_b(x)$, it will be the defected tooth surface.

The defected tooth surface will turn into the next step, and the system will judge the defect format.

3.3 Categorizer establishment among lesions

Based on the above description, we select texture and morphology parameter. Contrast value of defect part and the characteristic vector of eccentricity ratio and circularity. Same with the judgment of normal gear tooth surface and defect gear tooth surface, we need to make normalization processing before judgment.

$$m_{b1}=\frac{1}{5}\sum_{i=1}^{5}x_i^{b1} \qquad m_{b2}=\frac{1}{5}\sum_{i=1}^{5}x_i^{b2}$$

$$u_{b1}(x_i)=\frac{10/\|x_i-m_z\|^2}{\sum_{k=1}^{2}\sum_{j=1}^{10}(1/\|x_j-m_k\|^2)},i=1,2......10$$

$$u_{b2}(x_i)=\frac{10/\|x_i-m_z\|^2}{\sum_{k=1}^{2}\sum_{j=1}^{10}(1/\|x_j-m_k\|^2)},i=1,2......10$$

$$u_{b1}(x)=\frac{\sum_{i=1}^{3}u_{b1}(x_i)(1/\|x-x_i\|^2)}{\sum_{i=1}^{3}(1/\|x-x_i\|^2)}$$

$$u_{b2}(x)=\frac{\sum_{i=1}^{3}u_{b2}(x_i)(1/\|x-x_i\|^2)}{\sum_{i=1}^{3}(1/\|x-x_i\|^2)}$$

C. Error Analysis

This research use normal gear, wear gear, and scoring tooth surface. Each test set has 20 samples. Moreover, this research tests ten samples of normal tooth surface and wearing tooth surface. The testing evaluation is: the first categorizer has a better recognition effect. It has a higher analysis accuracy of normal sample, which reach to 80%. The second categorizer has lower accuracy that compares to the first one, the ratio is 80.6%. We judge the pitting accuracy rate is 80%, and the wearing accuracy rate is 75%.

Evaluation error will caused by the following reasons:

In the gear defect image, the different defect formats will express differences in the same defect through statistical characteristics. The individual sample will be erroneously judged.

The testing method analysis will create errors. The system recognition principle judges the standard sample data as the defected sample. The accuracy depends on the standard sample. The condition limitation makes hardly pure samples to be tested, which will lead bad recognition.

Image selection and pretreatment of upper image will have different effects of artificial and natural factors such as illumination, light intensity, and temperature.

The selected pattern recognition needs to be further improved.

(5) When we selecting samples, the sample quantity is limited. We use ppaper samples with helical gear. The helical gear has different defect format that compare with the normal gear in the image expression during the using process. This condition might cause decrease of accuracy rate.

D. Eexperimental Result

TABLE III. GEAR PARAMETER TABLE

Type	Modulus (m)	Tooth number (z)	Addendum coefficnent (h_a^*)	Bottom clearance prarameter (c^*)
1	4.5	45	1.0	0.25
2	2.5	26	1.0	0.25

The experimental process builds two identification modes. They are defect identification and normal sample identification, gear pitting and gear wearing sample identification. Each identification selects 20 samples. The result is in table 3.2, 3.3. When we are identifying normal tooth face and defect tooth fact, in order to ensure accuracy rate reliable, we will select tooth surface pitting and wearing of each sample. Part of the helical gears have helical angle $\beta=12^\circ$.

TABLE IV. DISTINGUISH THE ACCURACY RATE OF NORMAL TOOTH SURFACE AND DEFECT TOOTH SURFACE

	Normal identification	Defect identification
Normal	85%(17)	15%(3)
Defect	10%(2)	90%(18)

TABLE V. DISTINGUISH ACCURACY RATES OF TOOTH SURFACE PITTING AND TOOTH SURFACE WEARING

	Distinguish to pitting	Distinguish to wearing
Pitting sample	80%(16)	20%(4)
Wearing sample	25%(5)	75%(15)

V. SUMMARY

This paper uses computer image processing of MATLAB software. Moreover, it works many management to the selected gear defect form of pitting and attrition images such as graying process, smoothing process with the median filtering, and segmentation process with Canny operator. This paper takes the advantage defect pixel area after segmentation as characteristic value. Image texture characteristic uses gray co-occurrence matrix after normalization for extracting contrast as the characteristic. At the same time, contour tracing abstracts eccentricity ratio, circularity, and rectangularity. We can find out the eccentricity ratio and circularity can effectively distinguish pitting and attrition defect. We can set the diacritical point as pitting and attrition characteristic. At last, we use neighboring decision of pattern recognition to build the categorizer to distinguish normal tooth surface, defected tooth surface, pitting the tooth surface, and wearing tooth surface. We test and verify the recognition with ten samples, the results are ideal, but the accuracy needs to be improved.

Gear defects include pitting, attrition, and scoring. From the image expression, pitting and scoring, and attrition have obvious different characteristic value. However, attrition and scoring has the similar image characteristic that is hard to judge. How to use these methods to abstract image characteristic is the further research.

ACKNOWLEDGMENT

Natural science research of Henan Education Department (2008B210001).

REFERENCES

[1] Liu Xiaohui, Xu Lin, Xiao Chenjiang, Cao Jianming, Xiao Jiao. Defect identification of silicon solar cell based on

Matlab image processing, *Journal of Shanghai Jiaotong University*, No. 7, pp. 926-930, 2010
[2] Luo Junhui, Feng Peng, etc. MATLAB7.0 application during image processing,*China Machine Press*, 1007, 138-211.
[3] Cui Yanli, Cheng Pengfei, Dong Xiaozhi, Liu Zhihua, Wang Shuangxi. Image processing and characteristic value extraction research of plat disease in the greenhouse, *Transactions of the Chinese Society of Agricultural Engineering (Extra edition)*, No. 21, pp. 32-34, 2005
[4] Gao Shizhong. Fabric texture evaluation based on gray co-occurrence matrix, Computer Engineering and Design,Vol. 29, No. 16, pp. 4385-4388, 2008
[5] Zhang Yujin. Image processing techniques and evaluation basement, *Beijing: Higher Education Press*, pp. 243-245, 2002
[6] Xiao Junjian, Wang Huiying. Image processing application during gear defect detection, Journal of Mechanical Transmission, Vol. 33, No. 2, pp. 98- 100, 2005
[7] Ruan Qiuqi. Digital image processing, *Beijing: Electronic Industry Press*, 2001.
[8] Chen Lianqing, Gao Liguo, Shi Yongjie. Detection research of small plastic gear based on image identification, *China Mechanical Engineering*, Vol. 18, No. 13, pp. 1535-1539, 2007
[9] Liu Xian, Liu Yilun. Extraction research of defect automatic extraction research in X ray image, *Acta Optica Cinica*, No. 7, pp. 1016- 1020, 2006
[10] Shoujia Wang, Wenhui Li, Ying Wang, Yuanyuan Jiang, Shan Jiang, Ruilin Zhao,An Improved Difference of Gaussian Filter in Face Recognition, *Journal of Multimedia*, Vol. 7, No. 6, pp. 429-433,2012
[11] Qiang Deng, Yupin Luo,Saliency Detection by Selective Strategy for Salient Object Segmentation, *Journal of Multimedia*, Vol. 7, No. 6, pp. 420-428, 2012
[12] Qian Zhang,Reconstruction of Intermediate View based on Depth Map Enhancement, *Journal of Multimedia*,Vol.7, No.6, pp. 415-419, 2012
[13] Gang Wang, lin Xiao Gui,DRTEMBB: Dynamic Recommendation Trust Evaluation Model Based on Bidding, *Journal of Multimedia*,Vol. 7, No. 4, pp. 279-288, 2012
[14] Lu Lou, Xin Xu, Zhili Chen,Performance Analysis of a Novel LBS Application Using MBMS&TPEG in 3G Mobile Networks, *Journal of Networks*, Vol.7, No.4, pp. 699-706, 2012
[15] Sihai Zheng, Layuan Li, Yong Li,A QoS Routing Protocol for Mobile Ad Hoc Networks Based on Multipath, *Journal of Networks*, Vol. 7, No. 4, pp. 691-698, 2012
[16] Xiaobo Wang, Xianwei Zhou, Junde Song,Hypergraph based Model and Architecture for Planet Surface Networks and Orbit Access, *Journal of Networks*, Vol. 7, No. 4, pp. 723-729, 2012

Cheng Pengfei is mainly engaged in mechanical design and theory, image processing and pattern recognition technology research

Network Video Online Semi-supervised Classification Algorithm Based on Multiple View Co-training

Nanbin Zhang and Xiaoping Liu
Chongqing Three Gorges Vocational College, Wanzhou 404155, China
E-mail: znb1010@163.com, lxp928@163.com

***Abstract*-As information integration based on multiple modal has to problems like complexity calculation process and low classification accuracy towards network video classification algorithm, came up with a network video online semi-supervised classification algorithm based on multiple view co-training. According to extract the features in text view and visual view, to the feature vector in each view, uses graph as basic classifier and modeling, uses linear neighborhood belief propagation to make category labels propagation in each view, and gets category prediction outcomes in this view; in different views, uses co-training method to online extract unlabeled samples to expand the training set and to incrementally update basic classifier. To the integration of different model prediction outcomes, proposed an integration method aimed at category related. Finally made detailed experimental compare with support vector machine classification algorithm, the result showed, compared with support vector machine, the performance of learner increased greatly, more suitable for large-scaled network video online semi-supervised learning.**

***Keywords*-incremental online learning; text view; visual view; multiple model integration**

I. INTRODUCTION

Network video is a kind of significant data in the Internet, has the characteristics as large scale of data, heterogeneous information coexistence and many unmarked data [1]. Integrated uses heterogeneous information like visual, text and so on, is very important to make online semi-supervised classification for network video [2-4].

In recent years, many scholars used information integration based on multiple modal and made researches on network video classification, Yang Lin jun, etc, used the information like bottom-layered features, semantic features, audio features, additional text features and so on to classify, the result showed, multiple modal classification result was better than single modal, and the effect of support vector machine was the best [5-8]. Cui Bin, etc, extracted the visual features in training set to assist the calculation of word similarity in text message, avoided the extraction of visual feature in classification, increased the velocity and performance of classification [9-11]. Zhang Xu, etc, made characterizing definition used semantic concept model, used the semantic information contained in the text to increase the classification performance, and used incremental support vector machine to classify, but the calculation process was relatively complex [12-15]. To make better use of the relationship among data, Wu Xiao, etc, integrated used network video's header and label, relative video information, video uploader's personal preference and so on to improve the classification performance. Chen Zhi neng, etc, used the result feedback by search engine to assist network video classification [16-17]. The works all above place emphasis on supervised classification but cannot efficiently make online model update. The structure of this text is: the second part in introduced the system of network video online semi-supervised classification, the third part introduced the online semi-supervised classification model and detailed description of the online semi-supervised classification algorithm, finally, made a experimental simulation of the online video semi-supervised classification algorithm based on multiple view co-training.

This paper mainly made explosive and innovative work at the following aspects:

Aimed at the work of multiple modal information integration towards network video classification, placed emphasis on supervised classification but not online semi-supervised classification, cannot efficiently make online model updating, came up with the network video online semi-supervised classification algorithm based on multiple view co-training. Firstly extracted the features in text and visual, to the feature vector in each view, took views as the basis classifier and modeled, and used linear neighborhood propagation to make propagation of category labels of each view, then got the category predicting result in this view; in different views, online extracted unlabeled samples used co-training method to expanding the training set and incrementally updated the basis classifier. To the integration of different modal's prediction results, came up with a integration method aimed at category related. This method used local learning method to model and learns of the data, could get a better accuracy than global learning.

In order to further prove the accuracy and efficiency of network video online semi-supervised classification algorithm based on multiple view co-training, made

doi:10.4304/jmm.8.3.206-212

contrast experiment compared with the classification algorithm of support vector machine, made performance qualification to the category specific integration and the weight unified integration, the experiment result showed: on the condition of a small amount of labeled samples, the method in this text in text view was obviously superior to SVM algorithm, in visual view the two was approximately the same. After the two views integrated, the method in this text was superior to SVM algorithm about 8.3%; the overall classification accuracy of this algorithm reached 99%. The online network flow classification and the accuracy of classification prediction greatly increased, which fitted well the large-scaled online learning.

II. Proposed Scheme

As shown in Figure 1, this frame contains multiple modal feature extraction and online semi-supervised classification. The multiple modal feature extraction mainly contains the extraction to text features and visual features; in online semi-supervised classification, each modal is regarded as a view and uses graph as basic classifier, and then unline update the learner according to co-training method label and extracts unlabeled samples. The specific steps is as following: (a) in initial training set, uses linear neighborhood propagation to learn about the basic classification; (b) in the process of online learning, according to co-training method, uses text (visual) view to predict the unlabeled samples, extracts the samples with high confidence coefficient and with representativeness applies to the visual (text) view; (c) uses learned classifier to classify the text sample. In actual application, the two processes of online learning and classification can operate at the same time.

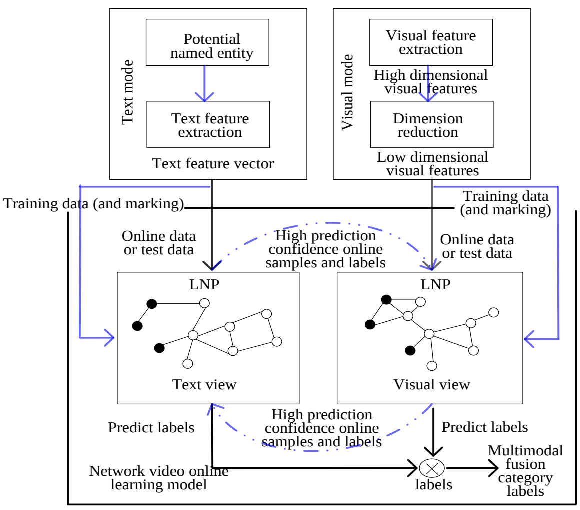


Figure 1. Automatic classification frame

A. Online Semi-supervised Classification Model

In each view, takes sample as apex, the similarity among samples as rims and makes the graph G=(V,W), there V contains initial i labeled samples: $L=\{(x_1,y_1),(x_2,y_2),\ldots(x_i,y_i)\}$, and u unlabeled samples, $U=\{x_{i+1}, x_{i+2,}x_{i+u,}\}$, and the dimension is e of input sample $x_i \subset r^e$; category label vector $y_i = [y_{ic}]_{1\times c \in R^e}$, if x_i belongs to c category, $y_{ic}=1$ and

$$y_{ic} = 0(c^{'} \neq c)$$

Supposes each point x_i in the graph can be linear reconstruction by K neighborhood point $x_{ij} \in N(x_i)$. The reconstruction weigh coefficient can be got by minimum error δ_i:

$$\delta_i = \left\| x_i - \sum_{j:x_{ij}} \in n(x_i) w_{ij} x_{ij} \right\|^2$$
$$s.t. \sum_{j:x_{ij}} x_{ij} \in n(x_j) w_{ij} = 1, w_{ij} \geq 0 \qquad (1)$$
$$s.$$

In the formula: w_{1ij} is reconstruction weight coefficient. Weight vector is w_i=[$w_{1i1}, w_{1i2,\ldots}w_{1ik}]^T$, weight matrix is $w=[w_{1i}]_{n\times n}$. The predicting value e of linear neighborhood propagation calculation by iteration input sample:

$$e_i^t = \sum_{j:x_{ij}} \in n(x_i) w_{ij} f_{ij}^{t-1} \qquad (2)$$

In the formula, e_j^{t-1} refers to the prediction value of x_j when t-1 times of iteration. Decomposes W as:

$$R = [R_{11}, R_{1e}, R_{e1}, R_{uu}]$$

In the formula, R_{11}, R_{1e}, R_{e1} and R_{uu} are four matrix. The prediction outcomes 0.0 in U convergent to

$$R_u = (1 - R_{uu})^{-1} R_u p \qquad (3)$$

In the formula, $p = [p_{ic}]_{n\times c}$ is label matrix.

In online learning, after the training set $x_0 = L_x uus$ adding new samples every time, formula (2) should be recalculated in all samples, the calculation complexity is high and unsuitable for online learning. So the weight vector w_i is revised as following:

$$\bar{w_i} = \phi_i w_i, \phi_i = diag(\theta_{11}, \theta_{12}, \ldots, \theta_{ik}) \qquad (4)$$

In the formula: $\bar{w_i}$ is respectively the weight vector before and after revise, $\theta_{ij} = \exp(d_{1ij} - d_{\min})^2 / p0^2$, d_{1ij} is the distance to its neighborhood point $x_{ij}, d_{\min}, \min_{i,j}, d_{1ij}$ and σ^2 is the variance of sample distance $\{d_{1ij}\}$, this text values $p=10$.

From formula (2) and (4), category label propagation and decay feature and brings the advantages as following: label prediction can expresses confidence coefficient, limits the noise brings by the error-labeled samples at a local area, and in favor of realizing the incremental update of model. Supposes the dataset only contains one label sample (x, y), the unlabeled sample according to label propagation $f_i = [f_{il}, f_{i2}, \ldots f_{ic}]$ can be known by formula (2) and (4):

$$f_{ic} = \sum_{ij:x_{ij} \in N(x_i)} \bar{w_{ij}} f_{ij} \leq (\max f_{ij}). \sum_{ij} \theta_{ij} w_{ij} = \xi_i f_{ic}^* \qquad (5)$$

In the formula:

$$\xi_i = \sum_{il} \theta_{ij} w_{ij} \prec 1, f_{ic}^* = \max f_{ij}$$

Repeats formula (5) can get the first node is $x_{si(1)} = x$

in the path s_i from x to x_i, the sequence number of m node is $s_i(m) = \arg\max_{kxk} \in N(xsi(m-1) f_{kc}$, in this path, the following inequality is true:

$$f_{ic} \le y_c . \coprod_{k \in s} \xi_k \le y_c .(\max \xi_k)^{|s_i|} \le y_c .(\max \xi_k)^{|s_1^*|} \quad (6)$$

In this formula: s_1^* is the shortest path from x_i to x, $|s_i|$ refers to the distance of path s_i, y_c is the c component of y. Formula (6) expresses label propagation takes exponential function as its upper bound, limits label propagation in local area $L_R(X)$ around x, when training set increases new labeled sample (x_a, y_a), this text makes incremental update using following formula:

$$f_{e,r} \left[f_{e,r}, f_n \right] + f_{e,r} \quad (7)$$

In the formula:

$$f_{e,r} = (1 - w_{uu})_t^{-1} \; w, w_{uu} y^{-e}$$

And calculates only in $L_R(x_e)$, f_a is the prediction to x_a based training set x_{t-1} at time t-1.

$$f_a = \sum_{i:x_{ij \in n(x_a)}} w_{ai} f_i$$

The characteristic of algorithm is: when training set added into new labeled sample, the label prediction only needs to make incremental calculation in $L_R(*)$. Figure 2 shows the calculated result in incremental update and all data is quite similar.

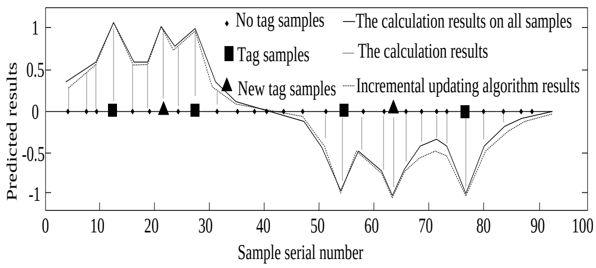


Figure 2. Predicting outcomes in 1-D data

Online semi-supervised classification

Network video online semi-supervised classification algorithm processes based on co-training frame. The collected unlabeled sample set during online operating process is setted as Up. The online semi-supervised learning process is:respectively predicts and labels the sample in Up according to the basic classifier in text and visual view; according to prediction result, extracts "good" unlabeled sample and its predicting label and adds into the opposite training set; makes basic classifier incremental update in new training set

a) The category label prediction in Up*, given* x_i U_p*, the label prediction value of view v(v=1,2) is :*

$$f_i^{(v)} = \sum_{ij} w_{ij} f_{ij}^{(v)} \quad (8)$$

In the formula, x_{ij} is the neighborhood point of x_i in training set X_t. According to formula (8), the calculation need not to change the structure of whole graph thus decreases the complexity of calculation.

b) The extraction of unlabeled sample

How to extracts unlabeled sample to expand training set is the essential problem of semi-supervised learning. The "good" unlabeled sample should be: ① high confidence coefficient of label prediction, the bringing of error prediction sample is equal to brings high noise; ② reasonable distribution in feature space, if the added unlabeled sample cannot better representative the whole sample distribution, the classifier preference improvement would be slow and has bad expansiveness.

From formula (6), if $f_i^{(v)}$ in formula (8) has bigger component than $f_{ic}^{(v)}$, it shows the sample is close to labeled sample. According to manifold assumption, the category of x_i is similar to sample x, so the label prediction result in formula (8) contains the confidence coefficient of label prediction result. For example, $f_{ic}^{(v)} = 0.96$ shows sample x_i has high proportion of being judged as c category in view v. On the other hand, the sample with big $f_{ic}^{(v)}$ alway close to labeled sample, so extracts certain biggest prediction samples from Up would not benefit for category information quickly expanding to the whole space. According to those analysis, extracts "good" sample in view 1 and offers to view 2 by following ways: orders from the biggest to the smallest of the prediction value $f_{ic}^{(1)}$ and $f_{ic}^{(2)}$ in Up, values $f_{ic}^{(1)}$ is the biggest and $f_{ic}^{(2)}$ not in the first 5% $m_{i,c}$ sample. Figure 3 is the schematic diagram of sample extraction, there the black and white node is the samples in $LxUUs$, grey node $X_4^{(1)}, X_4^{(2)}, X_4^{(1)}, X_5^{(1)}$ is the node in . In view 1, the prediction value is pretty big but in view 2 the predicting value is big. From geometric angle, brings $X_4^{(2)}$ in view 2 can actually get makes training set's distribution more suitable for the sample's real distribution than brings $X_5^{(2)}$. The "good" sample extracts in view 2 has similar method of extracting for view1.

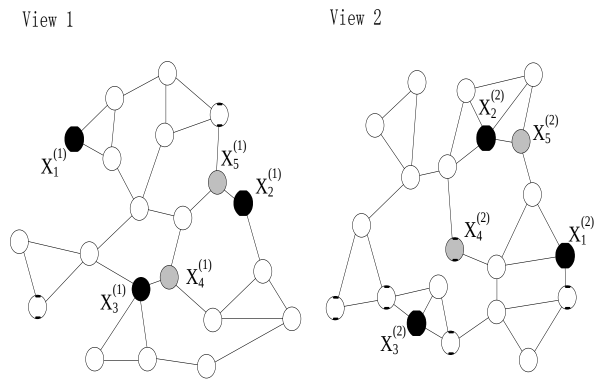


Figure3. Schematic diagram of sample extraction

The online learning algorithm describes as following:

(a) G_1 and G_2 are respectively the text and visual view's corresponding basic classifiers, the label sample is L, unlabeled sample is $U, x_0 = L?U$ is initial training set; R_1, R_2 are online sample set respectively extracted by G_1,G_2, at initial time, $R_1 = R_2 = \theta$;

(b) Respectively forms graphs by the text and visual views of training data X_0, calculates rim weight, makes prediction for sample category label according to linear neighborhood propagation, gets basic classifier G_1 and G_2;

(c) Online gets p video data $U_p = \{x_1,...x_p\}$, respectively calculates the text and visual features' vector sets, sets as $u_5^{(1)}$ and $u_p^{(1)}$;

(d) Uses G_1 to predict $u_p^{(1)}$ and gets the result $f^{(1)} = \left[f_1^{(2)}, f_2^{(2)},...f_p^{(2)}\right]$, uses G_2 to predict $u_p^{(2)}$ and gets the label $f^{(2)} = \left[f_1^{(2)}, f_2^{(2)},...f_p^{(2)}\right]$, meanwhile, according to the category label in chapter2.4 inte 明天（6.5 号）中午 12 点半在我办公室开•) of model updating in each view; in $L_R(\cdot)$, makes incremental updating basic classifier according to formula (7).

Category related multiple modal integration

Different modals has different classification abilities, the modal with strong ability of classification has relatively high predicting confidence coefficient to category label. Meanwhile, the same modal has different classification ability to different category data. So, integration weight coefficient should be category related. This text confirms this coefficient based on F_1, F_1=2pr/(p+r), there p and r are respectively refers to the accuracy and recall rate.

Given validation set $V=(V_1,V_2)$, V_1 and V_2 respectively corresponding to text view and visual view. Makes prediction of V_1 and V_2 by basic classifier G_1 and G_2, the result are $F1_1 = \left(F_1^1S, F_2^1S,...F_C^1S\right)$, $F1_2 = \left(F_1^2S, F_2^1S,...F_C^2S\right)$, d_{1ij} $f_1^1 s$ and $f_1^2 s$ are respectively the value of F1 matches by i category classifies by G_1 and G_2. Sets the weight coefficient vector of G_1 and G_2 are respectively $w_1 = \left[w_1^1, w_2^1,...w_c^1\right]$ and $w_2 = \left[w_1^2, w_2^2,...w_c^2\right]$.

This text confirms this weight coefficient by the following formula

$$w_c^1 = \frac{fs_c^1}{fs_c^1 + fs_c^2}; w_c^2 = \frac{fs_c^2}{fs_c^1 + fs_c^2} \tag{9}$$

And makes the modal with high classification accuracy accounts big weight. For test sample x_1, if G_1 makes prediction for $U_p^{(1)}$ and gets the result as $\bar{Y}_i = \left(f_{il}^{(1)}, f_{il}^{(2)},...f_{ic}^{(1)}\right)$, the predicting result of G2 is $\bar{y}_i^{(2)} = \left(f_{il}^{(2)}, f_{il}^{(2)},...f_{ic}^{(2)}\right)$, the predicting result after integration is :

$$\bar{Y}_i = \left(w_c^1 f_{il}^{(1)} + w_1^2 f_{il}^{(2)},..., w_c^i f_{ic}^{(2)} + w_c^2 f_{ic}^{(2)}\right)$$

The belonging category is:

$$\arg\max_{1\le c\le c}\left(w_c^1 f_{ic}^{(1)} + w_c^2 f_{ic}^{(2)}\right)$$

III. EXPERIMENTAL RESULTS

A. Experiment Data Set

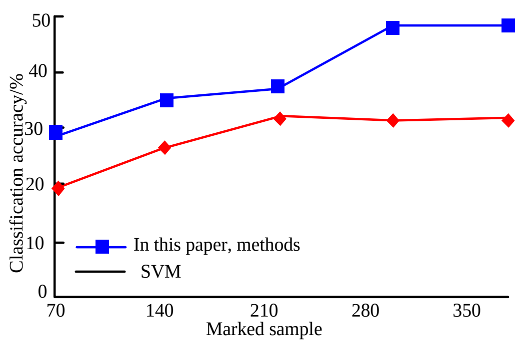
MCG-WEBV is the structured network video dataset based on Youtube, its visual features contains color histogram, color moments, edge histogram descriptor and so no. The text message is expressed by bag of words model. This text extracts 7 categories, 29437 network videos as experiment data. The detailed information shows in table 1.

TABLE1. EXPERIMENT DATASET

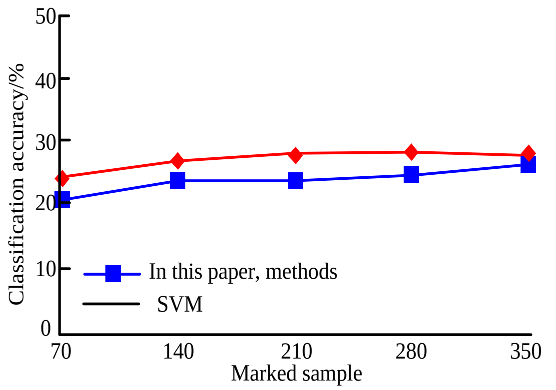
Number	Video category	Sample size	Scale(%)
1	MUSIC	6897	24.09
2	Entertainment	5187	18.45
3	News and Politics	4713	15.91
4	Sports	3980	13.26
5	Pets Animals	3210	11.02
6	Autos Vehicles	3209	10.89
7	Gaming	2348	7.97

B. Classification Results

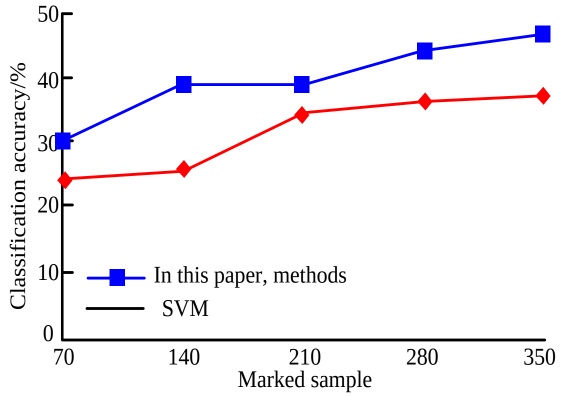
This chapter makes experimental compare of SVM and the method in this text in the environment of less sample. The data divided into labeled sample set and unlabeled sample set, the unlabeled sample set has 2964 data, the scale of labeled sample set are 70, 140, 210, 280 and 350. The two method both training in labeled sample set, predicting in unlabeled sample set. SVM realizes based on Libsvm toolkit, to text data, uses linear kernel, to visual data, uses radial basis function kernel function, extracts parameter to make classification performance the best. According to experiment, local neighborhood $L_R(*)$ is 3.5K. When forms the graph based on formula (1), text view takes included angle cosine as distance measure, visual view takes euclidean distance. The experiment result shows as figure 4. From the result, in the environment of less labeled sample, the text method performance in text view is greatly better than SVM algorithm, in visual view, the two performance are similar. After the two views integration, the method in this text is better than SVM algorithm about 8.3%.



(a) Text view



(b) Visual view



(c) The result after integration

Figure 4. The classification performance's compare result of the method in this text and SVM

C. Online Learning Accuracy

This text mainly analyzes on the classifier online updating effect on classification performance and unlabeled sample extraction method's effect on classification performance.

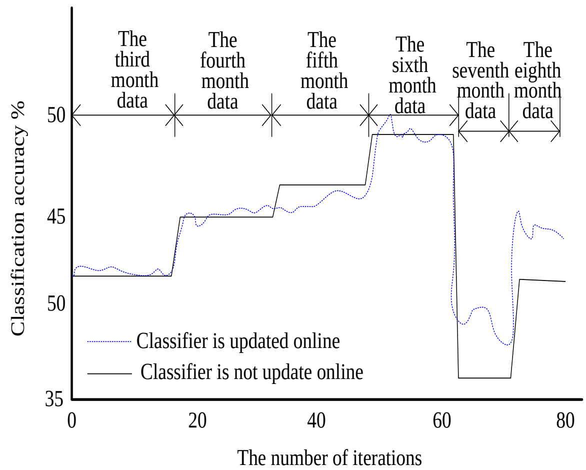


Figure5. The classifier online updating result after multiple model integration

This experiment compares the classifier under the online updating and off-line no updating to online data's prediction performance. Divides experiment data by months into training data and online data. Training data package contains the data in the first and second months, online data package contains all the data in the third to the eighth months, there the online data of each month randomly divides a part using as test set, to test the prediction performance of the classifier to the month (ie: online data) in learning. Online learning 80 times, the experiment result showed as figure 5.

From the experiment result, classifier online updating has little differences with off-line on updating in first several months keeps the same in certain degree, with time goes, the performance differences of the two is obvious in later several months. For the performance change trade, the classifier online updating gradually expresses better classification performance later.

b. Unlabeled sample extraction method's effect on classification performance

In multiple view co-training, how to predict unlabeled sample label in multiple view and to extract unlabeled sample to expand training set has significant effect on semi-supervised learning performance. This experiment analyzes this following three method's performance effect on online learning: when online learning, randomly increase labeled sample into sample set; when online learning, extracts unlabeled sample with high predicting confidence coefficient and its predicting label to expand training set; when online learning, randomly extracts the training set of unlabeled sample. The second method is the online learning method in this text, the first method is equal to supervised learning method, the third method is equal to the classifier totally not predict, randomly extracts unlabeled sample to expand training set.

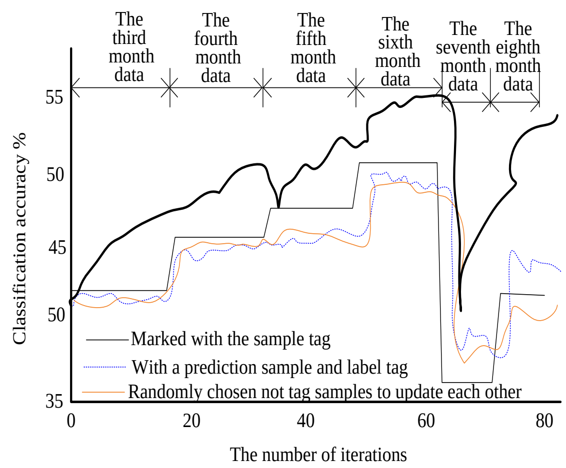


Figure 6. Performance change of the three sample extraction method under multiple modal integration

In the experiment, online learning 80 times, the result is shown as in figure 6. At the beginning of online learning, because the classifier and training data under each method has little differences, so the performance differences is little; with online updating processing, the differences of classifier's performance gradually becomes bigger. From figure 5, the performance of the second method among the first and the third, in m5 and m6, method 2 increases a lot compares with method 3, this shows the method in this text can online increase the performance of classifier. The increase of performance has close relationship with basic classifier: if the basic classifier has high accuracy and good generalization performance, then online semi-supervised learning has good performance improvement.

D. Category Related Multiple Model Experiment

The experiment in this chapter compares the result in similar weight coefficient and category related weight coefficient. The experiment uses 6157 training set, which includes 2894 labeled sample and 3263 unlabeled sample, uses for calculating specific categories' integrate weight coefficient testing set contains 2895 labeled sample, 24842 sample in test set. Calculates in test set and specific categories combine weight coefficient and integrate result shows in table 3. Category unified integrate weight confirms by search on grid (10%, 20%, 60%, 90%), makes the classification performance is the highest after integration, finally makes sure visual and text weight are both 50%.

TABLE 2. COMPARE OF SPECIFIC CATEGORY INTEGRATE WITH UNIFY WEIGHT INTEGRATION PERFORMANCE

Video category	Categories of uniform weight			Category specific weight		
	w_1	w_2	F_1	w_1	w_2	F_1
Music	60	60	35.98	51.32	49.80	37.03
Entertainment	60	60	54.09	54.23	47.34	54.04
News Politics	60	60	41.23	59.08	53.78	53.45
Sports	60	60	64.34	55.43	46.23	64.09
Gaming	60	60	43.05	56.21	44.23	45.23
Autos Vehicles	60	60	65.02	59．45	42.64	65.04
Pets Animals	60	60	46.69	53.04	48.07	47.79
Counterpoise	60	60	49.87	55.02	46.34	50.34

W_1 and W_2 in table 2 are respectively weight of text and visual view. From the experiment result, the result of the two integrated method has large improvement than single view's classification result (seen in the last line in table 3), the performance increase range all higher than 10%. Secondly, specific category weight integrate method has certain improvement than unify weight, from table 2, the former higher 0.8% than the later. Besides, the result shows text modal has better classification ability than visual modal, accounts more weight in integration. In realistic application, specific category weight's integrate method needs the predicting outcomes of single modal to extracts possible several sets of weight, finally extracts the category with the biggest possibility after integrating as final result.

IV. CONCLUSION

This paper proposed multiple view co-training method to solve the problems of multiple modal network video online semi-supervised classification. This model takes view as basic classifier and uses linear neighborhood propagation with decay to predict sample. In this text, the classifier corresponding to visual and text modal makes prediction, extraction, inter-label and expansion training set to online unlabeled sample by co-training method, makes the classifier can learn online and keeps unification with online data distribution. The multiple view co-training has following advantages: (a) As local learning always has better effect than global learning, so the method in this text can get better result; (b) Uses local learning into co-training can simply and efficiently makes incremental learning, pretty suitable for large-scale date's online learning. The experiment result shows, in the environment of less labeled sample, the method in this text is better than wildly-used SVM algorithm, compares it with the environment of classifier off-line no updating, classifier online learning gradually shows the increase on performance. Besides, specific category multiple information integration realizes more efficiently multiple modal information integration.

REFERENCES

[1] N. Omidvar, B. H. Khalaj. A game theoretic approach for power allocation in the downlink of cognitive radio networks. IEEE., 16th CAMAD, 2011:158-162

[2] D. Xu, Z. Y. Feng, Y. Z. Li, et al. Fair Channel allocation and power control for uplink and downlink cognitive radio networks. IEEE., Workshop on mobile computing and emerging communication networks, 2011:591-596

[3] W. Q. Yao, Y. Wang, T. Wang. Joint optimization for downlink resource allocation in cognitive radio cellular networks. IEEE., 8th Annual IEEE consumer communications and networking conference, 2011:664-668

[4] S. H. Tang, M. C. Chen, Y. S. Sun, et al. A spectral efficient and fair user-centric spectrum allocation approach for downlink transmissions.IEEE., Globecom.,2011:1-6

[5] K. Ruttik, K. Koufos, R. Janttir. Model for computing aggregate interference from secondary cellular network in presence of correlated shadow fading.IEEE, 22nd International symposium on personal, indoor and mobile radio communications, 2011: 433-437

[6] J. Naereddine, J. Riihijarvi, P. Mahonen. Transmit power control for secondary use in environments with correlated shadowing. IEEE, ICC2011 Proceedings, 2011:1-6

[7] Muhammad J. Mirza, Nadeem Anjum.Association of Moving Objects Across Visual Sensor Networks.Journal of Multimedia, Vol 7, No 1 (2012), 2-8

[8] Haiping Huang, Hao Chen, Ruchuan Wang, Qian Mao, Re nyuan Cheng.(t, n) Secret Sharing Scheme Based on Cylin der Model in Wireless Sensor Networks.Journal of Networ ks, Vol 7, No 7 (2012), 1009-1016

[9] Xin Huang, Xiao Ma, Bangdao Chen, Andrew Markham, Qinghua Wang, Andrew William Roscoe.Human Interactive Secure ID Management in Body Sensor Networks. Journal of Networks, Vol 7, No 9 (2012), 1400-1406

[10] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects across Visual Sensor Networks. Journal of Multimedia, Vol 7, No 1 (2012), 2-8

[11] Zhao Liangduan, Zhiyong Yuan, Xiangyun Liao, Weixin Si, Jianhui Zhao.3D Tracking and Positioning of Surgical Instruments in Virtual Surgery Simulation. Journal of Multimedia, Vol 6, No 6 (2011), 502-509

[12] Kuo-Feng Huang, Shih-Jung Wu, Real-time-service-based Distributed Scheduling Scheme for IEEE 802.16j

Networks. Journal of Networks, Vol 8, No 3 (2013), 513-517
[13] Kasman Suhairi, Ford Lumban Gaol, The Measurement of Optimization Performance of Managed Service Division with ITIL Framework using Statistical Process Control. Journal of Networks, Vol 8, No 3 (2013), 518-529
[14] Guang Yan, Zhu Yue-Fei, Gu Chun-Xiang, Fei Jin-long, He Xin-Zheng, A Framework for Automated Security Proof and its Application to OAEP. Journal of Networks, Vol 8, No 3 (2013), 552-558
[15] R. Berangi, S. Saleem, M. Faulkner, et al. TDD cognitive radio femtocell network (CRFN) operation in FDD downlink spectrum. IEEE, 22nd International Symposium on Personal, Indoor and Mobile Radio Communications, 2011: 482-486
[16] W. Ahmed, J. Gao, S. Saleem, et al. An access technique for secondary network in downlink channels. IEEE, 22nd International Symposium on Personal, Indoor and Mobile Radio Communications, 2011: 423-427
[17] D. L. Sun, X. N. Zhu, Z. M. Zeng, et al. Downlink power control in cognitive femtocell networks. IEEE., International conference on wireless communications and signal processing, 2011: 1-5
[18] M. J. Brusco, “An exact algorithm for a workforce allocation problem with application to an analysis of cross-training policies,” IIE Transactions, 40(5), pp.495-508, 2008
[19] V. L. Huber, and N. L. Hyer, “The human factor in cellular manufacturing,” Journal of Oper ations Management, 5(2), pp.213-228, 1985
[20] S. Topaloglu, and I. Ozkarahan, “An implicit goal programming model for the tour scheduling problem considering the employee work preferences,” Annals of Operations Research, 128(1-4), pp.135-158, 2004
[21] M. Sabar, B. Montreuil, and J. M. Frayret, “ Competency and preference based personnel scheduling in large assembly lines,” International Journal of Computer Integrated Manufacturing , 21(4), pp.468-479, 2008
[22] M. Sabar, B. Montreuil, and J. M. Frayret, “A multi-agent-based approach for personnel scheduling in assembly centers,” Engineering Applications of Artificial Intelligence, 22(7), pp.1080-1088, 2009

Zhang Nanbin, born in NanChong Sichuan in 1969 male, Han, master of engineering, associate professor, main research direction: computer application technology.

Liu Xiaoping, born in Chongqing Wanzhou in 1971, female, Han, bachelor's degree, senior engineer, main research direction: computer control technology.

Cell Segmentation in Cervical Smear Images using Polar Coordinates GVF Snake with Radiating Edge Map

Xu Chuanyun
College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China
College of Computer Science, Chongqing University, Chongqing, China
xuchuanyun0@qq.com

Zhang Yang
College of Computer and Information Science, Chongqing Normal University, Chongqing, China
495461428@qq.com

Wang Sen
College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China

***Abstract*—In order to segment nucleus and cytoplasm of single cervical cell in the semear images accurately, a novel Gradient Vector Field (GVF) Snake is proposed by improving edge map of GVF Snake (PGVF) to enhance the capacity of the active contour model of gradient vector field to locate the boundary of nuclei and cytoplasmic. In our method, the image is converted into polar coordinates and the Sobel operator is used to calculate horizontal boundary, and then Sand Inhibition Method is designed to inhibite the influence of interference elements on boundary location, and finally cervical cells are segmented with GVF. At the end, experiments performed on the Herlev dataset, which contains 917 images show it is more efficiency of the proposed algorithm than RGVF and nearly same accuracy.**

***Index Terms*—Cervical Cytology Image, Image Segmentation, Active Contour, Polar Coordinates, GVF Snake Model**

I. Introduction

The purpose of cervical cell image segmentation is to divide elements from the cell images to make a basis for the follow-up feature extraction, recognition, diagnosis, so many scholars committed to the research about cervical cell image segmentation problem.

The main features of global segmentation are color, texture, shape, intensity and so on, and the cluster strategies mainly used are supervised clustering, hierarchical clustering & unsupervised clustering [1-5]. In the case that the number of segmentation target cannot be determined, non parametric density estimation can provide relatively good segmentation results [6, 7]; the paper [8] used the segmentation method based on graph theory; the paper [9] made a comparative study on supervised segmentation machine learning methods and unsupervised machine learning methods for the application of images; the paper [10] did a tissue image segmentation using a supervised Bayesian frame.

The commonly used algorithms local cell segmentation contain adaptive threshold segmentation [11], watershed algorithm [12], active contour and level set [13], edge detection [14], and these methods behave more effective in the segmentation of nucleus and cytoplasm of individual cells.

The paper [15] gave us a idea that the cell color depth in cervical smear images have a characteristic that the nucleus is the most dark, the cytoplasmic second as well as the background bright, and put forward Radiation gradient vector field RGVF. By improving this thinking, we brought forward a sort of gradient vector field active contour model (PGVF: Polar coordinates GVF Snake) which solved the problem that equality in radiation edge was mistakenly inhibition, and improved the efficiency of segmentation as well as reduced the effect of noise [16-18].

The basic process includes five steps: 1. to change the image from Descartes coordinate to polar coordinate; 2. to calculate the edge map under the polar coordinate by use of the Sobel operator; 3. to Inhibit influence of interference elements on boundary location by using "Sand Inhibiting" algorithm; 4. to change the edge map from polar coordinate to Descartes coordinate; 5. to finish the image segmentation with GVF Snake model.

II. Polar Coordinate Transformation of Images

The images are usually shown in the Descartes coordinate, but they can be changed from Descartes coordinate to polar coordinate for the needs of the advantageous to handle, and the figure 1 shows this process of image coordinate conversion.

doi:10.4304/jmm.8.3.213-219

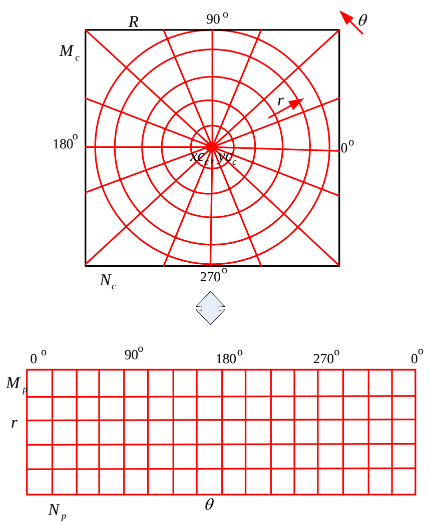


Figure.1 Transformation between Cartesian Coordinates and Polar Coordinates

Seen from Fig. 1, the conversion between Descartes coordinate and polar coordinate is not complete equivalence, in which there are deformation, interpolation and image loss. However during the cell segmentation process, the smaller transform image loss the better, so we give a conversion mode with the smallest Coordinate system transformation loss.

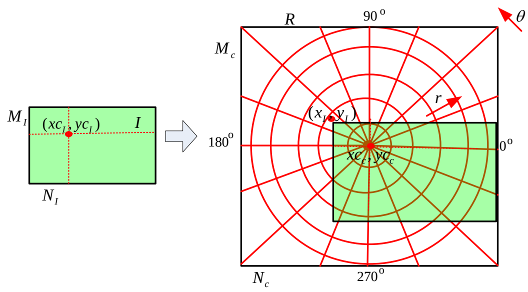


Figure.2 Conversion from rectangular image to square image

Images commonly displayed in Cartesian coordinate system, and also can be transformed from the Cartesian coordinates to polar coordinates according to the processing needs. However the conversion from Cartesian coordinate system to a polar coordinate system is not complete equivalence existing deformation and interpolation. Therefore we put forward a kind of method with minimal loss of images in the cell segmentation process.

Because the original images are mostly rectangular rather than square as well as the origin of polar coordinates (xc_c, yc_c) is not necessarily in the center of original image, we should put the rectangular image into a square image, and then do a polar system transformation, shown as Fig. 2.

In Fig. 2, (xc_I, yc_I) is the origin of original images for the polar coordinate transformation, and the size of the right image is

$$M_c = N_c = \max\{xc_I, yc_I, N_I - xc_I, M_I - yc_I\} \quad (1)$$

The location of the original image in the extended image is

$$x_I = xc_c - xc_I \quad (2)$$

$$y_I = yc_c - yc_I \quad (3)$$

xc_c, yc_c is the center of the extended image.

The extended image needs to intercept off the extension part after the inverse transform.

The coordinate transformation formula from Cartesian coordinates to polar coordinates is

$$\theta = \arctan(\frac{y - yc_c}{x - xc_c}) \quad (4)$$

$$r = \sqrt{(y - yc_c)^2 + (x - xc_c)^2} \quad (5)$$

$$y_p = r \quad (6)$$

$$x_p = \frac{\theta}{2\pi} N_p \quad (7)$$

where r, θ correspond respectively to polar and polar angle of coordinates polar, and x_p, y_p corresponds to the coordinates (x, y) of transformed image.

An example of the image after polar coordinate transformation is given in Fig. 3.

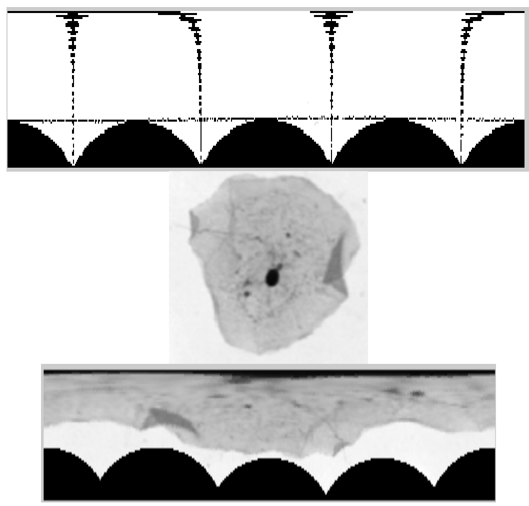


Figure. 3 Example of Image transformed

III. Improved GVF Active Contour Model

Based on the basic radiation gradient vector flow active contour model RGVF-Snake, we give a new edge map (EM) calculation method which replaces the edge map in the GVF model with a new edge map to get a new GVF-Snake model. We call this improved active contour model Polar coordinates-based Gradient Vector Flow Snake (PGVF Snake).

The Snake model is also called parametric active contour model. Snake is a continuous closed contour, and can be parametric represented to be $r(s) = (x(s), y(s))$ where $r(s)$ is the coordinate of contour curve S in the two dimensional plane and s is the normalized length of arc with value range $0 \le s \le 1$. Driven by the energy function, the contour curve S has an ability to attract the image features (usually the target edge), and its energy function can be defined to be

$$E^*_{snake} = \int_0^1 E(r(s))ds = \int_0^1 [E_{\text{int}}(r(s)) + E_{ext}(r(s))]ds \quad (8)$$

where $E(r(s))$ is the energy of the point s on the curve and is made up of the internal energy $E_{\text{int}}(r(s))$ & the external energy $E_{ext}(r(s))$.

The internal energy E_{int} makes the contour contracting as well as smoothing, and the internal energy function is defined as

$$E_{\text{int}}(r(s)) = \frac{1}{2}\left[\alpha |r_s(s)|^2 + \beta |r_{ss}(s)|^2\right] \quad (9)$$

where α & β control respectively the elasticity and smoothness of the contour, and $r_s(s)$ & $r_{ss}(s)$ are respectively the first-order derivative and the two order derivative of the profile curve.

The external energy E_{ext} drives the contour moving to the target boundary, and the external energy in this paper drives contour moving to the edge of the cytoplasm or nucleus. In the basic Snake model, the external energy function is defined as

$$E_{ext}(r(s)) = -|\nabla I(x,y)| \quad (10)$$

where ∇ is the gradient operator.

Under the common action of internal energy & external energy, the contour curve drives the contour moving, and when the contour reaches the target boundary the energy of contour curve E^*_{snake} becomes the minimum.

According to the variational principle, in order to make the minimum energy, the curve $r(s)$ has to meet the following Euler equations

$$\begin{cases} \alpha x_{ss} - \beta x_{ssss} - \dfrac{\partial E_{ext}}{\partial x} = 0 \\ \alpha y_{ss} - \beta y_{ssss} - \dfrac{\partial E_{ext}}{\partial y} = 0 \end{cases} \quad (11)$$

The solution of the Euler equation above is just the contour curve with the minimum energy.

Based on the basic Snake model, we give a gradient vector flow model (GVF-Snake) to use the gradient vector flow field as the external energy in order to increase the function range of the external energy. By this way, even if it is far from the target boundary the contour can be attracted to the target boundary, and also the contour can be attracted to the depression of the boundary.

The gradient vector flow (GVF) field is a vector field

$$\mathbf{v}(x,y) = (u(x,y), v(x,y))$$

and minimize the energy function by

$$E_{GVF} = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 \, dxdy \quad (12)$$

where μ is the important degree to control the first and second elements in the integral equation, and ∇f is the gradient of edge map.

The gradient of the image can be chosen as the edge map $f = |\nabla I|$. The optimal solution of $\mathbf{v}(x,y)$ can be obtained by the following Euler equations

$$\begin{cases} \mu\nabla_u^2 - (u - f_x)(f_x^2 + f_y^2) = 0 \\ \mu\nabla_v^2 - (v - f_x)(f_x^2 + f_y^2) = 0 \end{cases} \quad (13)$$

where ∇^2 is the Laplasse operator.

By using the gradient vector flow as the external force in the Snake model, we can get the dynamic equations of GVF-Snake model

$$\begin{cases} \alpha x_{ss} - \beta x_{ssss} + u = 0 \\ \alpha y_{ss} - \beta y_{ssss} - v = 0 \end{cases} \quad (14)$$

IV. CALCULATING EDGE MAP

EM (Edge Map) is an important factor in the GVF active contour model for it is an important external force to promote the activity profile close to the object boundary. So EM should be help to promote the activity profile close to the object boundary without external interference effect, and we can improve GVF model by optimizing the edge map.

Base on RGVF-Snake, we propose a new calculation method of Edge Map, in which the edge map in the GVF model would be replaced by a new edge map to get a new GVF-Snake model, and we call it PGVF Snake (Polar coordinates-based Gradient Vector Flow Snake). PGVF Snake can improve the calculation efficiency as well as suppress noise.

After being changed to the polar coordinate, the original radiation of the image $l_{xb,yb}$ turns into the line from the upper edge to the lower edge, shown in Fig.4.

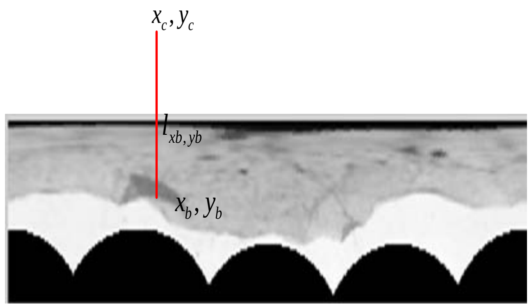


Figure.4. Radiating line on polar coordinates

Because the difference is very sensitive to noise, in order to improve the ability of restraining noise, the radiation difference of image under polar coordinates can be calculated by using the Sobel operator along the y direction with the following calculation formula:

$$RD_p = G_y \otimes I_p \quad (15)$$

where RD_p is the radiation difference of the image in the polar coordinates, I_p is the image in the polar coordinates , $\otimes$ is the convolution operators, G_y is the Sobel convolution Template:

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (16)$$

The radiation gradient RG_p can be got by the following formula:

$$RG_p = \frac{|RD_p| \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}}{2} \quad (17)$$

Seen from the formula (16), G_y only calculates the gradient change in the y direction but don't response to the gradient change in the x direction, which is do the effect to achieve. Because according to statistic analysis, the contour of cytoplasm and nuclear are generally convex polygon, and the nucleus is in the internal convex polygon. After the polar coordinate transformation, the original contour of the convex polygon would be eluded to a horizontal line, and the cell shape would be extracted by using the Sobel operator in the formula (16).

The radiation gradient got in the polar coordinates should be transformed back to the Descartes coordinate system, and Fig.5 shows the effect of radiation gradient transform to Descartes coordinate system.

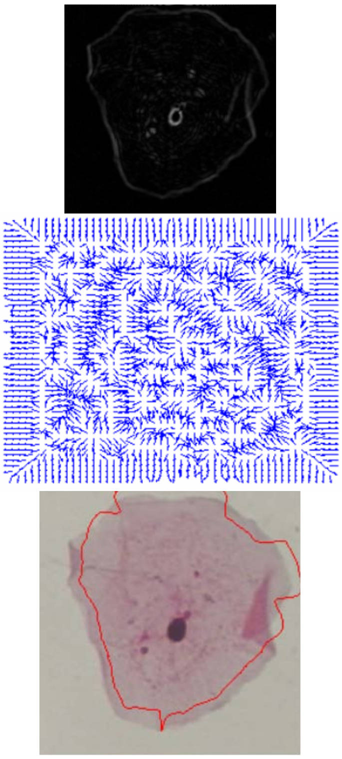
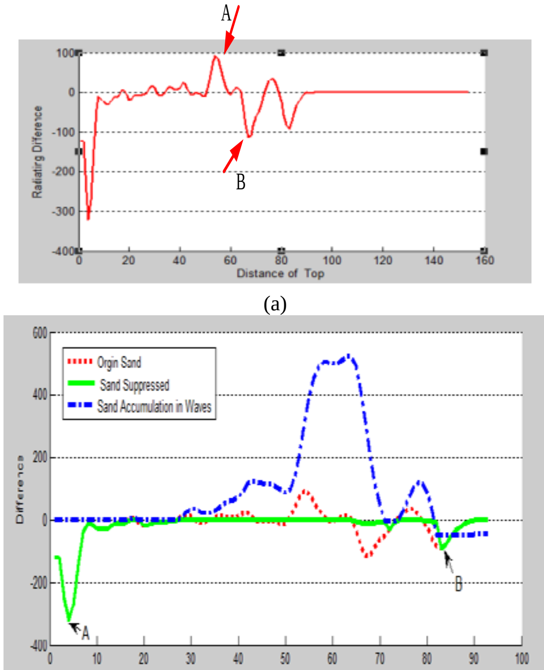


Figure 5. Fake cell contour

Known from the edge map in the figure 5, the cell contour gradient value is larger and the edge gradient in the radiation direction is well controlled.

V. Sand Inhibition Method for Edge Map Optimization

The target of edge graph optimization is to interference internal & external suppression of cells and edge enhancement in the vicinity of the cell contour. The basic starting point of optimization is the fact that the brightness of cytoplasm is darker than the background, which means that there would be a negative gradient at the cell periphery map which is needed to be enhanced in the edge image, shown as Fig. 6(a).



(a)

(b)

Figure 6. Edge map optimization (a) before optimization (b) after optimization

In Fig. 6(a), A and B correspond to a difference of interference. In the starting point of interference, the difference is positive; in the ending point of interference, the difference is negative. But at the edge of the cell, the difference is always negative because the cytoplasm is darker than the background. Based on these, we bring forward a new method named Sand Inhibition Method to eliminate interference, which is inspired by the wave washed over the beach. In this scenario, when the wave washed over the beach, the higher sand will be swept out and the sand would be changed into a flat sandy beach; once encountering a low pit, the sand in waves will be deposited to fill the pit, but the pit will not be filled if the waves have no sand. The inhibition effect is shown in Fig.2(b). Sand Inhibition Method could be described by the following formula.

$$S_y = \sum_{0}^{y} fp(I_r) \quad (18)$$

$$fp(I_r) = \begin{cases} I_r & S_{r-1} > and\ I_r < 0 \\ I_r & I_r > 0 \\ 0 & otherwise \end{cases} \quad (19)$$

$$I_{sy} = \begin{cases} I_y \times Ratio & S_{y\text{-}1} > 0\ and\ I_y < 0 \quad 0 \le Ratio \le 1 \\ I_y \times Ratio & I_y > 0 \quad 0 \le Ratio \le 1 \\ I_y & otherwise \end{cases}$$

(20)

The formula (18-20) describe the content of sand at the point y, where I_r is the height of sand at the point r , I_{sy} is the height of sand after the suppression at the point y, and $Ratio$ is the suppression intensity

The realization algorithm Sand Inhibition Method is given as below.

```
// Algorithm: Sand Inhibition Method
//Input: line - data need inhibition
// radio - suppression intensity
//Output: data after inhibition
    Y = size(line);
// get the content of sand in raw waves
    sy = 0;
    for i = 1:Y
            // the current point
        point = line(i);
            // If the value is greater than 0, the sand more
than 0 contained would be put in the wave.
        if point > 0
            sy = sy + point ;
                    // change the sand into the ground
            line(i) = line(i) * radio;
        elseif point < 0 && sy > 0
            sy = sy + point;
            // fill pit with sand
            line(i) = line (i) * radio ;
        end
    end
    return line;
```

$$E_{PGVF} = \iint \mu\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + \left|\nabla REM\right|^2 \left|\vec{v} - \nabla REM\right|^2 dxdy \qquad (21)$$

We can obtain the segmentation results with the energy formula (21). The edge maps & segmentation results after optimization by Sand Inhibition Method are shown as Fig. 7.

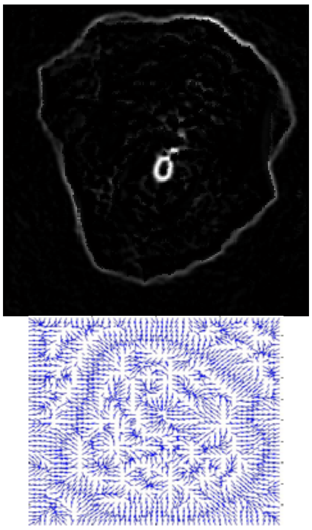


Figure 7. Edge map after optimization and segmentation results of sample

VI. Experimental Verification

We did some experiments by using Herlev cervical cell image data set which has 917 images to verifying the accuracy and efficiency of our method in the cervical cell image segmentation of nucleus & cytoplasm.

The experiments were divided into two classes -cell nuclei segmentation and cytoplasmic segmentation, and the contents contain the accuracy & efficiency of image segmentation. As for the of image segmentation, we make a comparison with manual segmentation, RGVF Snake segmentation and Sobel Snake segmentation about the cytoplasmic segmentation accuracy & cell nucleus segmentation accuracy; and as for the efficiency of image segmentation, we make a comparison with RGVF Snake segmentation and Sobel Snake segmentation on Herlev data set.

We used ZSI (Zijdenbos similarity index) to measure the accuracy, and ZSI was defined as

$$ZSI_{12} = 2\frac{\#\{A_1 \cap A_2\}}{\#\{A_1\} + \#\{A_2\}} \qquad (15)$$

where A_1 and A_2 were the domains needed to be compared, and $A_1 \cap A_2$ is the same domain of two domains.

We used STP (Seconds per Thousand Pixel) which is the time to split the image containing one thousand pixels end to measure the efficiency, and STP was defined as

$$STP = \frac{\sec onds}{\#\{A\}} \times 1000 \qquad (16)$$

where $\sec onds$ is the time for segmentation and $\#\{A\}$ is the pixels for segmentation.

The accuracy comparison of cervical cell nucleus segmentation was shown in Table 1. $\mu_{ZSI_{RH}}$ is the ZSI comparison between RGVF Snake segmentation results with manual segmentation results on Herlev data sets where R is the results of RGVF Snake segmentation, H is the results of manual segmentation, P is the results of PGVF Snake segmentation, S is the results of Snake segmentation using Sobel edge extraction, and so on. $\delta_{ZSI_{RH}}$ is the standard deviation of ZSI .

TABLE I.
ACCURACY COMPARISON NEULEUS SEGMENTATION

Class Name	$\mu_{ZSI_{RH}} \pm \delta_{ZSI_{RH}}$	$\mu_{ZSI_{SH}} \pm \delta_{ZSI_{SH}}$	$\mu_{ZSI_{PH}} \pm \delta_{ZSI_{PH}}$	$\mu_{ZSI_{RP}} \pm \delta_{ZSI_{RP}}$
	RGVF	Sobel GVF	PGVF	RGVF- PGVF
Nucleus	0.855±0.104	0.702±0.133	0.843±0.076	0.981±0.011
Cytoplasm	0.521±0.114	0.470±0.245	0.538±0.246	0.966±0.013

From Table 1, the accuracy of Sobel segmentation is very low, as well as RGVF and PGVF segmentation accuracy is close. The accuracy of cell nucleus segmentation is higher than the cytoplasm, because the boundary of nucleus in Herlev data set is clearer than the cytoplasm which makes it easier to split.

In order to reduce the influence of the quality of Herlev data set on the accuracy of segmentation, we calculated *ZSI* of RGVF and PGVF to get that *ZSI* of cell nucleus is 0.981 and *ZSI* of cytoplasm is 0.966, and found that the segmentation results of RGVF and PGVF would be close to each other.

TABLE II.
EFFICIENCY COMPARISON SEGMENTATION

Class Name	*SPI*		
	PGVF	RGVF	Sobel GVF
Efficiency	0.16608	0.90195	0.18445

From Table 2, STP of PGVF is 0.16608 second, and STP of Snake algorithm based on Sobel is 0.184459 second, and STP of RGVF is 0. 90195 second. The efficiency of PGVF Snake segmentation is slightly higher but not very obvious than Snake algorithm based on Sobel.

Figure 8. Comparison of Average SPI in the Herlev dataset

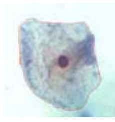
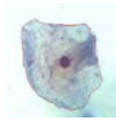
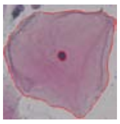
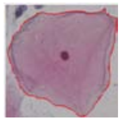
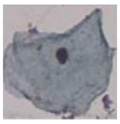
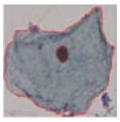
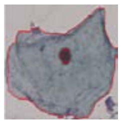
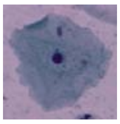
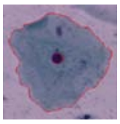
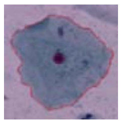
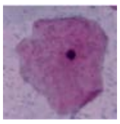
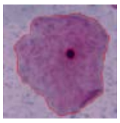
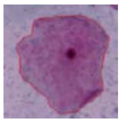
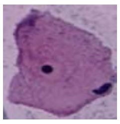
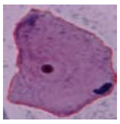
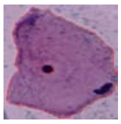
The efficiency of PGVF Snake segmentation can be 5.4 times than RGVF Snake segmentation, this is because that the gradient computation time can be greatly shortened after changing the coordinate system as well as the efficiency of Sand Inhibition Method to eliminate interference is higher than the positive inhibition & negative suppression based on stack.

At the end of the experimental resluts contrast, we give the samples of cervical smear images segmented shown in Table 3.

VII. CONCLUSION

Based on RGVF, we proposed a PGVF algorithm which transformed the image to polar coordinates system

TABLE III. SAMPLES OF CERVICAL SMEAR IMAGES SEGMENTED

No.	Original	RGVF Snake	PGVF Snake
1			
2			
3			
4			
5			
6			
7			

to extract the edge map, and then finished the edge graph optimization by using "sand inhibition method".

The experiments prove that PGVF can make an accurate positioning about the contour of cytoplasmic and nuclear, and has a high efficiency of segmentation. Because PGVF firstly changes the cell image from the Descartes coordinate to the polar coordinate system in order to calculate the radiation difference, and then takes the radiation gradient we got as the edge map of the PGVF model, and finally uses "Sand inhibition algorithm" to optimize the edge map of RGVF for the elimination of the interferencing of the cell's internal impurities on edge map. The specific experimental results show that PGVF is closer to RGVF in the segmentation accuracy, but has an obvious advantage in the segmentation efficiency about 5.4 times improvement.

The next step is the segmentation of the cell group formed by multiple cells. Aimed at the cell groups containing the overlapping and adhering cells, we use the separation point detection method based on the curve curvature extremum to detect the separation point of the overlapping cell, which can accurately position the cell overlap contact point with strong anti jamming ability as well as distinguish the concave region of the Cell overlap from the concave region producing by the cell division to a certain extent; and we use the overlapping cell separation method based on the elliptic curve fitting to separate the overlapping cells, which can well keep the original form of the cell.

ACKNOWLEDGEMENT

This research is supported by the National Natural Science Foundation of China under Grant No. 61202348, the National Natural Science Foundation of China under Grant No. 61173184, the Key Science and Technology Project of Chongqing under Grant No. CSTC2012GG-YYJS10027 and the Natural Science Foundation Project of Chongqing under Grant No. CSTC2012jjA1549.

REFERENCES

[1] Zhang Yang and Ma Yan. Image Semantics Recognition Based on Scene. Journal of Convergence Information Technology. vol. 7, pp. 343-349, 2012.
[2] Xu Chuan_yun, Zhang Yang and Yang Dan. Ontology based Image Semantics Recognition using Description Logics. International Journal of Advancements in Computing Technology. vol. 3, pp. 1-8, 2011.
[3] Md. Atiqur Rahman Ahad, Joo Kooi Tan, Hyoungseop Kim and Seiji Ishikawa. Analysis of Motion Self-Occlusion Problem Due to Motion Overwriting for Human Activity Recognition. Journal of Multimedia. vol. 5, pp. 36-46, 2010.
[4] Thanawin Rakthanmanon, Qiang Zhu and Eamonn J. Keogh. Efficiently Finding Near Duplicate Figures in Archives of Historical Documents. Journal of Multimedia. vol. 7, pp. 109-123, 2012.
[5] Xu Chuanyun, Zhang Yang, Wang Sen and Bin Shu. Image Recognition Method based on Scene Semantics. Journal of Advancements in Computing Technology. vol. 4, pp. 42-49, 2012.
[6] Xu Chuan_yun, Yang Dan and Zhu Lei. Solution to Grouping Problem with Constraints using Improved Genetic Algorithm. Advances in Information Sciences and Service Sciences. vol. 3, pp. 114-123, 2011.
[7] Grant Schindler, Frank Dellaert. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. Journal of Multimedia. vol. 7, pp. 124-131, 2012.
[8] Nicola Montecchio, Emanuele Di Buccio, Nicola Orio. An Efficient Identification Methodology for Improved Access to Music Heritage Collections. Journal of Multimedia. vol. 7, pp. 145-158, 2012.
[9] Bing Li, Scott T. Acton. Active Contour External Force Using Vector Field Convolution for Image Segmentation. IEEE TRANSACTIONS ON IMAGE PROCESSING. vol. 16, pp.2096-2106, 2007.
[10] Mohiuddin Ahmad, Muhammad Muinul Islam, Lutfunessa Lata, Taohida Taslima, Masayoshi Yamada. CLG Optical Flow Approach and Frequency Analysis of Cranial Ultrasonogram Image Sequences. Journal of Multimedia. vol. 5, pp. 47-54, 2010.
[11] P. Soille. Morphological Image Analysis:Principles and Applications, Second Edition. Springer-Verlag, 2008.
[12] Kuan Li, ZhiLu, WenyinLiu, JianpingYin. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. Pattern Recognition. vol. 45, pp. 1255–1264, 2012.
[13] Khalil Zebbiche, Fouad Khelifi, Ahmed Bouridane. Robust Multibit Decoding and Detection of Multiplicative Watermarks for Fingerprint Images. Journal of Multimedia. vol. 5, pp. 167-180, 2010.
[14] A. K. Jain. Data clustering: 50 years beyond k-means. Pattern Recognition Letters. vol. 31, pp. 651-666, 2012.
[15] Muhammad Sheikh Sadi, Md. Nazim Uddin, Abdul Ahad, Ariful Haque. An Efficient Approach to Recognize Fingerprints. Journal of Multimedia. vol. 7, pp. 327-331, 2012
[16] J. Kong, O. Sertel, H. Shimada, K. L. Boyer, J. H. Saltz, andM. N. Gurcan. Computer aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. Pattern Recognition. vol. 42, pp.1080-1092, 2009.
[17] Caroline M. Mendes, Luciano Silva, Olga Regina Pereira Bellon. IMAGO Visualization System: An Interactive Web-Based 3D Visualization System for Cultural Heritage Applications. Journal of Multimedia. vol. 7, pp. 205-210, 2012.
[18] MM Hafizur Rahman, Yukinori Sato, Yasushi Inoguchi. Dynamic Communication Performance Enhancement in Hierarchical Torus Network by Selection Algorithm. Journal of Networks. vol. 7, pp. 468-479, 2012.
[19] Robert Niese, Ayoub Al-Hamadi, Axel Panning, Bernd Michaelis. Emotion Recognition based on 2D-3D Facial Feature Extraction from Color Image Sequences. Journal of Multimedia. vol. 5, pp. 488-500, 2010.
[20] Anttila A, Pokhrel A, Kotaniemi-Talonen L. Cervical cancer patterns with automation-assisted and conventional cytological screening: A randomized study. Int J Cancer. vol. 128, pp. 1204–1212, 2011.

Object Tracking Approach based on Mean Shift Algorithm

Xiaojing Zhang
Department of Software Engineering, School of Software, Harbin University of Science and Technology, Harbin, China
Email: xjz81@163.com

Yajie Yue, and Chenming Sha
School of Software, Harbin University of Science and Technology, Harbin, China
Email:jielu81@yahoo.com.cn, shacm@163.com

Abstract—**Object tracking has always been a hotspot in the field of computer vision, which has a range of applications in real world. The object tracking is a critical task in many vision applications. The main steps in video analysis are: detection of interesting moving objects and tracking of such objects from frame to frame. Most of tracking algorithms use pre-defined methods to process. In this paper, we introduce the Mean shift tracking algorithm, which is a kind of important no parameters estimation method, then we evaluate the tracking performance of Mean shift algorithm on different video sequences. Experimental results show that the Mean shift tracker is effective and robust tracking method.**

Index Terms—**Object tracking, computer vision, Mean shift, no parameters estimation**

I. INTRODUCTION

Visual tracking in a video sequence on the target of interest the effective tracking has always been a typical problem in the field of computer vision. These two issues are reflected in the performance of the tracking algorithm is real-time and robustness. With the continuous development of computer vision, pattern recognition technology and digital video technology, many scholars its in-depth research, a lot of new ideas and methods. Of which, how to take into account real-time and soundness of the system is always the forefront of research. Combined with the current situation, how to further improve the robustness of the system on the basis of real-time tracking algorithm is the main purpose of this paper.

Object tracking has always been a hotspot in the field of computer vision. Its application includes video monitoring, human computer interaction, vehicle tracking and so on [1-5]. Many of these applications which require reliable object tracking techniques should meet in real-time constraints. To track the object in a video sequence should be defined the dynamic entities which constantly change under few of several influence factors. The challenges in a robust tracking algorithm are caused by the presence of noise, occlusion, background clutter, varying viewpoints and illumination changes. Recently, a few of algorithms are proposed to overcome these difficulties [6-9]. In a tracking video sequence, an unknown target should be defined as anything which is interesting to analysis. For example, one people walking on a road, cars in the road, planes in the air, hand or a face in motion and so on. Recently, the form and appearance representations are classified into three families as the representation by points, representation by bounding boxes and the object silhouettes and contour [10-11].

Recently, to select these right features, which play an important role in the tracking object in a video sequences [12-13]. Feature selection is closely similarly to the object and target representation. We always analysis the object motion to focuses on simple characteristics such as color, texture, shape, geometry and so on. Generally, most of tracking algorithms always use a combination of these features to track an object. In general, these features are selected manually by the user, which depend on the many fields and application domains. Yet, the problem of automatic feature selection has brought more and more attention in computer vision and pattern recognition, namely the detection of objects and then to track object in video sequences. As we all known, the object tracking is usual the first step in activities analysis system, including of interactions and relationships between objects of interest. Most of tracking algorithms have been proposed and improved. Object tracking is an estimation question or the trajectory analysis for an object when moving through a sequence of images or video. These usual tracking methods consist of block-matching, KLT, the Kalman filter [14-15], Mean shift [16-17], Camshift [18-19] and so on.

Object tracking algorithm based on color histogram for Mean shift which has obtained a wide range of applications, because of the method is simple and good real-time, it can deal with target deformation and some shelter. In these algorithms, a lot of things used statistical model for target tracking with neighbor pixel domain expression, often use a reference model through the nonlinear estimation for the parameters of the moving target model [20-21]. But in actual problems, we can't get the current image related to probability density distribution of information, and the lack of probability density distribution of prior knowledge, these parameters

doi:10.4304/jmm.8.3.220-225

not easy to determine the limit and estimate methods in the application of target tracking. Therefore, no parameters estimation method is used in the estimation of target best position information, also is in the continuous point probability density value available near the point of observation samples probability estimation. Mean shift algorithm is a kind of important no parameters estimation method, which has been successfully applied in target tracking and related computer vision field [11-12].

The aim of this paper is concerned with object tracking in video streams, object tracking algorithm based on Mean-shift is introduced in this paper. Mean-shift tracker which is based on color histogram and it is a kind of important no parameters estimation method, then we evaluate the tracking performance of Mean shift algorithm on different video sequences. Experimental results show that the Mean shift tracker is effective and robust tracking method.

II. MATERIALS

In the visual tracking, a method based on gradient matching algorithm is adopted to find the image pattern corresponding to the target image in the current frame efficiently. This kind of method is the key to obtain the similarity function in the current frame of probability density distribution once get the probability density distribution, we can accord to its gradient change direction matching search for the best path, in order to improve the efficiency of the pattern matching, and meet the requirement of real-time tracking algorithm. Because we can't obtain probability density distribution information of the current frame and it is lack of the prior knowledge of the overall probability distribution. So we only use non-parameter estimation method to get the gradient of probability density. Non-parameter density estimation is based on the following thought that continuous point's probability density value can be estimated by using points of observation of the area nearby samples set.

A. *Kernel Density Gradient Estimation*

Suppose $x_1, x_2, \cdots x_N$ are N independent distribution of n dimension vector space. The kernel density estimate is proposed as following [25]:

$$\hat{p}_N(x) = (Nh^n)^{-1}\sum_{i=1}^{N} K(\frac{x - x_i}{h}) \tag{1}$$

where $K(x)$ is a scale function, it must meet:

$$\sup_{y \in R^n} |K(y)| < \infty \tag{2}$$

$$\int_{R^n} |K(y)|\, dy < \infty \tag{3}$$

$$\lim_{\|y\| \to \infty} \|y\|^n K(y) = 0 \tag{4}$$

$$\int_{R^n} K(y)dy = 1 \tag{5}$$

The parameter must meet:

$$\lim_{N \to \infty} h(N) = 0 \tag{6}$$

To ensure estimated progressive unbiasedness. Minimum-variance estimation continuity by type guarantee:

$$\lim_{N \to \infty} Nh^n(N) = \infty \tag{7}$$

Probability meaning of uniform continuity by the following equation:

$$\lim_{N \to \infty} Nh^{2n}(N) = \infty \tag{8}$$

We use density gradient estimates as a probability density estimation gradient:

$$\hat{\nabla}_x p_N(x) \equiv (Nh^n)^{-1}\sum_{i=1}^{N} \nabla_x K(\frac{x - x_i}{h}) \tag{9}$$

$$\begin{aligned} \hat{\nabla}_x p_N(x) &\equiv (Nh^n)^{-1}\sum_{i=1}^{N} \nabla_x K(\frac{x - x_i}{h}) \\ &= (Nh^{n+1})^{-1}\sum_{i=1}^{N} \nabla K(\frac{x - x_i}{h}) \end{aligned} \tag{10}$$

where:

$$\nabla K(y) \equiv (\frac{\partial K(y)}{\partial y^1}, \frac{\partial K(y)}{\partial y^2}, \ldots \frac{\partial K(y)}{\partial y^n})^T \tag{11}$$

(1) Progressive bias

If function K(x) satisfies the formula from (2) to (4) and also to satisfy the equation (6), so we can have:

$$\int_{R^n} |g(x)|\, dx < \infty \tag{12}$$

The sequence of functions as followings:

$$g_N(x) \equiv h^{-n}(N)\int_{R^n} K(h^{-1}(N)y) g(x - y)dy \tag{13}$$

There are all points from the function g(x) will be convergence as:

$$\lim_{N \to \infty} g_N(x) = g(x)\int_{R^n} K(y)dy \tag{14}$$

According to the above analysis, the function will be progressive and unbiased estimates meet the criteria below:

$$\lim_{\|x\| \to \infty} p(x) = 0 \tag{15}$$

In the practical application of the probability density function is basically satisfy this condition. Therefore, we can obtain as followings:

$$\lim_{N \to \infty} E\{\hat{\nabla}_x p_N(x)\} = \nabla_x p(x) \tag{16}$$

(2) Continuity

if the second moment of the continuous point $\nabla_x p(x)$ is continuous for the function $\hat{\nabla}_x p_N(x)$ as:

$$\lim_{N \to \infty} E\{\left\|\hat{\hat{\nabla}}_x p_N(x) - \nabla_x p_N(x)\right\|^2\} = 0 \tag{17}$$

To satisfy the following criterions [26]:

$$\lim_{N \to \infty} h(N) = 0 \tag{18}$$

$$\lim_{N \to \infty} Nh^{2n}(N) = \infty \tag{19}$$

$$\sup_{y \in R^n} |K_i'(y)| < \infty \tag{20}$$

$$\int_{R^n} |K_i'(y)|\, dy < \infty \tag{21}$$

$$\lim_{\|y\|\to\infty}\|y\|^n K_i'(y)=0 \tag{22}$$

Where:

$$K_i'(y)\equiv\frac{\partial k(y)}{\partial y^i} \tag{23}$$

(3) Uniform continuity

For $\varepsilon>0$, if :

$$\lim_{N\to\infty}\Pr\left\{\sup_{R^n}\left\|\hat{\nabla}_x p_N(x)-\nabla_x p(x)\right\|>\varepsilon\right\}=0 \tag{24}$$

So the gradient estimation $\hat{\nabla}_x p_N(x)$ is called as uniformly continuous. If the function $\hat{\nabla}_x p_N(x)$ meets the following equations, the gradient estimation will be uniformly continuous.

$$\lim_{N\to\infty}h(N)=0 \tag{25}$$

$$\lim_{N\to\infty}Nh^{2n+2}(N)=\infty \tag{26}$$

$$G(W)=\int_{R^n}\exp(jW^TX)\nabla_x K(x)dX \tag{27}$$

So $\hat{\nabla}_x p(x)$ is uniformly continuous.

B. *Mean Shift Theory*

To give the kernel function $K(x)$ and the kernel function bandwidth h , probability density estimation of the point x is represented as following:

$$\hat{f}(x)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \tag{28}$$

Generally, Epanechnikov kernel is a common kernel function:

$$K_E(x)=\begin{cases}\frac{1}{2}c_d^{-1}(d+2)(1-\|x\|^2), & if\ \ \|x\|<1\\ 0, & if\ \ \|x\|\ge 1\end{cases} \tag{29}$$

where c_d is the sphere volume.Another is often used kernel function is Gaussian kernel:

$$K_N(x)=(2\pi)^{-d/2}\exp\left(-\frac{1}{2}\|x\|^2\right) \tag{30}$$

To define profile function of kernel function K to meet as followings $k:[0,\infty)\to R$. So the corresponding k (x) for Epanechnikov kernel function is described as:

$$k_E(x)=\begin{cases}\frac{1}{2}c_d^{-1}(d+2)(1-x), & if\ \ x<1\\ 0, & if\ \ x\ge 1\end{cases} \tag{31}$$

The equation (3) corresponding k(x) is:

$$k_N(x)=(2\pi)^{-d/2}\cdot\exp\left(-\frac{1}{2}x\right) \tag{32}$$

Fig. 1 shows the profile of Gaussian kernel.

Using the above definition, density estimation formula can rewrite as following:

$$\hat{f}(x)=\frac{1}{nh^d}\sum_{i=1}^{n}k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{33}$$

To define:

$$g(x)=-k'(x) \tag{34}$$

where the first order derivative of k(x) is existed in the interval except few points by hypothesis. Define the kernel function:

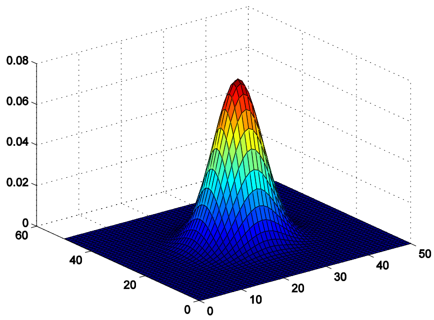


Fig.1 Gaussian profile

$$G(x)=Cg(\|x\|^2) \tag{35}$$

where C is the normalized constant. Then we use th $16\times16\times16$ bins. Fig.5, Fig.6, Fig.7, Fig.8 and Fig.9 show the tracking results on different image sequences by Mean shift tracker. From Fig.5 we can see that the mean shift based tracer proved to be robust tracking results.

From Fig.6we can obtain that the mean shift tracking algorithm will be robust under these conditions with varying viewpoints and illumination changes. In the David image sequence, the illumination is varying from weak to be strong, as well as the head has a great angle rotation. So udder such conditions, the tracker still has a good tracking performance and it is very robust.

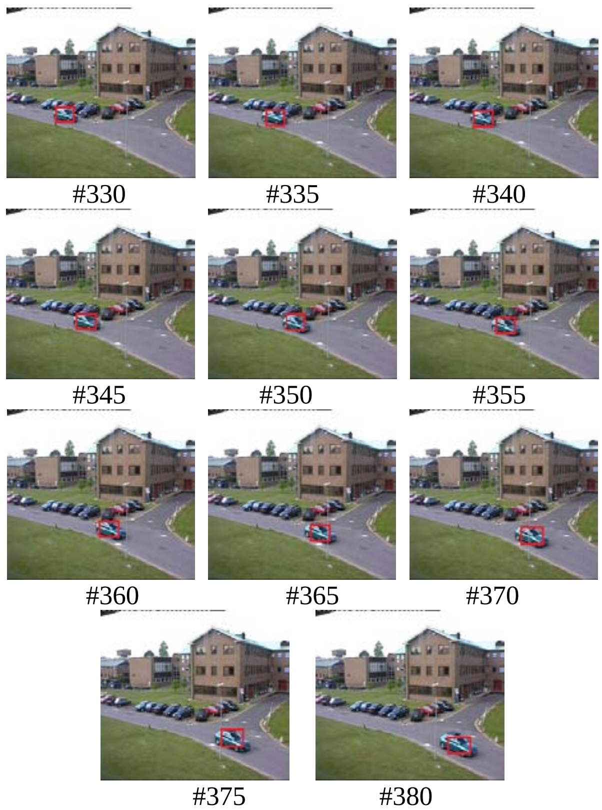


Fig.5 Results of Mean shift tracking

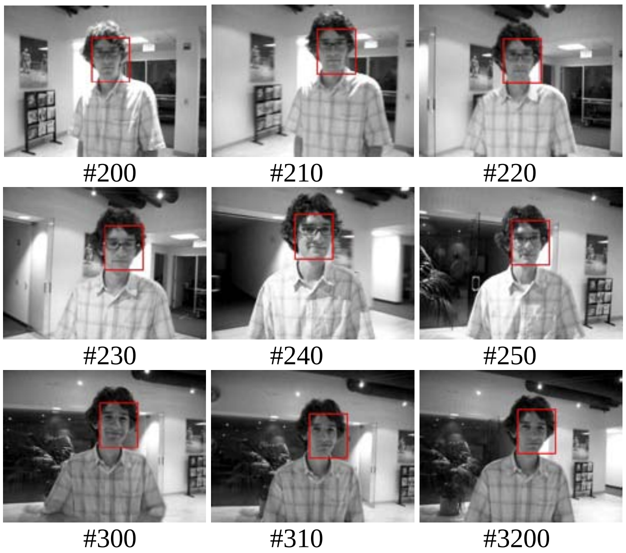


Fig.6 Results of Mean shift tracking on David image sequence

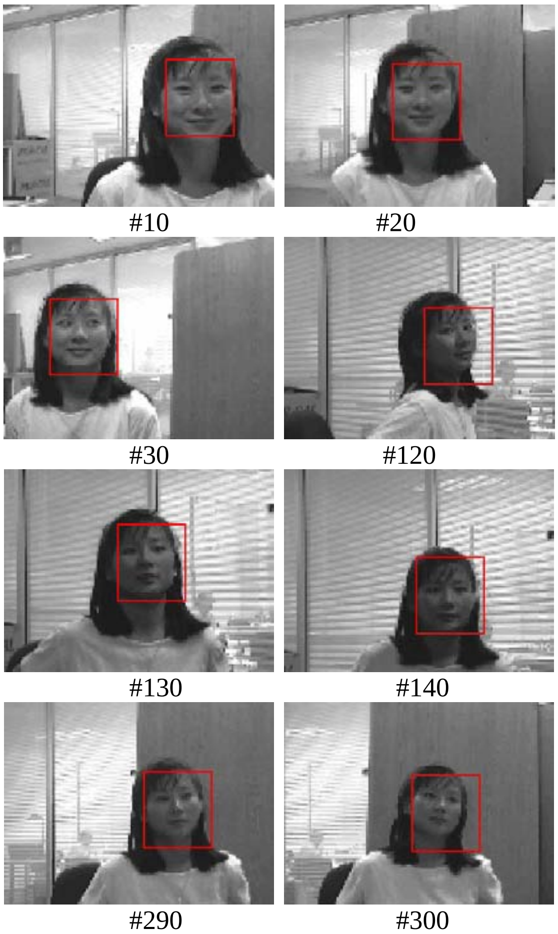
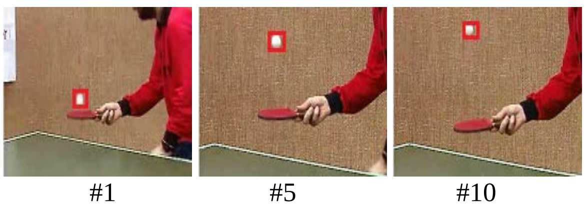


Fig.7 Results of Mean shift tracking on Face image sequence



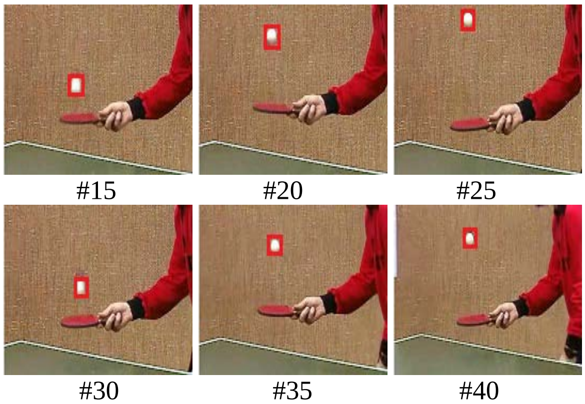


Fig.8 Results of Mean shift tracking on the ball sequence

Fig.9 tracking result of *girl* sequence

To demonstrate the efficiency of the mean shift tracking algorithm, Fig.8 presents the tracking results under the object in a fast motion condition. On the ball sequence, the ball has a greater speed and do back and forth movement up and down. The Mean shift tracking algorithm can obtain a better tracking performance and it can be robust to track the ball with a fast movement.

V. CONCLUSION

Visual tracking is a hot research topic in the field of computer vision, have important applications in the field of intelligent video surveillance, robot navigation, intelligent transportation, and defense security. In order to meet the actual needs, the current visual tracking methods need to deal with a lot of practical problems: how to establish an effective appearance model, while reasonably describe the target can better distinguish background; how to select a fast and efficient tracking inference algorithm to meet the real-time requirements; multi-target tracking correctly associated target. In order to solve these problems, visual tracking involves a number of academic disciplines, including: image processing, pattern recognition, probability theory and mathematical statistics, optimization theory and control theory. Visual tracking is a both theoretical and practical significance of the research theme

The Mean shift tracking algorithm basing on the Mean shift use the color histogram to track object, because of its simpleness perfect real-time performance, the ability to process the target deformation and the block situation and other different situation, and also the accuracy, so the Mean shift video tracking algorithm has been widely used. Experimental results show that the Mean shift tracking algorithm is effective, robust and can be used for tracking in different scenes.

REFERENCES

[1] Xue Mei, Haibin Ling, "Robust visual tracking using L1 minimization", 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1436-1443.

[2] Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift. Computer Vision and Image Understanding, 2009, 113(3): 345-352.

[3] Krainin M, Henry P, Ren X, et al. Manipulator and object tracking for in-hand 3d object modeling. The International Journal of Robotics Research, 2011, 30(11): 1311-1327.

[4] Grabner H, Matas J, Van Gool L, et al. Tracking the invisible: Learning where the object might be//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 1285-1292.

[5] Ning J, Zhang L, Zhang D, et al. Robust object tracking using joint color-texture histogram. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(07): 1245-1263.

[6] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011, 33(8): 1619-1632.

[7] A. Yimaz, O. Javed and M. Shah, "Object tracking: A survey," 2006, Vol. 4, pp. 13-20.

[8] Saravanakumar, S. Vadivel and A. Saneem Ahmed, "Multiple human object tracking using background subtraction and shadow removal techniques", The International Conference on Signal and Image Processing ,2010, pp. 15-17

[9] Meng G, Jiang G, "Real-time illumination robust maneuvering target tracking based on color invariance", Proceedings of the 2nd International Congress on Image and Signal, 2009, pp. 1-5

[10] Gammeter S, Gassmann A, Bossard L, et al. Server-side object recognition and client-side object tracking for mobile augmented reality//Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010: 1-8.

[11] N. A Gmez. "A Probabilistic Integrated Object Recognition and Tracking Framework for Video Sequences", Universitat Rovira I Virgili, PHD thesis, Espagne, 2009.

[12] Collins,R. T, Yanxi Liu and Leordeanu, M, "Online selection of discriminative tracking features", IEEE Transactions on Pattern Analysis and Machine Intelligence , 2010, Vol. 10, pp. 1631-1643.

[13] Ying-Jia Yeh, Chiou-Ting Hsu, "Online Selection of Tracking Features Using AdaBoost", IEEE Transactions on Circuits and Systems for Video Technology, 2009, Vol. 3, pp. 442-446.

[14] X. Sun, H. Yao and S. Zhang, "A Refined Particle Filter Method for Contour Tracking", Visual Communications and Image Processing, 2010, Vol. 8, pp. 1-8.

[15] D. Comaniciu, V. Ramesh and P. Meer, Kernel-Based Object Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, Vol. 14, pp. 564-577

[16] D. Chen, F. Bai, B. Li and T. Wang, "A Real-time Tracking Method for Scout Robot", IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009, pp. 1001-1006.

[17] L. Ido, L. Michael and R. Ehud, "Mean Shift tracking with multiple reference color histograms", Computer Vision and Image Understanding, 2010, pp. 400-408

[18] Wang Z, Yang X, Xu Y, et al. Camshift guided particle filter for visual tracking. Pattern Recognition Letters, 2009, 30(4): 407-413.

[19] Yin M, Zhang J, Sun H, et al. Multi-cue-based CamShift guided particle filter tracking. Expert Systems with Applications, 2011, 38(5): 6313-6318.

[20] Ess A, Schindler K, Leibe B, et al. Object detection and tracking for autonomous navigation in dynamic environments. The International Journal of Robotics Research, 2010, 29(14): 1707-1725.

[21] Saleemi I, Hartung L, Shah M. Scene understanding by statistical modeling of motion patterns//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 2069-2076.

[22] Chu D M, Smeulders A W M. Color invariant surf in discriminative object tracking//Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2012: 62-75.

[23] Wu Z, Thangali A, Sclaroff S, et al. Coupling detection and data association for multiple object tracking//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 1948-1955.

[24] Zulkifley M A, Moran B. Robust hierarchical multiple hypothesis tracker for multiple-object tracking. Expert Systems with Applications, 2012.

[25] Cacoullos T., Estimation of a multivariate density, Ann. Inst. Statist. Math., Vol.18, pp. 179-189.

[26] Fukunaga K., Introduction to statistical pattern recognition, Second Ed., Academic Press, New York, 1990.

Xiaojing Zhang received the B.S. degree in Communication Engineering from Northeast Dianli University in 2002. And she received the M.S. degree in Software Engineering from Yunnan University in 2005. Her primary research area focused on

Embedded System. Currently she is a Lecturer of the School of Software, Harbin University of Science and Technology, Harbin, China.

Yajie Yue received the B.S. degree in automation from Harbin University of Science and Technology in 2004. And she received the M.S. degree in Control Theory and Control Engineeringfrom Harbin University of Science and Technology in 2007. She has been working in Harbin University of Science and Technology from 2007. Currently her primary research area focused on IC design and Computer Architecture.

Chenming Sha was born in Heilongjiang Province, China, in 1981. She received the B.S. degree in Computer Science and Technology from Beijing Normal University, China in 2004, and M.S. degree in Computer application technology from Beijing University of Technology, China in 2007. Now she is working at Harbin University of Science and Technology, Harbin, China as a Lecturer.

Improved Signal Processing Algorithm Based on Wavelet Transform

Xiang Li and Haiyan Yao
Departments of Electronic Information and Electrical Engineering, Anyang Institute of Technology, Anyang, China
Email: whz101033@163.com, lix3972@163.com

***Abstract*—Wavelet analysis is a rapidly developing emerging discipline, at present; it has been widely used in practice. Study wavelet's new theory, new method and new application have important theoretical significance and practical value. Based on the problem of poor performance of using traditional wavelet transformation algorithm in signal denoising, this paper puts forward a improved scheme based on tradeoff between soft and hard threshold, the scheme is established on the basis of traditional soft threshold, hard threshold, and the obtained estimated wavelet coefficients value of this scheme are between soft threshold and hard threshold methods, so call it the tradeoff between soft and hard threshold method. Implementation steps of the improved scheme based on soft and hard threshold tradeoff are, firstly establishing wavelet coefficient estimator of soft and hard threshold tradeoff method, and adding factor in threshold estimator, so as to adjust the size of the estimated wavelet coefficients. Simulation experiments show that the proposed soft, hard threshold tradeoff based signal threshold improved deniosing method shows a strong effect of the practical signal denoising, the obtained reconstructed signal SNR has been tremendously improved than the traditional denoising method.**

***Index Terms*— improved wavelet transform; Signal denoising; Signal processing; Threshold denoising**

I. INTRODUCTION

As a new theory, wavelet analysis has indeed caused a great disturbance in science and technology. To mathematicians, wavelet analysis is a new branch of mathematics, it is the perfect crystal of functional analysis, Fourier analysis, spline analysis, harmonic analysis, and numerical analysis; in application fields, especially in signal processing, image processing, voice analysis and nonlinear science field, it is considered another efficient time-frequency analysis method after Fourier analysis. In principle, where there can be used Fourier analysis traditionally where we can take the place of wavelet analysis [1]. Wavelet analysis has good localization characteristics both in time domain and frequency domain, which overcome the disadvantages of traditional Fourier analysis [2], and because it uses the gradually fine time domain step to high frequency, thus it can focus any details of the analyzed signal, so the wavelet analysis has the laudatory title of "mathematical microscope" [3].

In recent years, wavelet theory has been further developed, people construct wavelet has many excellent properties at the same time, such as multiple function wavelet [4], M wavelet [5], etc., also to loosen the orthogonal wavelet conditions from another point of view, to study more general non orthogonal vectors, such as filter group, orthogonal wavelet translation, etc., making the wavelet theory more perfect. With the constant improvement of wavelet theory, its application field is becoming more and more widely.

Nowadays, the application field of wavelet analysis is very wide, it includes: many subjects in the field of mathematics; Signal analysis and image processing; Quantum mechanics, theoretical physics; Military electronic countermeasure and intelligent weapons; Computer classification and recognition; Artificial synthesis of music and language; Medical imaging and diagnostic; Seismic data processing; Large machinery fault diagnosis, etc; In mathematics, for example, it has been used in numerical analysis, structural fast numerical method, curve surface structure, differential equation solving, cybernetics, etc. In the field of signal analysis, it has been used in filtering, noise, compression, transmission, etc. In the field of image processing, it has been used in image compression, classification, identification and diagnosis, decontamination, etc. And in the field of medical imaging, it has been used in B ultrasonic reduction, computed tomography (CT), imaging time of nuclear magnetic resonance (NMR), improving resolution, etc.

(1) The application in signal and image compression is an important aspect of wavelet analysis. It is characterized by high compression ratio, fast compression speed, remaining invariance of the signal and image features, and anti-jamming during the delivering. There are many compression method based on wavelet analysis, in which the successful ones include wavelet packet with best base method, texture model method in wavelet domain, the wavelet zero tree compression, vector compression of wavelet transform, etc.

(2) The application of wavelet in signal analysis is also very extensive. It can be used in border processing and filtering, time-frequency analysis, signal-noise separation and extraction of weak signals, fractal index, the signal recognition and diagnosis and multi-scale edge detection, etc.

doi:10.4304/jmm.8.3.226-232

(3) The application in engineering technology includes computer vision, computer graphics, curve design, turbulence, remote cosmic research and biomedical research.

In recent years, wavelet theory has been further developed, people construct wavelet has many excellent properties at the same time, in 1994, Xu [6] proposed a spatial correlation based noise removal method, filtering based on the wavelet coefficients of signal and noise between neighboring scales correlation, although this method is not precise enough, but it's very direct and easy to implement. In the process of implementation of the algorithm, the estimation of noise energy is critical. Pan [7] elicited the theoretical calculation formula of noise energy threshold and gave an efficient method of estimating the signal noise variance, making spatial correlation filtering algorithm is adaptive. In 2000, Chang [8] combined the adaptive threshold with translation invariant denoising ideas, put forward image based spatial adaptive wavelet threshold denoising method, the selected threshold can make adaptive change according to the statistical characteristics of image itself. Oktem [9] proposed a transform domain method of Film – grain [10] noise removes and noise image compression.

In a word, in recent years, there are a lot of literatures about wavelet de-noising, some of them have achieved many good results, based on the analysis of these results, this paper made a lot of improvements, proposed some new algorithms, and compared with the original method to illustrate the advantage of the improved algorithm.

II. Wavelet Transform Algorithm

A. Overview of Wavelet Transform Algorithm

Apparently, time signal is only a rule that a physical quantity size changes with time, it can only directly reflect pieces of information of this physical quantity, a lot of important information contained within it [16].

Signal processing is a fabrication of a certain signal [18]. For digital signal processing, the software is especially important, which is closely related to the mathematical method of signal analysis. Frequency domain method of digital signal processing is using discrete Fourier transform (DFT) to convert the discrete time signal sequence into the frequency domain, getting the signal spectrum. So it often called spectral analysis. Now spectrum signal processing has become one of the most basic and most sophisticated signals processing method. The reason include: first it’s because in1965, J.W.C ooley and J.W.T ukey put forward the fast Fourier transform (FFT) algorithm, which reduces the Fourier transform time by several orders of magnitude, the second is because the rapid improvement of computer performance.

However, Fourier transform has the following obvious flaws:

(1) It is only suitable for the stable signal analysis, and useless for non-stationary signal.

(2) In order to get frequency domain feature of a time domain signal, it needs to use all the information of signal in time domain, even the future information.

(3) If signal only changes in a small neighborhood at a particular moment, then the entire frequency spectrum of the signal will be affected, and for the spectrum changes, it is unable to calibrate the time position and acuteness degree from fundamentally, that is to say, the Fourier transform is not sensitive to signal singularity.

(4) In engineering application, the measured signal often contains both high frequency information, and low frequency information, which requires for high frequency information, the time interval should be relatively small, for the purpose of giving accurate high frequency information; For low frequency information, the time interval should be relatively wide, for the purpose of giving a complete information of one cycle. Fourier transforms is useless on this point.

In order to overcome the shortage of Fourier exchange algorithm, it appears a kind of wavelet exchange algorithm.

As a analysis tool who can automatically adjust along with the change of frequency, Wavelet transform has been greatly developed since mid 1980s, and applied in areas of signal processing, computer vision, image processing, speech analysis and synthesis.

The emergence of wavelet analysis method can be traced back to 1910 years; Haar put forward the Haar standard orthogonal basis. And in 1938, Littlewood – Paley established the L-P theory to Fu cover series. In order to overcome the shortage of the traditional Fourier analysis, in the early eighty s, scientists have used the concept of "wavelet" for data processing, in which the famous is in 1984 when a French geophysicists Morlet adopted wavelet concept to store and express the seismic signal of petroleum exploration. Exploration in mathematics is mainly the “atom" and "elements" theory founded by R.C oifman and G.W eiss, these "atom" and "molecular" constitutes the part of the base in different function space. L. C arleron used alike "wavelet” functions to construct the unconditional basis in space H^1 of Stein and Weiss. Until 1986, the French mathematician Meyer successfully constructed a smooth function ψ with certain attenuation, the binary flexibility and translation $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2t^{-j} - k)\}$ constructed the standard orthogonal basis of $L^2(R)$.

Lemarie and Battle also put forward respectively the wavelet function with exponential decay after Meyer. In 1987, Mallat uses the concept of multi-resolution analysis, unifies the various specific wavelet constructions, and puts forward the current widely used Mallat fast wavelet decomposition and reconstruction algorithm. In 1988, Daubechies constructed the orthogonal wavelet basis with compactly support. Coifman, Meyer et al., in 1989 introduced the concept of wavelet packet. Spline function based single orthogonal small basis was put forward by Cui Jin taihe and Wang Jianzhong in 1990. In 1992, A.C ohen, I. D aubechhies et al. constructed biorthogonal wavelet base with compactly support. The same period, the relation between wavelet transforms and filter group also got further research. The basic theory of wavelet analysis was basically established.

In recent years, a simple effective wavelet constructing method - Lifting Scheme get a great attention and development. Using lifting scheme can decompose all the existing compactly supported wavelet into basic steps, in addition, it also provides a powerful means for nonlinear wavelet construction, and therefore, using the lifting scheme to construct wavelet is considered to be the second generation wavelet. Wavelet theory and its application is still in development, and it will get more deepen research in multi-scale method, the nonlinear wavelet structure on the non-stationary set and, non-uniform, time-varying signal processing etc in the future.

B. Constant Wavelet Transform

$\forall f(t) \in L^2(R)$, $f(t)$'s constant wavelet transform(sometime it is called integral wavelet transform), defined it as:

$$WT_f(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt, a \neq 0 \quad (1)$$

Or use the inner product form:

$$WT_f(a,b) = \left\langle f, \psi_{a,b} \right\rangle \quad (2)$$

In the formula:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad (3)$$

If making the inverse transformation exist, $\psi(t)$ should satisfy the admissibility condition:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (4)$$

In the formula, $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$.

Here, the inverse transformation is

$$f(t) = C_\psi^{-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_{a,b}(t) WT_f(a,b) db \frac{da}{|a|^2} \quad (5)$$

C_ψ , this constant limits the class of function ψ belongs to $L^2(R)$ which can serve as "basic wavelet (or mother wavelet) , especially require ψ as a window function, then ψ should belongs to $L^1(R)$, that is

$$\int_{-\infty}^{\infty} |\psi(t)| dt < \infty \quad (6)$$

So $\hat{\psi}(\omega)$ is a continuous function in R. we can get from formula (4) that $\hat{\psi}$ must be zero at origin, that is

$$\hat{\psi}(0) = \int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (7)$$

It can be found from formula (6) that the wavelet function has oscillatory.

The transform of constant wavelet has the following characters:

Property 1 (linear): assuming $f(t) = \alpha g(t) + \beta h(t)$, then

$$WT_f(a,b) = \alpha WT_g(a,b) + \beta WT_h(a,b) \quad (8)$$

Property 2 (translation invariance): if $f(t) \Leftrightarrow WT_f(a,b)$, then $f(t-\tau) \Leftrightarrow WT_f(a, b-\tau)$.

Translation invariance is a very good nature, in practice, although discrete wavelet transform used more widely, but in the case of needing translation invariance, discrete wavelet transform cannot be used directly.

Property 3 (flexible degeneration): if $f(t) \Leftrightarrow WT_f(a,b)$,then

$f(ct) \Leftrightarrow \frac{1}{\sqrt{c}} WT_f(ca, cb)$, in which $c > 0$.

Nature 4 (redundancy): it exist redundancy of information expression in the continuous wavelet transform. Its performance is unique of the reconstruction formula of continuous wavelet transform to restore the original signal, the wavelet kernel function $\psi_{a,b}(t)$ exist many possible choices. Despite the existence of redundancy can improve the signal stability of the calculation, but increase the difficulties of analysis and interpretation of wavelet transform results.

C. Discretization of Continuous Wavelet Transform

Because it exist redundancy in continuous wavelet transform, so it is necessary to clarify that for reconstructing signal, what kind of discretization should be taken for the variance a , b to eliminate the redundancy in transform, in practice, usually taking $b = \frac{k}{2^j}$, $a = \frac{1}{2^j}$; $j, k \in Z$, here

$$\psi_{a,b}(t) = \psi_{\frac{1}{2^j},\frac{k}{2^j}}(t) = 2^{j/2}\psi(2^j t - k) \quad (9)$$

Nano : $\psi_{j,k}(t)$.

Transformation form: $WT_f\left(\frac{1}{2^j}, \frac{k}{2^j}\right) = \left\langle f, \psi_{j,k} \right\rangle$

In order to reconstructing signal $f(t)$, it requires $\{\psi_{j,k}\}_{j,k \in z}$ is the Riesz basis of $L^2(R)$.

Definition 1: a function $\psi \in L^2(R)$ called a function R , if $\{\psi_{j,k}\}_{j,k\in z}$ is a Riesz basis in the following sense: $\{\psi_{j,k}\}_{j,k\in z}$'s linear span is dense in $L^2(R)$, and it exist positive constant A and B , $0 < A \le B < \infty$, making

$$A \| \{c_{j,k}\} \|_{l^2}^2 \le \| \sum_{j\in-\infty}^{\infty} \sum_{k\in-\infty}^{\infty} c_{j,k}\psi_{j,k} \|_2^2 \le B \| \{c_{j,k}\} \|_{l^2}^2 \quad (10)$$

For all the double infinite square sum sequences, $\{c_{j,k}\}$ established, namely $\| \{c_{j,k}\} \|_{l^2}^2 = \sum_{j\in-\infty}^{\infty} \sum_{k\in-\infty}^{\infty} |c_{j,k}|^2 < \infty$'s $\{c_{j,k}\}$ established.

Assuming ψ is a function R, then it exist a unique Riesz basis $\{\psi^{j,k}\}_{j,k\in z}$ of $L^2(R)$, the sense is

$$\left\langle \psi_{j,k}, \psi^{l,m} \right\rangle = \delta_{j,l}\delta_{k,m}, j,k,l,m \in Z \quad (11)$$

The above is antithesis with $\{\psi_{j,k}\}$, here each $f(t) \in L^2(R)$ has the only series expression like (7):

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \left\langle f, \psi_{j,k} \right\rangle \psi^{j,k}(t) \tag{12}$$

Particularly, if $\{\psi_{j,k}\}_{j,k \in z}$ constructs the orthonormal basis of $L^2(R)$, there has $\psi_{j,k} = \psi^{j,k}$, reconstruction formula is :

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \left\langle f, \psi_{j,k} \right\rangle \psi_{j,k}(t) \tag{13}$$

III. Signals Denoising based Improved Wavelet Transform Algorithm

A. Estimate the Soft and Hard Threshold Method

Assuming there is the following observation signal

$$f(t) = s(t) + n(t) \tag{14}$$

where $s(t)$ as the original signal, $n(t)$ is the whiter Gaussian noise of variance σ^2, obey $N(0, \sigma^2)$. It is difficult to extract all the signal $s(t)$ directly from the observed signals $f(t)$, we must use other transformation methods as a tool. In recent years, using theory of wavelet transform for signal denoising provides powerful tools and ways to overcome the limitation of using traditional method to process non-stationary signal.

For 1 dimensional signal $f(t)$, firstly we carry out the discrete sampling, we got N point discrete signa $f(n), n = 0,1,..., N-1$, the wavelet transform is

$$WT_f(j,k) = 2^{-j/2} \sum_{n=0}^{N-1} f(n) \psi(2^{-j} n - k) \tag{15}$$

$WT_f(j,k)$ is the wavelet coefficient, in practice use, it is complex by using formula (15) to calculate, and $\psi(t)$ does not display the expression in general, so we should depend on double scale equation, so as to get the recursion implement method of wavelet transform

$$SF(j+1,k) = SF(j,k) * h(j,k) \tag{16}$$

$$WT_f(j+1,k) = SF(j,k) * g(j,k) \tag{17}$$

In which h and g is respectively the low and high-pass filter corresponding to scaling function $\varphi(x)$ and wavelet function $\psi(x)$, $SF(0,k)$ is for the original signal $f(k)$, $SF(j,k)$ is the scale coefficient, $WT_f(j,k)$ is the wavelet coefficient. Accordingly, Wavelet Transform reconstruction formula is

$$SF(j-1,k) = SF(j,k) * \tilde{h}(j,k) WT_f(j,k) * \tilde{g}(j,k) \tag{18}$$

For convenience, Wavelet Transform coefficient $WT_f(j,k)$ recorded as $w_{j,k}$. After taking the discrete wavelet transform for observation signal $f(k) = s(k) + n(k)$, it can be seen by wavelet transform's linear nature, the wavelet coefficient $w_{j,k}$ got by decomposition still composed by two parts, one part is the wavelet coefficient $Ss(j,k)$ corresponding to $s(k)$, marked as $u_{j,k}$, another part is $n(k)$'s corresponding wavelet coefficient $Sn(j,k)$ marked as $v_{j,k}$.

The basic thought of wavelet threshold de-noising is

(1) firstly carry out wavelet transform to noising signal $f(k)$, got a group of wavelet coefficient $w_{j,k}$;

(2) through carrying out the threshold process on $w_{j,k}$, got the estimated wavelet coefficient $\hat{w}_{j,k}$, making $\| \hat{w}_{j,k} - u_{j,k} \|$ possibly small;

(3) using $\hat{w}_{j,k}$ to carry out the wavelet reconstruction, got the estimated signal $\hat{f}(k)$, namely the information after de-noising.

This paper mainly discusses how to estimate wavelet coefficient.

After taking out several times of wavelet decomposing for $f(k)$, due to the space asymmetry distribute signal $s(k)$'s corresponding wavelet coefficient $w_{j,k}$ at varies scale have relatively the bigger value at some certain position, these points corresponding to the change position and important information of original signal $s(k)$, and the other position, $w_{j,k}$'s value is small, for white noise $n(k)$, its corresponding wavelet coefficient $w_{j,k}$ distributed uniformity at each scale, and along with the increase of the scale, $w_{j,k}$ coefficient's size value decreases . therefore, the common de-noising method is to find a proper number λ as the threshold(limit), setting the wavelet coefficient $w_{j,k}$ lower than λ (mainly cased by noise $n(k)$)as zero, while for the $w_{j,k}$ higher than λ (mainly caused by $s(k)$), then reserve it or contract it, getting the estimated wavelet coefficient $\hat{w}_{j,k}$, it can be seen that it is caused basically by signal $s(k)$, then carry out reconstruction of $\hat{w}_{j,k}$, then we can reconstruct the original signal. The wavelet coefficient estimating method is :

Taking $\lambda = \sigma \sqrt{2 \log(N)}$, define

$$\hat{w}_{j,k} \begin{cases} w_{j,k}, | w_{j,k} | \geq \lambda \\ 0, \quad | w_{j,k} | < \lambda \end{cases} \tag{19}$$

It is called the hard threshold estimating method ; the soft threshold estimating defined as

$$\hat{w}_{j,k} \begin{cases} sign(w_{j,k})(| w_{j,k} | - \lambda), | w_{j,k} | \geq \lambda \\ 0, \quad\quad\quad\quad\quad\quad\quad | w_{j,k} | < \lambda \end{cases} \tag{20}$$

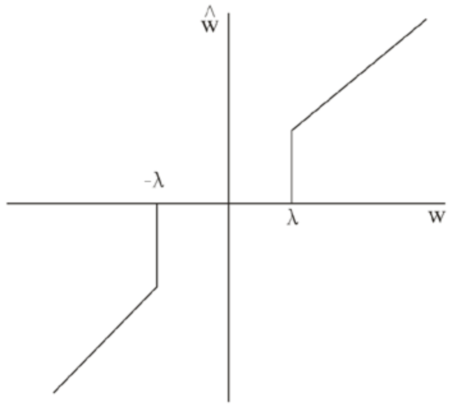


Figure 1. Hard threshold

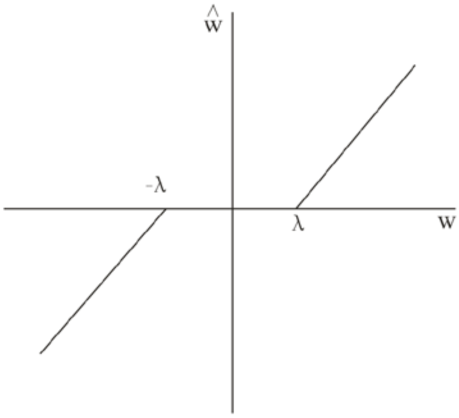


Figure 2. Soft threshold

Figure 1 is the hard threshold method, and Figure 2 is the soft threshold method. Although these two methods are widely applied in the practice, which also gets a good result, but this method has some potential defects. Such as in hard threshold method, $\hat{w}_{j,k}$ is not continuous at λ, it may cause some oscillation by using $\hat{w}_{j,k}$ to reconstruct obtained signal; although the $\hat{w}_{j,k}$ estimated by soft threshold method has a good continuity, but when $|w_{j,k}| \geq \lambda$, $\hat{w}_{j,k}$ and $w_{j,k}$ always in the constant deviation, which directly affects the approach degree of reconstructed signal and real signal.

B. Improvement scheme

In view of the existing defects of wavelet estimating, this paper designs a soft and hard compromise scheme, which overcomes the defects of soft and hard threshold.

Define 2;

$$\hat{w}_{j,k} \begin{cases} sign(w_{j,k})(|w_{j,k}| - \alpha\lambda), & |w_{j,k}| \geq \lambda \\ 0, & |w_{j,k}| < \lambda \end{cases}, (0 \leq \alpha \leq 1) \quad (21)$$

We call the above formula as the soft and hard threshold compromise wavelet coefficient estimator. Particularly, when α takes 0and 1, the above formula becomes the estimated method of hard and soft threshold. For general $0 < \alpha < 1$, this data $\hat{w}_{j,k}$ estimated by this method is between soft and hard method in the size, so we call it the soft and hard threshold compromise method. The model is shown in Figure 3.

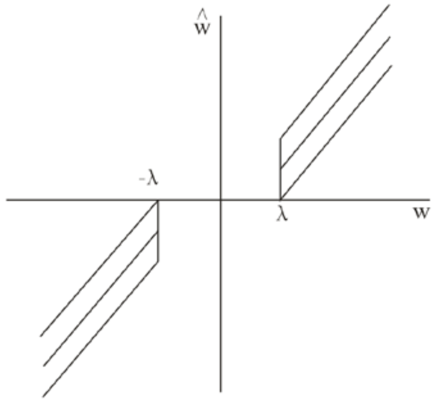


Figure 3. Soft and hard threshold compromise law

The idea of this method is quite simple, also easy to understand, but it is good in de-nosing.it can be fund by careful analysis, the $\hat{w}_{j,k}$ estimated by simple soft threshold method , its absolute value will always small in λ than $w_{j,k}$, so it needs to reduce the deviation: but if reduce this deviation into zero(to hard threshold) is not possible the best, because $w_{j,k}$ is composed by $u_{j,k}$ and $v_{j,k}$, it may makes $|w_{j,k}| > |u_{j,k}|$ due to the effect of $v_{j,k}$ (for most $w_{j,k}$), and our purpose id to make $\| \hat{w}_{j,k} - u_{j,k} \|$ minimum, thus order $|\hat{w}_{j,k}|$'s value between $|w_{j,k}| - \lambda$ and $|w_{j,k}|$, then the estimated wavelet coefficient $\hat{w}_{j,k}$ become more close to $u_{j,k}$. Based on this thought, we add factor α in the threshold estimator, adjust properly the size of α between 0and 1, we can get a better de-noising effect. In the experiment, taking $\alpha = 0.5$.

IV. Simulation Result

We use the proposed improved method to carry out the de-nosing test for a period of noise signals. For comparison, the article also compared with traditional wavelet transform algorithm. When carrying out Wavelet decomposition, the biggest decomposition scale j for 3, taking different threshold λ on every scale, namely $\lambda_j = \sigma\sqrt{2\log(N)} / \log(j+1)$, in which σ^2 is for the variance of noise, N is for the length of the discrete sampling signals, j is for the decomposition scale.

This article uses the proposed improved method to carry out de-noise for a noise signal of SNR is 8.226270, then compare the results, Tab 1 is the recovered SNR after varies de-nosing methods and compared with corresponding RMSE, it can be seen from table 1that the soft and hard threshold compromise method is obviously better than simple soft threshold methods in de-nosing, so as to verify the effectiveness of this improved method.

TABLE I.
SNR AND RMSE EXPERIMENT

Estimator	Soft thresholding	Hard threshold method	Soft and hard threshold compromise law
SNR	15.27.	14.33	15.58
RMSE	0.172	0.192	0.166

Because the selection of λ is quite random, and it may affect the performance of varies methods, therefore, the paper tried another group of threshold λ_j, and we got the result as Table II.

TABLE II. SNR AND RMSE EXPERIMENTAL II

Estimator	Soft thresholding	Hard threshold method	Soft and hard threshold compromise law
SNR	15.27.	15.73	16.31
RMSE	0.172	0.163	0.152

It can be seen from table 2, the soft and hard threshold compromise method shows a better de-noising effect than pure soft threshold denoising method.

It should be pointed out that the selection λ_j is not the best, if properly selected λ_j, the method can better reflect its superiority. In addition, for different λ_j, the comparison results of above method may have some differences, but according to the author's large number of trials, generally speaking, pure soft threshold and hard threshold methods' stability is poorer, the dependence of λ_j is strong, and for a given λ_j, at least one effect is not very good in soft threshold and hard threshold method; for soft and hard threshold method, no matter how to choose λ_j, it is always better than a simple soft threshold and hard threshold method.

In addition, in order to illustrate the superiority of the soft and hard threshold compromise method, this article also takes a de-nosing test for a noise Heavisine signal, figure 5 for the primitive Heavisine signal, figure 6 is for the original signal fixed with gaussian white noise signal, figure 7 is the obtained reconstructed signal after taking the proposed soft and hard threshold method to de-noising.

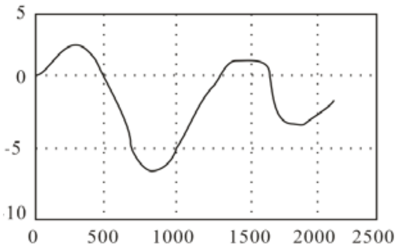


Figure 4. Original Heavisine signal

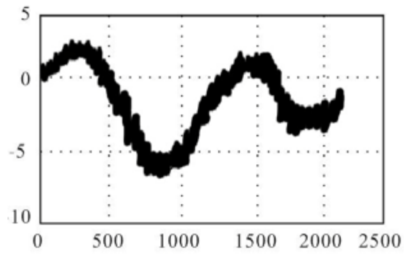


Figure 5. Containing noise Heavisine signal

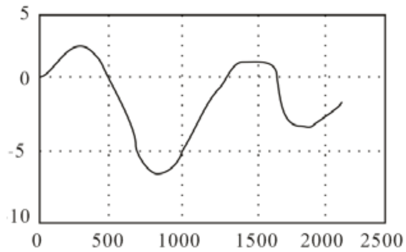


Figure 6. Reconstructed signal obtained after de-noising

Set another example, for a noise information whose SNR is 6.8098 in Matlab:

$$x = 30\sin(t) + 25\sin(2t) + rand(n)$$

Respectively using soft threshold, hard threshold, soft threshold tradeoff method, and the proposed improved methods to carry out the simulation experiment. The Wavelet base used in the simulation is the db3 wavelet and decomposition layer is for 3, the results as shown in table III.

TABLE III. DENOISING PERFORMANCES OF THE FOUR METHODS

Threshold value denoising method	SNR	Mean square error
Soft threshold method	6.9721	0.82825
Hard threshold method	7.6533	0.42006
Soft and hard threshold compromise method a=0.05	7.6640	0.41555
Soft and hard threshold compromise method a=0.10	7.6684	0.41365
Soft and hard threshold compromise method a=0.15	7.6665	0.41438
Improvement method a=2000,b=13	7.7321	0.38824
Improvement method a=4000,b=13	7.7336	0.38765
Improvement method a=6000,b=13	7.7341	0.38744
Improvement method a=8500,b=13	7.7345	0.38732

It can be seen from figure 7, the improved threshold de-noising method proposed in this paper shows a good effect on the actual signal denoising, the obtained reconstructed signal SNR has a significantly improvement than the traditional denoising method.

VI. CONCLUSIONS

In recent years, with the continuous development of wavelet theory, there produced some new wavelets with composite performance index, they can have some excellent properties which traditional wavelet is unlikely to have, researching the structure, characteristics and implementation of these new type of wavelet algorithm and analyzes the relationship between them with the orthogonal basis is a very meaningful work. Based on application state quo of traditional wavelet nose ring algorithm in signal processing, improves the estimation method of wavelet coefficient, puts forward a kind of soft and hard compromise wavelet coefficient estimation, so as to improve the performance of signal denoising. The experimental results show that the proposed improved threshold de-noising method shows a good effect to the actual signal denoising, the obtained reconstructed signal SNR has a significantly improvement than the traditional denoising method.

REFERENCES

[1] Pan Quan, Zhang Lei, Meng Jin-li, Zhang Hong-cai. Wavelet filtering method and application [M]. *Bei Jing: Tsinghua University Press*, 2005.9, pp. 8-9 .

[2] Taleb-Ahmed, A.Hamad, D.Tilmant. Vehicle license plate recognition in marketing application Intelligent Vehicles Symposium. *IEEE*, 9-11 June 2003 90-94.

[3] H. M. Qi, Z. Q. Chen, L. M. Zhang, “Towards end-to-end delay bounds on WMNs based on statistical network calculus,” *in Proceedings of the 9th International Conference for Young Computer Scientists*, pp. 493-497, Nov. 2008.

[4] Nicola Calabretta, Hyun-Do Jung, Harm Dorren. All-optical Label Swapping Techniques for Optical Packets at Bit-rate Beyond 160 Gb/s. *Journal of Networks*. Vol 5, No 11 (2010):1343-1349.

[5] Yanamura Y, Goto M.,Nishiyama D, Soga M, Nakatani H, Saji H. Extraction and tracking of the license plate using Hough transform and voted block matching. *IEEE IV2003 Intelligent Vehicles Symposium*. 2003: 243-246.

[6] Zhou Wei. Wavelet analysis of senior technical with MATLAB [M]. *Xi an: Xidian University Press*, 2005.12, pp.1.

[7] K. Xiong, Z. D. Qiu, H. K. Zhang, C. S. Li, H. C. Chao, “A scalable fast forwarding approach for IP networks,” *International Journal of Internet Protocol Technology*, vol. 3, no. 2, 2008, pp. 119-127.

[8] Y. M. Jiang, “Delay bound and packet scale rate guarantee for some expedited forwarding networks,” *Computer Networks*, vol. 50, no. 1, 2006, pp. 15-28.

[9] Y. Jia, J. Wang, C. Zhang, X. -S. Hua. Finding image exemplars using fast sparse anity propagation. *In: Proc. ACM SIGMM International Conference on Multimedia,* 2008, pp. 639 – 642.

[10] L. M. Zhang, D. Z. Huang, Z. G. Chen, “Models of bounds on end-to-end delay of long-range dependence traffic based on fractal leak bucket,” *Journal on Communications*, vol. 29, no. 7, 2008, pp. 32-43.

[11] J. G. Ma, M. X. Xi, Y. M. Lin, Y. P. Li, “Chinese Internet routerlevel hop count measurement and analysis,” *Application Research on Computer*, vol. 25, no. 27, 2008, pp. 2112-2114

[12] Ke Xiong, Yu Zhang, Zheng-ding Qiu, Dong-mei Sun, Ya-juan Qin. QoS-enhanced TNPOSS Network Model and its E2E Delay Bound for Multimedia Flows. *Journal of Networks*. Vol 4, No 8 (2009):795-802.

[13] 3GPP TS 36.211 V9.1.0, “Evolved universal terrestrial radio access (E-UTRA); physical channel and modulation," *www.3gpp.org*, Mar. 2010.

[14] Young-il Min, Hwan-min Park, Jun-hee Jang, Keun-dae Kim, Hyung-jin Choi. The Enhanced Decoding Method for QO-SFBC System in Frequency Selective Fading Channel Environment. *Journal of Networks*. Vol 6, No 3 (2011):378-386.

[15] P.Seddighian et. al., ‘All-optical swapping of spectral amplitude code labels for packet switching,’ Proc. Photonics in Switching 2007, *San Francisco, USA*, pp. 143-144, 2007.

[16] Y. Liu et al.,’Error-free 320 Gb/s all-optical wavelength conversion using a semiconductor optical amplifier,’ *Journal of Lightwave Techn*. vol. 25, pp. 103-108, 2007.

[17] H.W. Wang, D.W. Lin, and T.H. Sang, “OFDM signal detection in doubly selective channels with whitening of residual inter-carrier interference and noise,” *IEEE VTC*, pp. 1-5, May, 2010.

[18] H. M. Park, Y. I. Min, J. S. Han, and H. J. Choi, “Enhanced detection method for 3GPP LTE QO-SFBC system in frequency selective Rayleigh fading channel environment,” *ICUIMC*, pp. 558–562, Feb. 2010.

[19] J. Kim, S. L. Ariyavisitakul, and N. Seshadri, “STBC/SFBC for 4 transmit antennas with 1-bit feedback,” *IEEE ICC*, pp. 3943–3947, May 2008.

[20] H. Y. Kong, and H. V. Khuong, “Performance analysis of space-time block coded cooperative wireless transmission in Rayleigh fading channels,” *KICS JCN*, vol. 8, pp. 306–312, Sept. 2006.

[21] Haihui GAO, Jun MA, Haichun YANG, Chaoran CUI Construction of Tag Cloud for Query Result Summarization and Navigation. 2013 Vol. 9 (6) : 2509-2516.

[22] I. S. Dhillon, Co-clustering documents and words using bipartite spectral graph partitioning, *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 2001, pages: 269 -274.

Xiang Li received the B.S degree in school of Automation and Electrical Engineering of Tianjin Tianjin University of Technology and Education, China, in 2001 and achieved the M.S. degree in School of Electrical and Information Engineering at Jiangsu University, China, in 2008.He is currently Lecturer in College of Electronic Information and Electric Engineering at the Anyang Institute of Technology, China. His research interests are mainly in Signal Processing and cognitive radio.

Haiyan Yao received the B.S. degree in School of Information Engineering at Zhengzhou University, China ,in 2004 and achieved the M.S. degree in School if Electronic and Optical Engineering at Nanjing University if Science and Technology, China, in 2012. She is currently Lecturer in College of Electronic Information and Electric Engineering at the Anyang Institute of Technology, China. Her research interests are mainly in signal processing and wireless communication.

Speech Enhancement Algorithm Based on Combining Local Characteristic-Scale Decomposition and Difference Spectrum of Singular Values

Nan Li, Shu Deng and Quanbin Tang
College of Information Engineering, Northeast Dianli University, Jilin, P. R. China
whz101111@163.com

***Abstract*—De-noising of speech signal polluted by background noise, which is of great practical significance for effective transmission and accurate recognition of sound, is important. Local characteristic-scale decomposition is introduced, and it can divide the speech into low- and high-frequency parts without a loss in useful speech. The algorithm accelerates the speed of convergence, as well as pretreatment, and has less illusive component. In addition, two different criteria are proposed to select the reasonable noise reduction order based on the difference spectra of singular values. The proposed approach has improved the problem of noise reduction order selected by experience in singular value decomposition and enhanced de-noising effects. Simulation experiments verify the validity of the algorithm from subjective and objective evaluations.**

***Index Terms*—Local Characteristic-Scale Decomposition, Singular Value Decomposition, Difference Spectrum, Speech Enhancement**

I. INTRODUCTION

In the real communication environment, speech signals are inevitably subject to various interferences from transmission medium, communication equipment, and other speakers. The received speech influenced by interference becomes a noise-contaminated signal. Sometimes, in an extremely noisy harsh environment, speech information is covered, resulting in raw information that can hardly be recovered. Consequently, de-noising of noisy speech signals is important.

Speech enhancement algorithm is currently categorized into several transform domain methods, including time and frequency domains. The time domain methods, such as those based on parameters and model [1-3] and the subspace method [4, 5], can accurately estimate a speech model from noisy speech and obtain enhanced speech. However, estimating an accurate speech model with low signal-to-noise ratio (SNR) is difficult and requires a highly complex algorithm. The frequency domain methods, such as spectral subtraction [6] and minimum mean-squared error estimation [7], can estimate the original speech from noisy speech by using the strong relativity of speech short-time spectrum and weak relativity of noise. However, it mostly trends to produce musical noise so that perception results become poor. Therefore, researchers gradually concentrate on other transform domain methods. At present, the time-frequency domain method is the most frequently studied process.

The time-frequency domain methods in speech enhancement are mainly wavelet method [8] and empirical mode decomposition (EMD) [9-13]. However, a wavelet method lacks adaptability; a suitable wavelet basis and threshold are thus selected before using the wavelet method. Although EMD can overcome this drawback, this method must combine with the threshold at de-noising. Then, EMD produces a mass of illusive components and the problem of marginal effect when signals are decomposed by EMD. The local characteristic-scale decomposition (LCD) [14] is presented by Cheng Jun-sheng et al. to remedy these problems. This algorithm improves marginal effect and exhibits less illusive components in addition to fast calculation speed. The present study utilizes LCD to decompose speech signals and proposes two different criteria to reduce noise from noisy speech signals by analyzing the difference spectrum of singular values. This algorithm can remedy the selection of threshold and provide results with improved SNR and increased intelligibility.

In order to solve the problem, that it is difficult to select thresholds and unsatisfactory de-noising effect, a new algorithm is presented combining local characteristic scale decomposition algorithm and singular value decomposition (SVD). The singular value decomposition (SVD) algorithm is an efficient nonlinear filtering method, and it can decompose the signal into a series of singular value and its vector corresponding to time-frequency subspace. In the application of singular value decomposition, selection of effective singular value number is a difficulty. Selection criteria of effective singular values are proposed through analyzing the LCD characteristics of speech signal and difference spectrum characteristics of singular value. The speech signal is decomposed using LCD algorithm as different frequency signals, and the difference spectrum of singular value of these decomposition are analyzed. According to the ratio

doi:10.4304/jmm.8.3.233-240

of noise and useful speech signals after local characteristic scale decomposition, the effective singular values are determined by different criteria. When frequency signals contained mostly noise, the effective singular values numbers is determined by maximum criterion of difference spectrum and when frequency signals contained mainly speech signals, the effective singular values numbers is determined by terminal for zero criterion of difference spectra. This algorithm can remedy the selection of threshold and provide results with improved SNR and increased intelligibility.

II. Basic Theory

A. Local Characteristic-scale Decomposition

Based on the analysis of the frequency domain of a speech signal, the frequency of voice sound primarily focuses on the low band, whereas that of voiceless sound distributes in the high band. The noise signals mainly concentrate in the high frequency, and a small amount of noise information concentrates in the low band of speech. Hence, the low- and high-frequency components require different methods to reduce noise to achieve better sound effects.

Local characteristic-scale decomposition depends on the signal itself to divide signals. This process can obtain n-order intrinsic scale component (ISC) from high frequency to low frequency. An ISC must satisfy the following conditions. First, in the whole signal, the maximum value is positive, minimum value is negative, and any two adjacent maxima and minima must be monotonous. Second, a straight-line function can be obtained by any two adjacent maxima and minima. This condition can also determine the function value in the time between the two extreme points, so the ratio of function value and this extreme must be a constant.

LCD is performed as follows [15]:

1) All the extreme X_k of the whole signal and all corresponding time τ_k , $k=1,2,\cdots M$, where M is the number of extreme, are identified;

2) Straight-line function is determined by two adjacent maxima (minima) (τ_k, X_k) and (τ_{k+2}, X_{k+2}). Following the straight-line function, the function value A_{k+1} is obtained in the time τ_{k+1} , $k=1,3,\cdots M$;

3) Based on the formula $L_k = aA_k + (1-a)X_k$ ($k=2,3,\cdots M-1$), L_k is calculated, and the edges extreme (τ_0, X_0) and (τ_{M+1}, X_{M+1}) can be obtained by mirror symmetry continuation method proposed by Grilling. Then, A_1 and A_M are calculated to finally obtain L_1 and L_M ;

4) $L_1, L_2, \cdots L_M$ are connected with cubic spline lines as the mean value function m_1;

5) The first signal component can be calculated as $y(t) - m_1 = c_1$. Ideally, if c_1 satisfies the two conditions of ISCs, then it should be an ISC, as h_1. If c_1 does not satisfy the two conditions of ISCs, then c_1 is treated as the new data. This sifting procedure is repeated until the difference can satisfy the two conditions of ISCs, so that it can be treated as the first ISC component.

6) h_1 is separated from the rest of the signal by $y(t) - h_1 = r_1$. r_1 is treated as a new signal $y(t)$ and subjected to the same sifting process as described above, resulting in the second ISC h_2 , the third ISC $h_3, \ldots$.

The termination standard of this algorithm is the standard deviation (SD) calculated as $SD = \sum_{t=0}^{T}\left[\frac{\left|h_{ik}(t) - h_{i(k-1)}(t)\right|^2}{h_{i(k-1)}^2(t)}\right]$, where T is the length of time. To ensure an ideal ISC, SD should be less than 0.5.

Using LCD, noisy speech signals can be decomposed as

$$y(t) = \sum_{i=1}^{n} h_i + r_n \tag{1}$$

where h_i is the i^{th}-order ISC, which represents the different frequency band signal components in noisy speech. The residue r_n is a monotonic function, which reflects the average tendency of original signals.

B. Difference Spectra of Singular Values

Singular value decomposition (SVD) is a traditional de-noising method. First, a Hankel matrix is created from a noisy speech signal. Second, singular values are obtained by SVD, and the significant singular values to reduce noise are determined. Finally, enhanced speech is reconstructed from the useful singular values.

For a real matrix $A \in R^{m\times n}$, its SVD is defined as follows [16]:

$$A = UDV^T \tag{2}$$

where U and V are the orthogonal matrices $U \in R^{m\times m}$ and $V \in R^{m\times m}$, D is the diagonal matrix. $D = [diag(\sigma_1, \sigma_2, \cdots \sigma_q),0]$ $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_q > 0, \sigma_i (i=1,2,\cdots q)$ are called the singular values of matrix A. q is the rank of matrix A, usually $q \le \min(m,n)$.

The properties of the singular values are given in Ref. [17]

$$\sigma_t(A+B) \le \sigma_t(A) + \sigma_t(B), 1 \le t \le q \tag{3}$$

Supposing that $y(t)$ is the noisy speech, $s(t)$ is the pure speech, $n(t)$ is the noise signal, then A_y , A_s , and A_n are the Hankel matrices created from $y(t)$, $s(t)$, and $n(t)$.

The signal is decomposed by SVD, and the following inequality is obtained:

$$\sigma(A_y) \le \sigma(A_s) + \sigma(A_n) \tag{4}$$

$$\sigma(A_y) \le (\sigma_{s1} + \sigma_n, \cdots \sigma_{sk} + \sigma_n, \sigma_n, \sigma_n \cdots \sigma_n) \quad (5)$$

Based on expression (5), to extract the original speech signal $s(t)$, the key problem is to determine the sudden change point k in the singular values of the noisy speech signal. Then, the enhanced speech is reconstructed by the front k singular values. To describe its sudden change status, the concept of difference spectra is introduced. The forward difference of singular values is defined as follows: $b_i = \sigma_i - \sigma_{i-1}, i = 1,2,\cdots,q-1$, then the sequence $B = (b_1, b_2, \cdots, b_{q-1})$ is called the difference spectrum of singular value, which reasonably describes the change status between adjacent singular values.

The simulation signal is $f(t) = \sin(2t) + \sin(6t)$, sampling 512 points in the interval [0, 2π], and the added noise follows Gaussian distribution $N(0,1)$. The waveform comparison charts of the input signal with noise and original signal are shown as Figure 1.

Difference spectrum sequence of the signal with noise shows in Figure 2(a) is analyzed. The maximum of sudden change point appeared in the four order, so pure signal is restored by the first four orders. The final de-noising signal is shown in Figure 2(b), which is restored by difference spectrum maximum.

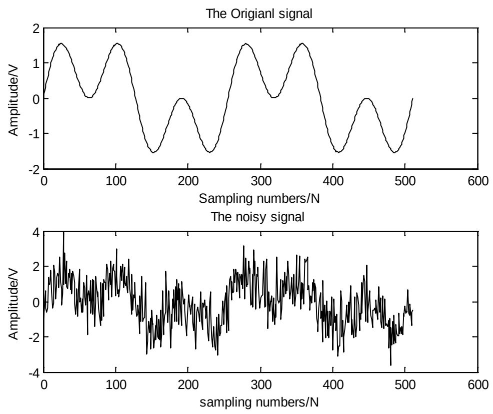


Figure 1. The original signal and noise signal time domain waveform

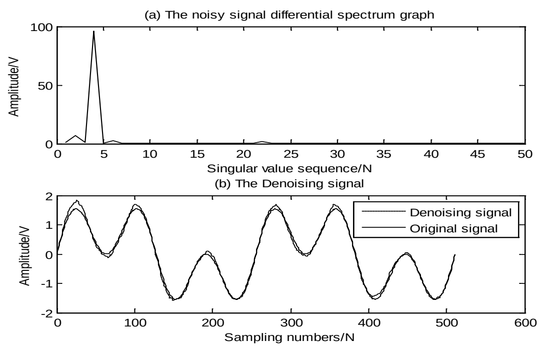


Figure 2. The noisy signal differential spectra and signal, and Time-domain waveform after singular value de-noising

C. LCD Characteristics of the Speech Signal with Noise

The noisy speech signal is decomposed by local characteristic-scale decomposition, and then it can obtain a lot of ISCs. Analyzing with the correlation coefficient of each ISC, the noise signal and the clean speech signal, it can determine the relationship between noise and speech signals in the ISCs.

The correlation coefficient is defined as

$$\rho_{XY} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \quad (6)$$

ISC values of voice “one” and “ten” with noise are showed in Table I and ISC values of pure signal are showed in Table II (The top eight ISCs are listed in tables) .Known from the Table I and Table II, the top two ISCs contain the large amount of noise components, because the similarity of ISC1, ISC2 and noise signal is the largest. However the similarity of component signal behind ISC3 and noise signal is diminishing, the similarity of component signal behind ISC3 and clean speech signal is increasing. The similarity is usually greater than 0.5.

TABLE I.
THE CORRELATION COEFFICIENT OF ISCS AND NOISE SIGNALS

SNR=0 dB	ISC1	ISC2	ISC3	ISC4	ISC5	ISC6	ISC7	ISC8
One	0.67	0.40	0.28	0.21	0.13	0.04	0.04	0.02
Two	0.67	0.40	0.31	0.13	0.09	0.08	0.04	0.05
Three	0.67	0.42	0.30	0.17	0.09	0.04	0.05	0.05
Four	0.67	0.43	0.27	0.13	0.07	0.07	0.05	0.05
Five	0.67	0.36	0.22	0.17	0.13	0.07	0.04	0.05
Six	0.69	0.42	0.28	0.15	0.09	0.02	0.04	0.04
Seven	0.67	0.43	0.26	0.17	0.09	0.05	0.04	0.05
Eight	0.69	0.42	0.24	0.11	0.10	0.07	0.05	0.05
Night	0.67	0.36	0.30	0.23	0.12	0.05	0.02	0.03
Ten	0.71	0.41	0.30	0.18	0.09	0.04	0.04	0.04

TABLE II.
THE CORRELATION COEFFICIENT OF ISCS AND CLEAN SPEECH SIGNAL

SNR=0 dB	ISC1	ISC2	ISC3	ISC4	ISC5	ISC6	ISC7	ISC8
One	0.01	0.12	0.13	0.23	0.33	0.66	0.51	0.10
Two	0.03	0.02	0.07	0.60	0.55	0.36	0.26	0.03
Three	0.01	0.04	0.10	0.33	0.56	0.52	0.39	0.10
Four	0.01	0.06	0.27	0.50	0.52	0.32	0.31	0.14
Five	0.03	0.19	0.34	0.25	0.29	0.44	0.46	0.23
Six	0.01	0.03	0.11	0.39	0.46	0.58	0.45	0.06
Seven	0.01	0.04	0.21	0.36	0.38	0.59	0.42	0.16
Eight	0.01	0.02	0.08	0.51	0.63	0.41	0.34	0.05
Night	0.02	0.16	0.14	0.13	0.28	0.64	0.50	0.16
Ten	0.02	0.03	0.07	0.33	0.55	0.63	0.28	0.05

D. Selection Criteria of Effective Singular Values

The results of SVD largely depend on the reasonable selection of noise reduction order. In signal processing, if too many singular values are selected and some noise will mix into the final signal, then it cannot desirably reduce noise. However, if too few singular values are selected, some useful signals will be lost and will result in waveform distortion of the signal. Thus, the selection of an effective singular value order is crucial. In Ref.[18], a method called singular entropy is presented to attempt to solve this concern. However, no obvious feature is found to determine the effective singular value, because the shape of a singular entropy sequence is similar to an inverse singular value sequence. The selection of an effective singular value based on user experience appears to be more frequently used than by singular entropy [19].

Thus, following the characteristic of difference spectra, two different criteria are proposed to select the reasonable noise reduction order:

1) Maximum criterion

If the position of the maximum k exists in the sequence of the difference spectra, then this point is the sudden change point. The maximum sudden change point represents the border between the singular value of the useful signal and noise. To determine the sudden change point, the singular values of noise can be reasonably detected, and then it can reduce noise.

2) Terminal for zero criterion

Given the equality of the singular values of noise, zeros exist in the sequence of the difference spectra. The noise reduction order is determined by searching for the first zero in the terminal of the difference spectrum [20].

III. Algorithm Implementation

As introduced in the basic theory of algorithms in section B, the implementation of the proposed algorithm will be discussed in this section. It mainly includes the signal pretreatment, selection of noise reduction order, and reconstruction of speech signal.

A. *Signal Pretreatment*

Signal pretreatment is an important step prior to implementation, and it contains the signal frame process, frequency division, and elimination of false components.

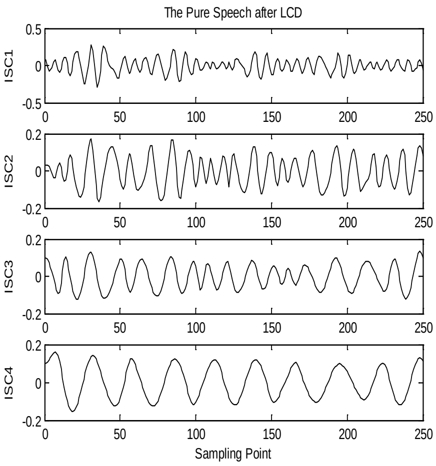
Sub-frame processing divides the signal, in which the long speech signal is truncated, improving the speed of operation. The unvoiced and voiced frames are obtained after this process, so that it effectively extracts the noise, unvoiced, and voiced data during frequency division. Then, LCD is used to divide the frame signal to produce a small amount of false ISCs. If the correlation coefficient is greater than 0.2, then it is considered as an effective ISC. Otherwise, the value is discarded. The time-domain waves of the pure and noisy speech frames after LCD are illustrated in Figure 3.

B. *Selection of Noise Reduction Order*

The effective ISCs remove noise by using SVD. The first and second order ISCs select the maximum criterion of the difference spectrum to determine the noise reduction order, while the other ISCs select the terminal for zero criterion.

The frequency spectrum of the ISC shown in Figure 4 is analyzed. Noise component is mainly found in the low-level ISCs, and a few useful speech signals are found. The high-level ISCs are mainly speech signals. However, a small amount of noise exists in this level. Therefore, different frequencies of ISCs must select different order of noise reduction using SVD.

Simulation experiment shows that the low-input SNR speech frame signal can remove more noise components and preserve most of the speech components in the maximum criterion of the difference spectrum when the input SNR of the speech frame is higher. These phenomena indicate that the content is mostly speech signal, and the maxima of the difference spectra cannot represent the border between the singular values of the useful signal and noise, which can influence the selection of order. Therefore, these signals should effectively remove noise by choosing the terminal for zero criterion.



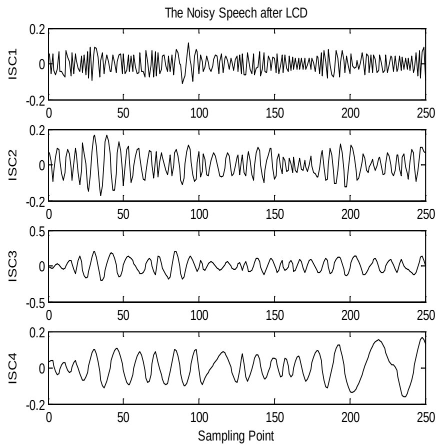
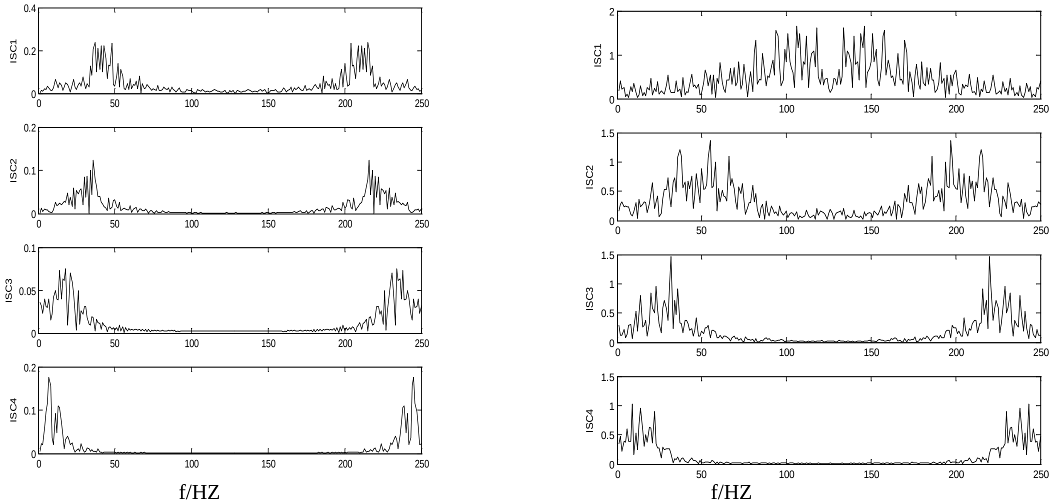


Figure 3. Time-domain waveforms of pure and noisy speech after LCD decomposition



(a)The spectrum of pure Speech frames after LCD (b)The spectrum of noise Speech frames after LCD

Figure 4. Frequency spectrum of pure and noisy speech frames after LCD

C. Reconstruction of Speech Signal

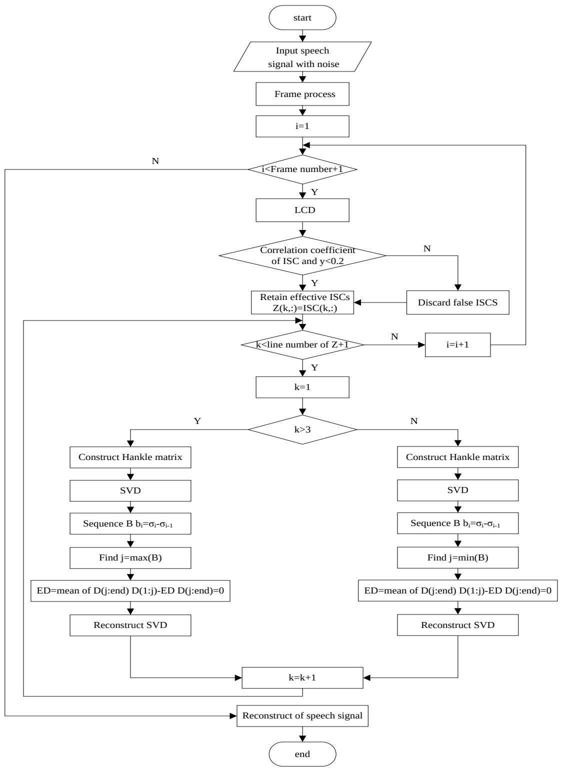


Figure 5. The algorithm flow chart

After the noise reduction order k is determined, $\sigma_k, \sigma_{k+1}, \cdots, \sigma_q$ are singular values of noise. Then, the average singular values of noise can be calculated as $\sigma_{mean(n)} = \frac{1}{(q-k-1)} \sum_{i=k+1}^{q} \sigma_i$. The sequence of singular values subtracts the average singular values of noise. Then, the k-point after the singular values are set to zero, is $\sigma_k' = \sigma_k - \sigma_{mean(n)}$, $\sigma_{k+1} = \sigma_{k+2} = \cdots = \sigma_q = 0$.

Finally, the de-noising speech is recovered from a new singular value sequence $Y' = (\sigma_1', \sigma_2', \cdots \sigma_k', 0, 0 \cdots, 0)$.

According to the above procedure, the algorithm flow chart is shown in Figure 5.

IV. Experimental Results

The time-domain waveform for an enhanced speech, input and output SNRs, the speech distortion measurement value, and the mean opinion score are selected as the evaluation performance indices to assess the performance of the proposed speech enhancement algorithm. Therefore, this performance can be analyzed objectively and subjectively.

A. Simulation

In this experiment, a female sound “seven” is chosen as a pure speech signal, with sampling frequency of $10kHz$ with 16 bits. The frame length is $W = 0.025 \times f_s$, and the frame shift is $50\% \times W$. In the case of input SNR ranging from -8 dB to 8 dB, the comparative results of the output SNRs in the different speech enhancement methods are shown in Figure 6. The different speech enhancements contain the traditional wavelet method, spectral subtraction, soft threshold in EMD, and the proposed algorithm.

Figure 6 shows that the highest output SNR of the enhanced speech was found in the proposed algorithm. The wavelet method and soft threshold in EMD results are nearly equal, because they essentially use the threshold to reduce noise.

The spectral subtraction results in the least effective noise reduction, because it depends on the estimation of the noise model to eliminate noise. If noise elimination is inaccurate, the final de-noising results are undesirable. Sometimes spectral subtraction introduces musical noise, which can influence the reduction of auditory effects.

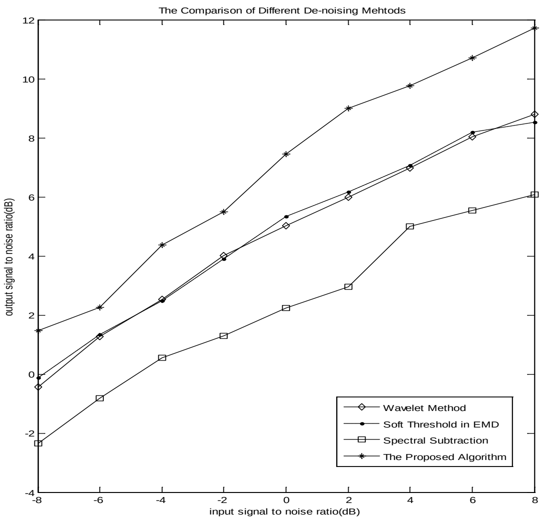


Figure 6. Comparison chart of output SNR using different algorithm

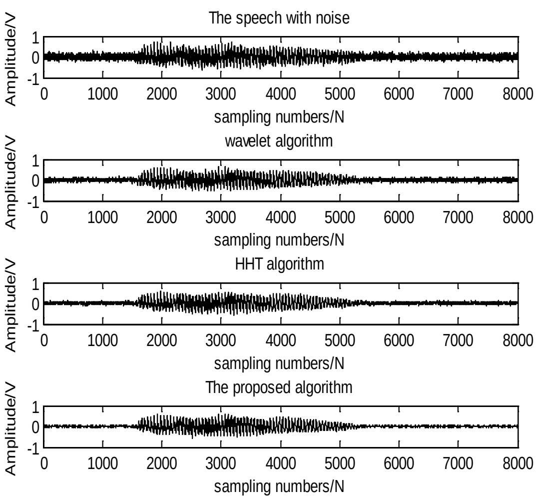


Figure 7. Three algorithms simulation comparison chart

In this paper, the subjective mean opinion score (MOS) acts as an objective evaluation criterion to assess the intelligibility of an enhanced speech algorithm. The noise of voice "seven" is reduced by four algorithms, and the results of MOS are the average of the scores achieved by 10 students to assess the enhanced speech. These results are tabulated in Table III, which shows that the intelligibility of enhanced speech by the proposed algorithm is better than the other algorithms.

When input SNR is -2dB, the three algorithms is used for the voice "seven". They are wavelet algorithm, HHT algorithm and improved algorithm. It is obvious that the improved algorithm can remove a lot of noise from Figure 7.

In this experiment, sound database of one to ten are chosen as speech signals. Mean square error (MSE) is acted as a criterion of measuring the enhancement speech distortion degree. Firstly, one to ten sounds in the input SNR ranging from $-8dB$ to $8dB$ are calculated their MSE. Secondly, each sound average its MSE is acted as final voice distortion measurement values. These results are tabulated in Table IV, which compares with the different speech enhancement methods.

TABLE III.
COMPARISON OF SUBJECTIVE MOS VALUES

The Score of MOS				
Input SNR(dB)	Wavelet Method	Soft Threshold in EMD	Spectral Subtraction	Proposed Algorithm
-8	1.8	2	1.2	3
-6	1.9	2.1	1.4	3.1
-4	1.9	2.1	1.8	3.4
-2	2	2.2	2.2	3.5
0	2.2	2.4	2.4	3.9
2	2.4	2.5	2.6	4.1
4	2.6	2.8	2.9	4.4
6	2.7	2.9	3	4.5
8	2.9	3	3.2	4.6

TABLE IV.
DISTORTION MEASUREMENT VALUES IN INPUT SNR FROM -8DB TO 8DB

DMV / voice	traditional wavelet method	soft threshold in EMD	spectral subtraction	proposed algorithm
one	0.0085	0.0084	0.0168	0.0081
two	0.0048	0.0050	0.0076	0.0043
Three	0.0050	0.0049	0.0092	0.0041
four	0.0084	0.0090	0.0100	0.0050
five	0.0027	0.0031	0.0032	0.0015
six	0.0011	0.0010	0.0014	0.0006
seven	0.0029	0.0025	0.0043	0.0018
eight	0.0025	0.0023	0.0033	0.0011
nigh	0.0030	0.0034	0.0050	0.0028
ten	0.0020	0.0017	0.0031	0.0020

The table IV shows that the least distortion measurement values (DMV) was appeared in the proposed algorithm, it explained the enhancement speech signal used by the proposed algorithm is the most similar to original signal .Therefore, the effect of the proposed algorithm is better than the traditional wavelet method, spectral subtraction and soft threshold in EMD.

B. Simulation Experiment in Different Noisy Conditions

In this experiment, a female sound "seven" is selected as a pure speech signal, with sampling frequency of $10kHz$ with 16 bits. The frame length is $W = 0.025 \times f_s$, and the frame shift is $50\% \times W$. The four noise files from NOISEX-92 database consisting of pink, aircraft, factory, and babble noises are added to the speech files. The noises of the four noisy speeches are removed by the proposed algorithm, and the results of the output SNR and speech distortion values for the various SNRs are tabulated in Table V. The mean square error (MSE) is selected as the evaluation criterion of speech distortion. Table V shows that the output SNRs are improved significantly and the speech distortion values are smaller. Therefore, the proposed algorithm to reduce pink, aircraft, factory, and babble noises is practical.

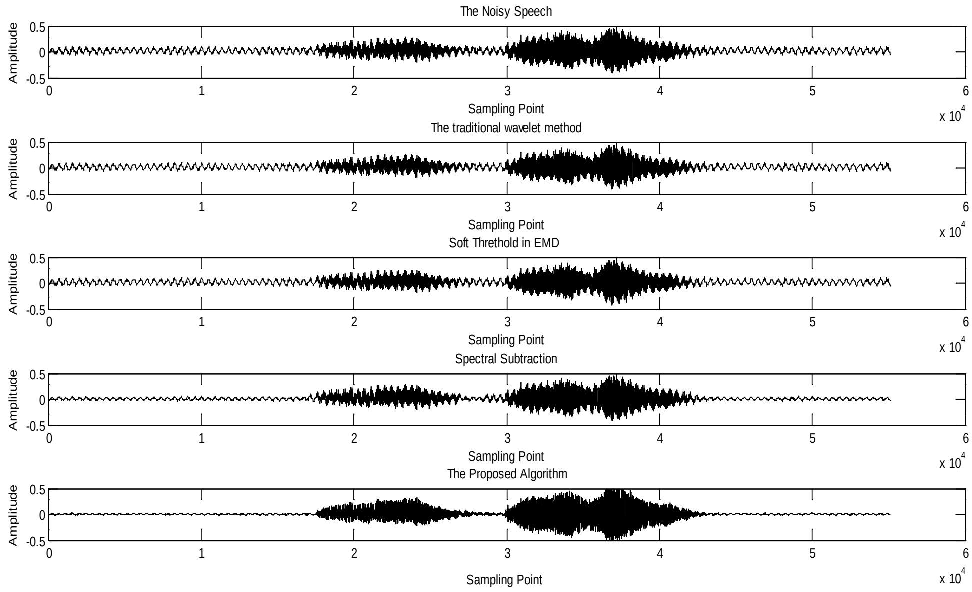


Figure 8. Waveforms of the enhanced speech in the different algorithms

TABLE V.
COMPARISON OF DE-NOISING RESULTS OF THE DIFFERENT NOISE CONDITIONS

Comparison	Pink Noise	Aircraft Noise	Factory Noise	Babble Noise
SNR_{input} (dB)	4.861	1.763	3.782	2.139
SNR_{output} (dB)	8.425	7.696	8.148	5.923
MSB	0.0009	0.0010	0.0009	0.0015

C. Simulation Experiment in De-noising Real Speech

In this experiment, a female sound “ni hao” is selected, with sampling frequency of $22.05kHz$. This voice is transcribed by Windows recording software in a real environment. The noise resources are derived from the sound of old fan blades and an old computer system. The waveforms for the enhanced speech using the traditional wavelet method, soft threshold in EMD, spectral subtraction, and the proposed algorithm are demonstrated in Figure 8. The proposed algorithm can considerably eliminate background noise from practically noisy speech, and the enhanced speech does not lose as much useful speech component as the other algorithms do.

V. CONCLUSIONS

A new speech enhancement algorithm based on LCD and difference spectrum of singular value is proposed. Based on the different frequency distributions of useful speech and noise signals after LCD, the two criteria of difference spectra can be used to determine the noise reduction orders so that a mass of noise can be removed. The proposed algorithm overcomes the drawbacks of the basis functions and threshold selection irregularity. However, this algorithm cannot introduce musical noise after noise reduction. The simulation experiments verify the effectiveness of the proposed algorithm.

REFERENCES

[1] Kim J B, Lee K Y and Lee C W.,” On the applications of the interacting multiple model algorithm for enhancing noisy speech”, *IEEE Transaction on Speech and Audio Processing*, vol. 8, no. 3, pp. 349-352, 2000.

[2] Zhao Huan, Zhao Kai, Liu He, et al. Improved MFCC feature extraction combining symmetric ICA algorithm for robust speech recognition. *Journal of Multimedia*, vol. 7, no.1, pp.74-81, 2012.

[3] Jindong Zhang, Guihe Qin, Ye Liu. Speech Separation in the Vehicle Environment Based on FastICA Algorithm. *Journal of Multimedia*, Vol. 7, no.1, pp.33-40,2012.

[4] Huan Zhao, He Liu, Kai Zhao, et al. Robust Speech Feature Extraction Using the Hilbert Transform Spectrum Estimation Method. *International Journal of Digital Content Technology and its Applications*, vol.5, no. 12,pp. 85-95, 2011.

[5] Huan Zhao, Yufeng Xiao. A Novel Robust MFCC Extraction Method Using Sample-ISOMAP for Speech Recognition. International Journal of Digital Content Technology and its Applications, vol.6, no. 19, pp. 393 - 400, 2012.

[6] Huan Zhao, Xiujuan Peng, Lian Hu, et al. “An Improved Speech Enhancement Method Based on Teager Energy Operator and Perceptual Wavelet Packet Decomposition”. *Journal of Multimedia*, vol. 6, no. 3, pp. 308-315,2011.

[7] Kuldip Paliwal, Belinda Schwerin, Kamil Wofcicki, “Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator”. Journal Speech Communication, vol.54, no.2,pp. 282-305, 2012.

[8] LI Ye, Wu Ya-feng, LIU Xue-fei.”Research on speech enhancement based on perceptual wavelet transform”,

Application Research of Computers, vol. 26, no. 4, pp. 1313-1315, 2009.

[9] LIU Li-wei, LI Jin-bao, ZHAO Kong-xin, DING Tie-fu, " The application of Hilbert-Huang Transform in speech Enhancement", *Microcomputer information*, vol.23, no.10, pp. 208-210,2007.

[10] Xiaojie Zou, Xueyao Li, and Rubo Zhang, "Speech Enhancement Based on Hilbert-huang Transform Theory", *Proceeding of the First International Multi-Symposiums on Computer and Computational Sciences*, vol.1, pp.208-213, 2006.

[11] LOU Li-xia, ZENG Yi-cheng, XIONG Min-quan , "An approach of speech enhancement combining EMD and adaptive filter", *Computer Engineering and Applications*, vol.48, no.9, pp. 104-107,2012.

[12] Zhao hua Wu,Norden E.Huang."A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method". *Proceedings of the Royal Society of London*, vol. 460, no. 2046, pp. 1597-1611, 2004.

[13] Norden .E.Huang et al,"The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proc. R. Soc. Lond. A*, vol. 454, no.1971, pp. 903-995, 1998.

[14] Cheng Jun-sheng, ZHENG Jin-de, YANG Yu, "A nonstationary signal analysis approach-the local characteristic-scale decomposition method", *Journal of Vibration Engineering*, vol.25, no.2, pp. 215-220, 2012.

[15] HU Mou-fa, DONG Wen-juan, WANG Shu-hong, CHEN Zeng-ping, "Singular Value Decomposition Band-Pass-Filter for Image Background Suppression and Denoising", *Acta Electronica Sinica*,vol.36,no.1, pp. 111 -116, 2008.

[16] Xuezhi Zhao, Bangyan Ye, "Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock", *Mechanical Systems and Signal Processing,* Vol.25, no.l5, pp.1617-1631,2011.

[17] WANG Tai-yong, WANG Zheng-ying, XU Yong-gang, LI Rui-xin, "Empirical Mode Decomposition and Its Engineering Applications Based on SVD Denoising", *Journal of Vibration and Shock*, vol.24, no.4, pp. 96-98, 2005.

[18] Li Ruixin,Wang Dongfeng,Han Pu,etal."On the applications of SVD in fault diagnosis".*Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol.4, pp.3763-3768, 2003.

[19] N.Virag,Single Channel.Speech Enhancement Based on Masking Properties of Human Auditory System. *IEEE Transactions on Speech and Audio Processing*, vol.7, no.2, pp. 126-137, 1999.

[20] P. Lockwood, J. Boudy, M. Blanchet." Experiments with a Nonlinear spectral subtraction (NSS), Hidden Markov models and projection, for robust recognition in cars". *Journal Speech Communication*, vol. 1, no.2-3, pp.215-228, 1992.

Nan Li achieved the MS. in Power electronics and Power transmission in 2005. He is currently Assistant Professor at the Faculty of Information Engineering of Northeast Dianli University, P.R.China. Her research interests are mainly in weak signal detection, nonlinear systems application in signal processing. She published 10 papers on journals and 6 papers on international conferences.

Analysis on Traffic Conflicts of Two-lane Highway Based on Improved Cellular Automation Model

Xiru TANG and Yanyan CHEN
Beijing Transportation Engineering Key Laboratory, Beijing University of Technology, Beijing 100124, China
E_mail: lilihuo@139.com, cdyan@bjut.edu.cn

***Abstract*—Based on microscopic traffic characteristics of two-lane highway and different driving characteristics for drivers, the characteristics of drivers and vehicle structure are introduced into Cellular Automation model for establishing new Cellular Automation model of two-lane highway. Through computer simulation, the paper analyzes the effect of the promotion of different vehicles, drivers and arrival rates on traffic conflicts of two-lane highway, which gets the relationship between the parameters such as road traffic and velocity variance and collision. The results indicate that the frequency of traffic conflicts has close relationship with the product of traffic flow and velocity variation. When the traffic flow and velocity variation are great, the frequency of the conflict is the greatest, and when the traffic flow and velocity variation are little, the frequency of the conflict is the least.**

***Index Terms*—traffic flow, traffic conflict, cellular automaton model, numerical simulation**

I. Introduction

The concept of traffic conflicts was first proposed by Perkins and Harris who defined a traffic conflict as any potential accident situation leading to the occurrence of evasive actions such as braking or swerving. This definition has since been refined and an internationally accepted definition is ‘an observable situation in which two or more road users approach each other in space and time for such an extent that there is a risk of collision if their movements remain unchanged’ [1]. The essence of traffic conflict is the manifestation of unsafe factors of traffic behavior. It not only can cause traffic accidents, but also can avoid traffic accidents because of taking safety measures. So accident is similar to conflict. The unique difference between them is if there is direct damaging consequence. In the second international conference about traffic conflict, all traffic engineering scholars considered that traffic conflict technique is an effective measure to identify accident black-spots [2-3].

The application of traffic conflict techniques in assessing the safety of a road entity (intersection, road segment, etc.) has been continuously gaining attention among safety researchers and practitioners. Several studies have demonstrated the feasibility of collecting conflict data using three ways: field observers [4], simulation models [5], and video-camera [6-9].

There are also several studies demonstrated how to assessing the safety of a road entity by using traffic conflict. Lord and Mannering summarized the various types of statistical crash frequency (count) models [10]. Gary et al. discussed an alternative approach to linking traffic conflicts with accidents by estimating the probability of unsuccessful avoidance maneuver by given the initial conditions [11]. Andrew P. Tarko discussed recent developments in safety modeling oriented towards the use of microscopic observations of vehicle interactions and better representation of crash causality [12].

Look through the researches on traffic conflict at home and abroad, they are focus on the relationship between conflict and accident, discriminated method and prediction on conflict. But in practical application, simple distance discrimination method, time criterion method and even rough ocular estimation are used to obtain conflict data [13].

These methods are suitable for fixed point observation on specific objects. Obtaining large-scale conflict data not only is difficult, but also has high cost, which is not applicable for observing large-scale road network with highway network as the representative.

The cellular automat model began to receive wide attention from the traffic and transportation community only after the simple formulation by Nagel and Schreckenberg, which is known as the NaSch model [14]. Daganzo presented how under some specific assumptions, a CA is equivalent to the kinematic traffic flow model and the car-following model [15]. Furthermore, the CA has been generalized to slow-to-start phenomena, traffic with high speed vehicles, signalized intersection, multilane multi- class traffic flow, inhomogeneous mixed traffic flow [16] and large traffic networks. The CA model flexibly introduces various parameters describing real traffic condition and can effectively simulate microscopic motion of vehicles in traffic flow, which is

Manuscript received January 3, 2013; revised June 1, 2011; accepted April 20, 2013.
Project number: 2012CB723303.
Corresponding author: Xiru TANG.

doi:10.4304/jmm.8.3.241-248

easy for studying interaction mechanics between vehicles. So it is widely applied in the studies of traffic flow theory.

Recent years, many jobs have been done to modify the CA model to make it more realistic. On the study of the modification cellular automaton model, Nagatani and Takashi presented a deterministic cellular automaton model to simulate the traffic flow in a two-lane roadway. It is shown that the exchange of cars has an important effect on phase transition between maximal velocity phase and the high-density phase [17]. Nagel et al. summarized different approaches to lane changing and classified the multitude of possible lanechanging rules for freeway traffic. They used a cellular automaton model for two-lane traffic to generate the density inversion, and found in many European countries, the densities somewhat below the maximum flow density [18]. Li et al. proposed a symmetric two-lane cellular automata model to investigate the aggressive lane-changing behavior of fast vehicle and the effect of different lane-changing probability [19]. Tang et al. proposed a car-following model by considering the anticipation effect of the lane changing probability of the leading vehicle on the car-following behavior of the following vehicle on single lane [20]. Wei Lv et al. extended the continuous single-lane models (OV model and FVD model) to simulate the lane-changing behavior on an urban roadway that consists of three lanes. The simulation results indicate that lane-changing behavior is not advisable on crowded urban roadway [21]. Peng et al. presented a new lattice model, and found that the consideration of lane changing probability in lattice model can stabilize traffic flow [22-23].

The objective of this paper is to explore the potential traffic conflict on different traffic condition of normal highway especially two-lane highway, which involved different type of drivers, different kinds of traffic structure, and different vehicle arrival rate. And then new CA rules with road traffic characteristics parameters and traffic conflict parameters are proposed. Using the new CA model, we set up different traffic scenarios, simulate the road traffic characteristics and analyze the attributive characters of highway traffic conflicts for realizing the highway risk criterion with the conflicts as the index.

II. Improved Cellular Automat Model of Two-lane Highway

In two-lane highway, the vehicles can be divided into cart and car according to climbing performance. And the divers can be divided into the radical and the conservative according to the characters of drivers in the past studies. Compared with the conservative drivers, the radical drivers are easy to overtake under the condition of congestion and may speed up as long as there is enough space between vehicles. But the conservative drivers keep great distance with the vehicle, and they nearly overtake. According to the above conditions, the rules of the radical drivers and the conservative drivers can be differentiated in the driving evolution rules, which can get CA rules of two-lane highway.

A. Evolution rules of positive driving vehicles

(1) The conservative vehicles are on the basis of NaSch model evolution rules, and the evolution rules of positive driving vehicles are:

(a) at $t+1$, the velocity of vehicles is:

①If $V_i(t) < gap_i(t)$, $V_i(t) < V_{i\max}$,

$$V_i(t+1) = \begin{cases} V_i(t) & \textit{Uniform probability}\,\rho_u \\ V_i(t)+1 & \textit{Accelerate probability}\,\rho_a \\ V_i(t)\text{-}1 & \textit{Deceleration probability}\,\rho_d \end{cases};$$

②If $V_i(t) < gap_i(t)$,

$$V_i(t) = V_{i\max},\ V_i(t+1) = \begin{cases} V_{\max}(t) & \textit{Uniform probability}\,\rho_u \\ V_{\max}(t)\text{-}1 & \textit{Deceleration probability}\,\rho_d \end{cases};$$

③If $V_i(t) \geq gap_i(t)$ satisfies the rule of overtake, it is obeyed. And if it doesn't satisfy the overtake rule,

$$V_i(t+1) = \begin{cases} \max(gap_i(t)-1,0) & \textit{probability}\,\rho \\ gap_i(t) & \textit{probability}\,1-\rho \end{cases};$$

(b) at $t+1$, the position of vehicles is:

$$X_i(t+1) = X_i(t) + V_i(t+1).$$

(2) The radical vehicles are on the basis of FI model evolution rules [24], and the evolution rules of positive driving vehicles are:

(a) at $t+1$, the velocity of vehicles is:

①If $V_i(t) < gap_i(t)$, $V_i(t) < V_{i\max}$,

$$V_i(t+1) = \begin{cases} V_i(t) & \textit{Uniform probability}\,\rho_u \\ V_i(t)+a & \textit{Accelerate probability}\,\rho_a \\ V_i(t)\text{-}1 & \textit{Deceleration probability}\,\rho_d \end{cases},$$

a is the maximum acceleration of vehicles;

②If $V_i(t) < gap_i(t)$, $V_i(t) = V_{i\max}$,

$$V_i(t+1) = \begin{cases} V_i(t) & \textit{Uniform probability}\,\rho_u \\ V_i(t)\text{-}1 & \textit{Deceleration probability}\,\rho_d \end{cases};$$

③$V_i(t) \geq gap_i(t)$, if it satisfies the overtaking rule, the following car overtakes. If it doesn't satisfy the overtaking rule, $V_i(t+1) = gap_i(t)$.

(b) at $t+1$, the position of vehicles is:

$$X_i(t+1) = X_i(t) + V_i(t+1).$$

B. Overtaking Rules

In the overtaking process, when the overtaking velocity is less than the maximum velocity of the vehicle, the radical drivers directly use the maximum velocity of vehicle, and the conservative drivers speed up 1 every hour and drives with the maximum velocity. Meanwhile, if the vehicles with low velocity satisfy the overtaking rules in the overtaking process, they can follow the overtaking vehicles. The overtaking process is shown in Figure 1.

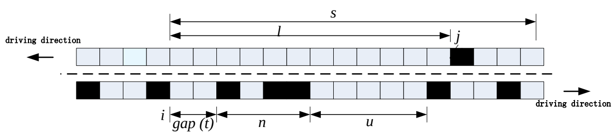


Figure 1 Overtaking Process

And the overtaking rules are as follows:
(1) When the car in front is hindered by the low-velocity cars, $gap_i(t) \le V_i(t)$ and $V_{i+1}(t) < V_{ihope}$;
(2) When the low-velocity cars have enough space within overtaking sight distance, $u \ge T_1 V_{i+m}(t) + 1$;
(3) When the neighboring trace has enough space, $T_2 > T_1$;
(4) If the above three conditions are satisfied, the car overtakes by changing track with the probability ρ_T

Speed of changing lane:
(1) The conservative vehicles:
$V_i'(t+1) = \min(V_i(t)+1, V_{i\max})$
(2) The radical vehicles: $V_i'(t+1) = V_{i\max}$

Position of changing lane: $X_i'(t+1) = X_i(t) + V_i'(t+1)$

V_{ihope} is the expected speed of the car, $X_i'(t+1)$ is the position of the i car entering the opposite lane at $t+1$.

C. Overtaking Retrograde Evolution Rules

In the overtaking process, the overtaking car and the following car make uniform motion within the overtaking sight distance, and the following car can't overtake it. When the car driving in the opposite direction may hinder the overtaking car, it should slow down.

And the overtaking retrograde evolution rules are:
Retrograde velocity: $V_i'(t+1) = V_{i\max}$
Location of vehicles update: $X_i'(t+1) = X_i'(t) + V_i'(t+1)$

D. Overtaking Conditions

As the opposite lane is needed to overtake in two-lane highway, which belongs to retrograde motion, it is not good to stop for too long time in the opposite lane. So the time of vehicles retaining in the opposite lane has a time upper limit and time lower limit.

(1) The shortest time of overtaking retrograde vehicle in the opposite lane, time upper limit T_1:

When $\gamma > 0$, if $\delta \ge \dfrac{m - 0.5\gamma(\gamma+1)+1}{\gamma}$,

$$T_1 = \mathrm{ceil}\left\{\left[\left(\delta+\frac{1}{2}\right)^2 + 2(m+1)\right]^{1/2} - (\delta+\frac{1}{2})\right\}.$$

And if $\delta < \dfrac{m - 0.5\gamma(\gamma+1)+1}{\gamma}$,

$$T_1 = \mathrm{ceil}\left[\frac{0.5\gamma(\gamma-1)+m+1}{\delta+\gamma}\right];$$

When $\gamma = 0$, $T_1 = \mathrm{ceil}\left[(m+1)/(V_{i\max} - V_{i+m}(t))\right]$, in which $\delta = V_i(t) - V_{i+m}(t)$, $\gamma = V_{i\max} - V_i(t)$, $m = n + gap_i(t)$, n is the length of the cars in front, and $ceil(x)$ is greater than the smallest positive integral of x.

(2) The longest time of overtaking retrograde vehicle in the opposite lane, time lower limit T_2:

When $l \le \dfrac{1}{2}\gamma^2 + \gamma\left(\eta + \dfrac{1}{2}\right)$,

$$T_2 = \mathrm{floor}\left[\sqrt{\left(\eta+\frac{1}{2}\right)^2 + 2l} - \left(\eta+\frac{1}{2}\right)\right],$$

When $l > \dfrac{1}{2}\gamma^2 + \gamma\left(\eta + \dfrac{1}{2}\right)$,

$$T_2 = \mathrm{floor}\left[(\frac{1}{2}\gamma(\gamma-1)+l)/(\eta+\gamma)\right].$$

$\eta = V_i(t) + V_j(t)$, when there is no vehicle in the opposite lane within the sight distance, $V_j(t) = 0$. Function $floor(x)$ is less than the smallest positive integral of x.

E. Identification of Overtake Completing

When the vehicle reaches T_1 on the opposite lane, the overtaking is completed and the vehicle returns the original lane.

The past researches indicated that the relationship between accident and conflict can be described by the severity of conflict. According to the severity, the traffic risk events can be divided into non-disruptive pass, non-serious conflict, and serious conflict and accident, and the quantitative relation between them distributes like a tower. There is direct relationship between serious conflict and traffic accident, so the paper focuses on studying the attributive character of serious conflicts.

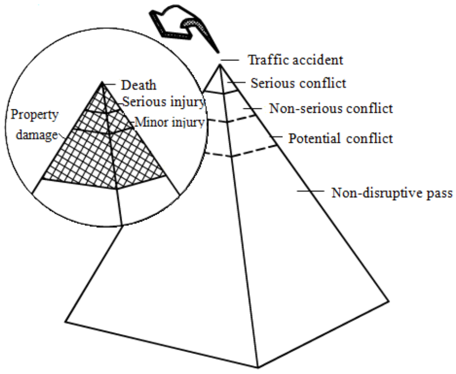


Figure 2 Classification Relationship Between Traffic Conflict and Traffic Accident

According to braking process formula of car and national standard, the incomplete braking distance of different vehicles with different velocities [11], as shown in Table 1.

TABLE 1 INCOMPLETE BRAKING DISTANCE WITH DIFFERENT SPEED

Speed (km/h)	***Critical distance Ds(s)***		
	Small car	***Midsize car***	***Large car***
5	0.3	0.4	0.47
10	0.87	1.12	1.21
15	1.72	2.14	2.22
20	2.83	3.47	3.49
25	4.21	5.11	5.04
30	5.86	7.06	6.96
35	7.78	9.31	8.95
40	9.98	11.88	11.31
45	12.44	14.75	13.94
50	15.17	17.93	16.84
55	18.17	21.42	20.01
60	21.44	25.22	23.44

65	24.99	29.33	27.15
70	28.8	33.75	31.13
75	32.88	38.47	35.38
80	37.23	43.51	39.9
85	41.86	48.85	44.69
90	46.75	54.5	49.75
95	51.91	60.46	55.08
100	57.35	66.37	60.68
105	63.05	73.31	66.55
110	69.02	80.19	72.69
115	75.26	87.39	79.1
120	81.78	94.89	85.78

III. Numerical Simulation

A. Parameter Calibration

In consideration of the type of vehicles in highway and great difference for length between carts and cars, in order to really describe the practical situation, the basic conditions of the model are set as follows: the length of cellular is 3.5m, the maximum speed of cars on two-lane highway and is $V_{max}=8$, and the maximum speed V_{max}' of carts is the speed of vehicles driving with uniform velocity, and the simulated 1 time step is 1 second. The other simulated parameters include:

The length of lane is 1000 cellular (3.5km). Open boundary conditions are used, and the lane under the initial condition is empty. The first cellular at the entrance of the road is set as the grid, and the arrival of vehicles follows Poissonian distribution, and the arrival coefficient is λ .

The initial speed of vehicles entering the road section: large cars satisfy the positive distribution that the mean value is 1 and variance is 1.3, and the small cars satisfy the positive distribution that the mean value is 4 and variance is 1.84. (The data is from S230 road section in Pinggu Beijing).

Output results of model: the average density, speed and flow of lane continue to use the macro definition of model. At time *t*, the average density of lane is $\rho_t = N_t / L$, the average speed is $\overline{V}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} V_i(t)$, and the average flow is $J_t = \rho_t \overline{V}_t$. In addition, the paper introduces variance of lane speed and the calculation formula is $Var_t = \sqrt{\frac{1}{N_t - 1}\sum_{i=1}^{N_t}(V_i(t) - \overline{V}_t)^2}$, which is used to describe the fluctuation of travel speed on the driveway.

Number of conflicts: the vehicles satisfying table 1 is recorded as one conflict, and the number of all conflicts on the road section is as the number of conflicts on the road section at time t. In order to eliminate the random influence of the system, the mean value of the model from 4000 to 8000 is selected, and 4000 mean values are for system average based on the time.

B. Symmetry of Lane Flow

The arrival coefficient λ of vehicles is from 0.1 to 1.0. The highway traffic characteristics with the proportions of different vehicles and different drivers are simulated according to the growth of 0.1 for figuring out the times of conflicts.

On simulating highway traffic characteristics with the proportions of different vehicles, the proportion of drivers is calibrated according to the proportion that the conservative and radical respectively accounts for 50%. On simulating highway traffic characteristics with the proportions of different driver, cars and carts respectively accounts fro 50%. As the arrival rates of vehicles on two lanes are same, and the traffic characteristics of lanes are similar, the traffic parameter of 1 for lanes is selected to be analyzed. And the analysis results are shown in Figure 3 and Figure 4.

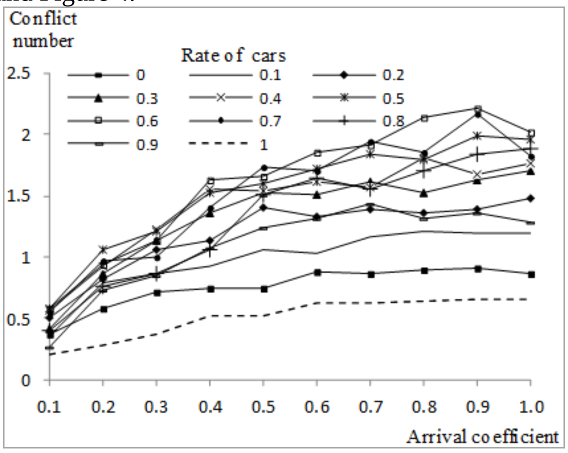


Figure 3. Relationship between the Number of Conflicts and Arrival Coefficient – Vehicles

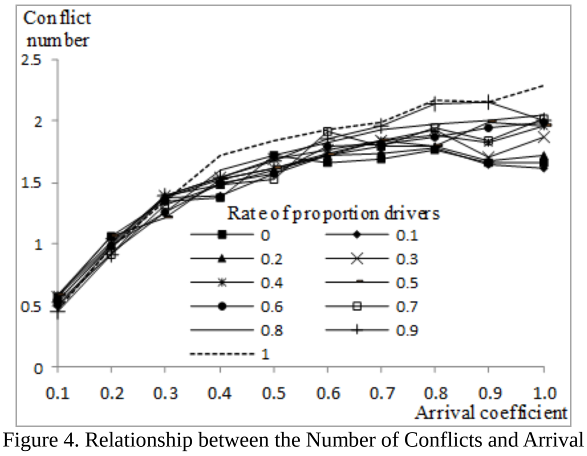


Figure 4. Relationship between the Number of Conflicts and Arrival Coefficient – Drivers

Figure 3 and Figure 4 is respectively the relationship between the number of conflicts and arrival coefficient with the proportion of different vehicles and different drivers.

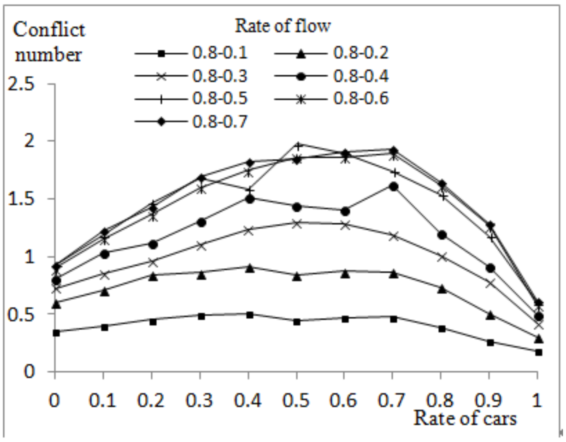
In Figure 3, the number of conflicts increases with the increase of arrival coefficient, and the curve for the number of conflicts of different small cars is very evident. In the situation of single vehicle, the number of conflicts is the smallest, and the number of conflicts is the greatest when the proportion of vehicles is 0.6 and 0.7, which shows that the probability of traffic accidents when the vehicles are not stable is greater than that when the vehicles are stable. With the increase of arrival coefficient, the vehicles increase, the times of vehicles influencing increases, and the times of conflicts increases, which conforms to the reality.

In Figure 4, the number of conflicts increases with the increase of arrival coefficient. For the proportion of different drivers, the curve begins to diverge when the arrival rate is greater than 0.3. And when the proportion of radical drivers is 1, the number of conflicts is the greatest and increases with the rapidest speed with the increase of arrival rate.

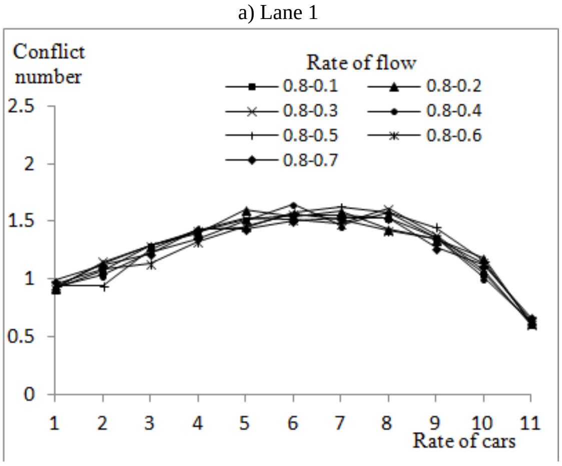
C. Asymmetry of Lane Flow

The arrival rate λ of lane 2 keeps 0.8, and the arrival rate λ of lane 1 is from 0.1 to 0.7. The highway traffic characteristics with the proportions of different vehicles and different drivers are simulated according to the growth of 0.1 for figuring out the times of conflicts.

On simulating highway traffic characteristics with the proportions of different vehicles, the proportion of drivers is calibrated 70% according to the proportion of the radical drivers, and the conservative drivers account for 30%. On simulating highway traffic characteristics with the proportions of different driver, small cars account for 70%, and large cars accounts for 30%. And the analysis results are shown in Figure 5 and Figure 6.

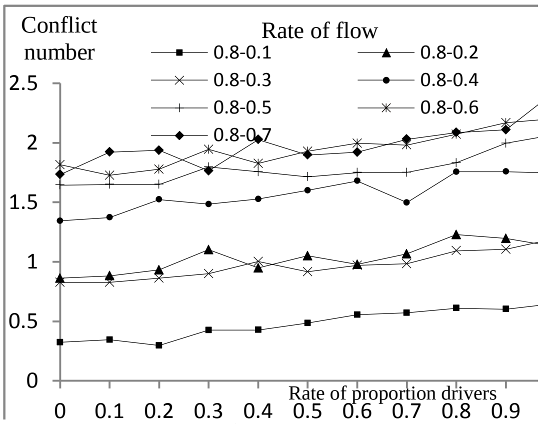


a) Lane 1

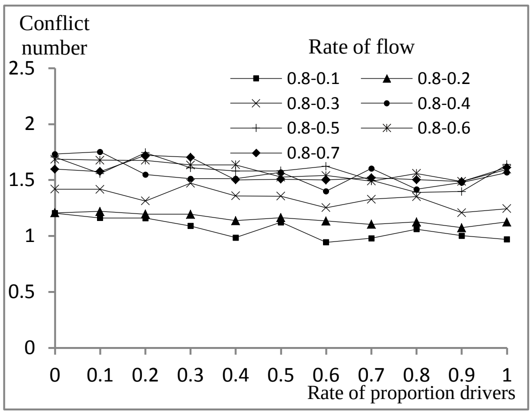


b) Lane 2

Figure 5. Relationship between the proportion of vehicles and the number of conflicts



a) Lane 1



b) Lane 2

Figure 6. Relationship between the proportion of vehicles and the number of conflicts

Figure 5 and Figure 6 is respectively the relationship between the number of conflicts and arrival coefficient with the proportion of different vehicles and different drivers.

In Figure 5, for the lane 1 with low arrival rate, the number of conflicts increases with the increase of arrival coefficient, and the curve is like parabolic. For the curve whose arrival rate is greater than 0.3, the proportion of small cars corresponding to the vertex of parabola is 0.7. And when the arrival rate is less than 0.3, the proportion of small cars corresponding to the vertex of parabola increased from 0.5 to 0.7. In addition, with the increase of arrival rate, the increase of the number of conflicts is very evident. For the lane 1 with high arrival rate, the number of conflicts increases with the increase of the proportion of small cars, the curve is parabolic, the proportion of small cars corresponding to the vertex of parabola is 0.7, and the curves with different proportions of flow are similar. We can see that the proportion of small cars has evident influence on the number of conflicts under different proportions of flow, but has not evident effect on the road with high flow.

In Figure 6, for the lane 1 with low arrival rate, the number of conflicts increases evidently with the increase of the proportion of radical drivers. When the arrival rate is 0.1, there are few vehicles, although there are many people overtaking, the times of severe conflicts is lower than the lowest level. When the arrival rate is 0.2 and 0.3, the number of conflicts increases greatly compared with the lowest level, but two curves which are similar are in the medium level. When the arrival rate is greater than 0.3, the number of vehicles increases to certain extent, the mutual function of vehicles increases, it is difficult for vehicles to speed up, slow sown and overtake, and the number of conflicts increases again. And with the increased of radical drivers, the times of vehicles overtaking increases, the number of conflicts increases. Although the lane 1 has little traffic flow, the number of conflicts is higher evidently than that on lane 2. As for the lane 2 with high arrival rate, the fluctuation of curve is stable. With the change the proportion of flow and the increase of arrival rate on the opposite lane, the number of conflicts on the lane increases.

D. Coupling Relationship

From the above analysis, we can know that the number of conflicts on road not only is influenced by the traffic flow, but also has close relationship with the fluctuation of speed. The chapter focuses on analyzing the coupling relationship between the number of conflicts and other traffic parameters.

The number of conflicts, velocity variance and the mean flow of lane are drawn into scatter diagram, as shown in Figure7 and Figure8.

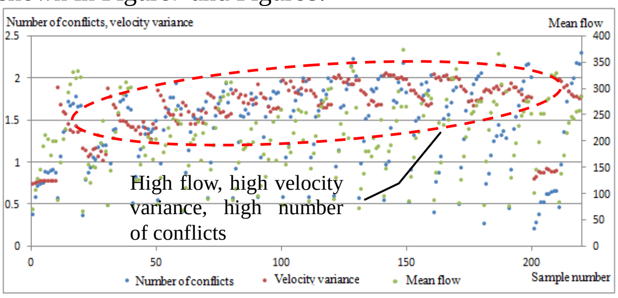


Figure 7. Scatter Diagram of Number of Conflicts, Velocity Variance and the Mean Flow of Lane under Symmetric Flow of Lane

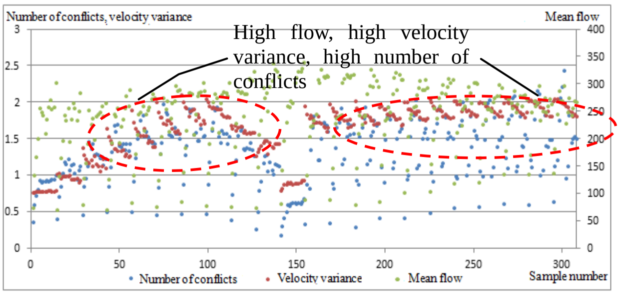


Figure 8. Scatter Diagram of Number of Conflicts, Velocity Variance and the Mean Flow of Lane under Asymmetric Flow of Lane

In Figure 8 and Figure 8, the number of conflicts is high under the condition of high speed variation and high mean flow of lane, which is consistent with the real situation that it is easy for traffic accidents when the fluctuation of velocity and the traffic flow are great.

We can deduce from the preamble analysis results that when the traffic volume or the speed fluctuation is great, it is easy for traffic conflicts. And we can summarize the logic relationship among the average flow of lane, speed variance and the number of conflicts, as shown in Figure 9. The paper introduces a new parameter, the product of the average flow of lane and speed variance. When the flow or speed fluctuation is great, the product of flow of lane and speed variance is great.

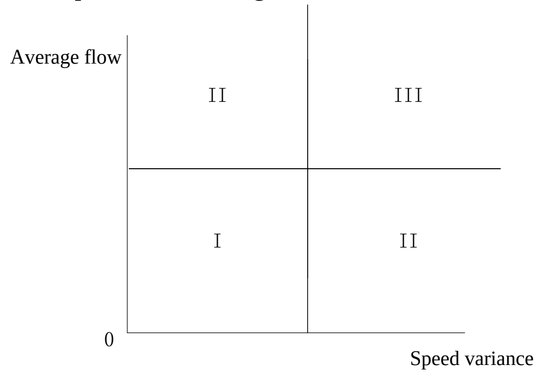


Figure 9. Relationship diagram among the average flow of lane, speed variance and the number of conflicts

In figure 9, I area means the condition of low speed variance and low average flow of lane, in which the number of conflicts is the smallest. II area is the condition of low speed variance and high average flow of lane, or the condition of high average flow of lane and low speed variance, in which the number of conflicts is greater evidently than that in I area. III area is the condition of high speed variance and high average flow of lane, in which the number of conflicts is the greatest.

The product of flow and speed variance and the number of conflicts are drawn into scatter diagram, as shown in Figure 10 and Figure 11.

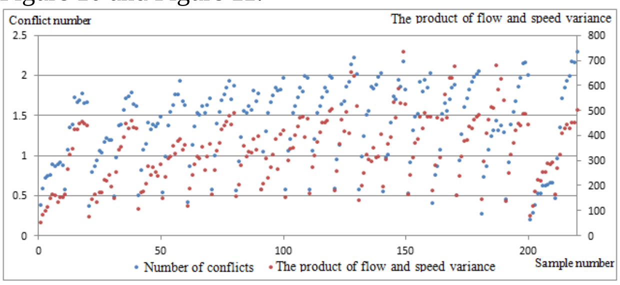


Figure 10. Scatter Diagram of the Product of Flow and Speed Variance and the Number of Conflicts under Symmetric Flow of Lane

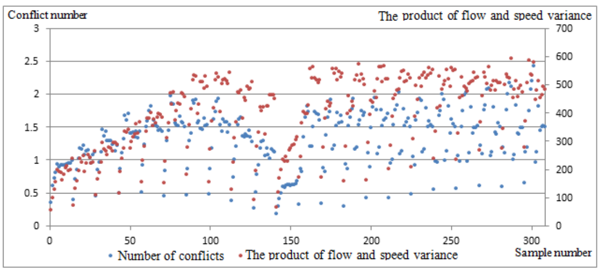
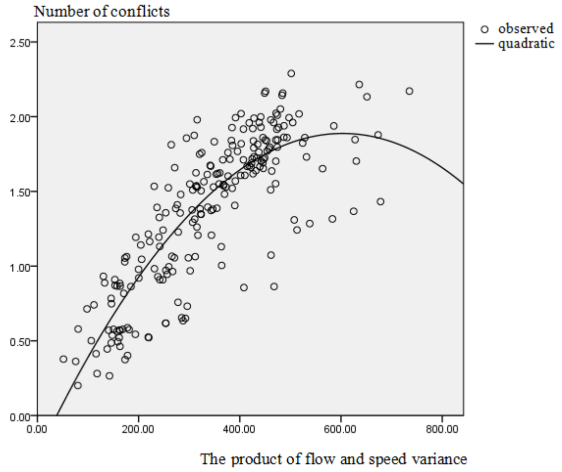
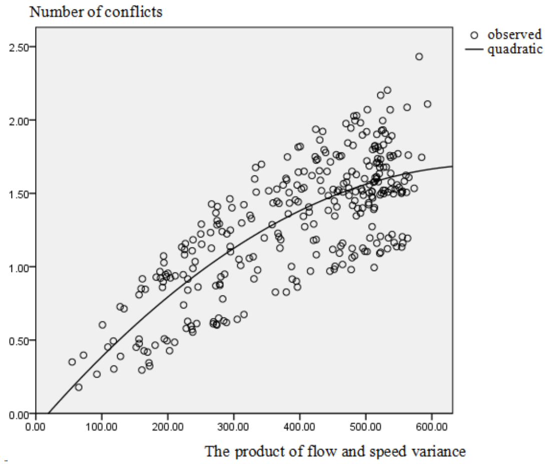


Figure 11. Scatter Diagram of the Product of Flow and Speed Variance and the Number of Conflicts under Asymmetric Flow of Lane

From Figure 10 and 11, we can see that there is close relationship between the product of flow and speed variance, and the number of conflicts. SPSS statistical data analysis software is used for correlation analysis, and the results are as follows:



a) Symmetric Condition of Traffic Flow



b) Asymmetric Condition of Traffic Flow

Figure 12. Regression Curve of the Product of Flow and Speed Variance and the Number of Conflicts

The goodness-of-fit index of curves (R^2) in Figure 12 is respectively 0.715 and 0.601. From the goodness- of-fit index, we can see that there is close relationship between the flow and speed variance and the number of conflicts, but the goodness-of-fit index under symmetric condition is higher than that under asymmetric condition, the reason for which is that the vehicles are influenced greatly by the vehicles on the opposite lane under asymmetric condition. And we can see that the number of conflicts is also influenced by other complicated factors.

IV. Conclusion

Based on microscopic traffic characteristics of two-lane highway and different driving characteristics for drivers, the characteristics of drivers and vehicle structure are introduced into CA model for establishing a new CA model of two-lane highway. Meanwhile, open boundary conditions are used for numerical simulation, which really realizes the traffic condition with the proportion of different characters and different drivers. The results indicate that the frequency of traffic conflicts has close relationship with the product of road traffic and velocity variation. When the road traffic and velocity variation are great, the frequency of the conflict is the greatest, and when the road traffic and velocity variation are little, the frequency of the conflict is the least.

Acknowledgements

The project as supported by the State Key Development Program for Basic Research of China (Grant No. 2012CB723303).

We also thank the anonymous reviewers for their suggestions and feedback that helped us to improve the quality of this paper.

Reference

[1] Amundson, F., Hyden, C., "Proceedings of first workshop on traffic conflicts", *Institute of Economics*, Oslo. 1977.

[2] Zhang Su, "Study on traffic conflict technology in China", *Southeast Jiaotong University Press*,1997.

[3] Cheng Wei. "Traffic conflict technology in city and application", *Science Press*, Beijing, 2006.

[4] Sayed, T., Zein, S., "Traffic conflict standards for intersections", *Transportation Planning and Technology*, vol.22, pp.309-323,1999.

[5] Mehmood, A., Saccomanno, F., Hellinga, B., "Simulation of road accidents by use of systems dynamics", *Transportation Research Record*,1746, pp37-46, 2001.

[6] Ismail, K., Sayed, T., Saunier, N., "Automated safety analysis using video sensors: technology and case studies", *Canadian Multidisciplinary Road Safety Conference*, Niagara Falls, Ontario, 2010a.

[7] Ismail, K., Sayed, T., Saunier, N., "Automated analysis of pedestrian-vehicle conflicts: a context for before-and-after studies". *Transportation Research Record*, 2198, pp.52-64, 2010b.

[8] Ismail, K., Sayed, T., Saunier, N., "Methodologies for aggregating traffic conflict indicators". *Transportation Research Record*, 2237, pp.10-19, 2011.

[9] Autey, J., Sayed, T., Zaki, M., "Safety evaluation of right-turn smart channels using automated traffic conflict analysis", *Accident Analysis and Prevention*, vol.45, pp.120-130, 2012.

[10]Lord, D., Mannering, F., "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives", *Transportation Research Part A*: Policy and Practice, vol.44,no.5,pp.291-305, 2010.

[11]Gary, A.D., Hourdos, J., Xionga, H., Chatterjeea, I., "Outline for a causal model of traffic conflicts and crashes", *Accident Analysis and Prevention*, vol.43,no.6, pp.1907-1919, 2011.

[12]Andrew P. Tarko, "Use of crash surrogates and exceedance statistics to estimate road safety", *Accident Analysis and Prevention*, vol.45,pp.230-240,2012.

[13]ZHANG Xue-liang, DENG Wei, GUO Tang-yi, "Development of traffic safety evaluation method based on TCT at grade crossings", *Journal of Transportation Engeering and Information*, vol.5,no.1, pp.85-90,2007.

[14]J. Esser, M. Schreckenberg. "Microscopic Simula- tion of Urban Traffic Based on Cellular Automata", *International Journal of Modern Physics C*, vol.8,no.5, pp.1025-1036, 1997.

[15]C.F. Daganzo, "In traffic flow, cellular automata kinematic waves", *Transportation Research B*, vol.40,pp. 396-403, 2006.

[16]L.W. Lan, C.W. Chang, "Inhomogeneous cellular automata modeling for mixed traffic with vehicles and motorcycles", *Journal of Advanced Transportation*, vol.39, pp.323-349,

2004.
[17]Takashi Nagatani, "Jamming transition in the traffic-flow model with two-level crossings", *Physical Review E*, vol.48, pp.3290-3294,1993.
[18]K. Nagel, D.E. Wolf, P. Wagner, P. Simon, "Two- lane traffic rules for cellular automata: a systematic approach", *Physical Review E*, vol.58, pp.1425-1437, 1998.
[19]X.-G. Li, B. Jia, Z.-Y. Gao, R. Jiang, "A realistic two-lane cellular automata traffic model considering aggressive lane-changing behavior of fast vehicle", *Physica A*, vol.367, pp.479-486,2006.
[20]T.Q. Tang, C.Y. Li, H.J. Huang, H.Y. Shang, "An extended optimal velocity model with consideration of honk effect", *Communications in Theoretical Physics*, vol.54, pp.1151-1155,2010.
[21]Wei Lv, et al., "Three-lane changing behaviour simulation using a modified optimal velocity model", *Physica A*, vol.390, pp.2303-2314,2011.
[22]G.H. Peng, X.H. Cai, C.Q. Liu, M.X. Tuo, "A new lattice model of traffic flow with the anticipation effect of potential lane changing", *Physics Letters A*, vol.376, pp.447-451,2012.
[23]Qian Yongsheng, et al., "Cellular Automation Models of Highway Traffic Flow Considering Lane-Control and Speed-Contro", *Communications in Theoretical Physics,* vol.56, no.04, pp.785-790, 2011.
[24]WU Ke-fei, KUANG Hua, KONG Ling-jiang, LIU Mu-ren, "Study of a Cellular Automaton FI-and-NS Mixed Model for Traffic Flow", *Journal of Guangxi Normal University(Natural Science Edition)*, no.4, pp.8- 12, 2005.
[25]Li Yang, Liu Jianxun, Shen Daji, Liu Xuan, Qudong, "Classification and identification for traffic conflicts in city", *Highways & Automotive Applications*, vol.6,pp. 50-53, 2009.

Xiru TANG, 1986.2-, Guilin, Guangxi province, China. Ph.D Candidate of Beijing University of technology. Xiru Received bachelor's degree in traffic engorging from Beijing University of technology, Beijing, China, in 2008 and Doctorate degree in transportation planning and management from Beijing University of technology, Beijing, China, in 2013.

Her areas of research include traffic safety, Intelligent Transportation and traffic planning. She has more than 10 paper publications in referred conferences/journals, chapters in one book and edited a book.

Yanyan CHEN, 1970.3-, Zhengzhou, He'nan province, China. Professor of Beijing University of technology. Yanyan reveiced Master's degree in structural optimization from Zhengzhou University, Zhengzhou, China, in 1994 and Doctorate degree in structural optimization from Harbin Institute of technology, Heilongjiang, China, in 1997.

Her areas of research include structural optimization, traffic safety, Intelligent Transportation, traffic planning, public transportation, road guidance and reliability research. She has visited several universities in India and abroad as Visiting Scientist and Professor. She has more than 120 paper publications in referred conferences/journals, edited five books.

Yanyan CHEN has served in various capacities in many academic activities, such as permanent members of China Communications and Transportation Association, permanent members of China Logistics Association.

Research on Nonlinear Characteristics of Image Measurement System for Instantaneous Concentration Field

WU Jing
School of Economic Mathematics, Southwestern University of Finance and Economics, Chengdu 710061, China
E-mail: 275068535@qq.com

***Abstract*—Quantitative measurement on instantaneous concentration field not only can provide scientific methods for people measuring environment wind tunnel, but also can provide important data for solving convection--diffusion problem in practical project. The established large environment and wind engineering wind tunnel needs to develop the measurement system of instantaneous concentration field in order to study concentration field of environmental pollution diffusion. Based on collecting, analyzing and selecting a large number of literatures, the paper comprehensively studies the image measurement of instantaneous concentration field, and develops the complete software and hardware system. And the developed measurement system is used to measure the results, and the nonlinear characteristics of instable concentration field are studied. Combined with experimental fluid mechanics, information technology, optical scattering and imaging theory, the paper makes quantitative calculation on instability of concentration field from an experimental point of view, which provides an important experimental result for using numerical method to explore the instability of concentration field.**

***Index Terms*—instantaneous concentration field, nonlinear characteristics, digital imaging processing**

I. Introduction

There have been mature and perfect numerical computing methods solving many steady-state flow problems so far [1]. But all calculation methods have the problems of stability, convergence, accuracy and convergence rate for unsteady flow and unsteady flow problems, which is one of the challenging research frontiers of computational fluid mechanics. Unsteady flow is common for natural and engineering problems such as shock wave, water wave motion, water flow, Hopf bifurcation, turbulent flow and widespread separated flow around a bluff body in various engineering equipments, which can be for mathematical description by using development-oriented differential equation. But as practical problems are very complicated, it is limited to solve the development-oriented differential equations and to solve approximate solution by using analytic methods [2, 3]. And it is difficult for numerical solution and numerical simulation to ensure the problem solving accuracy, so it is very important to make experiment flow measurement.

Experimental research methods are applied widely in fluid mechanics. Experimental study of fluid mechanics includes model tests or physical experiments which are made in laboratory equipments such as wind tunnel, shock tube, water tunnel, water tank, oil tunnel and hydroelectric analogue simulation. The advantages are that it can make observation experiments under the condition that it is the same to or similar to the researched problem. Therefore, the results of the experiments are reliable. But experimental methods are restricted by the model dimension. In addition, there are problems such as the influence of boundary. And some problems such as pneumatics, seepage of carbonate rock oil field, population and the fluidity of the war can’t be studied in the laboratory [4, 5].

For quantitative measurement of the flow, with the development of computer technique, obtaining displacement field, velocity field and concentration field from the recorded images have become the hot topic in many fields including mechanics, optics, medical science, video coding, machine vision, artificial intelligence and image processing. In recent decades, modern optics, the development of laser technology, computer technique and information processing technology not only has brought new vital force and vigor for the development of flow visualization， but also has made great progress in the ability of showing interior structure, quantitative extraction and analytical process of the flow information, which generates many whole flow-field technology and methods of qualitative display and quantitative measurement including Acoustic Doppler Flow Meter, ADFM, Acoustic Doppler Velocimetry, Tomographic Interferometry, Nuclerar-Magnetic-Resonace Image, Laser Doppler Velocimetry, Planar Laser-Induced Fluorescence, Particle-Tracking Velocimetry, Molecular-Tracking Velocimetry and article Image Velocimetry. These methods give fluxion structure, velocity distribution and concentration distribution with certain accuracy. But quantitative measurement on instantaneous concentration field has not been solved well.

doi:10.4304/jmm.8.3.249-255

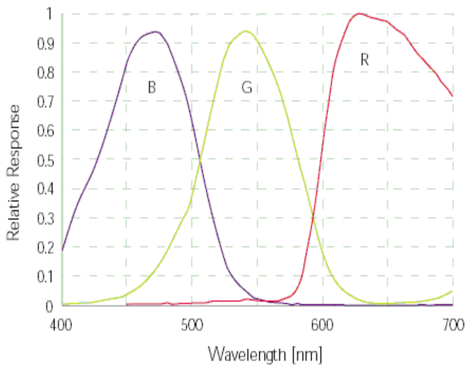
In order to study convection diffusion of complicated areas such as street canyon and residential environment, the paper develops a complete software and hardware system of image measurement for two-phase flow concentration field used in the wind tunnel. In the process of design and implementation, the measurement system firstly makes experiment on a small wind tunnel, and is directly used in large environment and wind engineering wind tunnel.

II. Proposed Scheme

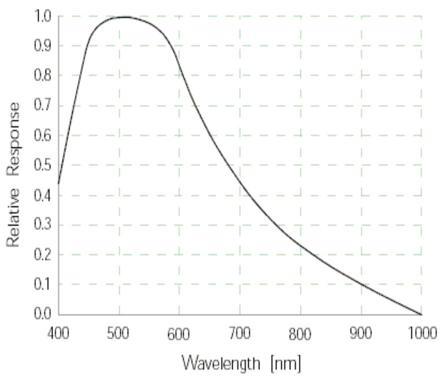
A. Image Acquisition System

Image acquisition system of PIVC is similar to that of PIV, and it consists of digital video camera, image collector and computer. And the camera determines the resolution, collection rate and digital-signal format of the system.

Optical imaging device is hardware equipment which transforms the signal of electromagnetic energy spectrum into electric signal. The main characteristic parameters of CCD include spectral response, photoelectric transfer characteristic, dynamic range, dark current, resolution and electronic shutter [2].



(a) Red, green and blue component of the response of color CCD



(b) Grey response

Figure.1 Schematic of spectral response sensitivity characteristics charts of BASLER A300 SERIES

When CCD is used to shoot flow fields, the selection of exposure time should satisfy the requirement that the displacement of particle movement couldn't exceed the minimum resolution of CCD under the control of synchronous generator, or the acquired flow images will have ghost image, which causes measuring error. In addition, in actual experiments, the density of incoming light is constant, and illumination of image plane of fixed optical system can realize the match of illumination of image plane and spectral response characteristic of CCD by choosing suitable relative aperture or exposure time of electronic shutter. In actual experiment, digital oscilloscope not only can be used to observe CCD signal, but also can be used to rectify the imaging distortion of CCD caused by various factors [3].

Spectral response of CCD has close relationship with wavelengths of light, and the wavelengths of light should correspond to peak response spectral characteristics of CCD. Generally speaking, spectral response range of CCD is 200nm～1100nm,and peak response wavelength is about 550nm. Figure 1 shows color response spectral characteristics and gray response spectral characteristics of array BASLER A300/A301/A302 SERIES. It is obvious that this category of CCD has very sensitive response to green ray (532nm), which must be considered on selecting light source.

CCD used for PIV has four working modes, continuous mode, control mode, triggering mode and double exposure mode. And each working mode and the relevant parameters can be set by software.

B. Image Processing Technique

The main measurement characteristics of images include luminous intensity and color. And the images of luminous intensity which are also called grey images or monochrome images which can be described by exposure and reflection model——two-dimensional intensity function:

$$f(x, y) = I(x, y)r(x, y) \tag{1}$$

x and y are space coordinates of image, $I(x, y)$ is incident function of incident light energy component depending on light source, $r(x, y)$ is the reflection function reflecting reflection characteristics of surface, and $0 < I(x, y) < +\infty$, $0 < r(x, y) < 1$. Incident function reflects external factors or environmental factors of image, and reflection function reflects the internal characteristics of image or the reflection characteristics of object on various lights.

The result of sample and quantization is an actual matrix. It is common that two main methods are used to express digital images. If an image $f(x, y)$ is sampled, the generating digital image has M lines and N columns. The value of coordinate (x, y) is discrete magnitude now. And integers are used to express the discrete magnitudes for convenience. The coordinate value of the origin is $(x, y) = (0, 0)$. The next coordinate value of the fist line of the image is expressed by $(x, y) = (0,1)$. And the upper left corner is the origin of the coordinates for coordinate representation of digital image.

In the digital process, M, N and discrete gray level L of each pixel need to be determined. There is no other requirements except for the requirement that M and N must be positive integers. However, for the consideration on process, storage and sampling hardware, the typical value of gray level should be integral power of 2, $L = 2^k$. If discrete gray level is equally spaced and is the integral within the interval $[0, L-1]$. Sometimes the value range of gray level is called dynamic range of image. And the images of all effective segments with gray level are called high dynamic range images. When a considerable number of pixels have the characteristic, the images have higher contrast. On the contrary, low dynamic range images seem like watered-down gray style. The bit number for storage is formula (2):

$$b = M \times N \times k \qquad (2)$$

Table 1 shows the bit number for storing square images when N and k is different common values. The gray level of each k is shown in parentheses. When the gray level of an image is 2^k, the image is called k-byte image. In the present flow visualization, the acquired digital images have 1024 possible gray levels, which is called 10-bit image.

Table 1. Bits for saving square images of common N s and k s

N/k	1 (L=2)	4 (L=16)	8 (L=256)	10 (L=1024)	11 (L=2048)	12 (L=4096)
32	1, 024	4, 096	8, 192	10, 240	11, 264	12, 288
64	4, 096	16, 384	32, 768	40, 960	45, 056	49, 152
128	16, 384	65, 536	131, 072	163, 840	180, 224	196, 608
256	65, 536	262, 144	524, 288	655, 360	720, 896	786, 432
512	262, 144	1, 048, 576	2, 097, 152	2, 621, 440	2, 883, 584	3, 145, 728
1024	1, 048, 576	4, 194, 304	8, 388, 608	10, 485, 760	11, 534, 336	12, 582, 912
2048	4, 194, 304	16, 777, 216	33, 554, 432	41, 943, 040	46, 137, 344	50, 331, 648
4096	16, 777, 216	67, 108, 864	134, 217, 728	167, 772, 160	184, 549, 376	201, 326, 592

Measurement of video quality is based on objective criterions of difference between the original signals and the processed signals, which is very important in applying video coding, the reason for which is that it not only can measure the distortion caused by compression, but also is a good criterion for motion estimation finding the best match (many PIV algorithms depend on it). In an ideal world, the measurement should relate to the difference between two video sequences, but it is difficult to find out the method. Although various quality measurements have been proposed, it is complicated for the method to calculating the correlation with visual perception of people. The design of most video processing systems makes the MSE between video sequence ψ_1 and ψ_2 minimal:

$$\mathrm{MSE} = \sigma_e^2 = \frac{1}{N}\sum_k \sum_{m,n} (\psi_1(m,n,k) - \psi_2(m,n,k))^2 \qquad (3)$$

N is the total number of pixels in each sequence. For color video, MSE of each color component is calculated respectively. PSNR replacing MSE with dB as unit is usually used for quality measurement of video coding, and the definition is:

$$\mathrm{PSNR} = 10\log_{10}\frac{\psi_{\max}^2}{\sigma_e^2} \qquad (4)$$

$\psi_{\max}$ is the maximum density value of video signal. For common each color 8-bit video, $\psi_{\max}$ is 255. For fixed peak value, PSNR is completely determined by MSE. As an important criterion, the quality of PSNR images higher than 40dB is very good, which means that it is close to the original image. The quality of PSNR images between 30dB and 40dB is good, and the quality of images less than 30dB is worse. In order to reduce the calculation, the measurement replacing MSE is MAD, and the definition is:

$$\mathrm{MAD} = \frac{1}{N}\sum_k \sum_{m,n} \left|\psi_1(m,n,k) - \psi_2(m,n,k))\right| \qquad (5)$$

For motion estimation, MAD is usually used to find out the best match of the known

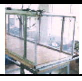
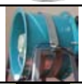
3. Overall design of measurement system

Based on the study on light scattering measurement methods of particles in concentration field and the image measurement technique, the paper develops complete image measurement system of instantaneous concentration field. The system not only has the function of instantaneous particle image field obtaining instantaneous concentration field, but also has the function of getting velocity field by calculating the collected PIV images. The system consists of hardware system and software system. And hardware system includes subsystems such as image capture, laser, CCD, particle exploder and control signal. Software system includes the modules such as image bus, concentration field display, velocity field analysis, curl field analysis, flow line analysis, memory management, other after processes and assistance.

Table 2 indicated the allocation of PIVC. The laser and digital camera of the system is the most expensive and the most important device. It's worth mentioning that the system can be as non-contact and solid-distorted measurement tool. The solid-distorted displacement field and velocity field can be obtained without laser, particle exploder and synchronous controller. On the other hand, the system can be expanded as motion estimation or motion monitoring system, which can be used for secure construction of buildings, warehouses and residential areas.

Figure 2 shows the hardware system architecture diagram of PIVC. Air, water and oil enter from the main entrance in the left corner. Some enter the tester through flow meter after branch entrance and particles generator

TABLE 2. CONFIGURATIONS OF THE PIVC SYSTEM

Legends	Parts	Parameters
	Laboratory table	Length × width × height=2200mm×700mm×700mm
	Laser	Model: Ambergreen-LFA050121 Name: 532nmpump end PuQuan solid-state green laser Maximum power: 1.02W (the current is 5.16A, and the stability is less than ±3%) Pumping source: Cw semiconductor laser medium: Nd:YVO4 Patch light: Beam diameter <1mm; Experimental section thickness:1mm, size: 700×700mm Divergence angle: <4mrad Cooling model: air cooling (natural cooling) Power input: 220V/50Hz/AC Overall dimensions: length × width × height =155mm×68mm×85mm Origin of product: Beijing cubic heaven and earth science and technology development co., LTD
	Camera	Model: SONY DSC-F828 Resolution: 640pixel×480pixel(30fps); 3264×2448 = 7990272 pixels(Single frame) Frame rate: 640pixel×480pixel 时 15fps~25 fps Bit rate: 25 MB/s Other parameters: Optical zoom 7.1; aperture F2.0-2.8; focal length f=7.1mm-51mm; 2/3 inches CCD; static and dynamic imaging; TIFF ape flac; exposure time 1/2000s Origin of product: SSGE
	Particle exploder	Model: FOGGER-AB-1500 Name: remote-control environmental tobacco oil heat sublimation smoke generator Particle: Particle size1µm ~500µm; environmental tobacco oil, Origin of produc: Beijing cubic heaven and earth science and technology development co., LTD
	Air blower	axial-flow fan 1800m3/h, 120w, wind speed of wind tunnel experimental section is 1m/s
	Computer	PC; CPU 3.0G; Mmory1.0G; Hard disc 80G
	PIVC software	Development; based on Windows XP; man-machine and dexterity
	Other necessary accessories	Auxiliary measuring equipment: camera tripod, laser tripod

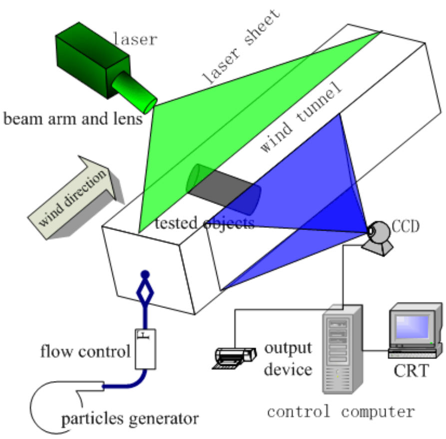


Fig. 2 Schematic of the PIVC hardware system

mixes. Meanwhile, the objective of the tester is illuminated by the patch light formulated by the laser on the right side. CCD shoots the flow field from the angle which is perpendicular to patch light, and the obtained information of flow filed is sent to the computer for process.

III. CFD NUMERIC CALCULATION OF CONCENTRATION FIELD

A. Control Equation

Continuity equation

$$\frac{\partial \bar{U}_i}{\partial x_i} = 0 \tag{6}$$

Momentum equation:

$$\frac{\partial \bar{U}_i}{\partial t}+\bar{U}_j\frac{\partial \bar{U}_i}{\partial x_j}=-\frac{1}{\rho}\frac{\partial \bar{P}}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\nu\frac{\partial \bar{U}_i}{\partial x_j}-\overline{u_i u_j}\right) \quad (7)$$

$k-\varepsilon$ equation:

$$\frac{\partial k}{\partial t}+\bar{U}_j\frac{\partial k}{\partial x_j}=\frac{\partial}{\partial x_j}\left(\frac{\nu_t}{\sigma_k}\frac{\partial k}{\partial x_j}\right)+\nu_t\left(\frac{\partial \bar{U}_i}{\partial x_j}+\frac{\partial \bar{U}_j}{\partial x_i}\right)\frac{\partial \bar{U}_i}{\partial x_j}-\varepsilon \quad (8)$$

$$\frac{\partial \varepsilon}{\partial t}+\bar{U}_j\frac{\partial \varepsilon}{\partial x_j}=\frac{\partial}{\partial x_j}\left(\frac{\nu_t}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x_j}\right)+\frac{\varepsilon}{k}\left(C_{\varepsilon 1}\nu_t\left(\frac{\partial \bar{U}_i}{\partial x_j}+\frac{\partial \bar{U}_j}{\partial x_i}\right)\frac{\partial \bar{U}_i}{\partial x_j}-C_{\varepsilon 2}\varepsilon\right)$$

(9)

In the equations, $\overline{u_i u_j}=2/3k\delta_{ij}-\nu_t(\bar{U}_i/\partial x_j+\bar{U}_j/\partial x_i)$, $\nu_t=C_u(k^2/\varepsilon)$; $C_u=0.09$; $C_{\varepsilon 1}=1.44$; $C_{\varepsilon 2}=1.92$; $\sigma_k=1.0$; $\sigma_\varepsilon=1.3$; ν_t is dynamic viscosity turbulence coefficient, and ε is turbulent dissipation term.

If gas and solid two- phase flow has no effect on flow field or the distribution of flow filed has no relation with solid phase (or liquid phase), it is true when the diameter of particle is small. And if there is no reaction and mass exchange between fluid and particle, particle diffusion is the result caused by fluid and particle. And advection-diffusion equation of particle is:

$$\frac{\partial C_i}{\partial t}+\frac{\partial (u_j C_i)}{\partial x_j}=\frac{\partial}{\partial x_j}\left(K\frac{\partial C_i}{\partial x_j}\right)+S_i \quad (10)$$

In the equation, C_i is the concentration of particle, K is turbulent diffusion coefficient, $K=\nu_t/\sigma$, $\sigma=0.7$.

B. Numerical Method

The transmission of fluid medium is the transmission of air and CO mixture, so it belongs to two-phase flow of air and CO. Numerical method uses grid computing method. According to the actual situation of wind tunnel experimental table and measurement of flow filed, meshing generation is firstly made, as shown in Figure 3, the size of the dimension of cylindrical radius with 1 is 160, the computational domain uses non-uniform grid, the grids near cylinder wall and CO entrance are very intensive, and other grids are sparse.

Reynolds number is

$$Re=\rho vl/\mu=vl/\nu=1\times 0.16/(1.512\times 10^{-5})\approx 10582$$

, and it is the same to Reynolds number in actual measurement experiment, which is good for comparison.

The equations from equation (6) to equation (10) use Finite Volume Method for discrete, the discrete format is QUICK and the time step is 0.005s.

In order to prevent the discrete and instability, under-relaxation method is used for momentum equation and scalar transport equation, and SIMPLE algorithm is used for coupling of pressure and velocity.

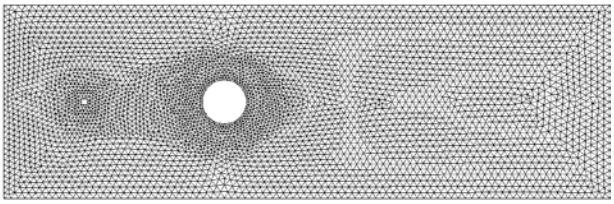


Figure 3. Meshes of computation regions

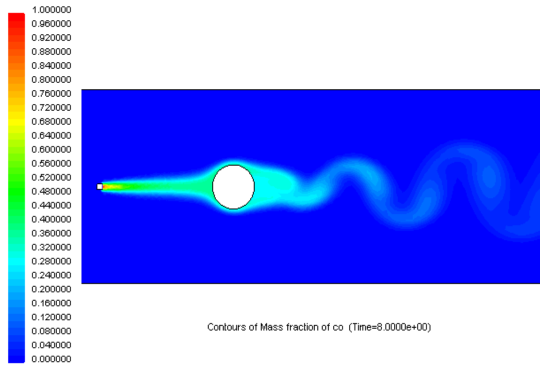
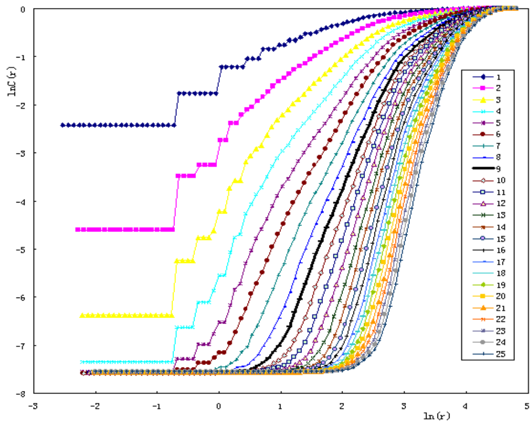
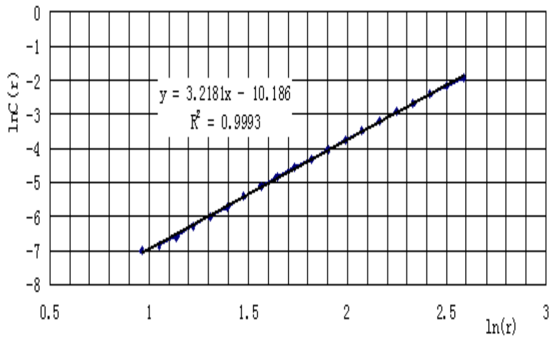


Figure 4. Computational velocity field and concentration field of CO ($t=8s$) in air and CO two-phase flow



(a) Calculation of correlation dimension



(b) Fitting result of linear segment

Fig. 5 Results of calculating the associated dimensions for pulsating concentrations with different embedded dimensions

Inflow boundary condition: velocity of dimension with 1 $u=1$, $v=0$.

Outflow boundary condition: derivative of each flow parameter along the flow line $\frac{\partial u_i}{\partial x} = \frac{\partial k}{\partial x} = \frac{\partial \varepsilon}{\partial x} = \frac{\partial p}{\partial x} = 0$.

Wall boundary condition: non-slipping condition is used for solid wall, and wall-function method is used nearby the wall.

Span direction boundary method: using free boundary condition, $\partial / \partial n = 0$.

The variables in the process of calculation should use double precision.

IV. Chaotic Characteristics of Time Series for Concentration Distribution

According to the judgment if the time sequence is mixed, the quantization calculation with the degree of controllability can be obtained by calculating correlation dimension, Kolmogorov entropy and maximal Lyapuno index of time sequence. Autocorrelation function method is used to select time delay $\tau = 5$ and calculate correlation dimension, which can get $\ln C(r) \sim \ln r$ curve graph in Figure 5(a). The linear segment with $m = 9$ sequence is selected for fitting, and the result is shown in Figure 5(b), which can get correlation dimension $D = 3.218$. Through further calculation, we can get Kolmogorov entropy $K_2 = 0.63691$ and the maximum Lyapunov index $L_{\max} = 0.138541$.

The analysis on time sequence of other points in field can get similar results, from which we can see that the concentration system has singular attractors ($D = 3.218$ is fractal dimension), which is chaotic motion (because $K_2 = 0.63691$ is positive number) and instable (because $L_{\max} = 0.138541$ is positive number).

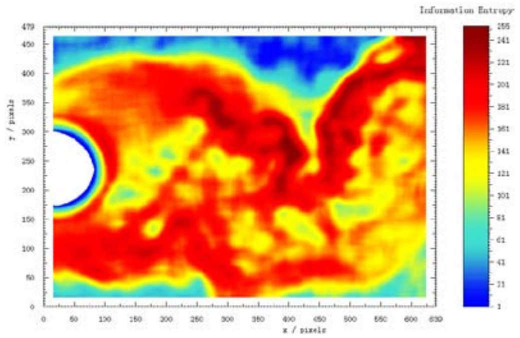


Figure 6. Computational result of the information entropies of the concentration field

Figure 6 is pseudo-color image of information entropy distribution of Figure 5. We can see from Figure 6 that the information entropy can represent uncertainty of information source for the system.

The time sequence of information entropy of concentration filed shown in Figure 7 can be got by exploring the change of information entropy with the time. And Figure 7 is the result of FFT. Obviously, the sequence has evident controlled composition and has certain ordering [4]. According to the judgment if the former time sequence is chaotic, the quantization calculation with the degree of controllability can be got by calculating the correlation dimension, Kolmogorov entropy and maximal Lyapuno index of time sequence [5].

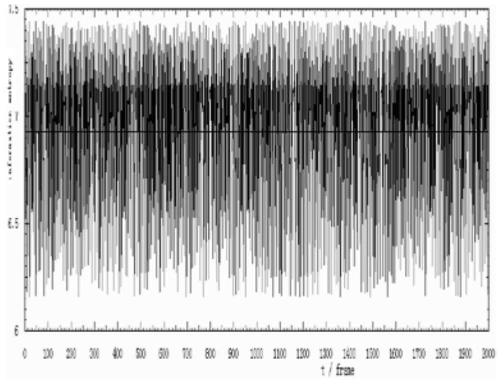
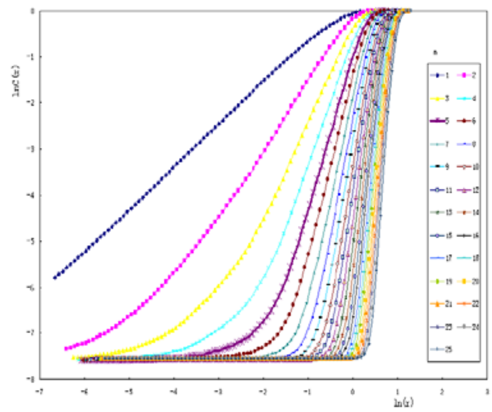
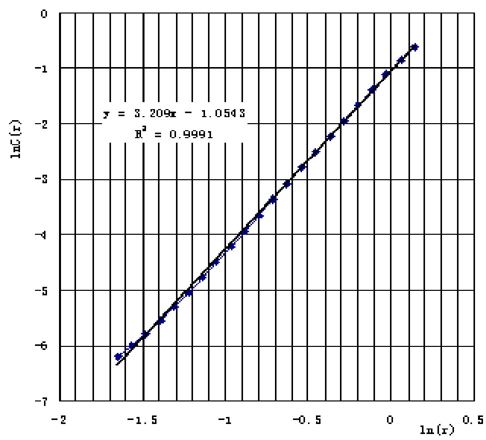


Figure 7. Time series of the information entropies of the concentration fields



(a) Calculation of correlation dimension with different embedded dimensions



(b) Fitting result of linear segment ($m = 5$ sequence

Figure 8. Non-linear characteristics analysis of the information entropies of the concentration fields

Autocorrelation function method is used to select time delay $\tau = 5$ and to calculate correlation dimension, which can get $\ln C(r) \sim \ln r$ curve graph in Figure 8(a). The linear segment with m = 5 sequence is selected for fitting, and the result is shown in Figure 8(b), which can get correlation dimension $D = 3.209$. Through further calculation, we can get Kolmogorov entropy

$K_2 = 0.511828$ and the maximum Lyapunov index $L_{\max} = 0.109229$.

The analysis on time sequence of other points in field can get similar results, from which we can see that the concentration system has singular attractors (because $D = 3.209$ is fractal dimension), which is chaotic motion (because $K_2 = 0.511828$ is positive number) and instable (because $L_{\max} = 0.109229$ is positive number).

V. CONCLUSION

Two-phase flow measurement is a study which is concerned all over the world, and there are still many problems needing to be solved. The quantitative measurement on concentration field not only can provide scientific methods for people measuring environment wind tunnel, but also can provide important data for solving convection--diffusion problem in practical project [6]. The established large environment and wind engineering wind tunnel needs to develop the measurement system of instantaneous concentration field in order to study concentration field of environmental pollution diffusion. Based on collecting, analyzing and selecting a large number of literatures, the paper comprehensively studies the image measurement of instantaneous concentration field, and develops the complete software and hardware system. And the developed measurement system is used to measure the results, and the nonlinear characteristics of instable concentration field are studied [7]. Combined with experimental fluid mechanics, information technology, optical scattering and imaging theory, the paper develops image measurement study on instantaneous concentration field.

REFERENCES

[1] Khalitov D. A., Longmire E. K. Simultaneous two-phase PIV by two-parameter phase discrimination. Experiments in Fluids, pp.252-268,2002, 32.

[2] Sato Y., Inaba S., Hishida K., etc. Spatially averaged time-resolved particle-tracking velocimetry in microspace considering Brownian motion of submicron fluorescent particles. Experiments in Fluids, pp.167-177,2003, 35,.

[3] Sakakibara J., Nakagawa, M., Yoshida M. Stereo-PIV study of flow around a maneuvering fish. Experiments in Fluids, PP.282-292,2004, 36,.

[4] Han Lei, Tao Leren, Zheng Zhiao, Yang Zhiqiang, Cheng Jian. Experimental study on influence of gas liquid on performance of rolling rotor compressor refrigeration system, Journal of Refrigeration, PP45, 2010, (04).

[5] Tao Hong, Tao Leren, Zheng Zhiao, Zhang Liqun. Experimental study on gas-liquid two phase flow oscillation flow pattern causing instability of refrigeration cycle, Journal of Refrigeration, PP.36,2009, (02).

[6] Si Shaojuan, Ouyang Xinping, Zhang Lianjie, Hong Siwen. Experimental study on thermal performance of T-type boiling enhanced heat exchange tube, Thermal Energy and Power Engineering, PP.36-37,2011,(04).

[7] Hong Siwen, Ouyang Xinping, Jia Chuanlin, Chen Jianhong. Experimental study on thermal performance of condensation enhanced heat exchange tube, Journal of Refrigeration, PP.67-68, 2009, (01).

A High-precision Hand-held Face Detection System

Chenghao Jiang, Jinhua Yang, Lijuan Zhang, and Xiaokun Wang
School of Optoelectronic Engineering, Changchun University of Science and Technology, Jilin 130022, China
E_mail: jiangch412@126.com

***Abstract*—This paper aims at portable high-precision surface defect detection equipment designed based on laser non-contact measurement, though a CCD camera and dual word line semiconductor lasers can achieve the measurement of small surface in a close distance. Resultant data is sent to PC via wireless transmission mode to achieve the separation of detection module with terminal display processing module. The detection system with a small-volume structure which can achieve fast and accurate scanning can be used to achieve the device workpiece surface defect measurement under various complex occasions. It is able to achieve real-time display of the measurement results through MATLAB graphical user interface. High measuring accuracy and reliability of the system is verified by experiments, and the depth measurement accuracy can reach 40μm. The results provide an important basis for the detection of surface blemish and improve product design and processing quality.**

***Index Terms*—Face detection, Three-dimensional reconstruction, Laser scanning, Machine vision, High-accuracy measurement**

I. Introduction

With the rapid development of science and technology, surface accuracy requirements of the precision instruments become more and more important. The measurement accuracy directly impact on the practical performance of instruments, however, some of very small cracks and bumps can not be discovered in time, which will cause serious consequences [1]. At present, the application of such detection equipment is mainly based on contact measurement, with low accuracy and poor stability. A small part of the non-contact laser measurement is mainly used for the detection of designated flatness detection and its bulk can not be achieved online portable measurement [2-5]. Today, with the increasing demand of surface high-precision detection, it is of great significance for equipment online operation monitoring, analysis of surface defects generation and development mechanism, and structural improvements to develop high-precision portable laser scanning online surface detector.

In order to meet the instruments surface detection requirements, this article has designed a high precision defect detection device based on laser triangulation measurement. It uses dual-beam measurement and wireless data transmission mode to achieve the data transmission between detection module and terminal processing display module. Except for reducing the volume, it also reduces the complexity of the post-date processing. The application of valid stripes proposed method to complete the measured information extraction is easier than the traditional two-camera measurement 3D stitching and reconstruction in three-dimensional information of the late match. The basis requirement of measurement depth range is 100cm, besides; the use of the extended structural design can not only accomplish close-range measurement, but also can get long distance and larger-scale measurement by replacing the lens and change the laser angle.

The key design requirements for the scanner are that it shall be small and fast, portable and reliable with high accuracy measurement results while the cost of measurement is reduced. The small 3D scanning device will be placed over the target to be detected, and the laser scans the area for flaws, and then sends the resultant data to PC via wireless transmission and high-resolution 3D surface maps are shown by measurement software. The measured surface defects' location, size and other parameters can be calculated by the analysis software. It can adapt to a variety of different environment measurement requirements and measurement results with high precision. The depth error is within 40μm, the equipment has good practical value.

II. System Overview

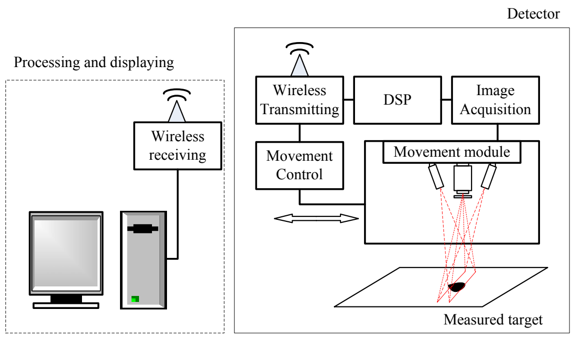


Figure 1. Framework Diagram of System

Laser triangulation is used as the basic measurement principle in this device. Laser scans the testing targets and gets the laser image by CCD camera, and then we can deduce the spatial coordinates of the measurement

doi:10.4304/jmm.8.3.256-261

target by analysis of the laser stripe position in image. Combined with practical application and direction development of the instrument, the following requirements are proposed: miniaturization and portable, low cost, high resolution and accurate measurement results. The designed system framework is shown in Figure 1:

The system consists of four parts: control system, image capture and processing, wireless transmission, and data analysis and display. Laser and CCD camera are placed in small connection board, and the board is fixed in small electronic control displacement platform. The control system controls CCD camera for image acquisition and synchronously movement electronic control displacement platform. After a high-speed digital processing chip DSP, we obtain information of the laser stripe by scanned images. Wireless transmission module will take processed stripe information to the PC terminal. After processing software, three-dimensional coordinate image of the measured surface will be real-time displayed.

In order to obtain three-dimensional spatial coordinates of the surface measured, it is required to establish spatial three-dimensional coordinate system with the center of the lens as coordinate origin. Any point of the laser stripe position (x, y, z) can be calculated by the point imaging in the CCD plane's position (x ', y', f).

The geometric relationship between object points and image points can be expressed in the following formula 1 [6-8]:

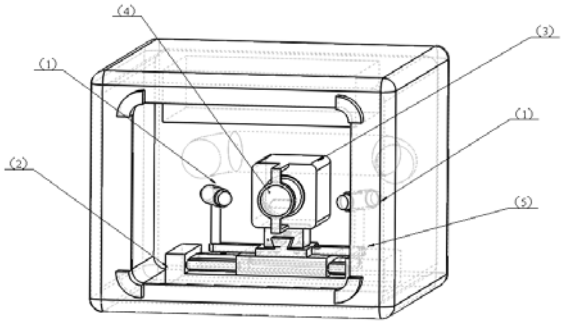
$$\begin{cases} x = \dfrac{bx'}{f\cot\theta - x'} \\ y = \dfrac{by'}{f\cot\theta - x'} \\ z = \dfrac{bf}{f\cot\theta - x'} \end{cases} \qquad (1)$$

where f is focal length, b is baseline, θ is laser beam angle.

III. Proposed Scheme

For rapid measurement of the test surface, multi-beam or coded light projection measurement methods are often used today. The entire surface measurement of the target can be achieved in one-time projection. This method not only can achieve rapid measurement, but also may increase the maximum utilization efficiency of CCD imaging plane. But this method have higher requirements for measurement environment, thus it is only applicable to the measurement of a continuous smooth surface and can't be used in the measuring complex and undulating surface. With a relatively high cost, the projection equipment also makes the volume too large, and can not adapt to different occasions and different accurate measurement requirements.

To avoid the drawbacks of the measurement method above-mentioned, dual-beam horizontal scan mode is used in this paper, with the measuring end structure as shown below:



1- Word lines semiconductor laser, 2- High-precision electronic control displacement platform, 3-CCD camera, 4-650nm narrowband filter, 5- Small connection board
Figure 2. Detection Device

In the system, two 650nm small word line lasers are placed in both sides of the CCD camera, and emission power is 5 ~ 15mw adjustable and line width is less than 1mm in 180mm measuring distance. The semiconductor lasers and CCD camera can achieve close-surface information measurement through a simple combination method. It not only reduces the size and cost of measurement device, but also can carry out an effective treatment for visual blind spot phenomenon generated by the propagation characteristics of light in the measurement process. Accurate extraction of the measured surface information by quickly and efficiently stripe extraction algorithm, makes it easier to achieve a three-dimensional stitching and reconstruction in the latter part of the three-dimensional information stitching than the traditional dual camera measurement. Because the measurement is carried out by use of two lasers and will get two 3-D maps. At any time, it can be drawn via equation 1 that two laser lines are in the same coordinate system with the center of the lens; for this, it is only required to insert the second image point coordinates in the first image coordinate system in 3D point cloud fusion to compensate respective vision dead angle. Because of the two lasers' different scanning position, it is able to further increase horizontal and vertical resolution of the scanner and increase the scanning area by this method.

In order to achieve 50mm × 50mm area face detection with 10cm distance, a monochrome CCD camera shall be chosen as optical receiver; the image plane resolution is 795 (horizontal) × 596 (vertical) pixels, with image acquisition speed of 50 frames/s and 5mm focal length of the lens. As measured simultaneously with a dual stripe, if the measurement surface is too large ups and downs, streaks may intersect with each other, which will interfere with the accuracy of the measurement results. In order to eliminate mutual interference between the stripes, combined with the overall design requirements of the instrument, it is required to select the semiconductor laser and camera lens whose center distance b (baseline) has a length of 45mm and the lens axis angle at laser direction is 30 °. By above structure design, it is 80mm length from the lens center; the two laser stripes coincide together in the optical axis of the lens. As the distance increases, the stripes gradually deviate on both sides, with imaging on both sides of CCD image plane. The front instrument housing is designed with a length of 130mm from the center of the lens, so as to achieve 80mm-180mm of depth measurement. To achieve their stripes information extraction only need to have an image separation process for the left and right 397 pixels. Camera and laser are uniformly fixed in small electronic control displacement platform; platform uses a ball screw with a positioning accuracy of less than 10μm and can achieve the fastest speed 40mm/s lateral movement. The size of entire instrument is 230mm × 230mm × 250mm can meet the requirements of the portable measuring.

In the actual measurements, in order to meet different needs of different target measurement requirements, we need to extend the measurement distance range. As shown in Figure 3, the design follows structure to connect CCD camera and semiconductor lasers together. It can achieve CCD camera forward and backward movement and rotation of the laser. Adjusting the CCD camera position can bring baseline changes within a certain range, and the laser and optical axis angle's change can change detection range expansion. With the choice of appropriate focal length lens, it is able to complete measurements under different fields of view. In order to rapidly obtain the new system parameters value, the system shall achieve precise calibration of the new parameters by a set of the rapid calibration apparatus in a very short time. The instrument fixed in the long-distance displacement of the table or placed in the pipeline can achieve a larger area and real-time three-dimensional detect.

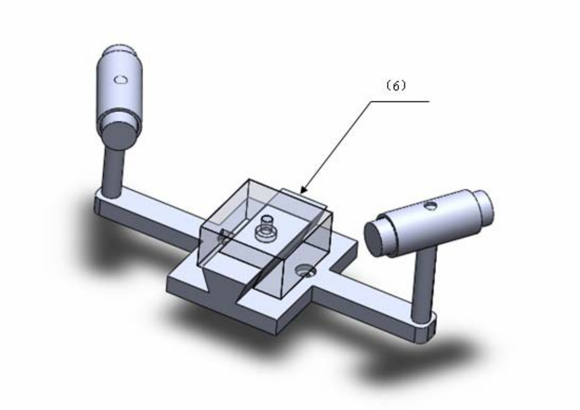


Figure 3. The Measurement Connection Frame

It can be found through large experiments that the light unevenly distributed in environment will bring more environmental elements inside, as shown in Figure 4(a). Due to the influence of ambient light, the laser stripe information extraction can not be obtained accurately. In order to eliminate the influence of ambient light on the imaging result, a 650nm narrowband filter is placed in front of imaging lens to reduce the interference caused by ambient light. As shown in Figure 4(b), the narrowband filter effectively filters out of the surrounding environment interference [9].

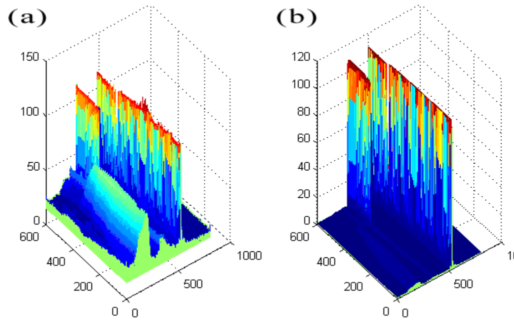


Figure 4. (a) Gray Surface with No Filter (b) Gray Surface Chart with Filter

The work process of scan processing module is shown in Figure 5.

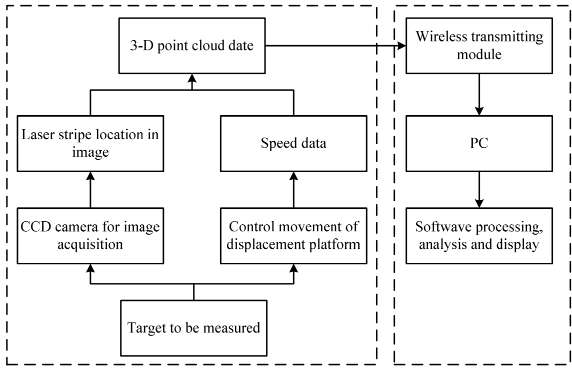


Figure 5. The Detection Device Work Flow Chart

Electronic control displacement platform carries camera and laser to scan the surface to be measured with the demand speed, and gets real-time image of the laser stripe by CCD camera. After this, we choose DSP as image processing chip to extract laser stripe position [10-11].

Firstly, we need to do filtering for images. The use of sensitive optoelectronic components in the system is susceptible to the interference of external factor and photosensitive element CCD will generate lots of noise by the factor of reflected light and stray light. In the laser stripe extraction process, if the image noise is too large, it will bring many errors to the laser stripe center extraction.

Wiener filter is to find an estimate f' so that the statistical error function (equation 2) minimum.

$$e^2 = E\{(f - f')^2\} \tag{2}$$

The expression can be expressed in the following formula 3 in the frequency domain:

$$F(u,v) = G(u,v)\left[\frac{1}{H(u,v)} \cdot \frac{|H(u,v)|^2}{|H(u,v)|^2 + S_\eta(u,v)/S_\eta(u,v)}\right] \tag{3}$$

where $H(u,v)$ is degraded function.

$$H(u,v)^2 = H*(u,v)H(u,v) \qquad (4)$$

where $H*(u,v)$ is the complex conjugate of $H(u,v)$, $S_\eta(u,v) = |N(u,v)|^2$ is the power spectrum of noise,.

Select the 5×5 template to do wiener filter for scanned image. After processing, image noise gets a better inhibition and obvious edge of stripe.

Secondly, we need one picture to take a rough calculation of laser stripe gray threshold and remove the pixels outside the threshold. The image gray average G_e and mean-square deviation σ can be calculated by all image pixels gray value. Take q as the summation of pixels' gray value; gray value is greater than S and S can be calculated by the following formula:

$$S = [G_e + (5\times\sigma)/6] \qquad (5)$$

The laser stripe threshold T can be calculated by this equation:

$$T = q/q' \qquad (6)$$

where q' is the total number of those pixels. Empty the pixels that gray value is less than T, and extract the laser line position roughly. Using the image threshold processing to get the result as shown in figure 6 below:

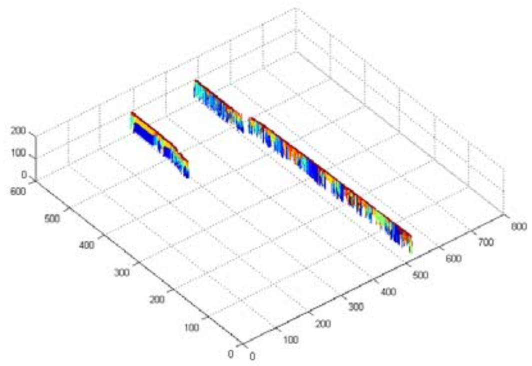


Figure 6. Extraction of Laser Light with Light Threshold Analysis

Finally, a least square fitting is used to calculate sub-pixel position of the laser stripe center. Least square method is to find the best function matching by summing the error of square. Based on given data point (x_i, y_i), i=1,2,…,n, assume the fitting function form as equation 7:

$$p(x) = \sum_{k=0}^{m} a_k \phi_k(x) \qquad (7)$$

where $\{\phi_k(x)\}_{k=0}^{m}$ is a known linear independence function, solve the coefficient $a_0, a_1, \ldots, a_m$ make the value of equation 8 minimum, P(x) is considered as a fitting function.

$$\begin{aligned} &\phi(a_0, a_1, \ldots a_m) = \\ &\sum_{i=1}^{n}[p(x_i) - y_i]^2 = \\ &\sum_{i=1}^{n}[\sum_{k=0}^{m} a_k \phi_k(x_i) - y_i]^2 \end{aligned} \qquad (8)$$

By a least square fitting, it is able to get the minimum error curve of the discrete points [12] and the extreme point measurement accuracy can reach to 0.1pixels. The fitting curve for a line is shown below:

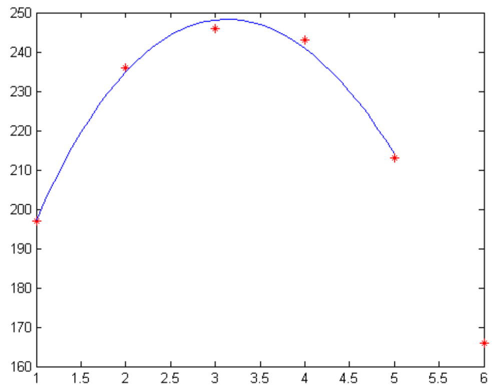


Figure. 7 Laser Pixel Curve Fitting

The results of stripe center position data is transferred to PC by wireless transmission module and then the surface three-dimensional information is displayed by processing software. Through processing and analysis by data processing software MATLAB, it is able to achieve calculation and evaluation of the surface measured.

IV. Experimental Results

The measurement requirements of instrument design are to meet a small area of measurement within a close range of 80mm to 180mm depth; testing equipment can be placed on the front of the tested surface. Upon running equipment, we can directly achieve the acquisition and processing of the measuring surface, as shown in figure 8.

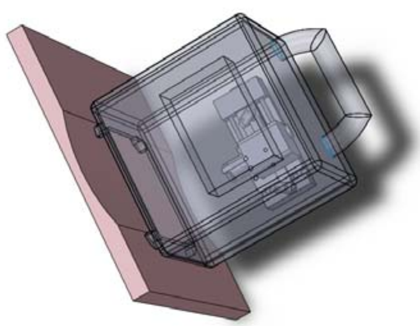


Figure 8. Target Detection

In order to verify the reliability of the equipment in actual work process, we do the following test experiment:

A. Partial Surface of Microscope Scanning Imaging

The partial surface of microscope shell is selected to do scan imaging. Through DSP image processing system, sub-pixel level of laser stripe position accuracy can be attained, and wireless transceiver module transmits data to PC.

In order to facilitate observation and analysis, GUI graphical user interface in MATLAB is used as the display interface for real-time data processing. The laser stripe position information and system parameters value after calibrated are taken into the formula 1 to obtain the spatial coordinate of laser stripe. The 3-D surface image will be displayed in the interface, with the measurement results shown in Figure 9:

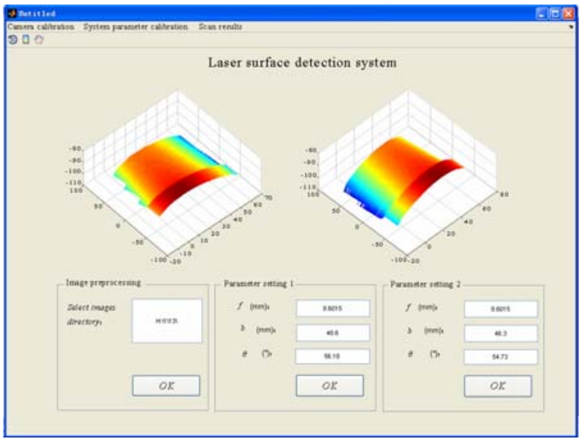


Figure 9. The Face Detection Device GUI Interface

Due to the use of two lasers measurement, we will get two different type scan images, as shown above. Analysis of the two images, target information on the left can be collected more full by the left laser imaging stripe, and the right side laser imaging stripe, just opposite. Integrate stack all data of two laser scanning information can well compensate for each other's lack of information caused by vision dead angle and a more comprehensive point cloud data can be obtained.

B. Plaster Mold Scanning Imaging and Data Fusion

Sometimes, we want to directly observe the surface properties of the measured target. For this purpose, the two images are fused into one image and get a complete 3D view. In this system, two three-dimensional maps have the same coordinate system can easily achieve the three-dimensional image fusion though MATLAB software. Finally, we will get a more complete three-dimensional information graph through a date fitting. The measured surface's micro-structural information is more directly reflected.

Select dome plaster mold as a target to be detected which high of 50mm. We get two three-dimensional maps through two different lasers which are shown in figure 10(a) and figure 10(b), their fusion result is shown in figure 10(c):

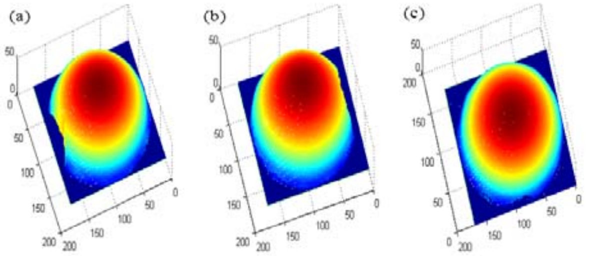


Figure 10. (a) The left laser scanning results (b) The left laser scanning results (c) Images fusion result

Though their fusion result we can get the dome plaster mold surface information directly and full-scale. The mold's volume and height can be accurately calculated.

C. Surface Analysis

For 3D data points of measuring surface, through MATLAB processing and analysis can quickly access information of the tested surface, judge the hollow position, and show its mark in the user interface. Select a tile which has defects as measured target and get measurement results as shown in Figure.11, through processing and analysis can quickly realize the surface defect and the visual dead angle due to the propagation characteristics of the light can be maximum eliminated.

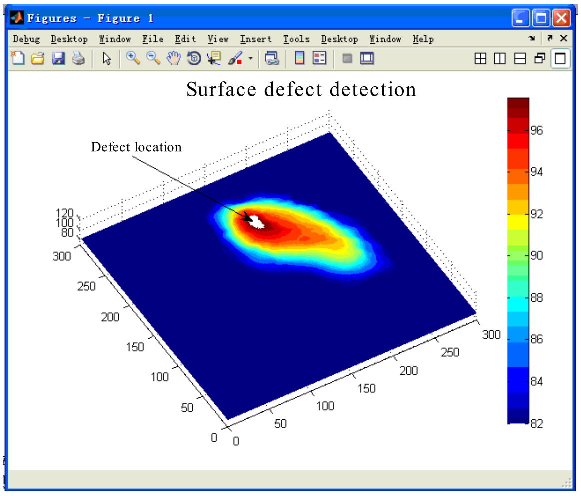


Figure 11. Detection For Ceramic

After a number of micro-movement measurement experiment and analysis can get the scanner uncertainty of measurement is 0.4 pixels. From the differential for equation 1, we can draw the factors affect the measurement accuracy [13-15]:

$$\delta z \approx \frac{z^2}{fb}\delta x' \tag{9}$$

System accuracy can be drawn from the equation above. System accuracy δz is proportional to f, b and inversely proportional to z^2 and $\delta x'$. Lens focal length and baseline length value is at a certain value when parameters were selected. Take system calibration parameters value into formula 2, we can get the device measurement accuracy can reach 40μm, the horizontal resolution can up to 10μm and vertical resolution up to

15μm. It meets to the accuracy requirements of engineering applications.

V. Conclusions

In this article, a small laser defect detection system was designed. It used a simple way to achieve high-precision measurement of the target surface. The system has the advantage of low-cost, portable measurement and can be adapted to a variety of complex measurement environment online- measurement. Through the device, the process of surface defects change will be recorded and the experiments also show its effectiveness.

In the measurement, the selection of CCD camera and line structured light generation have a direct impact on the image quality and will affect accuracy of measurement results. The more accurately the light source and CCD camera is useful to improve the measurement accuracy.

References

[1] Joseph P. Lavelle, Stefan R. Schuet, Michael Derouen. A Handheld, “wireless 3D laser scanner for shuttle tile inspection”, Buenos Aires, 2007:1~6.

[2] Yang Peng, Wu Fan, Hou Xi, “simulation analysis of absolute measurement for rotationally asymmetric surface error”, Opto-Electronic Engineering, vol. 38. no. 1, pp. 93-102, 2011.

[3] Xiong Zhiyong, Zhao Bin, “Inner hole range-finding sensor with laser triangulation based on trapezoid prism”, ACTA OPTICA SINICA, vol. 31, no. 12, 2011.

[4] Paul Viola, Michael Jones, “Fast and robust classification using asymmetric adaboost and a detector cascade”, Neural Information Processing Systems, pp. 1311–1318, 2002.

[5] Minh-Tri Pham, Tat-Jen Cham, “Online learning asymmetric boosted classifiers for object detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[6] Ali Peiravi, Behrai Taabbodi, “A reliable 3D laser triangulation-based scanner with a New simple but accurate procedure for finding scanner parameters”, Journal of American Science, vol. 6, no. 5, pp. 80-85, 2010.

[7] Nie Chenchen, Zhang Haibo, Tao Wei, Zhao Hui, “Integrative laser triangulation displacement sensor system based on FPGA and DSP”, Computer Measurement & Control, vol. 20, no. 6, 2012.

[8] Chen Feng, Chen Wuyi, Chen Zhitong, “Laser triangulation based measurement about the surface 3D”, Journal of Beijing University of Aeronautics and Astronautics, vol. 37, no. 10, pp. 1321-1325, 2011.

[9] Pang Miao, Yuan Xuewen, Gao Xueyan, Yan Gaoshi, “Spot distortion calibration in measurement of laser intensity distribution based on imaging by scattering”, ACTA OPTICA SINICA, vol. 30, no. 2, pp. 451-455, 2010.

[10] Vincent Paquit, Jeffery R.Price, “Near-infrared imaging and structured light ranging for automatic catheter insertion”, Medical Imaging Oak Ridge National Laboratory, Oak Ridge, Tennessee, 2006.

[11] LI Jie, “Detecting tire tread morphology based on laser triangulation”, pp. 26-28, 2009.

[12] Zhi-Kai Huang, Ling-Ying Hou, Xing-Wang Zhang, Jing-Jing Lian, "Using Implicit Shape Model and SVM for Face Detection", IJACT, Vol. 3, No. 11, pp. 130 ~ 137, 2011.

[13] Li Xiaojie, Zhao Kai, Zheng Xingming, “Development of surface roughness tester based on laser triangulation method”, Transactions of the Chinese Society of Agricultural Engineering, vol. 28, no. 8, pp. 116-121, 2012.

[14] Xie Jin, Wei Feng, TAMAKI Junich. “Laser measurement and evaluation for brittle/ductile cutting of micro/nano scale V-shaped groove”, Optics and Precision Engineering, vol. 17, no. 11, pp. 2771-2778, 2009.

[15] Wei Youning, Su Zhenwei, Li Haipeng, Gong Ke, “Inspection of corrosion defects on inside wall of compressed natural gas wells by laser triangle method”, Computer Measurement & Control, val. 18, no. 5, pp. 1001-1003, 2011.

Bus Passenger Recognition and Track of Video Sequence

Donghua Zhou
Southwest Petroleum University, Chengdu, Sichuan 610500, China
E-mail: Zhoudh_zxy@163.com

***Abstract*—Study on bus passenger recognition and track of video sequence is a research aiming at realizing making a statistics on passenger flow volume of bus, which not only helps traffic services center rationally schedule vehicles, but also can avoid overload. The system makes machine vision technology and digital image processing technique applied in customer counting of bus. The key questions in the study include effective collection of video image, effective extraction of moving object in sequence images, recognition of moving targets, trace and count of moving objects. The paper makes deep analysis on relevant technology from the perspective of the theory and practical application of the algorithm, and the study makes achievements. The theoretical algorithm for the study on bus passenger recognition and track of video sequence is distinctive. The practical experiments indicate that the system can satisfy the real-time requirements, and can accurately make statistics on the number of passengers, which has great application value.**

***Index Terms*—machine vision, frame difference, fuzzy clustering, Kalman filter, cost function**

I. Introduction

Intelligent transportation system emerges as the times require. Intelligent transportation system is a timed, accurate and efficient comprehensive transportation system which comprehensively applies advanced information technology, computer technology, data communication technology, sensor technology, electron control technique, automatic control theory, operation research and artificial intelligence to transportation, service control and vehicle manufacture for strengthening the relationship among vehicles, roads and users. The detection recognition and trace of moving objects have become a focus of machine vision research. Moving objects are common in daily life such as moving animals, driving cars and many other objects in nature. Acquiring and detecting the moving objects of sequence images is widely applied to traffic surveillance, banking service, homing guidance of aviation and military aircraft. So the study on detection, recognition and trace of multiple moving objects in sequential images has realistic significance [1].

The subject uses video image sequence to detect the flow of bus passengers. After the study on the existing data and the analysis on the whole system, the subject is divided into three parts, detection of moving objects, recognition of moving objects and trace of moving objects in which the recognition and trace of moving objects is difficult and important [2].

Recognition of moving objects is to identify the detected moving objects in moving regions. It is not only an important part of computer intelligent identification, but also is an important step of machines releasing people from heavy labor. It is mainly used for industrial detection and human-computer interaction now. The study on the recognition of human motion is popular at present, for example, the computer is used to identify the gestures of deaf-mute for achieving human-computer interaction, and the recognition and trace of industrial detection. The motion recognition in the subject is to identify how many people there are in the motion region. Some foreign scholars have made studies and proposed that the number of recognition motion is simple, but it is difficult to identify and segment the mixed moving objects. There is no relevant study in China. The paper makes detailed introduction from the perspective of simple recognition and accurate recognition [3].

Target tracking is equivalent to creation of continuous image frames based on the corresponding matching problems based on position, speed, shape, texture and colors. In the subject, the primary objective of tracing is to know the direction of motion and to count well. In the process of the study, there are two steps, prediction and tracing. Kalman filter is firstly used to predict, and then cost function is used for tracing. Lastly, count is based on objective chain.

According to the traffic status and the existing problems in large cities, in order to catch up with the world advanced level of transportation, it is necessary to combine with the actual condition of urban transportation in our country to make systematic and deep study on systematic theory model and implementation methods of intelligent identification and intelligent coordination scheduling of urban bus events, which not only has great scientific value and practical significance for solving the existing problems of urban transportation in our country, but also has wide application prospect [4].

The objective of moving object detection is to extract the region of variation from background image and sequential image. Effective segmentation of moving region is very important for target classification, tracing and identification, the reason for which is that the treatment process only considers the pixels of images corresponding to moving regions. The dynamic change of background image such as the influence of weather,

doi:10.4304/jmm.8.3.262-269

illumination, reflection and disturbance makes motion detection difficult. For moving object detection, the paper considers the problem of time complexity, and analyzes the influence of illumination, which achieves good effect.

According to the traffic condition and the existing problems in large cities in our country, in order to catch up with and surpass the advanced world level of traffic transportation, it is necessary to combine the urban actual traffic conditions in our country and makes systematic and deep study on urban bus events intelligence identification and intelligent coordination scheduling system theory model, which not only has great scientific value and realistic significance for solving the existing urban traffic problems in our country, but also has wide application prospect.

II. Extraction of Moving Objects

A. DSP Vision System Based on TMS320DM642

Traffic flow detection system includes image input device, image output device and software system. The subject uses CCD as image input device. In order to adapt to the operation of traffic flow detection system in the outdoor settings, TMS320DM642 with small volume, low power consumption and high reliability is selected as the core processor and is equipped with peripherals including image acquisition, power module, camera and LCD screen.

Real-time video signal processing algorithm is very complicated, has large calculation, and has high requirements on computing power of processor, and CPU of the common computers can't satisfy the demands. Therefore, it needs special digital signal processing chips to complete real-time process of multimedia signal. TMS320DM642 launched by Texas Instruments Incorporated applied the second generation VLIW framework, which is applicable for digital imaging processing. The basic frequency of TMS320DM642 can be configured as 600M, the execution speed of the highest instruction is up to 480MIPS, and it has high-speed and flexible controlling operation and matrix operation processing capacity. It has 64 general purpose registers with 32 bits and 8 independent arithmetic elements in which there are two multiplication units with 32 bits and six arithmetic logic units [5]. 8 arithmetic elements not only can run some new instructions which speed up the process on video and images, but also use parallel execution structure. In addition, TMS320DM64 has special on-chip peripherals and on-chip memory.

Three video ports of TMS320DM462 can be configured flexibly. They not only can be connected with common video CODEC device, but also support various resolution ratio and video standards such as CCIR601, ITU-BT.656 and BT1120. When three video ports are configured to be video acquisition and video display mode, every video port has two channels, Channel A and Channel B. There are 5120-byte acquisitions displaying cache. TMS320DM642 has complete development tools including C language compiler and assembly optimizer [6].

B. Motion Detection Methods

The objective of motion detection is to use sequential images processing technology to extract the moving foreground region from the background for reducing the calculation of searching body. As the processing objects of the following body recognition and tracing only considers motion region, segmentation quality of motion region has direct influence on the following processing effect. So motion detection is very important for body detection and tracing. However, dynamic change of background (the influence of weather, illumination and moving object) makes motion detection difficult. The extraction methods of motion objects are different, based on which the motion detection can be divided into inter-frame difference method, background subtraction and optical flow method.

As the system has certain requirements on timeliness, it is necessary to choose appropriate algorithm to satisfy the requirement of timeliness under the premise of guaranteeing accuracy. After comprehensively considering various conditions, we choose inter-frame difference method to extract the needed motion objects.

C. Image Difference

Video background and camera which were studied in the subject are relatively static. If we are in the k position, the camera continuously captures two frames, f1 and f2. Because the camera is relatively static and can be considered to be static in a very short time interval, the background should be identical in the field of view. Only motion objects move during the interval T. And we can make difference operation on adjacent two frames, discard the part which keeps invariant of images and retains the part which keeps changing, which can effectively retain waiting area of motion objects and eliminate most background region [7].

We defines difference algorithm as follows. If f1(x,y) is the pixel set of f1 image and f2(x,y) is the pixel set of f2, the difference image R(x,y) is defined as:

$$R(x,y)=\begin{cases}1 & \left|f_2(x,y)-f_1(x,y)\right|>\varepsilon \\ 0 & \left|f_2(x,y)-f_1(x,y)\right|<\varepsilon\end{cases} \quad (1)$$

After difference of the adjacent two frames, the pixel point of the achieved difference image which is greater than the given threshold is changed into 1, and the pixel point less than the threshold is changed into 0, which realizes binaryzation of difference images.

The common threshold methods include minimal error method, iteration method, Otsu method and best entropy method. These methods and application are analyzed and introduced in the paper.

1) Minimal error method

For minimal error method, gray level histogram of images is seen as probability density function of portfolio objectives and background which are distributed normally, and it is common that object and background are supposed to be distributed normally. So histogram is approximate to formula (2).

$$p(g)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-[(g-\mu_1)^2/2\sigma_1^2]}+\frac{1}{\sigma_2\sqrt{2\pi}}e^{-[(g-\mu_2)^2/2\sigma_2^2]} \quad (2)$$

In formula (2), $\sigma_1,\sigma_2,\mu_1,\mu_2$ mean standard deviation and mean value of two components. Two sides of the equation take logarithm, and we can get the following quadratic equation.

$$\frac{(g-\mu_1)^2}{\sigma_1^2}+\ln\sigma_1-2\ln p_1=\frac{(g-\mu_2)^2}{\sigma_2^2}+\ln\sigma_2-2\ln p_2 \quad (3)$$

We can find out the optimal threshold value by solving the equation. But the parameters of mixed distribution, σ,μ,p not only relate to the image to be segmented, but also are commonly unknown, which makes solution difficult. Kittler & Illingworth (1986) proposed a simple method of finding the best threshold value such as formula (4).

$$\begin{aligned}J(t)&=1+2p_1(t)\ln\sigma_1(t)+p_2(t)\ln\sigma_2(t)\\&-2p_1(t)\ln p_1(t)+p_2(t)\ln p_2(t)\end{aligned} \quad (4)$$

In formula (4),

$$p_1(t)=\sum_{g=0}^{t}h(g),\ p_2(t)=\sum_{g=t+1}^{l-1}h(g)$$

$$\mu_1(t)=\frac{\sum_{g=0}^{t}g\bullet h(g)}{p_1(t)},\ \mu_2(t)=\frac{\sum_{g=t+1}^{l-1}g\bullet h(g)}{p_2(t)}$$

$$\sigma_1^2=\frac{\sum_{g=0}^{t}h(g)(g-\mu_1(t))^2}{p_1(t)},\ \sigma_2^2=\frac{\sum_{g=t+1}^{l-1}h(g)(g-\mu_2(t))^2}{p_2(t)}$$

Function $J(t)$ is a criterion function which can describe the average correct classification performance of the whole image. For a known threshold value t, criterion function indirectly reflects overlapping amount of Gaussian distribution model with objects and background. The gray value which makes criterion function $J(t)$ a minimum is the threshold value with minimal error and is the best threshold value of image segmentation.

2) Iteration method

In 1978, Ridler and Calvard proposed iteration method of selecting threshold value, but it was time-consuming. Trussel simplified it, in which the histogram is randomly divided into two parts, the average grey value of every part is calculated, and the mean value of two mean gray levels is used as new segmentation threshold. Concrete steps are as follows:

The maximum and minimum gray value of images, Z_1 and Z_K are calculated. Z_1 and Z_K can be determined by the histogram of images, and the initial value of threshold is as shown in formula (5).

$$T_K=(Z_1+Z_K)/2 \quad (5)$$

(1) According to the threshold, T_K segments the images into two parts, object and background, and the mean gray values of two parts are figured out, Z_0 and Z_1.

$$Z_0=\frac{\sum_{Z(i,j)<T^K}Z(i,j)*\mathrm{N}(i,j)}{\sum_{Z(i,j)<T^K}\mathrm{N}(i,j)} \quad (6)$$

$$Z_1=\frac{\sum_{Z(i,j)>T^K}Z(i,j)*\mathrm{N}(i,j)}{\sum_{Z(i,j)>T^K}\mathrm{N}(i,j)} \quad (7)$$

In formula (6) and (7), $Z(i,j)$ is the gray value of (i,j), and $N(i,j)$ is the weight coefficient of (i,j). And $N(i,j)=1$.

New threshold is figured out.

$$T^{K+1}=\frac{Z_0+Z_1}{2} \quad (8)$$

(3) If $T^K=T^{K+1}$, the calculation is over. Or $T^K=T^{K+1}$, and it is converted into 2.

3) OTSU

In 1980, Dajin Exhibition of Japan proposed OTSU, and the method is also called Dajin method. OTSU is binarization method of automatically selecting threshold based on discrimination function and principle of least square. For an image, a threshold can be divided into background and object, based on which different thresholds can be selected to achieve different classification separation performances, and classification variance reflects the performance of classification separation. Automatic threshold method of classification variance means the method that uses classification variance as the judgment basis and selects the threshold making intra-class variance maximal and inter-class variance minimal as the optimal threshold value [8].

Concrete procedures of OTSU are as follows:

(1) If the gray scope of image is $G=\{0,1,2,...,L-1\}$, the threshold t is selected to divide it into C_0 and C_1.

(2) The probability and mean value of C_0 and C_1 are:

$$\omega_0=\sum_{i=0}^{t}p_i,\ \omega_1=\sum_{i=t+1}^{L-1}p_i=1-\omega_0 \quad (9)$$

$$\mu_0=\sum_{i=0}^{t}ip_i,\ \omega_1=\sum_{i=t+1}^{L-1}ip_i/\omega_1 \quad (10)$$

(3) Variance of C_0 and C_1 are:

$$\delta_0^2=\sum_{i=0}^{t}(i-\mu_0)^2p_i/\omega_0,\ \delta_1^2=\sum_{i=t+1}^{L-1}(i-\mu_1)^2p_i/\omega_1 \quad (11)$$

(4) Pattern recognition theory is used to calculate inter-class variance δ_W^2, intra-class variance δ_B^2 and population variance δ_T^2. When δ_B^2/δ_T^2 is maximal, the classification performance is optimal, and the threshold is the best threshold.

4) Maximum entropy method

Entropy is a uniform measurement. When entropy measures uniformity, it can derive the maximum entropy method. If t is the threshold value, gray level distribution of object is formula (12).

$$p_0 / P_t, p_1 / P_t, ..., p_t / P_t \qquad (12)$$

In formula (12), $P_t = \sum_{i=0}^{t} p_i$ 。

Gray level distribution of background is as formula (13).

$$p_{t+1} / (1-P_t), ..., p_{L-1} / (1-P_t) \qquad (13)$$

Entropy of the object is:

$$E_1(t) = \sum_{i=0}^{t} \frac{p_i}{P_t} \ln \frac{p_i}{P_t} \qquad (14)$$

Entropy of the background is

$$E_2(t) = \sum_{i=t+1}^{L-1} \frac{p_i}{1-P_t} \ln \frac{p_i}{1-P_t} \qquad (15)$$

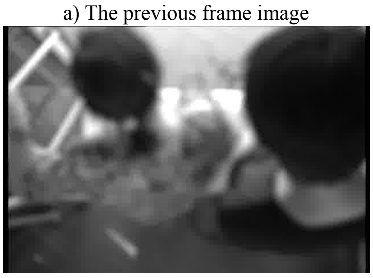
Entropy of the histogram is:

$$E(t) = E_1(t) + E_2(t) \qquad (16)$$

When $E(t)$ is maximal, it means the gray level distribution of background and object is identical. And t means the threshold value of segmenting two regions [9].

As the contrast of passengers and background is different in images, it is difficult to use a uniform threshold value to separate the passengers from the background. Different threshold values are used for segment according to local features of images. In practical process, we need to divide the image into a number of sub-regions for selecting threshold values according to concrete problems, or dynamically select the threshold of every point according to certain neighborhood region for image segmentation. The system uses self-adaptation of images to take the threshold value. Combined with the characteristics of images, the study comprehensively considers the advantages and disadvantages of various algorithms. And the processing effect of various algorithms is compared. The design selects threshold selection method of OTSU.

a) The previous frame image



b) The posterior frame image

c) Image after difference and binarization

Figure 1. Image frame difference method

Figure 1 is the images after binarization of filtering and difference. We can see that general outline of motion objects in sequential images can be extracted after difference and binarization of the previous frame image and the posterior frame image.

III. Target Identification

A. Identification of the Number of People

According to the photographic principle of camera, the biggest difference between one people and many people is the area. Therefore, making statistics on the area and finding area distribution rule of one people and many people can make further statistics on the number of people.

The focal length, photograph distance and shooting angle of cameral are different, the area of the acquired motion objects is different, and it is difficult to use empirical value or identical formula for calculation. So the paper uses statistical method. The paper makes statistics on the area of 50 people, and the statistical result shows that the mean value is 80000 and the variance is 2000. As the gesture sand directions of people change, the mean value of the area of two people is 150000, and the variance is 2500. By analogy, we can initially determine the relationship between the area s and the number of people.

The number of people is

$$\begin{cases} 1 & 40000 \le S \le 100000 \\ 2 & 130000 \le S \le 180000 \\ 3 & 200000 \le S \le 240000 \\ 4 & 250000 \le S \le 280000 \end{cases}$$

According to the above relationship, the number of people can be determined. If the number of people in the object is known, how to determine their concrete positions is the difficulty and emphasis of the subject. And the paper uses fuzzy clustering method to realize.

B. Determination of Center Point

After calculating the number of people included in each lump, we need to use a feature point to simulate a people, and the central point is the best feature point. But it is not easy to figure out the central point of every people. We use fuzzy clustering method of fuzzy mathematics in the paper. Firstly, the paper introduces membership function of fuzzy clustering method. And membership function can be used to express the membership relationship of sample x_k on subset X_i , as follows:

$$u_{X_i}(x_k) = u_{ik} = \begin{cases} 1 & x_k \in X_i \\ 0 & x_k \notin X_i \end{cases} \quad (17)$$

For fuzzy mathematics, the membership can't be fixed value, 0 and 1. It may be fuzzy values approximating to 1such as 0.99 and 0.98, and it might be fuzzy values approximating to 0 such as 0.01 and 0.02.

C.Fuzzy c Mean clustering algorithm

If $U = [u_{ik}]_{c \times n}$ is hard-dividing matrix, $p_i (i = 1, 2, ..., c)$ is the representative of the i kind or the central point. The objective function of hard clustering analysis is defined as:

$$J_1(U, P) = \sum_{i=1}^{c} \left(\sum_{x_k \in X_i} (d_{ik})^2 \right) \quad (18)$$

In formula (18), d_{ik} means the degree of distortion of sample x_k in the i kind and typical sample p_i in the i kind which can be measured by using the distance between two vectors. $J_1(U, P)$ is the error sum of squares of various samples and typical samples. $J_1(U, P)$ u_{ik} can be expressed as:

$$J_1(U, P) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} (d_{ik})^2 \quad (19)$$

Clustering criteria is seeking the best group (U,P)which makes $J_1(U, P)$ minimal under the condition of satisfying the requirements. The common method of solving the optimization problem is to use iteration method to solve the approximated value of $J_1(U, P)$.

Dunn promoted the objective function of hard clustering to fuzzy clustering. In order to avoid trivial solution and ensure the promotion is significant, Dunn used the membership square of the distance between each sample and various prototypes to weight, which expands intra-class error square and objective function into intra-class weighted squared error and objective function:

$$J_2(U, P) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^2 (d_{ik})^2 \quad (20)$$

Bezedek developed the objective function of Dunn as more general form, and gave general description based on function fuzzy clustering:

$$J_m(U, P) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2, \quad m \in [1, \infty) \quad (21)$$

In the formula, m is called weighted index or smoothing parameter. If the degree of membership is not weighted, the promotion form hard clustering objective function to fuzzy clustering objective function is effective.

In the above objective functions, the measurement distance between sample x_k and typical sample p_i of the i classification is defined as:

$$(d_{ik})^2 = \|x_k - p_i\| = (x_k - p_i)^T (x_k - p_i) \quad (22)$$

It is Euclidean distance.

Clustering criterion is the minimum of $J_m(U, P)$, $\min\{J_m(U, P)\}$. Each column in the matrix U is independent, so

$$\begin{aligned} \min\{J_m(U, P)\} &= \min\left\{ \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2 \right\} \\ &= \sum_{k=1}^{n} \min\{ \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2 \} \end{aligned} \quad (23)$$

The constraint condition of the above extreme value is the equation, $\sum_{i=1}^{c} u_{ik} = 1$, Lagrange multiplier method can be used for solution:

$$F = \sum_{i=1}^{c} (u_{ik})^2 (d_{ik})^2 + \lambda \left(\sum_{i=1}^{c} u_{ik} - 1 \right) \quad (24)$$

The necessary condition of the optimized order is $\frac{\partial F}{\partial \lambda} = \left(\sum_{i=1}^{c} u_{ik} - 1 \right) = 0$ and

$$\frac{\partial F}{\partial u_{jt}} = [m(u_{jt})^{m-1} (d_{jt})^2 - \lambda] = 0 \quad (25)$$

From formula (25), we can get

$$u_{jt} = \left[\frac{\lambda}{m(d_{jt})^2} \right]^{\frac{1}{m-1}} \quad (26)$$

Formula (26) is substituted into formula (25), and we can get

$$\begin{aligned} \sum_{l=1}^{c} u_{lt} &= \sum_{l=1}^{c} \left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} \left[\frac{1}{(d_{lt})^2} \right]^{\frac{1}{m-1}} \\ &= \left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} \left\{ \sum_{l=1}^{c} \left[\frac{1}{(d_{lt})^2} \right]^{\frac{1}{m-1}} \right\} = 1 \end{aligned} \quad (27)$$

So

$$\left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} = \frac{1}{\sum_{l=1}^{c} \left[\frac{1}{(d_{lt})^2} \right]^{\frac{1}{m-1}}} \quad (28)$$

Formula (28) is substituted into formula (26), and we can get

$$u_{jt} = \frac{1}{\sum_{i=1}^{c} \left[\frac{d_{jt}}{d_{lt}} \right]^{\frac{2}{m-1}}} \quad (29)$$

As d_{ik} may be 0, we define that when d_{ik} =0, $u_{ik} = 0 (i \neq k)$.

We can use the similar method to acquire the value of p_i when $J_m(U, P)$ is minimum. If $\frac{\partial}{\partial p_i} J_m(U, P) = 0$, we can get

$$\sum_{k=1}^{n}(u_{ik})^m \frac{\partial}{\partial p_i}\left[(x_k-p_i)^T(x_k-p_i)\right]=0$$

$$\sum_{k=1}^{n}(u_{ik})^m\left[-2(x_k-p_i)\right]=0$$

$$\sum_{k=1}^{n}(u_{ik})^m(x_k-p_i)=0$$

And we can get

$$p_i=\frac{1}{\sum_{k=1}^{n}(u_{ik})^m}\sum_{k=1}^{n}(u_{ik})^m x_k \tag{30}$$

If data set X, the number of clustering categories c and weight m are known, we can use formula (29) and (30) to determine the best fuzzy classification matrix and clustering center. And the optimization problem can be solved by using iterative algorithm.

Then we can get the concrete steps of fuzzy c mean value algorithm:

Initialization: The number of clustering categories is given, c, $2\le c\le n$, and n is the number of data. Iteration stopping threshold is set as ε. And the initialization clustering prototype mode is set to be $P^{(0)}$, and iteration counter is b=0.

Step 1: Formula (31) is used to calculate or update partition matrix $U^{(b)}$:

For $\forall i,k$, if $\exists d_{ik}^{(b)}>0$,

$$u_{ik}^{(b)}=\left\{\sum_{j=1}^{c}\left[\left(\frac{d_{ik}^{(b)}}{d_{jk}^{(b)}}\right)^{\frac{2}{m-1}}\right]\right\}^{-1} \tag{31}$$

If $\exists i,r$, $d_{ir}^{(b)}=0$,

$u_{ir}^{(b)}=1$, and $j\ne r, u_{ij}^{(b)}=0$

Step 2: Formula (32) is used to update Clustering prototype model matrix $P^{(b+1)}$:

$$P_i^{(b+1)}=\frac{\sum_{k=1}^{n}\left(u_{ik}^{(b+1)}\right)^m\bullet x_k}{\sum_{k=1}^{n}\left(u_{ik}^{(b+1)}\right)^m},\quad i=1,2,\ldots,c \tag{32}$$

Step 3: If $\left\|P^{(b)}-P^{(b+1)}\right\|<\varepsilon$, the algorithm stops and outputs partition matrix U and clustering prototype P. Or b=b+1, and turning to step 1.

IV. Target Numeration

A. Segmentation of Tracing Area

In order to make statistics on the number of people easy, tracing region needs to be segmented before counting multiple targets. As shown in Figure 2, target area is divided into three areas:

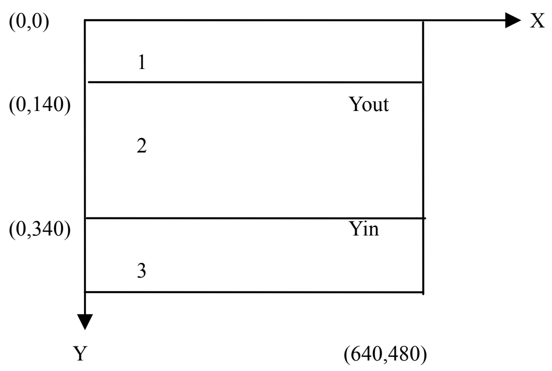


Figure 2. Division of target area

If someone gets on the bus and enters the area 1 firstly, he is traced from here to Area 2. When he crosses the yin line, counting begins. If someone gets off the bus, he should be traced continuously Area 2 and 3. And counting doesn't begin until he crosses Yout line.

As there is someone getting on the bus or getting off the bus at the front door and back door, the front door and the back door need to be tracked.

B. Improvement of Algorithm

In the process of people getting on and getting off the bus, the center of mass may shake. People get on or get off the bus slowly, which causes the center of mass move backward. Counting had begun from the first frame to the second frame in fact. It is because of the back of the center of mass that the second counting generates.

The solution is to use the second frame. When counting, not only the situation of the first frame needs to be observed, but also the situation of the second frame should be observed. If the second frame is less than Yout, the third frame is greater than Yout, and the fourth frame is less than Yout, we can't count. When the second frame is greater than Yin, the third frame is less than Yin, and the fourth frame is greater than Yin, we can't count, either.

In addition, people may stop suddenly because of resistance in the process of getting on the bus, which can't detect the breakdown of object-chain caused by motion target. The solution is as follows: we prepare two stacks, if the sudden stop makes that we can't find the center point and practical matching point after Kalman prediction, we need to press the point in a stack. And the points in the stack are popped up sequentially for matching until the next frame image comes. If there is some points not matching, the unmatched points need to be pressed in the other stack, which can ensure that the object-chain doesn't break.

C. Experimental Results

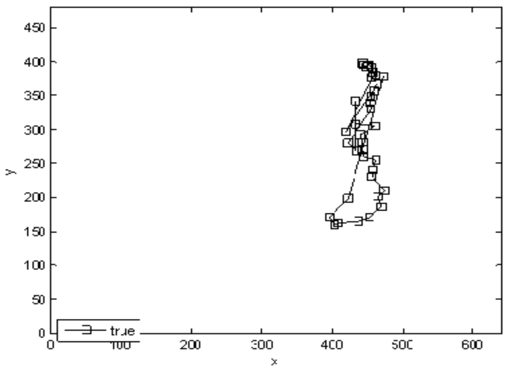
According to the above algorithm, we observe the follow-up passengers condition of a group of samples:

a)Original image

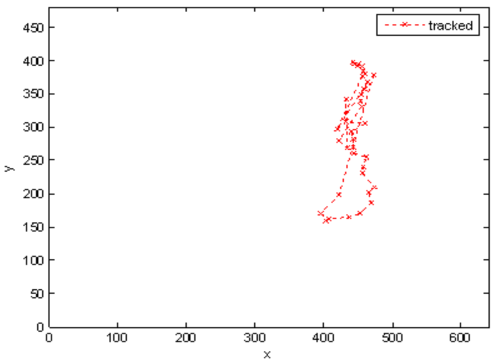
b) Center of mass image

Figure 3. Passenger tracking image

After achieving characteristic point of each motion object, the center point was for continuous centroid tracing. The tracing results are shown in the following figure.



c) Actual roadmap of center of mass



d) Prediction roadmap of center of mass

Figure 4. Moving object tracking

From the results, we can see that the prediction and tracing algorithm in the paper not only can trace the characteristic points of motion objects and judge if feature points cross the given counting line, but also can record the walking route of passengers getting on and getting off the bus, which can get ideal counting results. The algorithm is optimized to exclude some interference factors, which can make the smooth the tracing of passengers and statistics of the number of people smooth.

V. Summary

Passenger flow volume not only is an important factor of public transportation system, but also is important information. The foreign countries have made many studied on using image sequences for detecting passenger flow volume. Bu practical application system doesn't have high accuracy and has no strong adaptation to the environment. Many systems can't get the essence of the method because of intellectual property right protection of products. The study has great theoretical and practical significance.

The paper studies the algorithm based on passenger flow volume counting system of video. The algorithm can be divided into three parts. Firstly, frame difference method is used to detect motion targets. Then, fuzzy C mean clustering is used to get the center of mass of the object. Lastly, the body target is matched, traced and counted, which can get the number of people crossing counting line in real time. The experiment proves that the algorithm is easy and feasible, and the statistical data has higher accuracy.

The deep study for almost a year makes us get a systematic solution on passenger flow volume. The program has been used to realize the detection on passenger flow volume. 1. The measurement on body center of video sequential images is realized. It not only has high accuracy, but also can provide data for the following prediction and trace. 2. Detailed study on Kalman filter can realize predicting the position of the present center of mass in the next frame, which has credible accuracy. 3. Center of mass matching and counting function is realized, which has high accuracy for the video after experiment.

References

[1] Christian Sch'uldt and Ivan Laptev, Recognizing Human Actions: A Local SVM Approach, Pattern Recognition,"17th International Conferenceon (ICR'0 4), PP. 32-36, 2004, Volume 3.

[2] Aaron F.Bobick, James W.Davis, The Recognition of Human Movement Using Temporal Templates, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACH INE INTELLIGENCE, PP. 257-267, 2001, 23(3).

[3] N. Krahnst, over M. Yeasin, Towards a Unified Framework for Tracking and Analysis of Human Motion, Workshop on Detection and Recognition of Events in Video, PP. 1-8, 2001, 8.

[4] I. Haritaoglu, D. Harwood, L. Davis, W4: real-time surveillance of People and their activities. IEEE transactions on Patten Analysis and Machine Intelligence, PP. 809-830, vol. 22(8), Aug 2000.

[5] C. Stauffer, W. Grimson. Adaptive background mixture

models for real-time tracking. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, PP. 24-25, vol. 2, 1999.
[6] M, Kilger. A shadow handler in a video-based real-time traffic monitoring system. Proceedings of IEEE Workshop on Applications of Computer Vision, PP. 11-18, 1992, 30, Nov-2, Dec 1992.
[7] Chaba, Yogesh, Singh, Yudhvir, Aneja Preeti. Performance Analysis of Disable IP Broadcast Technique for Prevention of Flooding-Based DDoS Attack in MANET. Journal of Networks, pp. 45, 2009, 4(3).
[8] Shaosheng Fan, Hainan Wang. Multi-direction Fuzzy Morphology Algorithm for Image Edge Detection. Journal of Networks, pp. 56, 2011, 6(6).
[9] Li, Frederick W. B., Lau, Rynson W. H.. Emerging Technologies and Applications on Interactive Entertainments, Journal of Multimedia, pp. 47-81, 2011, 6(2).

Multimodal Medical Image Fusion Framework Based on Simplified PCNN in Nonsubsampled Contourlet Transform Domain

Nianyi Wang [1, 2]
1. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
2. School of Mathematics and Computer Science Institute, Northwest University for Nationalities, Lanzhou 730000, China,
Email: xbmusailor@126.com

Yide Ma*
School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
*Corresponding Author, Email: maydmayd@126.com

Kun Zhan, and Min Yuan
School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China,
Email: 105966809@QQ.com, minyuanlz@126.com

Abstract—**In this paper, we present a new medical image fusion algorithm based on nonsubsampled contourlet transform (NSCT) and spiking cortical model (SCM). The flexible multi-resolution, anisotropy, and directional expansion characteristics of NSCT are associated with global coupling and pulse synchronization features of SCM. Considering the human visual system characteristics, two different fusion rules are used to fuse the low and high frequency sub-bands respectively. Firstly, maximum selection rule (MSR) is used to fuse low frequency coefficients. Secondly, spatial frequency (SF) is applied to motivate SCM network rather than using coefficients value directly, and then the time matrix of SCM is set as criteria to select coefficients of high frequency subband. The effectiveness of the proposed algorithm is achieved by the comparison with existing fusion methods.**

Index Terms—**medical image fusion, spiking cortical model (SCM), nonsubsampled contourlet transform (NSCT), multimodal image fusion, Simplified PCNN**

I. INTRODUCTION

Medical imaging has become increasingly important in medical diagnosis, which enabled radiologists to quickly acquire images of human body and its internal structures with effective resolution. Different medical imaging techniques such as X-rays, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) provide different perspectives on human body. For example, CT scans provide dense structures like bones and implants with less distortion, MRI scans provide normal and pathological soft tissue within the body while PET scans provide better information on blood flow and flood activity with low spatial resolution. Therefore, an improved understanding of a patient's condition can be achieved through the use of different imaging modalities. A powerful technique used in medical imaging analysis is medical image fusion, where streams of information from medical images of different modalities are combined into a single fused image. Medical image fusion plays important role in clinical applications such as image-guided surgery, image-guided radiotherapy, noninvasive diagnosis, and treatment planning [1], [2].

So far, many effective theories and methods for medical image fusion have been proposed, such as FSD pyramid [3], gradient pyramid [4], Laplacian pyramid [5] DWT pyramid [6], SIDWT pyramid [7], morphological pyramid [8], ratio pyramid [9], and contrast pyramid [10]. All the above methods share one characteristic: each method is efficient for specific types of images and each approach has its own limits. For example, contrast pyramid method loses too much information from the source images in the process of fusion; ratio pyramid method produces lots of false information that does not exist in the source images; and morphological pyramid method creates many bad edges [11].

In this paper, we propose a new method of medical image fusion using nonsubsampled contourlet transform (NSCT) and spiking cortical model (SCM).

During last decade, medical image fusion algorithms that based on multiscale decomposition (MSD) become important methods [12]. NSCT [13] is one of popular MSD methods. It is proposed by Cunha, Zhou, and M.N.Do, and has been successfully used in image fusion fields and achieved satisfactory fusion effects.

Pulse coupled neural network (PCNN) is a visual

doi:10.4304/jmm.8.3.270-276

cortex-inspired network characterized by global coupling and pulse synchronization of neurons [14], and has been widely applied in intelligent computing. SCM is one of the simplified PCNN models that is mainly derived from Eckhorn's model and deduced from primate visual cortex, and also has been proved an effective image processing tool [15].

In recent years, researchers proposed several image fusion algorithms based on transform domain and PCNN. In Literature [16], a fusion algorithm based on Discrete Ripplet Transform (DRT) and Intersecting Cortical Model (ICM) for multimodal medical image is proposed. As another simplified PCNN model, ICM was used to obtain the fusion coefficients. Literature [17] presents a multi-source image fusion scheme based on lifting stationary wavelet transform (LSWT) and a novel dual-channel PCNN. Literature [18] discussed image fusion based on Shearlets and PCNN. In Literature [19], a Contourlet hidden Markov Tree (CHMT) and clarity-saliency driven PCNN based fusion approach is proposed for remote sensing images fusion. PCNN was first used in contourlet domain for visible and infrared image fusion in literature [20]. Qu, X. et al. proposed an image fusion algorithm based on spatial frequency (SF) motivated PCNN in NSCT domain [21]. The image fusion technique proposed by Xin, G. et al. based on dual-layer PCNN model with a negative feedback control mechanism in the NSCT domain has shown promising results in multifocus image fusion [22]. Literature [23] discussed fusion methods based on PCNN and NSCT in multimodal medical image fusion field.

However, in most of these PCNN and NSCT based algorithms, the value of single pixel in spatial or MSD domain is directly used to motivate one neuron. In fact, human's visual system in most time is sensitive to edges, directional features, etc. So, using single pixel in MSD domain purely is not enough. It is necessary to use spatial frequency, which stands for gradient energy in NSCT domain, to motivate SCM neurons [15], [21], [23].

The main purpose of this paper is to find an efficient image fusion algorithm for medical images of different modalities, based on the shift-invariance, multi-scale and multi-directional properties of NSCT along with human visual characteristics of SCM.

II. Spiking Cortical Model (SCM) and Nonsubsampled Contourlet Transform (NSCT)

A. Spiking Cortical Model

Like traditional PCNN models, each neuron in SCM consists of three parts: receptive field, modulation field, and pulse generator. For SCM, its two features make it more suitable for image processing. For one thing, SCM has been proved accords with Weber–Fechner law, since it has high sensitivity for low intensities of stimulus but low sensitivity for high intensities, and Weber–Fechner law is a logarithmic rule relating the level of subjective sense of intensity to the physical intensity of a stimulus. For another, time matrix of SCM can be recognized as human subjective sense of stimulus intensity, literature [15] discussed SCM's features and its application in image processing in detail [15]. The SCM neuron model is shown in Figure.1.

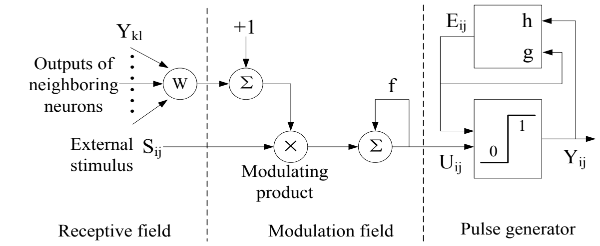


Figure 1. SCM model

B. Nonsubsampled Contourlet Transform

Figure 2 shows the decomposition framework of the NSCT. NSCT uses nonsubsampled pyramid filter bank (NSPFB) and nonsubsampled DFB (NSDFB) in its decomposition framework. The NSPFB is achieved by using two-channel nonsubsampled 2-D filter banks. The NSDFB is achieved by switching off downsamplers and upsamplers in each two-channel filter bank in DFB tree structure and upsampling filters accordingly [13], [24]. NSCT has the properties of shift-invariance which benefits designing fusion rules. The common NSCT-based image fusion approach consists of the following steps: Firstly, perform NSCT on source images to obtain lowpass subband coefficients and bandpass directional subband coefficients at each scale and each direction, NSPFB is used to complete multiscale decomposition and NSDFB is used to complete multi-direction decomposition. Secondly, apply some fusion rules to select NSCT coefficients of the fused image. Finally, employ inverse NSCT to the selected coefficients and obtain the fused image.

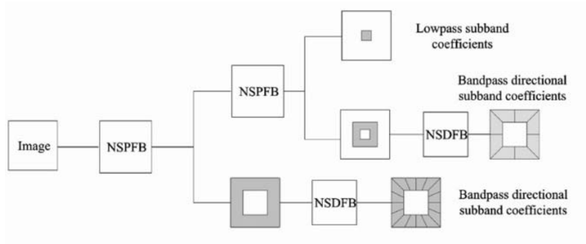


Figure 2. Decomposition framework of nonsubsampled contourlet transform. NSCT uses nonsubsampled pyramid filter bank (NSPFB) and nonsubsampled DFB (NSDFB) in its decomposition framework

III. Proposed Fusion Methods

A. Fusion Scheme

The notations used in this section are as follows: A, B, R represent two source images and final fused image, respectively. C= (A, B, R). LFS^{C} indicates the low frequency subband (LFS) of image C. $HFS^{C}_{g,h}$ indicates the high frequency subband (HFS) of image C at scale *g* and direction *h*. *(i, j)* denotes spatial location, thus LFS^{C} *(i, j)*, $HFS^{C}_{g,h}$ *(i, j)* denote coefficients located at *(i, j)* of low frequency and high frequency subband, respectively.

We use maximum selection rule (MSR) [6] to select low frequency coefficients of LFS^R from LFS^A and LFS^B as follow:

$$\mathrm{LFS}^{\mathrm{R}}(i,j)=\begin{cases}\mathrm{LFS}^{\mathrm{A}}(i,j), & \mathrm{LFS}^{\mathrm{A}}(i,j)\geq \mathrm{LFS}^{\mathrm{B}}(i,j)\\ \mathrm{LFS}^{\mathrm{B}}(i,j), & \mathrm{LFS}^{\mathrm{A}}(i,j)< \mathrm{LFS}^{\mathrm{B}}(i,j)\end{cases} \tag{1}$$

The coefficients of HFS of source images are selected by using SCM. As human vision system are sensitive to features such as edges, contours etc., so instead of using SCM directly, spatial frequency (SF) is considered as the gradient features of images to motivate SCM networks [25].

The SF is defined as

$$S_{i,j}^{g,h}=\sum_{i\in M,j\in N}(I_{i,j}^{g,h}-I_{i-1,j}^{g,h})^2+(I_{i,j}^{g,h}-I_{i,j-1}^{g,h})^2 \tag{2}$$

where $S_{i,j}^{g,h,}$ denotes the SF of the pixel that located at *(i,j)* on scale *g* and direction *h*, and $I_{i,j}^{g,h,}$ denotes the coefficients of the pixel that located at *(i,j)* on scale *g* and direction *h* respectively.

SF in each high frequency subbands are inputted to SCM to motivate neurons and generate pulse of neurons as follow:

$$U_{i,j}^{g,h}(\mathrm{n})=fU_{i,j}^{g,h}(\mathrm{n}-1)+S_{i,j}^{g,h}\sum_{k,l}W_{i,j,k,l}^{g,h}Y_{k,l}^{g,h}(n-1)+S_{i,j}^{g,h} \tag{3}$$

$$E_{i,j}^{g,h}(\mathrm{n})=gE_{i,j}^{g,h}(\mathrm{n}-1)+hY_{i,j}^{g,h}(n-1) \tag{4}$$

$$Y_{i,j}^{g,h}(n)=\begin{cases}1, & 1/\left(1+\exp\left(-\gamma U_{i,j}^{g,h}(\mathrm{n})-E_{i,j}^{g,h}(\mathrm{n})\right)\right)>0.5\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

$$T_{i,j}^{g,h}(\mathrm{n})=T_{i,j}^{g,h}(\mathrm{n}-1)+Y_{i,j}^{g,h}(n) \tag{6}$$

where SF $S_{i,j}^{g,h}$ is set as feeding input of SCM, $U_{i,j}^{g,h}(n)$ is internal activity, *n* denotes iteration times. $Y_{i,j}^{g,h}(n)$ is output, $E_{i,j}^{g,h}(n)$ is dynamic threshold, $W_{ij,kl}^{g,h}$ is synaptic weight matrix applied to the linking field, *f* and *g* are decay constants, and *h* is threshold magnitude coefficient. As a typical neuronal nonlinear transform function, the Sigmoid function [26] is applied in SCM to improve performance, which helps make output reachable. γ is a parameter of Sigmoid function. The nonlinearity of Sigmoid function can be used to generate pulse. Sigmoid curve has an "S" shape, with its slope increasing as γ increases. If $Y_{i,j}^{g,h}(n)$ is equal to *1*, it means the neuron will generate a pulse, or we can say one firing occurs. The sum of $Y_{i,j}^{g,h}$ in *n* iteration (namely the firing times) is defined as $T_{i,j}^{g,h}(n)$ to represent the image information. Rather than $Y_{i,j}^{g,h}(n)$, researchers often analyze $T_{i,j}^{g,h}(n)$, because neighboring coefficients with similar features representing similar firing times in a given iteration times. In this paper, we set firing times $T_{i,j}^{g,h}(n)$ as criteria to select coefficients of high frequency subbands.

B. Algorithm Step

1). Decompose source images A and B by NSCT to get low frequency and high frequency subbands coefficients of each image.

2). Select coefficients of LFS^R by using formula (1).

3). Calculate SF as described in formula (2) by using overlapping window on coefficients of HFS.

4). Input SF of each HFS into SCM and generate pulse of neurons with formula (3) ~ (5). And then compute the firing times $T_{i,j}^{g,h}(n)$ by formula (6).

5). Fuse coefficients of each HFS by the following rules:

$$\mathrm{HFS}_{g,h}^{R}(i,j)=\begin{cases}\mathrm{HFS}_{g,h}^{A}(i,j), & T_{i,j}^{g,h,A}(\mathrm{n})\geq T_{i,j}^{g,h,B}(\mathrm{n})\\ \mathrm{HFS}_{g,h}^{B}(i,j), & \text{otherwise}\end{cases} \tag{7}$$

6). Apply inverse NSCT on the fused LFS and HFS to get the final fused image.

It is important to perform statistical assessment of the quality of different fusion techniques along with the visual assessment due to the widespread use of multi-sensor and multi-spectral images in medical diagnosis. Therefore, image quality evaluation tools are required to compare the results achieved by different fusion techniques. In this paper, quantitative assessment of different image fusion algorithms are compared using following evaluation criteria, which have been proved to be effective to a great degree [2].

The notations used in this section are defined as : A and B are source images, R the final fused image, $m\times n$ the size of the image that has *L* grey levels; $f(i,j)$ denotes grey value of pixel (i,j). Then the above 4 indices are mathematically described as:

1. Mutual information (MI):

$$\mathrm{MI}=\frac{\sum_{i=0}^{L-1}\sum_{k=0}^{L-1}P_{A,R}(i,k)\log((P_{A,R}(i,k))/(P_A(i)P_R(k)))}{IE\_A+IE\_B}+\frac{\sum_{j=0}^{L-1}\sum_{k=0}^{L-1}P_{B,R}(j,k)\log((P_{B,R}(j,k))/(P_B(j)P_R(k)))}{IE\_A+IE\_B} \tag{8}$$

where *P(i)* indicates probability of pixels whose grey value amount to *i*; $P_{A,R}(i,k)$ and $P_{B,R}(j,k)$ are the normalized grey histogram between A and R and the normalized grey histogram between B and R, respectively. *IE_A* and *IE_B* denote the information entropy (IE) of image A and B. MI [27] can be used to measure amount of information transferred from source images to final fused image. Fusion performance would be better and better with MI increasing.

2. Standard deviation (SD):

$$SD=\sqrt{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}(f(i,j)-\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}f(i,j))^2} \tag{9}$$

SD indicates deviation degree between grey values of pixels and the average one of the fused image.

3. Energy of laplacian (EOL):

$$\begin{aligned}EOL=\sum_{i=2}^{m-1}\sum_{j=2}^{n-1}&(-f(i-1,j-1)-4f(i-1,j)-f(i-1,j+1)\\&-4f(i,j-1)+20f(i,j)-4f(i,j+1)\\&-f(i+1,j-1)-4f(i+1,j)-f(i+1,j+1))^2\end{aligned} \tag{10}$$

EOL is one of the useful indices to describe clarity of image.

4. $Q^{AB/F}$:

$$Q^{AB/F} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} (Q^{AF}(n,m)\omega^A(n,m) + Q^{BF}(n,m)\omega^B(n,m))}{\sum_{n=1}^{N} \sum_{m=1}^{M} (\omega^A(n,m) + \omega^B(n,m))} \tag{11}$$

$Q^{AB/F}$ [28] is proposed by C.S.Xydeas et al. as an objective image fusion performance measure. $Q^{AF}(n,m)$= $Q_g^{AF}(n,m)$ $Q_\alpha^{AF}(n,m)$. $Q_g^{AF}(n,m)$ and $Q_\alpha^{AF}(n,m)$ are the edge strength and orientation preservation values, respectively. $Q^{BF}(n,m)$ is similarly computed. $\omega^A(n,m)$ and $\omega^B(n,m)$ reflect the importance of $Q_g^{AF}(n,m)$ and $Q_\alpha^{AF}(n,m)$, respectively. The dynamic range of $Q^{AB/F}(n,m)$ is [0,1].

To evaluate the performance of the proposed image fusion approach, four different groups of human brain images are considered (see Figure. 3(a, b), Figure. 4(a, b), Figure. 5(a, b), Figure. 6(a, b)).The four groups of source images come from the website of the Atlas project of Harvard Medical School [29]. Figure.3 (a) and Figure.4 (a) are original CT image, Figure.3 (b) and Figure.4 (b) are original MRI image, Figure.5 (a) is original coronal FDG image, Figure.5 (b) is original coronal MR-T1 image, Figure.6 (a) is original MR-T1 image, Figure.6 (b) is original MR-T2 image. CT image shows structures of bone, while MR image shows areas of soft tissue. All images have 256-level gray scale. It can be seen that due to various imaging principle and environment, the source images with different modality contain complementary information.

For all these image groups, results of proposed fusion framework are compared with Averaging method, PCA method, discrete wavelet transform (DWT) with DBSS (2, 2), Laplacian pyramid (LP), morphological pyramid (MP). Parameters of these methods are set by: pyramid level = 4, selection rules: high-pass = select max, lowpass = average [11]. The visual comparison for fused images according to different fusion algorithms are shown in Figure.3 (c–h), Figure.4 (c–h), Figure.5 (c–h), Figure.6 (c–h).

IV. Evaluation and Analysis

From the fusion results displayed in Figure 3-6, it is clear that the Averaging and PCA algorithm give lose too many image details and provide poor fusion results compared with other four algorithms. This is because both of them have no scale selectivity. This limitation is modified in DWT, LP and MP. Morphological pyramid (MP) provides satisfactory fusion result, but it always brings fake image information and result in block effect. The remaining DWT, Laplacian pyramid (LP) and our proposed method achieved similar fusion effect. From subjective observation, we can see that the proposed algorithm is effective in multimodal medical image fusion. One of the reasons behind the performance of the proposed method is the human visual characteristics of SCM, which brings high contrast and more informative details to fused images.

Statistically, the objective performance evaluation and comparison among existing and proposed algorithms are depicted in Table 1. From experimental data we can see that the values of QABF of our proposed methods are the best in image group 1, group 2 and group 4, the values of MI of our proposed methods are the best in image group 2 and group 3. Though the other values of our method are not the best, it still got the second high values and outperformed the other 4 methods in each group. From table, we can easily conclude that the proposed algorithm can preserve high spatial resolution characteristics with less spectral distortion. The objective evaluation and comparison we discussed above verified that the proposed method is an effective fusion method for multimodal medical image fusion.

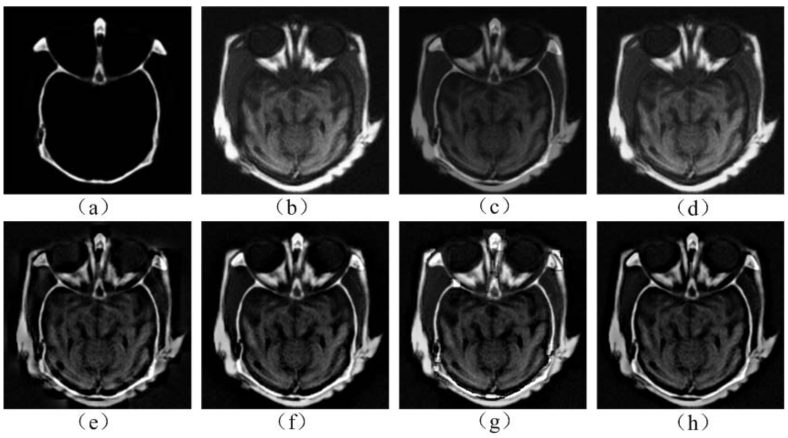


Figure.3. Fusion results of image group 1: (a) original CT image; (b) original MRI image; Fused image using (c) Averaging method, (d) PCA, (e) DWT, (f) Laplacian pyramid (LP), (g) morphological pyramid (MP), (h) our proposed method. All original images are from the website of the Atlas project of Harvard Medical School [29].

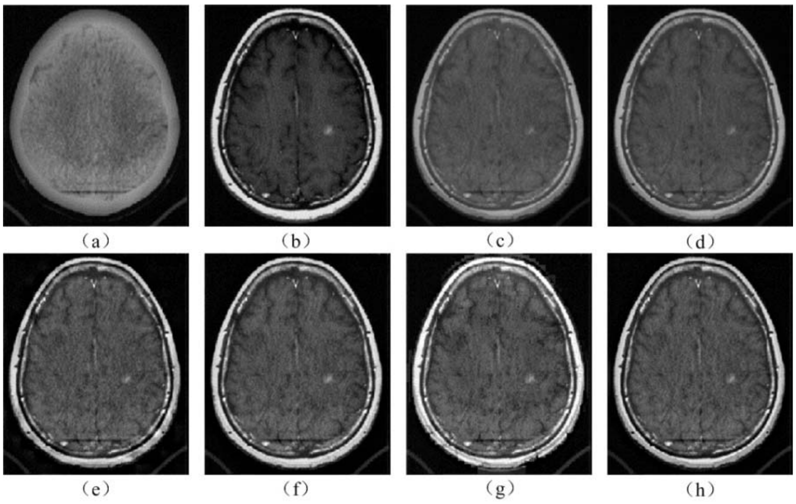


Figure.4. Fusion results of image group 3: (a) original CT image; (b) original MRI image; fused image using (c) Averaging, (d) PCA, (e) DWT, (f) LP, (g) MP, (h) our proposed method.

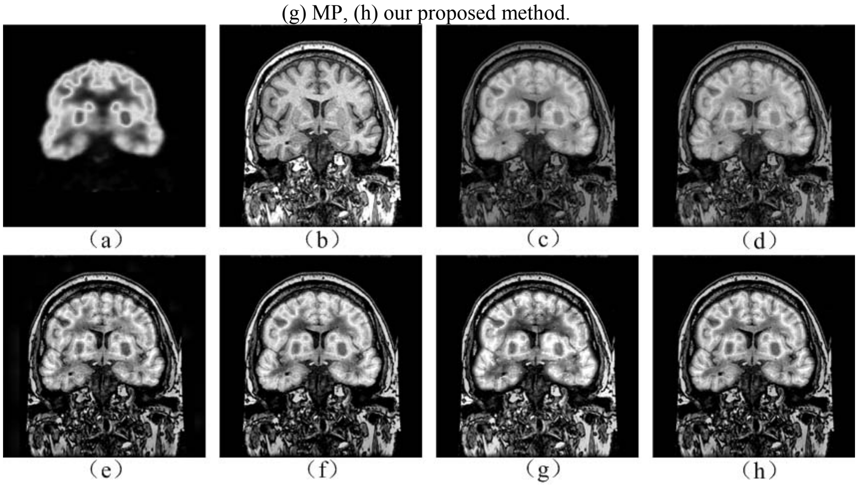


Figure.5. Fusion results of image group 2: (a) original coronal FDG image; (b) original coronal MR-T1 image; Fused image using (c) Averaging, (d) PCA, (e) DWT, (f) LP, (g) MP, (h) our proposed method.

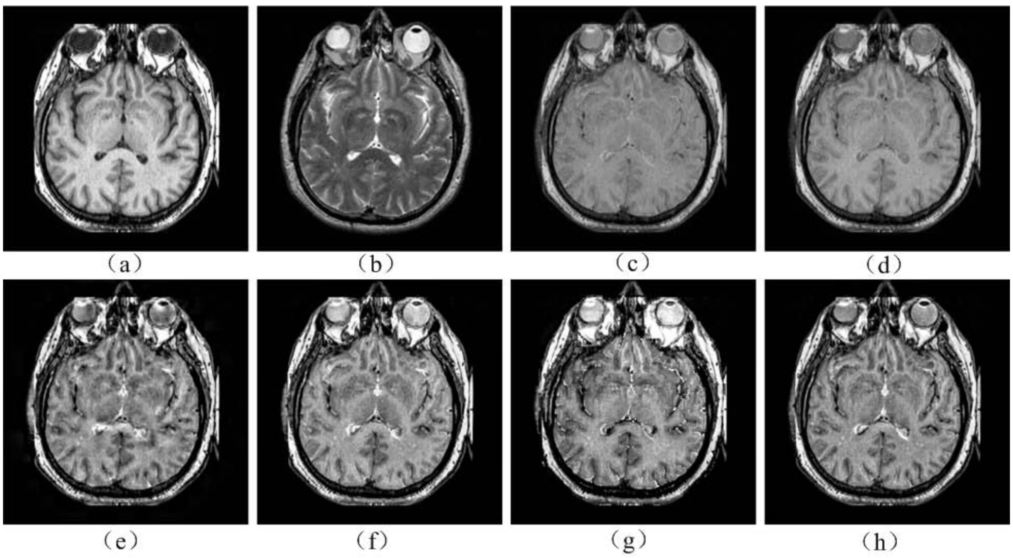


Figure.6. Fusion results of image group 4: (a) original MR-T1 image; (b) original MR-T2 image; Fused image using (c) Averaging method, (d) PCA, (e) DWT, (f) Laplacian pyramid (LP), (g) morphological pyramid (MP), (h) our proposed method.

Table.1. Comparison of six fusion method of three image groups

source Images	***evaluation indices***	***Averaging***	***PCA***	***DWT***	***LP***	***MP***	***Proposed***
image group 1	MI	3.5955	**4.2999**	1.3981	1.7041	2.2017	**3.8106**
	SD	7.8904	**8.4250**	7.9004	7.8479	7.9643	**8.0021**
	EOL	0.2306	0.3273	0.9520	**1.0039**	3.0433	**1.0028**
	$Q^{AB/F}$	0.4264	0.6549	0.6339	**0.7442**	0.7087	**0.7463**
image group 2	MI	3.1889	**3.1936**	1.4811	1.6714	1.9625	**3.2123**
	SD	9.3425	**9.6711**	9.4424	9.4346	9.6130	**9.6308**
	EOL	0.3805	0.5236	1.2423	1.2797	**1.4310**	**1.2955**
	$Q^{AB/F}$	0.4398	0.6486	0.6774	0.7020	**0.7051**	**0.7097**
image group 3	MI	2.4207	**2.6009**	1.9205	2.1461	2.3388	**2.6808**
	SD	**11.2629**	10.5719	10.4373	9.9959	9.4862	**10.6887**
	EOL	1.2571	1.5205	4.3872	**4.4333**	**5.5133**	4.0921
	$Q^{AB/F}$	0.3862	0.4805	0.5405	**0.5832**	0.5530	**0.5766**
image group 4	MI	2.6184	**2.8710**	2.2257	2.3604	2.3740	**2.6334**
	SD	**10.6775**	**10.6557**	10.4352	10.5081	10.2899	10.5290
	EOL	0.2330	0.2996	0.8110	0.9174	**1.2899**	**0.9977**
	$Q^{AB/F}$	0.3591	0.4359	0.4426	**0.4965**	0.4277	**0.5190**

V . CONCLUSIONS

In this paper, we proposed a new medical image fusion algorithm based on NSCT and SCM. The flexible multi-resolution, anisotropy, and directional expansion characteristics of NSCT are associated with global coupling and pulse synchronization features of SCM. Considering the human visual system characteristics, two different fusion rules are used to fuse the low and high frequency sub-bands respectively. Spatial frequency is applied to motivate SCM network rather than using coefficients value directly. The efficiency of the proposed algorithm is achieved by the comparison with existing fusion methods. The statistical comparisons prove the effectiveness of the proposed fusion method.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers and editors for their invaluable suggestions. This work was jointly supported by the National Natural Science Foundation of China (No. 61175012, 61162021, 61201422), Science and Technology Innovation Project of Ministry of Culture No [2011] 820, and Innovative Team Subsidize of Northwest University for Nationalities.

REFERENCES

[1] Wong, Alexander, and William Bishop. "Efficient least squares fusion of MRI and CT images using a phase congruency model." Pattern Recognition Letters, Vol. 29, No.3, pp. 173-180, 2008.

[2] Bhatnagar, Gaurav, Q. M. Jonathan Wu, and Zheng Liu. "Human visual system inspired multi-modal medical image fusion framework." Expert Systems with Applications, 2012.

[3] Anderson, C. H. U.S. Patent No. 4,718,104. Washington, DC: U.S. Patent and Trademark Office, 1998.

[4] Burt, Peter J. "A gradient pyramid basis for pattern-selective image fusion." Proceedings of the Society for Information Display, pp. 467-470, 1992.

[5] Burt, Peter, and Edward Adelson. "The Laplacian pyramid as a compact image code." IEEE Transactions on Communications, Vol.31, No.4, pp. 532-540, 1983.

[6] Li, Hui, B. S. Manjunath, and Sanjit K. Mitra. "Multisensor image fusion using the wavelet transform." Graphical models and image processing, Vol.57, No.3, pp. 235-245, 1995.

[7] Rockinger, Oliver. "Image sequence fusion using a shift-invariant wavelet transform." Image Processing, 1997. International Conference on Proceedings, Vol. 3, IEEE, 1997.

[8] Toet, A. "A morphological pyramidal image decomposition." Pattern Recognition Letters, Vol.9, No.4, pp. 255-261, 1989.

[9] Toet, Alexander. "Image fusion by a ratio of low-pass pyramid." Pattern Recognition Letters, Vol.9, No.4, pp. 245-253, 1989.

[10] Toet, Alexander, Lodewik J. Van Ruyven, and J. Mathee Valeton. "Merging thermal and visual images by a contrast pyramid." Optical Engineering, Vol.28, No.7, pp. 287789-287789, 1989.

[11] Wang Zhaobin, and Yide Ma. "Medical image fusion using< i> m</i>-PCNN." Information Fusion, Vol.9, No.2, pp. 176-185, 2008.

[12] Zhang, Zhong, and Rick S. Blum. "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application." Proceedings of the IEEE, Vol.87, No.8, pp. 1315-1326, 1999.

[13] Da Cunha, Arthur L., Jianping Zhou, and Minh N. Do. "The nonsubsampled contourlet transform: theory, design, and applications." IEEE Transactions on Image Processing, Vol.15, No.10, pp. 3089-3101, 2006.

[14] Johnson, John L., and Mary Lou Padgett. "PCNN models and applications." IEEE Transactions on Neural Networks, Vol.10, No.3, pp. 480-498, 1999.

[15] Zhan, Kun, Hongjuan Zhang, and Yide Ma. "New spiking cortical model for invariant texture retrieval and image processing." IEEE Transactions on Neural Networks, Vol.20, No.12, pp. 1980-1986, 2009.

[16] Kavitha, C. T., C. Chellamuthu, and R. Rajesh. "Multimodal Medical Image Fusion Using Discrete Ripplet Transform and Intersecting Cortical Model." Procedia Engineering, Vol.38, pp. 1409-1414, 2012.

[17] Chai, Y., H. F. Li, and J. F. Qu. "Image fusion scheme

using a novel dual-channel PCNN in lifting stationary wavelet domain." Optics Communications, Vol.283, No.19, pp. 3591-3602, 2010.

[18] Qiguang, Miao, Shi Cheng, and Xu Pengfei. "A novel algorithm of image fusion based on shearlets and PCNN." Neurocomputing, 2012.

[19] Yang, Shuyuan, Min Wang, and Licheng Jiao. "Contourlet hidden Markov Tree and clarity-saliency driven PCNN based remote sensing images fusion." Applied Soft Computing, Vol.12, No.1, pp. 228-237, 2012.

[20] Liu, Shengpeng, and Yong Fang. "Infrared image fusion algorithm based on contourlet transform and improved pulse coupled neural network." JOURNAL OF INFRARED AND MILLIMETER WAVES-CHINESE EDITION-, Vol. 26, No.3, pp. 217, 2007.

[21] Qu, Xiao-Bo, et al. "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain." Acta Automatica Sinica. Vol.34, No.12, pp. 1508-1514, 2008.

[22] Xin, Guojiang, et al. "Multi-focus image fusion based on the nonsubsampled contourlet transform and dual-layer PCNN model." Information Technology Journal, Vol.10, No.6, pp. 1138-1149, 2011.

[23] Das, Sudeb, and Malay Kumar Kundu. "NSCT-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency." Medical & biological engineering & computing. Vol.50, No.10, pp. 1105-1114, 2012.

[24] Do, Minh N., and Martin Vetterli. "The contourlet transform: an efficient directional multiresolution image representation." IEEE Transactions on Image Processing, Vol.14, No.12, pp. 2091-2106, 2005.

[25] Eskicioglu, Ahmet M., and Paul S. Fisher. "Image quality measures and their performance." IEEE Transactions on Communications, Vol.43, No.12, pp. 2959-2965, 1995.

[26] Forgac, R., and I. Mokris. "Feature generation improving by optimized PCNN." Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on. IEEE, 2008.

[27] Qu, Guihong, Dali Zhang, and Pingfan Yan. "Information measure for performance of image fusion." Electronics letters. Vol.38, No.7, pp. 313-315, 2002.

[28] Xydeas, C. S., and V. Petrovic. "Objective image fusion performance measure." Electronics Letters. Vol.36, No.4, pp. 308-309, 2000.

[29] http://www.med.harvard.edu/AANLIB

Nianyi Wang received the B.S. degree in computer science from Lanzhou University, Gansu, China in 2002, and received the M.S. degree in computer application technology from Lanzhou University in 2007. He is currently a lecturer in the School of Mathematics and Computer Science Institute, Northwest University for Nationalities. He is currently pursuing the Ph.D. degree in radio physics at Lanzhou University. His current research interests include artificial neural networks, image processing, and pattern recognition.

Yide Ma received the B.S. and M.S. degrees in radio technology from Chengdu University of Engineering Science and Technology, Sichuan, China, in 1984 and 1988, respectively. He received the Ph.D. degree from the Department of Life Science, Lanzhou University, Gansu, China, in 2001. He is currently a Professor in the School of Information Science and Engineering, Lanzhou University. He has published more than 50 papers in major journals and international conferences and several textbooks, including Principle and Application of Pulse Coupled Neural Network (Beijing: Science Press, 2006), and Principle and Application of Microcomputer (Beijing: Science Press, 2006). His current research interests include artificial neural networks, digital image processing, pattern recognition, digital signal processing, and computer vision.

Kun Zhan received the B.S. degree in electronic information science and technology from Lanzhou University, China, in 2005. He received the Ph.D. at the same university in 2010. Currently, He works in Lanzhou University. His main research interests are image processing and neural computation.

Min Yuan received the B.S. degree in electronic information science and technology from Lanzhou University, China in 2002, and received the M.S. degree in Communication and information system from Lanzhou University in 2005. She is currently a lecturer in the School of Information Science and Engineering, Lanzhou University. She is currently pursuing the Ph.D. degree in radio physics at Lanzhou University. Her research interests include compressed sensing, magnetic resonance imaging, and sparse signal and image processing, wavelets and other multi-scale transforms.

Accurate Image Retrieval Algorithm Based on Color and Texture Feature

Chunlai Yan
School of Mathematics and Computer, Panzhihua University, Panzhihua, 617000, China
E-mail: wingvestige@163.com

***Abstract*—Content-Based Image Retrieval (CBIR) is one of the most active hot spots in the current research field of multimedia retrieval. According to the description and extraction of visual content (feature) of the image, CBIR aims to find images that contain specified content (feature) in the image database. In this paper, several key technologies of CBIR, e. g. the extraction of the color and texture features of the image, as well as the similarity measures are investigated. On the basis of the theoretical research, an image retrieval system based on color and texture features is designed. In this system, the Weighted Color Feature based on HSV space is adopted as a color feature vector, four features of the Co-occurrence Matrix, saying Energy, Entropy, Inertia Quadrature and Correlation, are used to construct texture vectors, and the Euclidean distance for similarity measure is employed as well. Experimental results show that this CBIR system is efficient in image retrieval.**

***Index Terms*—image retrieval, feature extraction, similarity measures, gray level co-occurrence matrix), color, texture**

I. INTRODUCTION

Along with the development of multimedia technology, digital image library appeared in the early nineteen-nineties. The technology of text-based image retrieval was unable to meet the retrieval of image content in multimedia information. In order to solve this problem, the technology of content-based image retrieval CBIR [1] (Content-Based Image Retrieval) emerged accordingly. Different from the existing system, CBIR is capable to find the image with specified characteristics or contents in target image collection according to the description of image contents. Color is an important attribute of an image. It is stable to a certain extent for it is not sensitive to size, direction and geometry. In many cases, color is the simplest and most effective feature to describe an image. In addition, texture is also an important visual feature of an image. Texture analysis is generally done by two methods: statistical texture analysis and structural texture analysis. In recent years, the researchers studied many effective methods about the image retrieval based on color and texture features during color and texture research. In the aspect of color feature extraction, color histogram was the most commonly-used method studied by Swain and Ballad, in which color space was divided into a plurality of fixed subspaces, then the pixels of each subspace was counted and the similarity between images was measured by histogram intersection. For this method, the advantage was that it was simple and effective and is not sensitive to the rotation and scaling of the image, but its disadvantage was that the space distribution information of color was ignored. In order to overcome the defects of the color histogram which cannot reflect space information, the researchers tried to study a retrieval method based on local color features, such as dividing the image into the appropriate blocks artificially and then extracting the local color features from each block. However, the said method had less rotation invariance and translation invariance. Hsu proposed to choose some representative color from the image and divide the image into many rectangular regions. In each region, a single color was used, and the similarity between two images was calculated by the overlapping extent of the similar color regions. In the aspect of texture feature extraction, the most common method was the co-occurrence matrix based on the second order gray statistical features, which was proposed by Haralick. It could reflect the texture feature of an image, but too much computation was involved. Based on the HVS's study about visual perception, Tamura summarized six texture features, such as Coarseness, Contrast, Directionality, Line-likeness, Regularity and Roughness. These features were often used in the image retrieval system. In the structural analysis methods, Carlucci proposed a texture model with such elements as line segment, open polygon and closed polygon, of which arrangement rule was defined by a graph grammar structure. Lu and Fu used the tree grammar structure to represent the texture. They segmented the texture into 9 * 9 windows and expressed the spacial structure of each decomposition unit as a tree. The structural analysis method was usually applied in regular textures. Because the single feature was one-sided in the image description, the retrieval algorithm based on color and texture emerged as the times require, e.g. retrieval algorithm based on local accumulative histogram and co-occurrence, proposed in Reference. This method was more consistent with the visual features of the character, and the retrieval result is more satisfactory.

This paper aims to study an image retrieval algorithm, which combines color feature and texture feature. In the aspect of the color feature extraction, the weighted color feature based on HSV is used as the color feature vector. By weighting the main color of different block, the weighted color feature based on main color is generated.

doi:10.4304/jmm.8.3.277-283

In the aspect of the texture feature extraction, four characteristics of the gray level co-occurrence matrix, saying energy, entropy, moment of inertia and relevance [2-4], are used as the components of the texture feature vector. At last, the author combines color feature and texture feature with the weighted method. In addition to separate color and texture retrieval, the algorithm also supports color and texture retrieval synthetically. Moreover, the experimental results are satisfactory.

II. Image Retrieval Algorithm

In this algorithm, the right color space is selected in accordance with people's visual features firstly, color features are extracted, and the weighted color feature based on HSV is used to calculate the color similarity distance of two images; then the texture feature is extracted by four features of the gray level co-occurrence matrix; finally color and texture features of the image are combined by Gauss normalization method to generate synthetic image retrieval feature, and an experiment is carried out to show the effectiveness of the algorithm. The operating steps are as follows:

A. Color Feature Extraction Algorithm

Color histogram is the most commonly used method in color feature, and HSV color space is relatively consistent with the characteristics of a character's visual perception. So, the HSV color space is adopted in this system to extract color feature, including two sections: the conversion of RGB space to HSV space and the quantization of HSV space.

Through the nonlinear transform, the R, G, B values of RGB color space are converted to the H, S and V values of HSV space. It is hypothesized: $v=\max(r,g,b)$, and $r^{'},g^{'},b^{'}$ are defined as follows:

$$r^{'}=\frac{v^{'}-r}{v^{'}-\min(r,g,b)},\quad g^{'}=\frac{v^{'}-g}{v^{'}-\min(r,g,b)},$$

$$b^{'}=\frac{v^{'}-b}{v^{'}-\min(r,g,b)} \tag{1}$$

Then,

$$h=60\times h^{'},\ s=\frac{v^{'}-\min(r,g,b)}{v^{'}},\ v=\frac{v^{'}}{255} \tag{2}$$

$$h^{'}=\begin{cases}5+b^{'}, & r=\max(r,g,b)\ and\ g=\min(r,g,b)\\ 1-g^{'} & r=\max(r,g,b)\ and\ g\neq\min(r,g,b)\\ 1+r^{'} & g=\max(r,g,b)\ and\ b=\min(r,g,b)\\ 3-b^{'} & g=\max(r,g,b)\ and\ b\neq\min(r,g,b)\\ 3+g^{'} & b=\max(r,g,b)\ and\ r\neq\min(r,g,b)\\ 5-r^{'} & others\end{cases} \tag{3}$$

In the formula (3), $r\in[0,255]$, $b\in[0,255]$, and $h\in[0,360], s\in[0,1], v\in[0,1]$.

After the above conversions, the color information of each pixel in the image can be represented by H, S and V. Through many times of color model analysis, three components of HSV can be quantified as unequal intervals according to a character's color perception. The hue H is divided into 16 parts, and saturation S and brightness V into 4 parts by a character's visual resolving power. Then, the formula about the quantization can be gotten as below:

$$S=\begin{cases}0 & s\in[0,0.15)\\ 1 & s\in[0.15,0.4)\\ 2 & s\in[0.4,0.75)\\ 3 & s\in[0.75,1)\end{cases},\quad V=\begin{cases}0 & v\in[0,0.15)\\ 1 & v\in[0.15,0.4)\\ 2 & v\in[0.4,0.75)\\ 3 & v\in[0.75,1)\end{cases},$$

$$H=\begin{cases}0 & h\in[345,360]\ or\ h\in[0,15)\\ 1 & h\in[15,25)\\ 2 & h\in[25,45)\\ 3 & h\in[45,55)\\ 4 & h\in[55,80)\\ 5 & h\in[80,108)\\ 6 & h\in[108,140)\\ 7 & h\in[140,165)\\ 8 & h\in[165,190)\\ 9 & h\in[190,220)\\ 10 & h\in[220,255)\\ 11 & h\in[255,275)\\ 12 & h\in[275,290)\\ 13 & h\in[290,316)\\ 14 & h\in[316,330)\\ 15 & h\in[330,345)\end{cases} \tag{4}$$

In the three vectors of HSV, the character's eye is mainly based on the hue H, which is followed by the saturation S, and the last is the brightness V when the color is classified. At the same time, a one-dimensional vector L can be obtained by the quantitative series and frequency bandwidth of HSV after the combination.

At the same time, L series H, S, V and the band width of the desirable combination can be quantified on the basis of the one-dimensional vector:

$$L=HQ_SQ_V+SQ_V+V \tag{5}$$

where Q_S and Q_V are quantified as series S and V, $Q_S=4$, $Q_V=4$, and the formula can be expressed as:

$$L=16H+4S+V \tag{6}$$

The algorithm of color feature extraction can be summarized as follow:

Step 1: Main color extraction

First, the image is divided into 9 * 9 blocks, and it is assumed that m blocks are gotten totally. After the quantitative synthesis of each pixel in the block, a color vector with eighty-one one-dimensional color components is obtained. Second, the color with the most pixels is extracted as the main color. At last, the integrated main color feature vector of the image is gotten as $L = (l_{1,p}, l_{2,p} ..., l_{m,p})$ ($L_{k,p}(k = 1, 2, ..., m)$ is the main color of No. k block).

Step 2: Get the weight of each block by the center-weighted method

When watching an image, we always clap eyes on the area near the center. So, a bigger weight is used for the central area and a smaller weight for the marginal area when the feature of the image is extracted. Moreover, the standard normal distribution function is used as the weight to extract the features (e.g.: $\phi(x) = e^{\frac{-x^2}{2}}$ $(x \ge 0)$).

In the algorithm, x is the distance between each block and the central block. Then, the comprehensive weighted main color feature vector is gotten from the image: $LL = (L_{1,p}, L_{2,p}, ...L_{m,p})$.

$L_{k,p} = \phi(x_k) \times l_{k,p} (k = 1, 2, ..., m)$ is the weighted main color feature of No. k block.

***Step 3*:** Similarity distance calculation

It is hypothesized that $LL' = (L'_{1,p}, L'_{2,p}, ..., L'_{m,p})$ and $LL'' = (L''_{1,p}, L''_{2,p}, ..., L''_{m,p})$ are the comprehensive weighted color feature vectors of two images, then the similarity distance between the key image and the retrieval image is calculated as follows:

$$D_1(A, B) = \sum_{i=1}^{m} \left| L'_{i,p} - L''_{i,p} \right| \tag{7}$$

B. Texture Feature Extraction Algorithm

Currently, the gray level co-occurrence matrix (GLCM) [5-7]is an important texture analysis method, which has been recognized by the world. In this paper, GLCM is used to extract the texture features, mainly including three parts of grayhound, GLCM and its feature calculation, and feature vector normalization.

Because the extraction method of image texture features based on GLCM is built on the image gray value. So, it is firstly needed to convert RGB space to YUV space, and the value Y is the gray value of the image. The formula for calculating the pixel gray value of the image is as follows:

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \tag{8}$$

It is hypothesized that the total number of pixels in rectangular image at X-axis direction is N_x and that at Y-axis direction is N_y. In order to reduce the amount of calculation, the image gray must be merged. It is also hypothesized that the highest gray level is N_g, $L_x = \{1, 2, ..., N_x\}$, $L_g = \{1, 2, ..., N_g\}$, and then it can be understood that the image, which will be texture-analyzed, is a mapping from space $L_x \times L_y$ to G. For each point in $L_x \times L_y$, there will be a gray level to match in G. The GLCM id defined as $[p(I, J, D\theta)]$ [8-9], of which the direction is θ and the interval is D. $[p(I, J, D\theta)]$ is the element at the row i and the column j of GLCM. It represents the frequency of occurrence, of which a pair of pixels' gray levels is i and j, and the interval is D in all direction of θ.

The value of GLCM can be calculated by the formula below:

$$p(i, j, d, \theta) = \frac{p(i, j, d, \theta)}{\sum_i \sum_j p(i, j, d, \theta)} \tag{9}$$

The symbiotic matrix $[p(I, J, D\theta)]$ reflects synthetical information of the image gray distribution about direction, magnitude of changes and local neighborhood, but it cannot provide characteristics to distinguish the texture directly but further extract texture features from the matrix. There are totally fourteen texture features summarized in GLCM at the present. Gotlieb and Kreyszig have proved by experiments that energy, entropy, moment of inertia and dependency are the most effective features. Normally, under normal circumstances, a better result can be achieved by a small value of d.

$$ASM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i, j)]^2 \tag{10}$$

This is a measure of image gray distribution. When the distribution of co-occurrence matrix elements is concentrated on the principal diagonal, the corresponding ASM value is large, or otherwise, the ASM value is small.

$$ENT = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i, j) \log p(i, j)] \tag{11}$$

When the co-occurrence matrix elements are distributed uniformly, the ENT value is large, and conversely, the ENT value is small.

$$CON = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [(i - j)^2 p(i, j)] \tag{12}$$

For a coarse texture, the gray level difference is small and the corresponding CON value is small, because the co-occurrence matrix elements are focused around the principal diagonal. Conversely, for a fine texture, the corresponding CON value is large.

$$COR = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_X)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y} \tag{13}$$

where, μ_x and σ_x are the mean and variance of the set $\{px(i) : i = 1, 2, ..., N_g\}$, and μ_y and σ_y are the expectation and variance of the set $\{py(j) : j = 1, 2, ..., N_g\}$. The correlation is used to describe the similarity about the row or column elements of the matrix.

After the calculation of the four texture features as above, a four dimensional texture feature vector can be constructed.

The feature vector normalization is designed so that each component of feature vectors has the same weight. Because different components of the feature vectors may have completely different physical meaning and their variations are also different, there will be a large deviation if the similarity distance is calculated directly without normalization. So, it is necessary to normalize each component of the feature vectors to a consistent range.

In this paper, Gauss normalization method [10-12] is adopted. And it is hypothesized that the feature vector is V and the component is v_i. Then, Gauss normalized formula is as below:

$$v_i = \frac{|v_i - \mu|}{3\sigma} \tag{14}$$

Where, μ and σ are the expectation and variance of each component v_i of the feature vector. After the normalization of feature vectors, the probability that a component in the vectors is between the interval 0 and 1 is 99%. For the value more than 1, it can simply correspond to 1. The advantage of Gauss normalization method is that, even though there are some very large or small values in the feature vectors, the importance of components v_i is not deviated when the similarity distance in vectors is calculated.

C. Similarity Measure Algorithm

Euclidean distance [13-14] is the simplest distance formula, which is widely applied in the content based image retrieval system. So, in this paper, the Euclidean distance is followed to measure the similarity of color and texture features. The formula is as follows:

$$D_2(A,B) = \sqrt{\sum_{i=1}^{N} (A_i - B_i)^2} \tag{15}$$

where, A_i and B_i are the components of the texture feature vectors in the key image and the retrieval image.

Because the relationship between the dimensions of the vector is not considered in the Euclidean distance and each dimension must be equally important, its application scope and effectiveness are affected to some extent. However, the Euclidean distance has better effect just applied separately in the similarity measure of the color histogram and texture feature.

D. Retrieval Algorithm based on Color and Texture

In this paper, the algorithm combines color similarity distance from retrieval based on color feature and texture similarity distance from retrieval based on texture feature linearly as a synthetical similarity distance by user's color and texture weight (color weight plus texture weight is equal to 1), and the distance calculation method is as follows [15-19]:

Synthetical similarity distance = Color similarity distance × Color weight + Texture similarity distance × Texture weight [15] (16)

The retrieval based on color and texture, is actually to get a single color and texture similarity distance by matching color and texture features, and then the synthetical similarity distance is calculated as the final similarity measurement result. Compared with the color retrieval or texture retrieval, synthetical retrieval is more effective and can be changed according to the different requirements of customers on the importance of color or texture. For the same key image, the retrieval result of a single color or texture often has greater difference, but the retrievals based on color and texture can be replenished mutually. If unsatisfied with the retrieval results for the first time, the users can adjust the weight of color and texture gradually to get a more satisfactory result.

E. The process of image retrieval

Step 1: Open the key image: the user selects a key image to be retrieved (this system only supports BMP images of 24 bits, and it will give a prompt if you try to open the images of other kinds).

After the key image is opened, the system will calculate and display the color features and the texture features (four features of GLCM: energy, entropy, moment of inertia and dependency) automatically.

Step 2: Open the image library: the user selects an image folder on the local computer as an image library for retrieval.

Step 3: Retrieval: After selecting the key image and the image library, the user can choose retrieval based on color, retrieval based on texture or retrieval based color and texture.

When choosing the retrieval based on color or texture, the system traverses each image in the image folder, calculates its color or texture characteristics, measures the similarity with the key image and finally calculates the similarity distance. When the system chooses the retrievals based on color and texture synthetically, the user firstly sets the color and texture weight; then the system traverses each image in the image folder, calculates its color and texture characteristics at the same time and measures the similarity with the key image; finally, it combines color similarity distance and texture similarity distance as a synthetical similarity distance according to the weight set by the user early.

Step 4: Display of results: after the retrieval, if there is similarity distance between the image retrieved and the key image, the paths of image files is listed out. The images are sorted from small to big according to the similarity distances. A smaller similarity distance suggests that the similar degree between the key image and the image retrieved is bigger. If the similarity distance is 0, it means that the key image is retrieved. The user can click on the item in the list to view and compare the weighted color feature based on HSV and texture feature of each image. At the same time, the system displays 10 images with the highest similarity and reports to the user how many images have been retrieved and whether the key image has been retrieved.

III. EXPERIMENTAL RESULTS

For the algorithm proposed in this paper, Visual C++ 6 is used as the development environment, and the MFC library is adopted to verify its effectiveness. In the experiment, four groups of images, such as natural scenery, cartoon dolls, buildings and flowers, are gotten from the network library, and each group includes 100 images. Then, three methods, such as the weighted color feature based on HSV, the gray level co-occurrence matrix and the comprehensive weighted color feature based on HSV and gray level co-occurrence matrix, are used to verify the effectiveness of the algorithm. figure 2 to figure 6 show that the retrieval results of comprehensive method is much better, and the similar images in the list are more ahead.

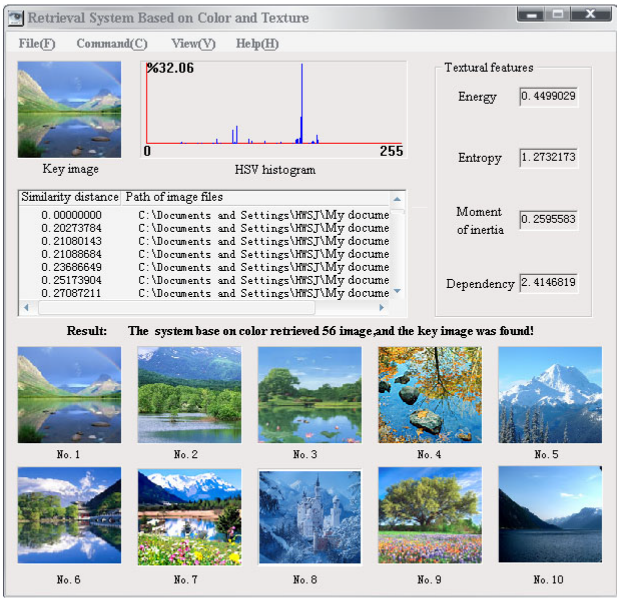


Figure 1. The result of retrieval algorithm based on color in natural scenery

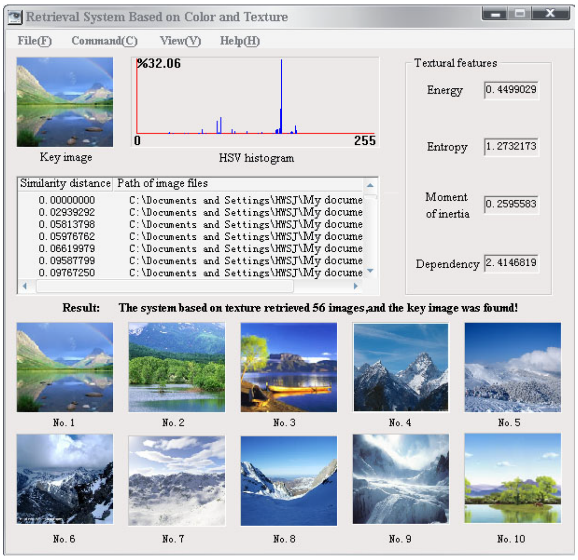


Figure 2. The result of retrieval algorithm based on texture in natural scenery

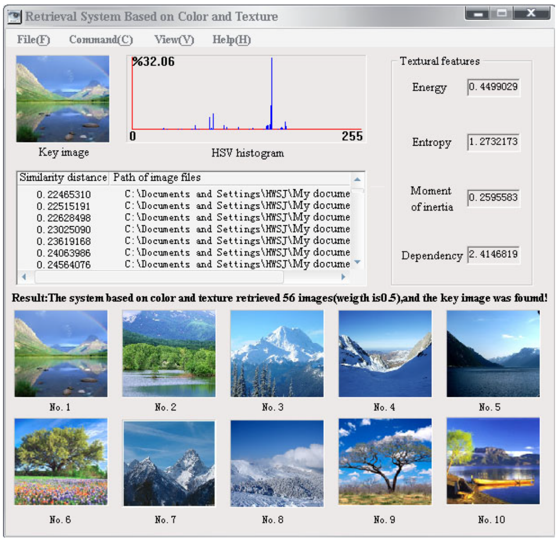


Figure 3. The result of retrieval algorithm based on color and texture in natural scenery

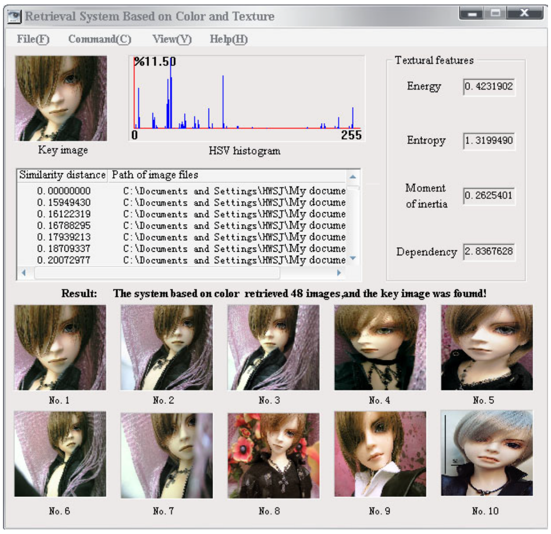


Figure 4. The result of retrieval algorithm based on color in cartoon dolls

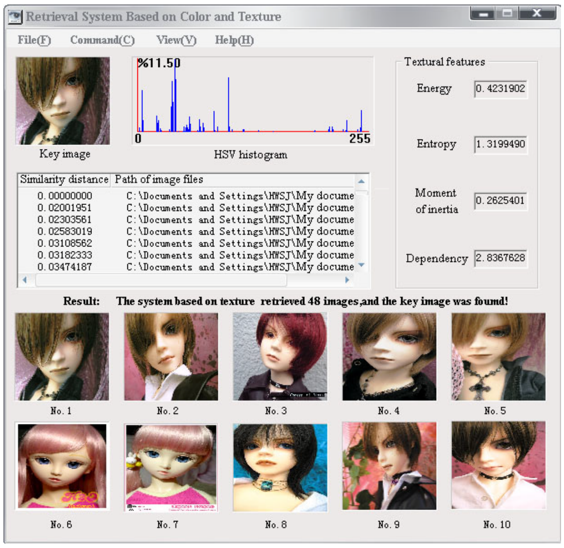


Figure 5. The result of retrieval algorithm based on color in cartoon dolls

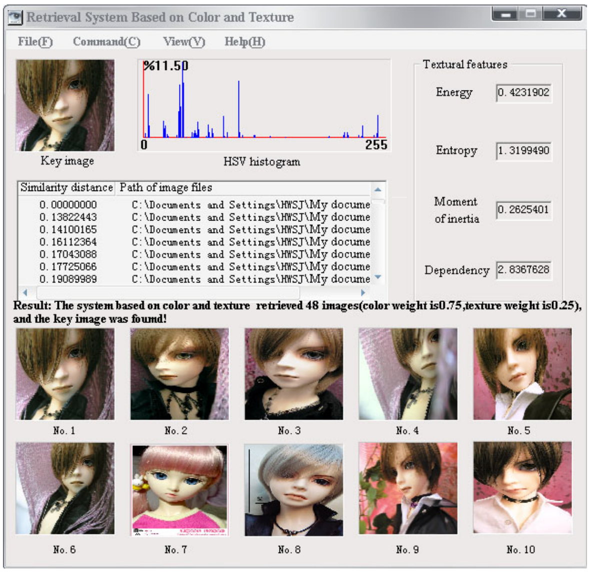


Figure 6. The result of retrieval algorithm based on color and texture in cartoon dolls

In addition, the following average precision ratio formula is also adopted to evaluate and compare various methods [20-22]:

$$precision = \frac{Retrieved \cap Relevant}{Retrieved} \quad (17)$$

For example, 10 cartoon dolls are selected from the image library as the key image and the top 10 images in each retrieval list are chosen, to compute the average precision and retrieval time. From Table 1, it can be found that the precision based color and texture is much higher if the average retrieval time is ignored.

TABLE 1 AVERAGE PRECISION AND RETRIEVAL TIME BASED ON THREE METHODS

Retrieval algorithm	Average precision (%)	Average retrieval time (s)
The weighted color feature based on HSV	55.27	0.8
The gray level co-occurrence matrix	70.65	0.9
The comprehensive algorithm in this paper	85.58	1.8

TABLE 2. THE SELECTION OF WEIGHT IN THE COMPREHENSIVE ALGORITHM OF THIS PAPER

Color and weight(w_1, w_2)	Natural scenery (%)	Cartoon dolls (%)	Buildings (%)	Flowers (%)
0.1,0.9	62.36	59.29	51.32	71.35
0.2,0.8	64.63	60.85	57.02	72.28
0.3,0.7	67.78	63.34	72.57	75.53
0.4,0.6	78.24	72.25	63.15	85.56
0.6,0.4	68.35	83.14	57.69	74.48
0.7,0.3	63.43	59.74	51.48	69.55
0.7,0.3	61.68	58.75	50.36	68.85
0.1,0.9	60.62	57.46	50.28	67.65

From table 2, we can find that, for different weights of color and texture, the results of the retrieval based color and texture have different effects. If unsatisfied with the retrieval result based on color or texture, the user can set the color and texture weights according to the specific situation. For example, for the image with bright color and unobvious texture, a larger color weight and a smaller texture weight can be set; But for the image, of which the texture is more obvious and the image color is not bright, a larger texture weight and a smaller color weight can be set. In a word, the retrieval based on color and texture can usually have a better effect in the process of color and texture weight adjustment.

IV. CONCLUSION

First of all, this algorithm can be used to retrieve the key image in the image library accurately. Secondly, from a large number of tests, it can be found that single color retrieval is more accordant with the visual and psychological characteristics of the character while compared with the texture retrieval, which indicates that good results can usually be obtained if the weighted color feature based on HSV is combined with the Euclidean distance. For images with obvious texture feature, the results of the texture retrieval are satisfied, but for the others, the results are not very ideal and are quite different from those of the retrievals based on color. Also, it shows that the texture feature of the gray level co-occurrence matrix is not consistent with the character's visual characteristics, and it is necessary to combine the color feature with other features to get more ideal results. The experimental results verified the effectiveness of the method in this paper.

REFERENCES

[1] Arnold W. M. Smeulders, Senior Member, Marcel Worring. Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(12): 1349-1380.
[2] Gong M,Fu Z.Zhang Y S.Image Retrieval Based on Integrate BTC Color Moment and GLCM . Video Engineering, 2012, 36(11): 30-33,40.
[3] Zhang G,Ma Z M.An approach of using Gabor wavelets for texture feature extraction. Journal of image and graphics, 2010, 15(2): 247-254.
[4] Chang Z, Hou Y Q,Li M L ,Cheng T, Liu L N. Image Retrieval Based on Combined Color with Te- xture Feature. Journal of Chinese Computer Systems, 2011, 32(1): 161-164
[5] Kwatra V, Essa I, Bobiek A, et a1. Texture Optimization for Example-based Synthesis// Proc Of SIGGRAPH'00 Angeles, USA: ACM Press. 2005: 795-802.
[6] Datta R, Joshi D, Li J, Wang J Z. Image retrieval: ideas, influences, and trends of the new age . ACM Computing Surveys, 2008, 40(2): 1-60.
[7] Wu J, Lu M, Liu C. SVM - feedback scheme within hybrid learning framework for image retrieval. Acta Electronica Sinica, 2010, 38(9): 2101-2106.
[8] Liu Y, Zhang D, Lu G. Region-based image retrieval with high level semantics using decision tree learning. Pattern Recognition, 2008, 41(8): 2554-2570.
[9] Wan X. Content based image retrieval using manifold-ranking of blocks [A]. Proc of IEEE Int Conf on

Multimedia and Expo (ICME'07). Beijing: IEEE Computer Society, 2007. 2182-2185.

[10] Li F, Dai Q, Xu W, Er G. Correlated probabilistic label propagation for region-based image retrieval [A]. Proc of IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP'07). Hawaii: IEEE Computer Society, 2007: 789-792.

[11] S H Feng, D Xu, X Yang. Attention-driven salient edge(s) and region(s) extraction with application to CBIR . Signal Processing, 2010, 90: 1-15.

[12] Rahmani R, Goldman S A. MISSL: multiple- instance semi-supervised learning. Proc of Int Conf on Machine Learning (ICML'06). Pittsburgh: Association for Computing Machinery, 2006, 705-712.

[13] Wang Q,Peng J Y,Guo S.han. Approach for image retrieval based on hybrid features kernel. Computer Engineering and Applications, 2012, 48(15):168:171.

[14] Wang X Z,Wang X F.Image retrieval approach integrated color and multi-scale texture features by Adaboost algorithm. Electronic Design Engineerint, 2012, 20(4): 165-169.

[15] Song Z,Yan Y F.Image retrieval based on combined features of interest point. Journal of Computer Applications, 2012, 32(10): 2840- 284, 2847.

[16]Cinque L, Levialdi S, Pellicano A. Color-based Image Retrie Using Spatial Chromatic Histograms//Proceeding of l999 IEEE International Conference on Multimedia Computing and Systems: IEEE Computer Society, 1999:969-973.

[17]Kwatra V, Essa I, Bobiek A, et a1. Texture Optimization for Example-based Synthesis//Proc Of SIGGRAPH’00 Angeles, USA：ACM Press．2005:795-802.

[18]Datta R ,Joshi D, Li J,Wang J Z.Image retrieval:ideas,influences,and trends of the new age. ACM Computing Surveys, 2008, 40(2): 1-60.

[19]Wang J Z,Li J,Wiederhold G. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(9):947-963.

[20]Rahmani R,Goldman S A,ZhangH,Cholleti S R,Fritts J E.Localized content based image retrieval. IEEETranson Pattern Recognition and Machine Intelligence, 2008, 30(11): 1902-1912.

[21Rahmani R, Goldman S A,Zhang H, Krettek J,Fritts J.Localized content based image retrieval[A].Proc of ACM Int Conf on Multimedia Information Retrieval (ACMMIR'05). Singapore: Association for Computing Machinery, 2005, 227-236.

[22]A. Shinya, K. Hiroaki. Content-based Image Retrieval System by Multi-dimensional Feature Vectors. International Conference on Digital Printing Technologies, 2003, 841-845.

Visual and Artistic Images Denoising Methods Based on Partial Differential Equation

Zhenhe Ye [1], Xin Li [2], and Ying Li [2]
1. College of Mechanical and Electrical Engineering, Agricultural University of HeBei, Baoding, 071000, China
2. College of Arts, Hebei Normal University of Science & Technology, Qinhuangdao, 066000, China
Email: yezhenhe@163.com

Abstract—**Partial differential equation has a remarkable effect on image denoising, compression and segmentation. Based on partial differential equations, the denoising experiment is carried out on those artistic images requiring high degree of visual reduction through the application of 3 image-denoising algorithm models including thermal diffusion equation, P-M diffusion equation and the TV diffusion equation. By this experience, the respective characteristics in image-denoising of these 3 methods can be analyzed so that a better way can be chosen in adapting to digitization of artistic images or in dealing with distant signal.**

Index Terms—**Denoising, thermal diffusion equation, P-M diffusion equation, TV diffusion equation, artistic images**

I. Introduction

The visual art sets a high demand on images' reduction in brightness, color and texture. However, in digital display and distant transmission, the obtained images are often affected by various noises inevitably, which will certainly affect the accurate artistry in the subsequent images presentation [1, 2, 3, 4]. Therefore, those questions, which have been widely researched, are how to eliminate noise effectively, to improve image quality and to denoise images which are also necessary courses in image processing [5, 6]. Partial differential equation is one of the important parts in mathematical analysis and at present it has been widely used in many fields such as image processing and computer visual digitalization. In its application in image processing, the first step is to build an image partial differential equations model. Assuming that the image is a 2-D function in real number field about x, y, since the image is a continuous piecewise linear (PWL) function, it is feasible that we use function to approach the original image in continuous field except for the edge so as to accurately express intense local content and information of the visual images and represent the subject as well as the visual and mental information which the drawer wants to express [7, 8]. Therefore, this paper will focus on the discussion and improvement of key problems in image denoising process so that this method can be well applied in processing the visual art images [9, 10] .

II. Theoretical Basis of Partial Differential Equation

A. Basic Conception

In a differential equation, if the unknown function to be solved only contains one argument, the equation is called ODE (ordinary differential equation); if the unknown function to be solved contains at least one argument, and there appears partial derivative of each order of multivariate function for different independent variables in the equation, the equation is called partial differential equation [11, 12, 13].

The partial differential equation on unknown function $u(x_1, x_2, \cdots, x_n)$ is a relation similar to

$$F(x, u, Du, \cdots \frac{\partial^m}{\partial x_1^{m_1} \partial x_2^{m_2} \cdots \partial x_n^{m_n}}) = 0 \qquad (1)$$

where $x = (x_1, x_2, \cdots, x_n), Du(u_{x_1}, u_{x_2}, \cdots, u_{x_n})$, F is the known function of independent variable x , unknown function u and several finitude partial derivatives on u .F can be showed that it doesn't contain the independent variable x and the unknown function u , but it must contain u 's partial derivative. The order $m = m_1 + m_2 + \cdots m_n$ of the highest derivative is called order of the partial differential equation.

B. Advantages of Partial Differential Equation

The method of partial differential equation has remarking advantages in both theory and calculation, which are revealed as follows: Firstly, it has a stronger local adaptability. Fourier transform method has no localization at all, so it is only applicable to stationary signals processing. However, the images are usually non-stationary. Partial differential equation is based on serial image models, and it makes the changes of the value of a certain pixel only rely on an "infinitesimal" neighborhood of the pixel point at the current time t . In this sense, it can be seen that the partial differential equation method in image processing has "infinite" ability of local adaptability. Secondly, it has high flexibility. If a basic model is set up successfully, and after some modification or improvement for it, a processing method can be obtained with more perfect performance and more wide application. And the modification and improvement are usually direct and easy. For example, two-dimension is upgraded into three-dimension, and a single-valued image processing method into vector diagram. Finally, the partial differential equation provides a analytical mode of

doi:10.4304/jmm.8.3.284-290

the image in continuous space as well as a better image processing effect[14] [15]. What's more, the existence, uniqueness and stability of the algorithm solution can be proved available in the unique analysis-theory framework of the partial differential equation. The model is unrelated with the grid size (corresponding with the image pixel) of digital images. When the size of the grid's meshes tends towards zero, the discrete filter can be understood as the approximation of continuous differential operators in the partial differential equation.

III. Proposed Scheme

A. Thermal Diffusion Equation

Thermal diffusion equation is set up based on the physical model, and the model is the following: the heat of a partial heated iron tube will spread gradually as the heat conduction process until the temperature of the whole iron tube is consistent [16, 17]. When the thermal diffusion equation is applied to image denoising, the noises of the image are equal to the catastrophe point of the temperature. Thermal diffusion will make the gray values of the whole image tend to be consistent, and then the purpose of denoising can be achieved.

Suppose the original image is $u(x,y,0)$, $u(x,y,t)$, the smooth image of the time t, and then the partial differential equation of the thermal diffusion is

$$\frac{\partial u(x,y,t)}{\partial t} = \Delta u(x,y,t) \qquad (2)$$

where $\Delta u(x,y,t)$ is the Laplace operator of the image, with initial condition being $u(x,y,0)$, and its solution is

$$u(x,y,t) = G_t * u(x,y,0)$$

$$G_t(x,y) = \frac{1}{4\pi t}\exp\left(-\frac{x^2+y^2}{4t}\right) \qquad (3)$$

Here $*$ is the convolution. It can be seen from the above that: The original image and the convolutions of Gaussian filters of different scales are equivalent to the solution of the thermal diffusion equation which is isotropic, and has the same diffusion intensity to all the areas of the image. With the strengthening of smoothing effect, particulars such as edges of the original image, will disappear gradually.

The discrete form of Equation (2) is

$$u_s^{t+l} = u_s^t + \frac{\lambda}{|n_s|}\sum_{p\in n_s}\nabla u_{s,p} \qquad (4)$$

where u_s is sampling discrete image, s is pixel coordinate, t is iteration, λ is the weighting of distribution coefficient reflecting the smoothness, n_s means all the consecutive points of the pixel s, and $|n_s|$ represents the number of consecutive points (those four points, directly adjacent to the pixel, are often adopted from the up, down, left and right with edge excepted.)

$$\nabla u_{s,p} = u_p - u_s',\ p \in n_s \qquad (5)$$

Namely the gradient value of image along the direction (s,p). When Equation (5) is substituted into Equation (4), the template of Equation (2)'s action on image can be obtained:

$$\begin{pmatrix} & \lambda & \\ \lambda & 4-4\lambda & \lambda \\ & \lambda & \end{pmatrix} / 4 \qquad (6)$$

All the weight of pixel s acting on each directions are λ. Regardless of the location of pixel s, it has the same smoothness intensity to all pixels. When λ is 1, it is the denoising filter of mean value. In this way, the characteristic of image is blurred while being denoised.

B. P-M Diffusion Equation

The shortcoming of thermal equation of energy is that the coefficient is a constant and cannot be diffused differently according to different directions [18, 19]. Focusing on the shortcoming of thermal diffusion equation, Perona and Malik improved the model of thermal diffusion. The basic thought is that image is separated into domains of different scales in scale space, and is smoothed in the internal domains of all the scales while weakening the smoothing effect on the edge of domains, namely edge [20, 21, 22].

Perona and Malik put forward nonlinear diffusion in 1990 and introduced function $c(x,y,t)$ [23] [24], which satisfied the properties put forward above and improved the thermal diffusion equation:

$$\frac{\partial u(x,y,t)}{\partial t} = div[c(x,y,t)\nabla u(x,y,t)] \qquad (7)$$

The initial condition is $u(x,y,0)$, u is the input image, div is the divergence operator, ∇ represents the gradient, and c is the diffusion coefficient. Under ideal conditions, the corresponding value of $c(x,y,t)$ is 1 in the internal domain of image, and Equation (7) degenerates into the isotropic thermal diffusion Equation (2) which will do isotropic diffusion in the internal domain of the image; in the edge that is the edge of the image domain, the value of $c(x,y,t)$ is 0, and does not perform the diffusion. The diffusion coefficient of equation is decided by the spatial location of the image, namely the gray change of it. The coefficient is anisotropic so that the edge feature of the image can be reserved when the diffusion occurs.

However, in the actual application of image processing, the edge of image domains, namely the edge, cannot be known in advance. So function $E(x,y,t)$, a vector-function, is introduced as evaluation of the edge points. This thesis adopts gradient operator to evaluate the edge points, namely

$$c(x,y,t) = g(|E(x,y,t)|) = g(|\nabla u(x,y,t)|) \qquad (8)$$

In accordance with the strategy of smoothing mentioned above, $g(\bullet)$ sets the degree of diffusion, and satisfies the following conditions: I $g(s)$ is a smooth and non-increasing function. II $g(0)=1$, and $g(s)\ge 0$. When III $s\to\infty$, and $g(s)\to 0$, the gradient of image is large at the edge, $g(s)$ is small, and the degree of diffusion becomes smaller; on the contrary, the gradient of image is small, $g(s)$ is large, and the degree of diffusion becomes larger. As a result, the image can be smoothed alternatively— smoothed less at the edge and more in the smooth areas. The functional figure 1 of $g(s)$ is shown as follows:

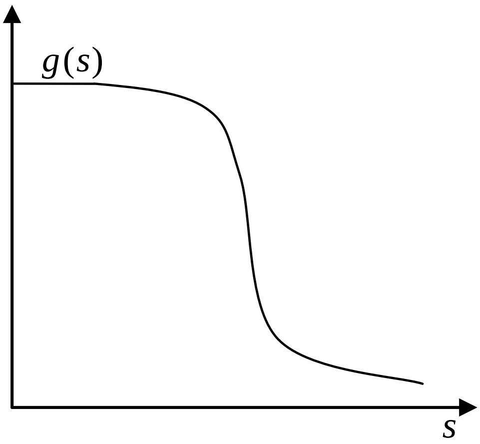


Figure1.Note how the caption is centered in the column.

Here the function of diffusion coefficient $c(x,y,t)$ becomes a monotonic decreasing function which adopts $\nabla u(x,y,t)$ as the variable. From the above, the following can be obtained:

$$\begin{cases} \dfrac{\partial u(x,y,t)}{\partial t} = div[g\left|\nabla u(x,y,t)\right|\nabla u(x,y,t)] \\ u(x,y,0) = u_0(x,y) \end{cases} \qquad (9)$$

This is the traditional P-M diffusion model equation. And the two functions of diffusion coefficient are

$$g(|\nabla u|) = \frac{1}{1+\left(\frac{|\nabla u|}{k}\right)^2} \qquad (10)$$

or

$$g(|\nabla u|) = \exp\left(-\left(\frac{|\nabla u|}{k}\right)^2\right) \qquad (11).$$

where k is the threshold parameter of gradient controlling the process of diffusion. The larger its value is, the better smoothing effect it will have. And the smaller k value can strengthen the edge. The effects these two functions of diffusion coefficients act on the diffusion of image are different. The previous function will strengthen the high-contrast edges of image and weaken the low-contrast edges of it, and the latter function will reserve the large areas in the image and remove the small ones in it.

Generally, g is the gradient function of image. And with the increasing of gradient, it will decrease monotonically and its value range is restricted to [0,1], such as the formulas (10), (11). The diffusion coefficient decides the way of diffusion and provides a diffusion-controlling strategy of local adaptability. It makes the diffusion process the location of noise as much as possible, but stops at the edge of image.

The discrete version of P-M equation is

$$u_s^{t+l} = u_s^t + \frac{\lambda}{|n_s|}\sum_{p\in n_s} c(\nabla u_{s,p})\nabla u_{s,p} \qquad (12)$$

As a result, the template of P-M equation's action on the image is

$$\begin{pmatrix} & c(u_{p1}-u_s) & \\ c(u_{p2}-u_s) & 4-\sum_{p\in n_s} c(\nabla u_{s,p}) & c(u_{p4}-u_s) \\ & c(u_{p3}-u_s) & \end{pmatrix} / 4 \qquad (13)$$

Here, the weight value of pixel s is no longer a fixed value, but is related to the spatial location of pixel s. If the pixel is at the location with large gradient, the relevant weight value of pixel will be small and the smoothing degree will be weak, because $c(x,y,t)$ is a monotonically decreasing function with $|\nabla u(x,y,t)|$ as its variable. On the contrary, if the pixel s is at the location with small gradient, the smoothing degree will be intense. Usually, the gradient is larger at the edge of image, and smaller at other areas. So the P-M equation can denoise the image while reserving or even strengthening the edge.

C. TV Diffusion Equation

A discrete image can be regarded as a real function above R2 in the bounded closed region, and the essence of denoising those noisy images is to recover the real images from them, so the characteristic information (such as the edge) should be kept as much as possible so it will not be lost during denoising. Let's suppose $u^0(x,y)$) as the original images and $u^0(x,y)$ as noisy images.

$$u^0(x,y) = u(x,y) + n(x,y) \qquad (14)$$

$n(x,y)$ is a additive noise with its 0 average and σ^2 variance, and it is a gaussian noise here. The processing of denoising images is to estimate the original images $u(x,y)$ according to the prior knowledge we have and the statistics of noise, which is to work out $u(x,y)$ by equations. But usually it is infeasible to solve (14) directly in reality, because usually there are errors in actual statistics, and solving directly often result in worse consequence. The way of solving this problem is to adopt regularization, which can solve the equation by restricting its scope. Then the problem of denoising is turned into the problem of optimization.

$$\begin{aligned} &\int_\Omega u(x,y)d\Omega = \int_\Omega u^0(x,y)d\Omega \\ &\int_\Omega \left|u(x,y)-u^0(x,y)\right|^2 d\Omega = \sigma^2 \end{aligned} \qquad (15)$$

The corresponding form without constrain conditions is:

$$E_\lambda(u) = \int_\Omega |\nabla u(x,y)|d\Omega + \frac{\lambda}{2}\int_\Omega \left|u(x,y)-u^0(x,y)\right|^2 d\Omega \qquad (16)$$

λ is a Lagrange multiplier, which is a constant given before-hand and decides the intensity of smooth denoising, therefore, the value of it depends on the noise level. Here the function u need to be obtained while the value of $E_\lambda(u)$ being the minimum, which is a functional extremum problem, namely variation.

Through Euler equation, extremum condition of Equation (16) can be obtained as follows:

$$-\nabla\left(\frac{\nabla u}{|\nabla u|}\right)+\lambda(u-u^0)=0 \tag{17}$$

The following is obtained by solving Equation (17) with gradient descent method

$$\frac{\partial u}{\partial t}=\nabla\left(\frac{\nabla u}{|\nabla u|}\right)+\lambda(u^0-u) \tag{18}$$

In actual calculation, such situation may occur that ∇u becomes zero in flat areas of image, when the denominator on the right of Equation (18) needs to be adjusted. Generally a small constant a^2 will be added to the denominator, and

$$|\nabla u|_a=\sqrt{|\nabla u|^2+a^2} \tag{19}$$

The performance of TV model will not be affected as long as the value of a^2 is small enough.

Here the following can be obtained:

$$\begin{aligned} &E_{\lambda,a}(u)=\int_\Omega |\nabla u|_a d\Omega+\frac{\lambda}{2}\int_\Omega |u-u^0|^2 d\Omega \\ &\frac{\partial u}{\partial t}=\nabla\left(\frac{\nabla u}{|\nabla u|_a}\right)+\lambda(u^0-u) \end{aligned} \quad (x,y)\in\Omega \tag{20}$$

The discrete version of Equation (20) is:

$$\frac{u_{i,j}^{n+1}-u_{i,j}^n}{\tau}=\sum_{(k,l)\in N(i,j)}\frac{g_{k,j}^n+g_{i,j}^n}{2h^2}(u_{k,l}^n-u_{i,j}^n)+2\lambda(u_{i,j}^0-u_{i,j}^n) \tag{21}$$

Thereinto, τ is time step, and h is pixel space $g_{i,j}^n=\dfrac{1}{|\nabla u_{i,j}^n|_a}$. α or β are used to represent the coordinates of pixel: (i,j) or (k,l) , and

$$w_{\alpha\beta}^n=\frac{g_\alpha^n+g_\beta^n}{2h^2} \tag{22}$$

After the arrangement of the above equation, we can obtain

$$u_\alpha^{n+1}=\tau\sum_{\beta\in N(\alpha)}w_{\alpha\beta}^n u_\beta^n+\tau\lambda u_\alpha^0+u_\alpha^n\left(1-\tau\sum_{\beta\in N(\alpha)}w_{\alpha\beta}^n-\tau\lambda\right) \tag{23}$$

There are two parameters in the above equation: Lagrange multiplier λ and time step τ . Reference [3] provides the optimal estimation of approximation for λ

$$\lambda=\frac{1}{\sigma^2}\frac{1}{|\Omega|}\sum_{\alpha\in\Omega}\sum_{\beta\in N(\alpha)}w_{\alpha\beta}(u_\beta-u_\alpha)(u_\alpha-u_\alpha^0) \tag{24}$$

$|\Omega|$ is the area of image zone. From Equation (24), we can see that λ is in inverse proportion to variance of noise. When smoothing images first, we can choose $\lambda=\dfrac{1}{\sigma^2}$, and update the value of λ with Equation (24) after a period of iterations.

Time step can either be constant or calculated according to the data of each step. It affects the convergence of iterative equation directly and Equation (25) provides the calculation method of current iteration step size in accordance with the previous iteration results.

$$\tau=\frac{1}{\sum_{\beta\in N(\alpha)}w_{\alpha\beta}+\lambda} \tag{25}$$

Putting Equation (25) into Equation (23), we can obtain:

$$u_\alpha^{n+1}=\sum_{\beta\in N(\alpha)}h_{\alpha\beta}u_\beta^n+h_{\alpha\alpha}u_\alpha^0 \tag{26}$$

$$h_{\alpha\beta}=\frac{w_{\alpha\beta}}{\lambda+\sum_{\gamma\sim\alpha}w_{\alpha\beta}}\qquad h_{\alpha\alpha}=\frac{\lambda}{\lambda+\sum_{\gamma\sim\alpha}w_{\alpha\beta}} \tag{27}$$

Thereinto,

$$h_{\alpha\alpha}+\sum_{\beta\in N(\alpha)}h_{\alpha\beta}=1 \tag{28}$$

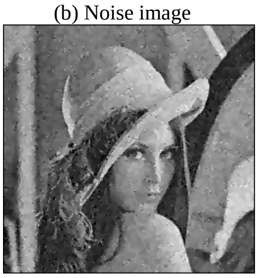
The iterative process of denoising image using TV model is equivalent to that of non-linear lowpass filtering. For edge pixels, these pixel values are preserved as the value of $|\nabla u|_a$ is bigger, and as a result, the value of $h_{\alpha\alpha}$ becomes smaller; for pixels in flat areas of image, it is the equal of lowpass filtering for the pixels as the value of $|\nabla u|_a$ is smaller while the value of $h_{\alpha\alpha}$ is bigger. The smooth intensity of TV model for different pixels is different as well as anisotropic. Although the number of iterations increases from 10 to 35, the difference of Lena image after TV diffusion is not as obvious as that after P-M diffusion. This is because TV diffusion equation (16) includes $\dfrac{\lambda}{2}\int_\Omega |u(x,y)-u^0(x,y)|^2 d\Omega$.

The iteration and denoising for image with TV model is equal to non-linear lowpass filtering, the diffusion degree in different points are different as well as anisotropic and stable. Since TV model is based on the “optimization” of the image's integral variation model and it has globally optimal solution, it can get stable and correct results even when the image noise is very large.

IV. Experiment

Select a 256*256 grayscale image, and add white Gaussian noise with mean value of 0 and variance of 25 to it. Denoise the noise image with thermal diffusion, P-M diffusion and TV diffusion respectively. And the iterations are 10 and 20, as shown in Figure2, Figure3.


(a) Original image (b) Noise image

(c) Denoising with thermal Diffusion equation by 10 iteration

(d) Denoising with t hermal diffusion equation by 20 iterations

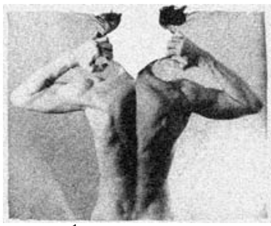

(e) Denoising with P-M diffusion equation by 10 iterations

(f) Denoising with P-M diffusion equation by 20 iterations

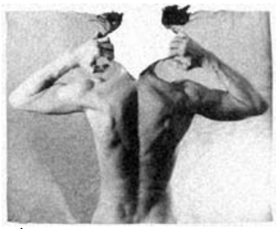
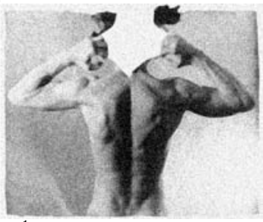
(g) Denoising with T-V diffusion equation by 10 iterations

(h) Denoising with T-V diffusion equation by 20 iterations

Figure 2. Denoising the image with three methods

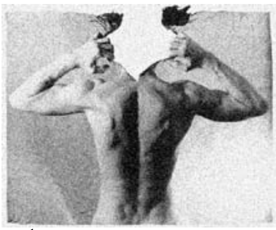
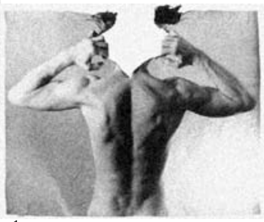



(a[1]) Original image (b[1]) Noise image

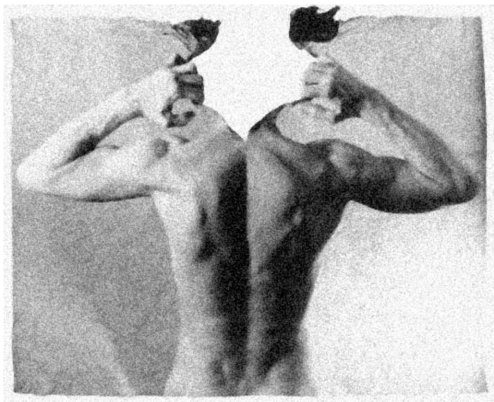
(c[1]) Denoising with thermal Diffusion equation by 10 iteration

(d[1]) Denoising with t hermal diffusion equation by 20 iterations

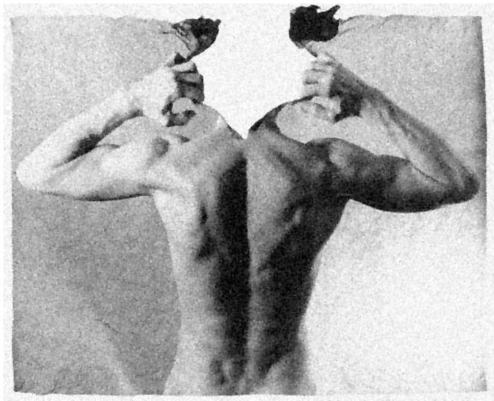



(e[1]) Denoising with P-M diffusion equation by 10 iterations

(f[1]) Denoising with P-M diffusion equation by 20 iterations



(g[1]) Denoising with T-V diffusion equation by 10 iterations



(h[1]) Denoising with T-V diffusion equation by 20 iterations

Figure3. Denoising the artistic image with three methods

Observing images denoised through isotropic thermal diffusion equation, we can see that: when the iteration number reaches 10, the edge of image is fairly clear, while there is still lots of noise on the image; when the iteration number reaches 20, the image becomes obscure seriously because the noise is basically eliminated, some important information such as the edge and details of the image gets missing for smooth, though. Isotropic thermal diffusion equation turns image-denoising into a partial differential issue, however, its denoising effect is pretty average for the severe edge information missing; and image-denoising effect of P-M diffusion equation is obviously better than that of thermal diffusion equation. The noise is mostly eliminated with maintenance of the edge to some extent, but obscurity appears on some details of the image. This is because that P-M diffusion equation can seize the changes with large image gradient very well, but it cannot seize the texture and petty edge caused by tiny changes, which is the reason of missing image information. While the noisy points are being reduced, there appear some bigger noisy points, which are caused by the misjudging of some noisy points as edge because of their rather large gradient values. As a result, those noisy points are not removed, on the contrary, are strengthened to some extent. And color block appears, that is, after the P-M diffusion, there appear a lot of areas with the same gray value which produces such visual effect as the image is composed of many areas with different brightness. This is because that P-M diffusion equation is smoothing the image regionally, and in areas with rather small gradient, the value of diffusion coefficient function $g(|\nabla u|)$ is larger with strong diffusion intensity, thus eliminating the noise effectively, decreasing the gradient values and at the same time inclining gray value into constant in those areas; in areas with rather big gradient, the value of diffusion coefficient function $g(|\nabla u|)$ is small with weakened diffusion intensity, and along with the diffusion, the gradient values of those areas will evolve towards infinity, finally resulting to a zero value of diffusion coefficient function $g(|\nabla u|)$ and inactive diffusion on these areas, after which the value in each area tends to be constant and since the diffusion cannot be conducted between areas, there will appear staircase effect for largely different gray values among areas although the edge feature of the image is preserved; while the image denoised through T-V diffusion has no such obvious distinction as that through P-M diffusion even when the number of iterations is increased from 10 to 20. This is due to the constraint $\frac{\lambda}{2}\int_{\Omega}\left|u(x,y)-u^0(x,y)\right|^2 d\Omega$ contained in TV diffusion equation, which controls the degree of deviation between denoised image and original image. Therefore, the distinction led by iterations is not as obvious as that of P-M diffusion, which is hence stable and pathological. But compared with P-M diffusion, the edge preserving effect of TV diffusion is slightly poorer. Although the function of bounded variation (BV) has a favorable ability of expression for image edge information, it cannot represent the information of image texture and details very well. Therefore, it is inevitable to drop out some information like texture and details of the image in the process of denoising.

From the denoised image with three methods mentioned above, it can be seen that the denoising effects after the iteration of 10 times are basically the same. However, after the iteration of 20 times, denoising effects using the method of P-M diffusion is obviously better than by the other two. Besides, when the noise is filtered, the edge can be protected. As for the method of thermal diffusion equation, with the increase of iteration, the noise is filtered, but some important information such as the edge and particulars of the image will be dropped because of the smoothing. Finally, the image has been blurred seriously. Although the isotropic equation of thermal diffusion changes image denoising into partial differentiation, its denoising effect is very common and edgeal information is dropped seriously. The effect of image denoising by P-M diffusion equation is obviously better than that by thermal diffusion equation. Most of the noise is removed and the edge is also protected. The impact of increasing iteration on the image denoised by TV diffusion is not obvious, and the image is stably non-pathological. However, compared with P-M diffusion, the effect of edge-maintaining is slightly inferior.

V. CONCLUSION

It is unavoidable for works of artistic images to be jammed by noise in collecting and displaying image information through visual hierarchical system and to lose its realistic artistry by shortage of visual effect of artistic images. The denoising of image is crucial and its effect influences directly the degree of accuracy of subsequent analysis and processing. The paper analyzed the application of partial differential equation on denoising algorithmic model of image and conducted contrast experiments for image-denoising through thermal diffusion equation, P-M diffusion equation and TV diffusion equation respectively. Compared with other denoising methods achieved on visual images, although each of these methods has their advantages and disadvantages, they reduce the adverse effects of noise on images and meet the different requirements in real-time processing of artistic images in visual system.

REFERENCES

[1] Gao Chaobang, Zhou Jiliu,"Image enhancement based on guaternion fractional directional differentiation" . *Acta Au-tomatica Sinica*, vol. 37, no. 2, pp. 150–159, 2011.

[2] Zaidi H,Abdoli M, Fuentes C L, et al, "Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma ",*European Journal_of Nuclear Medicine and Molecular Imaging*, vol. 39, no. 5, pp. 881–891, 2012.

[3] CHEN Qiang, ZHENg Yuhui, SUN Quansen, et al,"Patch similarity based anisotropic dillusion for image denoising ",*Journal of Compute r Research and Development,* vol. 47, no. 1, pp. 33–42, 2010.

[4] Song Huaibo, He Dongjian, Xin Xiangjun, “Unstructured road detection and obstacle recognition algorithm based on machine vision”, *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE),* vol. 27, no. 5, pp. 225–230, 2011.

[5] Rajan J, Kannan K, Kaimal M R,”An improved hybridmodel for molecular image denoising”, *Journal of Math-ematical Imaging and Vision,* vol. 31, no. 1, pp. 73–79, 2008.

[6] WinnemOller H, Olsen S C, Gooch B. Real-time videoabstraction [C]//Computer Graphics Proceedings, AnnualConference Series, ACM SIGGRAPH, Boston,pp. 1221–1226, 2009.

[7] BozchalooiI S, Liang M,”A smoothness index － guided approach to wavelet parameter selection in signal denoisingand fault detection”,*Journal of Sound and Vibration,* vol. 1, no. 2, pp. 246–247, 2007.

[8] Zhang Songhai, Chen Tao, Hu Shimin. “Automatic conversionfrom video clip to flash animation “. *Journal of Computer-Aided Design & Computer Graphics,* vol. 19, no.5, pp.667–671, 2007.

[9] Chen Yan, Zhang Man, Ma Wenqiang, et al,”Positioning method of integrated navigation based on GPS and machine vision”, *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, vol. 27, no. 3, pp. 126–130, 20011.

[10] Zhenhe Ye, Xin Li,”Model Evaluation Based on Emotional Furniture Industrial Design Elements “, *International Journal of Advancements in Computing Technology* , vol. 22, no. 4, pp. 73–78, 2012.

[11] Buades A, Coll B, Morel J M,”Self-similarity-based image denoising”, *Communications of the ACM,* vol. 54, no. 5, pp. 109–117, 2011.

[12] Badulescu P, Zaciu R, Removal of mixed-noise using order statistic filter and wavelet domain Wiener filter[A], In: Proceedings of the International Semiconductor Conference[C], Sinai Romania, p.301~304,1999.

[13] Zhenhe Ye, Qinghua Bai,”Agricultural Implementation Security Research Based on Ergonomics “, *Journal of Convergence Information Technology,*vol. 20, no.7, pp. 65–71, 2012.

[14] Li Jinjiang, Li Yewei, An Zhiyong, “Image denoising algorithm based on the nonsubsampled double density contourlet transform”,*Journal of Convergence InformationTechnology,* vol. 6, no. 9, pp. 105–1113, 2011.

[15] Didas S, Weickert J, Burgeth B, “Properties of higher order nonlinear diffusion filtering “, *Journal of Mathematical Imaging and Vision,* vol. 35, no. 1, pp. 208–226, 2009

[16] Chih-Chieh Yang, Pei-Ju Lin, Chih-Cheng Sun , “Product Form Design Using Virtual Hand and Deformable Models”, *International Journal of Digital Content Technology and its Applications,* Vol. 6, no. 11, pp. 8 ~ 17, 2012.

[17] Rahman M M,Arelfin MG,.Kumar M P K, et al,”Optimal nonlocal means algorithmfor denoising ultrasound image “,*Computer Engineering and Intelligent Systems,* vol. 3, no. 3, pp. 56–64, 2012.

[18] Pan Quan, Zhang Lei, Zhang Hongcai et al, ”A daptive wavelet based spatially de-noising”, In: Proceedings of the 4th International Conference on Signal Processing Proceedings[C], Beijing China, p.486~489,1998.

[19] Gilboa G, Sochen N, Zeevi Y Y,”Variational denoising of partly textured images by spatially varying constraints”,*IEEE Transactions on Image Processing,* vol. 15, no. 8, pp.2281–2289, 2006.

[20] Steidl G, Weickert J, Brox T, et al, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization and SIDEs, *SIAM Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004.

[21] Pan Quan, Zhang Lei, Zhang Hongcai et al, *Adaptive wavelet based spatially de-noising*, In: Proceedings of the 4th International Conference on Signal Processing Proceedings[C], Beijing China, 1998:486~489

[22] Zhang Xiao Ping, Desai M D, Nonlinear adaptive noise suppression based on wavelet transform[A], *In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*[C], Seattle USA, p.1589~1592, 1998.

[23] Chang S G, Yu Bin, Vetterli M, Spatially adaptive wavelet thresholding with context modeling for image denosing, *IEEE Trans. Image Processing,* vol. 9, no. 9, pp. 1522–1530, 2000.

[24] LI Yan-jun, SU Hong-yi, YANG Feng, eet al.Improved algorithm study about removing image noise . *Journal of Computer Engineering and Design,* vol. 30, no. 12, pp. 2995–2996, 2009.

Zhenhe Ye received his M.E. degree from Hebei University of Technology, China, in 2009. He is a key teacher in mechanical and Electrical Engineering College, Hebei Agricultural University, His research interests include image compression, Computer-graphic design, interaction design and its theories and Ergonomics etc.

Xin Li received her M.A. degree in Art from Hebei University of Technology,China in 2008 . She is a key teacher in Art Colle ge, Hebei Normal University of Science & Technology and her research interests include environmental art, visual communicati on and CAD, etc.

Ying Li received her M.A. degree in Art from Hebei Normal University, China, in 2005. Her is a key teacher in Art College, Hebei Normal University of Science & Technology, Her research interests include environmental art, visual communication, CAD, etc.

A New Method of Objective Speech Quality Assessment in Communication System

Weiwei Zhang, Yongyu Chang, Yitong Liu, and Leilei Xiao
Wireless Theories and Technologies (WT&T) Lab,
Beijing University of Posts and Telecommunications
Beijing, P. R. China, 100876
Email: zweiwei506@gmail.com

Abstract—**On the Quality of Experience (QoE) evaluation of communication system, the quality of speech is an important factor to evaluate the system. Perceptual evaluation of speech quality (PESQ) is a well known objective speech quality assessment method for the voice QoE evaluation. It is proposed by International Telecommunication Union (ITU) and is formed as the ITU-T P.862 Recommendations. PESQ applies Bark-scale frequency to evaluate the Mean Opinion Score (MOS) for speech quality of voice communication system. But through our research, we find that the PESQ algorithm has limitations for evaluating speech quality. In order to change these limitations, this paper proposes a new objective evaluation algorithm by using ERB-scale to take the place of Bark-scale frequency. The ERB-scale is more accurate than Bark-scale to describe the frequency selectivity of the human ear when frequency at lower domain. We call the new algorithm as NPESQ (New Perceptual Evaluation of Speech Quality) whose accuracy is tested in our experiment. Through experimental comparison against PESQ and NPESQ, the results demonstrate the improvement of the NPESQ. Therefore, it can be concluded that the new algorithm can improve the accuracy of the measurement.**

Index Terms—**Quality of Experience (QoE), Speech quality assessment, Perceptual Evaluation of Speech Quality (PESQ), Bark-scale frequency; ERB-scale frequency**

I. INTRODUCTION

The communication technology is undergoing a rapid develop. Voice communication is the main service in communication network. Service providers are faced with offering high quality voice communication system [1]. Effective evaluation of system performance is becoming critical. Most of the evaluation models are system-centric i.e. Quality of Service (QoS) model [2]. But, the customer perception is more important to evaluate the network. We need a user-centric model to evaluate communication system i.e. Quality of Experience (QoE) model [3]. QoE is a very important assessment mechanism. It's more closely related to human perception as compared to QoS. Network providers can evaluate subscriber's QoE using QoS parameters [4], [5].

On the voice QoE evaluation, the quality assessment of speech is divided into subjective evaluation and objective evaluation [6]. At present, the subjective evaluation is the most accurate method for next generation communication networks. However, it is time-consuming, expensive and high resources input, which is not suitable for real-time communication. The objective quality assessment [7], [8] has replaces the subjective method. And it has become the main quality assessment method of speech QoE.

The PESQ algorithm is an objective speech quality assessment method for end-to-end speech quality assessment of communication system and speech codecs. It is approved as ITU-T recommendation P.862 [9]. The PESQ can be used for evaluation of the different types of communication networks [10]. It takes into account the delay of network, and uses an model of auditory and cognitive modeling techniques. So it is robust to the communication delay and noise of environment [11]. Network operators and equipment manufacturers widely used PESQ to evaluate the voice QoE of communication system.

However, through our research, we find that the PESQ algorithm has limitations for evaluating speech quality. Some research report also shown that the limitations of PESQ for measuring voice quality in mobile and VoIP Networks [12],[13]. In order to change this limitation, this paper proposes a improved algorithm of PESQ. The new algorithm is called as NPESQ (New Perceptual Evaluation of Speech Quality). The NPESQ can improve the accuracy of the measurement.

The key of the objective evaluation is that we transform the reference signal and degraded signal to an internal representation. This representation is analogous to the psychophysical representation of audio signals in the human auditory system. We use perceptual frequency and loudness. The PESQ make use of the Bark-scale frequency to stimulate auditory perceptual system, whose loudness model is Zwicker model. Another such loudness model is Moore model [14], which perceptually motivated scale is the ERB-scale frequency [15]. With the purpose of improving the accuracy of PESQ, we propose the NPESQ algorithm which uses ERB-scale take the place of Bark-scale frequency of the PESQ.

In order to verify the accuracy of the NPESQ algorithm, we do lots of experiments. We compare the

doi:10.4304/jmm.8.3.291-298

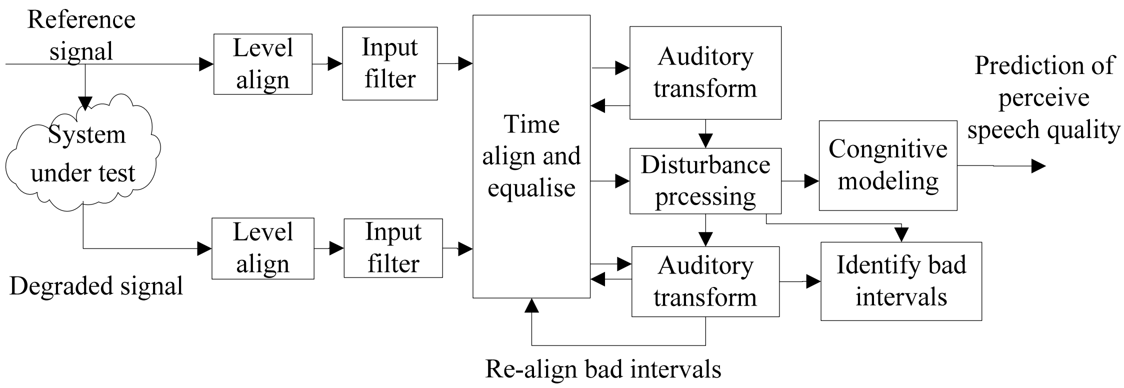


Figure 1.The structure of the PESQ algorithm

result scores of PESQ and NPESQ through testing the codecs of EVRC (Enhanced variable rate codec),AMR (Adaptive Multi-Rate Speech Codec) and EVRC-B (Enhanced variable rate codec-B). The Third Generation of Mobile Communications (3G) is a great revolution in the history of communications. Speech coding is a key technique in 3G and will play a very important role in data compression because of its strong ability of reducing the bit-rates of speech signals [16]. The EVRC, EVRC-B and AMR codec are the most popular and useful codecs in 3G communication systems [17],[18]. The EVRC and EVRC-B are used in CDMA2000 communication system. AMR is used in WCDMA communication system. These three speech codecs are popular used in current. So in our experiment we choose these speech codecs to verify the performance of the NPESQ algorithm.

The overview structure of paper is as follows. Section II outlines system model and expressions of the ITU P.862. and gives overview process of PESQ algorithm. Section III introduces the ERB-scale frequency and the comparison of Bark-scale frequency and ERB-scale frequency. Section IV gives the process of NPESQ algorithm. Section V give the experiment of subjective MOS test and objective test, and show the results of experiments and the work discussion. Section Ⅵ presents the conclusion.

II. Overview of PESQ Algorithm

A. PESQ Overview

The PESQ algorithm is an objective speech quality evaluation algorithm for speech quality evaluation of voice communication network and speech codec. The overview structures of PESQ are shown in the Figure1. Each block of Figure1 is explained below.

B. Process of Each Block

- Level align: Both the reference and degraded speech signal are necessary to be aligned to the same level.
- Time aligns: In a mobile network, the transmission delay is exists. The reference and degraded speech are time aligned, so degraded speech match the reference speech.
- Auditory transform model: The auditory transform of PESQ algorithm is a psychoacoustic model. This model includes several steps: time-frequency mapping, frequency warping, and compressive loudness scale.
- Disturbance processing and cognitive modeling: The disturbance signals are summed in the frequency scale. The disturbance densities and the asymmetrical densities are calculated and integrated along the frequency axis.
- Realignment of bad intervals: If there is a frame disturbance above a threshold in the consecutive frames. This frame is called bad interval. We need to realign it.
- Compute the PESQ score: The PESQ score is computed by combining the average disturbance value and asymmetrical disturbance value.

III. Comparison of BARK-scale and ERB-Scale

A Introduce of ERB-scale

From the perception view of the human auditory system, the relationship of sensing frequency and the actual frequency of the sound is a nonlinear mapping. The ERB (Equivalent Rectangular Bandwidth) gives an approximate of filter bandwidths in human auditory system. It is a measure in psychoacoustics where a perceptual scaling is needed to weigh perceptual importance of differences in fundamental frequency.

There is a differential equation in view of the ERB value $\boldsymbol{v}$ in Hz of a human auditory filter with a center frequency of f kHz[19].

$$\frac{d\boldsymbol{f}}{d\boldsymbol{v}} = 6.23 \cdot \boldsymbol{f}^2 + 93.39 \cdot \boldsymbol{f} + 28.52 \qquad (1)$$

Solving this equation gives approximately the following relation between the ERB value v and the frequency f in Hz:

$$\boldsymbol{v} = 11.17268 \cdot \log\left(1 + \frac{46.06538 \cdot \boldsymbol{f}}{\boldsymbol{f} + 14678.49}\right) \qquad (2)$$

Or the converse:

$$\boldsymbol{f} = \frac{676170.4}{47.06538 - \boldsymbol{e}^{0.08950404 \cdot \boldsymbol{v}}} - 14678.49 \qquad (3)$$

A newer approximation is [20]:

$$\frac{df}{dv} = 107.94 \cdot f + 24.7 = 24.7 \cdot (4.37 \cdot f + 1) \quad (4)$$

That is:

$$\Delta f = 24.7 \cdot (4.37 \cdot f + 1) \quad (5)$$

B Comparison of Bark-scale and ERB-scale

Both the Bark and the ERB are scales, which are more closely related to how sound is represented in the auditory system than a linear frequency scale. Both scale assume that the bandwidth of an auditory filer (critical band) corresponds to a constant length along the basilar membrane. Thus, on a linear frequency scale, the filter bandwidth increases with increasing center frequency for frequency is greater than a "cut-off" frequency. This cut-off frequency is assumed to be about 500 Hz for the ERB scale. For frequencies smaller than this cut-off frequency, the filter bandwidth is constant on a linear scale. However, both scales yield about the same filter bandwidth for frequencies between 1 and 5 kHz, and differ markedly below and above.

The following equations have been proposed to calculate the auditory filter bandwidth Δf (Hz) as a function of frequency f (kHz):

$$\Delta f = 25 + (1 + 1.4 \cdot f^2)^{0.69} \quad (6)$$

for the Bark scale, and

$$\Delta f = 24.7 \cdot (4.37 \cdot f + 1) \quad (7)$$

for the ERB scale. The filter bandwidths calculated from these two equations are plotted against center frequency of the respective filter.

We compare the filter bandwidths by using the ERB scale and Bark scale. The filter bandwidth in units of Hz, and each filter bandwidths has a center frequency. The ERB scale and Bark scales have similar filter bandwidths, when the frequency is above 1 kHz and below 5 kHz. But there are dissimilar markedly at lower and higher frequencies. ERB is more accurate to describe the frequency selectivity of the ear of human when frequency at lower domain. So we use ERB scale to take place of Bark scale to compute the loudness of auditory system.

IV. Description of NPESQ Algorithm

The NPESQ algorithm is a new method of objective method of speech quality assessment method of speech communication system.The whole algorithm consists of three parts: Level and time alignment pre-processing; Auditory transform and Cognitive modeling. Each part is explained below.

A. Level and Time-alignment Pre-processing

a) Level alignment: The input of the system is reference signal $X(t)$ and degraded signal $Y(t)$. The reference signal is the original speech signal. The degraded signal is the reference signal through the test system The test system is a voice system. The signal level of reference signal and degrade signal is different, when the original speech signal through the voice system. In order to compare the original and degrade signal, we need adjust the signal level to uniform level.

b) IRS filtering: It is assumed that the listening tests were carried out using an IRS receiver or a modified IRS receiver characteristic in the handset. The input of IRS filter are $X_S(t)$ and $Y_S(t)$,and the output are $X_{IRSS}(t)$ and $Y_{IRSS}(t)$.

c) The time-alignment process: Before performing auditory transform, it is needs to estimate the time delay between the degraded speech and the original speech. Because the parameter of the PESQ score is calculated frame-by-frame, and the speech signal is time-varying, each frame is not the same delay. If the time is not aligned between the frame and frame, it will cause large errors. The time alignment is including envelope-based alignment and the fine alignment.

B. Auditory Transform

The auditory transform is a physiological acoustic model, it mapped the speech to the time-frequency domain, to simulate the process of human ear receive speech. The process of the auditory transform consists of four steps: processing by Hanning window, Short-term Fast Fourier Transform, Frequency spectrum is mapped to ERB-spectrum and ERB-spectrum is mapped to loudness. The structure of the Auditory transform is shown in Fig.2.

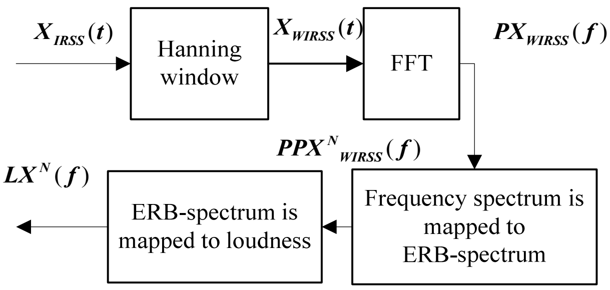


Fig.2. The Auditory transform of NPESQ

a) Processing by Hanning window: The size of Hanning window is 32 ms. For 8 kHz this amounts to 256 samples per window and for 16 kHz the window counts 512 samples while adjacent frames are overlapped by 50%.

The Hanning window function is:

$$W(n) = 0.5(1 - \cos\frac{2\pi n}{N}), 0 \le n \le N-1 \quad (8)$$

With n is the number of frame.

The reference and degraded signal become $X_{WIRSS}(t)_n$ and $Y_{WIRSS}(t)_n$ after being windowed.

$$\begin{cases} X_{WIRSS}(t)_n = W(n)X_{IRSS}(t) \\ Y_{WIRSS}(t)_n = W(n)Y_{IRSS}(t) \end{cases} \quad (9)$$

b) Short-term Fast Fourier Transform: The humanear performs a time-frequency transformation which is implemented in NPESQ by a short-term FFT with a Hanning window. The power spectra–the sum of the squared real and squared imaginary parts of the complex FFT components–are stored in separate real valued arrays for the original and degraded signals. Through FFT transform, the time domain into the frequency domain:

$$\begin{cases} X_{WIRSS}(t)_n \Rightarrow FFT \Rightarrow X_{WIRSS}(k)_n \\ Y_{WIRSS}(t)_n \Rightarrow FFT \Rightarrow Y_{WIRSS}(k)_n \end{cases} \tag{10}$$

Then calculate the power spectral density. The $PX_{WIRSS}(k)_n$ and $PY_{WIRSS}(k)_n$ are the power spectral of reference and degraded signal.

$$\begin{cases} PX_{WIRSS}(k)_n = (\operatorname{Re} X_{WIRSS}(k)_n)^2 + (\operatorname{Im} X_{WIRSS}(k)_n)^2 \\ PY_{WIRSS}(k)_n = (\operatorname{Re} Y_{WIRSS}(k)_n)^2 + (\operatorname{Im} Y_{WIRSS}(k)_n)^2 \end{cases} \tag{11}$$

c) Frequency spectrum is mapped to ERB-spectrum: The calculation process of ERB spectrum is mainly consists of two steps:The first step is to divide the entire auditory frequency range into a series of frequency band by using a certain length of the ERB scale

The second step is to calculate the center frequency and sound pressure level of each band by using the energy center of gravity method.

The formula of calculate the center frequency is:

$$f_c = \frac{\sum_{i=1}^{n} f_i \cdot Y_i}{\sum_{i=1}^{n} Y_i} \tag{12}$$

where $\boldsymbol{f_c}$ on behalf of the center frequency of the band, n is the total number of the spectrum in the band, $\boldsymbol{f_i}$ is the *i-th* spectrum within the frequency band, $\boldsymbol{Y_i}$ is the power spectrum amplitude of the *i*-th spectral line.

The formula of calculate the sound pressure level of each band is:

$$SPL_c = 10\lg\frac{\sum_{i=1}^{n} Y_i}{2P_{ref}^2} \tag{13}$$

where SPL_c is the sound pressure level, P_{ref} is the reference sound pressure.

The center frequency f_c and the frequency band SPL_c are composed of the ERB spectrum. ERB spectrum as the input of model can be carried out loudness calculation. The warping function that maps the frequency scale in Hertz to the pitch scale in ERB does not exactly follow the values given in the literature. The resulting signals are known as the pitch power densities $PPX^N{}_{WIRSS}(f)_n$ and $PPY^N{}_{WIRSS}(f)_n$.

d) ERB-spectrum is mapped to loudnes The ERB-spectrum is mapped to (Sone) loudness scale, including a frequency-dependent threshold and exponent. This gives the perceived loudness in each time-frequency cell.

The resulting two-dimensional arrays $LX^N(f)_n$ and $LY^N(f)_n$ are called loudness densities.

C. Cognitive Modeling

a) Calculation of the disturbance density: The difference of the two signals loudness density is calculated. When the difference is positive, the degraded signal to introduce some components. When the difference is negative, the loss of some of the components. This difference is known as the original disturbance density. After this process the value of the raw disturbance density is moved to The direction of the absolute value decreases, with the distance of the masking threshold.

Due to the masking effect of the human ear, we need to process each time-frequency component to get a disturbance density $D(f)_n$.

b) The processing of asymmetry: The asymmetry factor is calculated from a stabilized ratio of the ERB spectral density of the degraded to the reference signals in each time-frequency cell.

The asymmetrical disturbance density $DA(f)_n$ is calculated by multiplication of the disturbance density $D(f)_n$ with an asymmetry factor.

c) Aggregation of the disturbance densities: The disturbance values are aggregated using an L_p norm, which calculates a non-linear average using the following formula:

$$D_n^N = M_n \sqrt[3]{\sum_k \left(\left|D(f)_n\right| W_f\right)^3} \tag{13}$$

$$DA_n^N = M_n \sqrt[3]{\sum_k \left(\left|DA(f)_n\right| W_f\right)^3} \tag{14}$$

with M_n a multiplication factor, resulting in an emphasis of the disturbances that occur during silences in the original speech fragment, and W_f a series of constants proportional to the width of the modified ERB bins. These aggregated values, $D^N{}_n$ and $DA^N{}_n$, are called frame disturbances.

d) Realignment of bad intervals: Consecutive frames with a frame disturbance above a threshold are called bad intervals. Each bad interval is realigned and the disturbance is recalculated. For each frame, if the realignment results in a lower disturbance value, the new value is used. Then aggregation of the disturbance within split second intervals, to calculate the average symmetric disturbance d_{SYM}^N and the average asymmetric disturbance d_{ASYM}^N.

e) Computation of the NPESQ score: A combination of two parameters symmetric disturbance d_{SYM}^N and asymmetric disturbance d_{ASYM}^N shows a good balance between accuracy of prediction and ability of generalization. The final NPESQ score is a linear combination of the d_{SYM}^N and d_{ASYM}^N.

V. EXPERIMENT ANALYSIS

A. Experiment Parameter Settings

The experiments of this paper are carried out on the speech database, which is suitable to PESQ and NPESQ algorithm. The speech database includes reference and various degraded speech signal. The signal storing format is 16-bit linear PCM. The sample rate of speech is 8 kHz. The speech contained one pairs of sentences separated by silence, totaling 8 s in duration. The temporal structure has 3 parts: leading and trailing silence, utterances, and silent intervals between utterances. The length of leading and trailing silence is about 0.5 s. The length of interval silence is about 1s.

Our experiments test includes 24 different conditions. We use different codecs to test our algorithm. We choose three types of speech codecs, including Enhanced Variable Rate Codec (EVRC), EVRC-B and Adaptive Multi-Rate (AMR) speech codec, which all are used in third generation (3G) communication systems. The AMR supports 8 speech encoding modes, we choose the lowest and the highest rate mode, that is AMR at the rate of 12.2 kbps and 4.75 kbps. The EVRC-B including 8 different channel rates, we choose the 5.8 kbps, 6.6 kbps and 9.3 kbps. Total, we use six codec modes: EVRC, AMR at the rate of 12.2kbps and 4.75 kbps, EVRC-B at 9.3kbps, 6.6kbps and 5.8 kbps.

B. Subjective MOS Test

Our experiments include subjective MOS tests and objective PESQ and NPESQ test.

Subjective speech quality is defined as the result of a subjective test, where a number of test persons listen to and judge the test material. The subjective MOS tests use 64 listeners, and each listeners test 54 sentence pairs. The Subjective MOS tests are standardized by ITU in order to ensure repeatability of the experimental results. The MOS (mean opinion score) scale is used to assess the quality of voice transmission. The specific implementation is given in ITU-T P.800 [21]

In a subjective MOS test, the test persons listen to every speech sample. After listening to each speech sample, thetest person shall grade the sample according to the following scale of Table I.

TABLE I. THE SCORE GRADE OF MOS

Score($\boldsymbol{W_i}$)	Quality Level	Description
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

The finally MOS score of a subjective test is calculated by the formula:

$$MOS = \frac{1}{N}\sum_{i=1}^{p} W_i N_i \qquad (15)$$

With N is the number of total votes, $\boldsymbol{N_i}$ is the number of a certain score, $\boldsymbol{W_i}$ is the score of every vote, *i* is the score of every grade, *p* is represent the total score grade, *p* is 5. Through the formula we can get the subjective MOS scores.

C. Objective PESQ and NPESQ Test

The Objective test use PESQ algorithm and NPESQ algorithm. The PESQ scores are obtained based on ITU P.862. using the identical speech executables as subjective MOS test. The process is overview in the section II. The NPESQ scores were calculated according to the algorithm of NPESQ, which the process is the section IV of this paper.

D. Experiment Result of Test

Our experiments use two ways to reflect the performance of the objective speech quality evaluation method NPESQ algorithm.

The first way is use the absolute error that is we compare the score difference of subjective MOS tests and objective tests to verify the accuracy of the NPESQ. The smaller the difference that the higher the accuracy.

The second ways is that we use the correlation coefficient between the subjective MOS scores and the objective MOS scores to reflect the result. A good objective quality measure algorithm should have a high correlation with subjective experiments.

In experiment 1, we direct to compare the score difference of the subjective and objective test. We use clean speech to test. We use six codec modes: EVRC, AMR at the rate of 12.2kbps and 4.75 kbps, EVRC-B at 9.3kbps, 6.6kbps and 5.8 kbps.

The experiment result of the test is shown in the below Table II. The scores of PESQ and NPESQ have been mapped from the objective scores. In the table, the $\Delta 1$ is the difference of PESQ scores and MOS scores. The $\Delta 2$ is the difference of NPESQ scores and MOS scores. From the data, we can find the $\Delta 2$ is smaller than $\Delta 1$.That is to say, the NPESQ scores are more close to the subjective MOS scores. So the accuracy of NPESQ is better than the PESQ algorithm. This conclusion also displayed in the Figure 3. From the Fig. 3, we can find the score of NPESQ is more close to the score of subjective MOS.

TABLE II. SCORE COMPARISON OF TEST UNDER CLEAN SPEECH

Codec	**MOS**	**PESQ**	**NPESQ**	$\Delta 1$	$\Delta 2$
AMR4.75kbps	3.692	3.279	3.421	0.666	0.286
AMR12.2kbps	4.212	4.151	4.218	0.160	0.080
EVRC	4.292	3.835	3.987	0.457	0.113
EVRCB9.3kbps	4.385	3.807	4.122	0.585	0.336
EVRCB6.6kbps	4.308	3.471	3.842	0.737	0.453
EVRCB5.8kbps	4.041	3.267	3.633	0.894	0.511

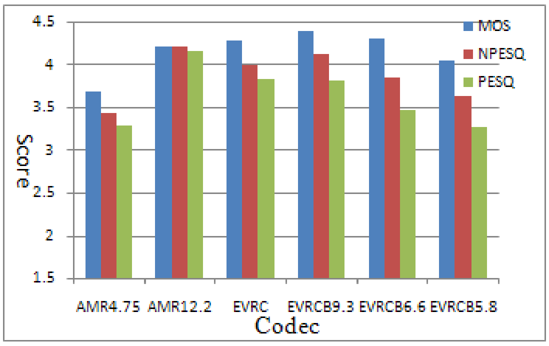


Figure 3. The score comparison of PESQ and NPESQ

In experiment 2, we use 24 conditions to test the speech, that is 6 codec modes, and 4 conditions of each codec. We use correlation coefficient as a comparison criterion. The correlation coefficient is calculated to measure the closeness of the fit between objective and subjective scores.

The tests are usually done under different condition. We calculate the correlation coefficient r on the score of each condition and the averaged score. The correlation coefficient is calculated with formula:

$$r=\frac{\sum_{i=1}^{N}(MOS_O(i)-\overline{MOS_O})(MOS_S(i)-\overline{MOS_S})}{\sqrt{\sum_{i=1}^{N}\sum_{i=1}^{N}(MOS_O(i)-\overline{MOS_O})^2\sum_{i=1}^{N}(MOS_S(i)-\overline{MOS_S})^2}} \tag{16}$$

In this formula, $MOS_S(i)$ is the condition subjective MOS score for the condition i, and $\overline{MOS_S}$ is the average over the MOS scores $MOS_S(i)$, $MOS_O(i)$ is the mapped PESQ and NPESQ score for condition i, and $\overline{MOS_O}$ is the average of the predicted objectives scores $MOS_O(i)$.

We compute the coefficient of PESQ and NPESQ under the test condition of AMR codec, EVRC codec and EVRCB codec. The result is shown in the TABLE Ⅲ below. From the table, we can find the correlation between the subjective MOS and objective NPESQ is higher than the correlation between the subjective MOS and objective PESQ. The correlation coefficient is respectively improved under the codec of AMR, EVRC and EVRC-B. So it is indicating that the accuracy of NPESQ algorithm is higher than the PESQ.

TABLE III. CORRELATION COEFFICIENT OF PESQ AND NPESQ

Codec	PESQ	NPESQ
AMR	r1=0.8565	r2=0.9335
EVRC	r3=0.8985	r4=0.9390
EVRC-B	r5=0.8978	r6=0.9125

The test condition and the scores of AMR codec, is shown in TABLE IV. In the table the BN is represent the background noise. The AMR codec has 8 encode rates; we choose the lowest and the highest rate. And choose the background noise under 0 dB, -6 dB and -16 dB. Figure 4 shown the score result of the subjective MOS test and objective test of AMR codec under different test conditions. From the Figure 4, we can find the score of NPESQ is more close to the score of subjective MOS.

TABLE IV. TEST CONDITION AND RESULT OF AMR

No.	Condition	MOS	PESQ	NPESQ
1	AMR4.75k+Clean Speech	3.692	3.279	3.421
2	AMR4.75k+BN_0dB	1.854	2.601	2.711
3	AMR4.75k+BN_-6dB	2.662	2.851	2.889
4	AMR4.75k+BN_-16dB	3.554	3.262	3.512
5	AMR12.2k+Clean Speech	4.212	4.151	4.218
6	AMR12.2k+BN_0dB	3.177	3.443	3.235
7	AMR12.2k+BN_-6dB	3.104	3.736	3.254
8	AMR12.2k+BN_-16dB	3.979	4.014	4.114

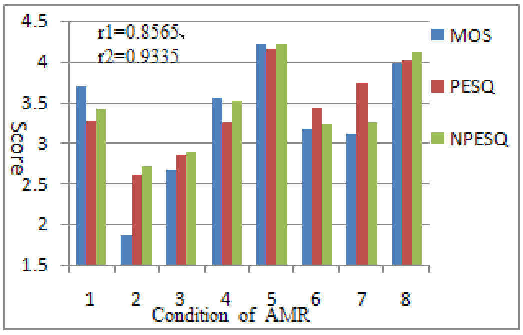


Figure 4. Score of subjecive and objective for AMR

The subjective MOS score and objective score of EVRC codec is shown in the TABLE Ⅴ. Figure 5 shown the score result under different 4 test conditions.

TABLE V. TEST CONDITION AND RESULT OF EVRC

No.	Condition	MOS	PESQ	NPESQ
1	EVRC+Clean Speech	4.292	3.835	3.987
2	EVRC+BN_0dB	2.156	2.337	2.516
3	EVRC+BN_-6dB	3.321	2.681	2.982
4	EVRC+BN_-16dB	4.191	3.187	3.459

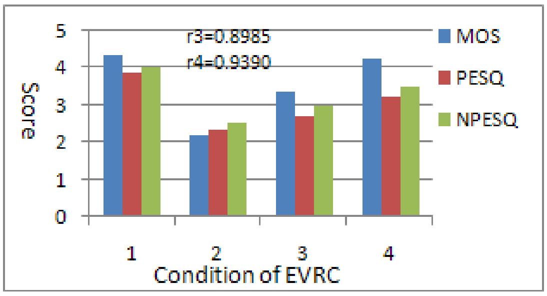


Figure 5. Score of subjecive and objective for EVRC

The EVRC-B is test under 12 differernt conditions. Each condition and the scores is shown in TABLE Ⅵ. We choose three rate of EVRC-B codec. Figure 6 shown the score result of the subjective MOS test and objective test of EVRC-B codec under different test conditions. From the Figure 6, we can find the score of NPESQ is more close to the score of subjective MOS.

In conclude, the absolute error scores of the NPESQ is lower than the PESQ. That is the smaller difference between NPESQ score and subjective MOS score. So the higher the accuracy of NPESQ is. The correlation coefficient of NPESQ is higher than the PESQ, under the

TABLE VI. TEST CONDITION AND RESULT OF EVRC-B

No.	Condition	MOS	PESQ	NPESQ
1	EVRCB_9.3kbps+Clean Speech	4.385	3.807	4.122
2	EVRCB_9.3kbps+BN_0dB	2.947	2.344	2.214
3	EVRCB_9.3kbps+BN_-6dB	2.927	2.633	2.741
4	EVRCB_9.3kbps+BN_-16dB	3.697	3.122	3.495
5	EVRCB_6.6kbps+Clean Speech	3.308	3.471	3.842
6	EVRCB_6.6kbps+BN_0dB	2.041	2.206	2.102
7	EVRCB_6.6kbps+BN_-6dB	2.979	2.509	2.721
8	EVRCB_6.6kbps+BN_-16dB	3.541	2.954	3.441
9	EVRCB_5.8kbps+Clean Speech	4.041	3.267	3.633
10	EVRCB_5.8kbps+BN_-0dB	1.895	2.204	2.018
11	EVRCB_5.8kbps+BN_-6dB	2.771	2.509	2.664
12	EVRCB_5.8kbps+BN_-16dB	3.593	3.185	3.397

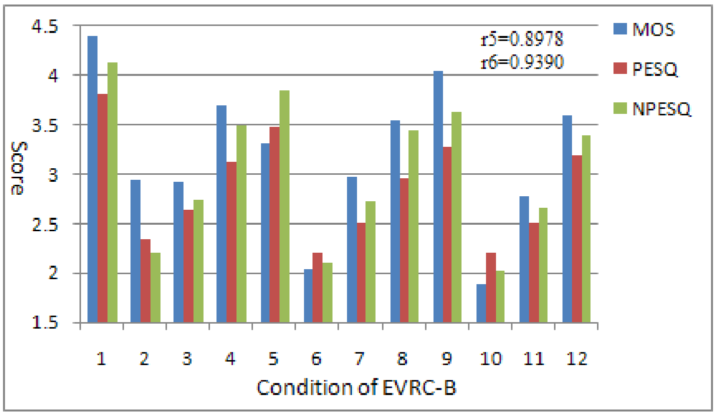


Figure 6. Score of subjecive and objective for EVRCB

codec of AMR, EVRC and EVRC-B. This is also explaining the accuracy of NPESQ is higher than the PESQ.

VI. Conclusion

On the study of voice QoE in communication system, the objective speech quality assessment is more useful and easier implement than the subjective speech quality assessment method. The PESQ algorithm is a well known objective speech quality assessment method for the voice QoE evaluation. But through our research, we find that the PESQ algorithm has limitations for evaluating speech quality. This paper introduces a novel algorithm for objective assessment of speech quality. The new algorithm use the ERB-scale frequency takes place of the Bark-scale frequency of PESQ. We give the calculation process of the NPESQ algorithm. And we do experiments to verify the new algorithm. Our experiments include subjective MOS tests and objective PESQ and NPESQ test. The experiment results indicate that the NPESQ scores are more close to the subjective MOS scores. Moreover, the correlation coefficient of NPESQ is higher than the PESQ. So the accuracy of NPESQ algorithm is better than the PESQ algorithm. We believe that NPESQ is suitable for many applications in assessing the speech quality of voice communication system.

References

[1] Abhishek Bhattacharya and Wanmin Wu “ Quality of experience evaluation of voice communication: an affect-based approach,” Human-centric Computing and Information Sciences, July 2012.

[2] ITU-T Rec. P.10/G.100 (incl. Amendment 2), “ Vocabulary for Performance and Quality of Service,” 2008.

[3] ITU-T Study Group 12, “ Definition of quality of experience (QoE),” International Telecommunication Union, January 2007

[4] Linfen Sun, E.C.Ifeachor, “ Voice quality prediction models and their application in VOIP networks,” IEEE Transaction on Multimedia, vol.8, pp.809-820, August 2006.

[5] T. Hoßfeld et al., “ Testing the IQX Hypothesis for Exponential Interdependency between QoS and QoE for Voice Codecs iLBC and G.711,” 18th ITC Specialist Seminar Quality of Experience, Karlskrona, Sweden, May 2008.

[6] P. Mowlaee, R. Saeidi, M.G. Christensen and R. Martin “ Subjective and objective quality assessment of single-

channel speech separation algorithms," ICASSP 2012 IEEE International conference on , pp.69-72
[7] J. Ma and P. C. Loizou, " SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech," Speech Communication, vol. 53, no. 3, pp. 340 – 354, 2011.
[8] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, " An algorithm for intelligibility prediction of time-frequency weighted noisy speech," Audio, Speech, and Language Processing, IEEE Transactions on, vol.19, no. 7, pp. 2125 –2136, Sept. 2011.
[9] ITU-T Recommendation P.862. " Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-To-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," February 2001.
[10] Whai-En chen, " Real-Time VoIP Measurement for Mobile Devices," IEEE, Systems Journal,Volume5, Issue 4,pp. 538-544,Dec.2011.
[11] S.Pennock, " Accuracy of the Perceptual Evaluation of Speech Quality (PESQ) algorithm," in Proc. Measurement of Speech and Audio Quality in Networks, Prague, Czech Republic, May 2002.
[12] Ditech Networks whitepaper, " Limitations of PESQ for Measuring Voice Quality in Mobile and VoIP Networks," 2007.
[13] ITU-T study group 12, " PESQ Limitations for EVRC Family of Narrowband and Wideband Speech Codecs," QUALCOMM Inc.,2008.
[14] Moore B C J, Glasberg B R. " A model of loudness applicable to time-varying sounds," J. Audio Eng. Soc, 2002,50(5):331-341.
[15] B.R. Glasberg and B.C.J. Moore, " Derivation of auditory filter shapes from notched-noise data", Hearing Research, Vol. 47, Issues 1-2, p. 103-138, 1990.
[16] J. Zhou, X. She, and L. Chen, " Source and channel coding adaptation for optimizing voip quality of experience in cellular systems, " in Wireless Communications and Networking Conference (WCNC), 2010 IEEE, April 2010, pp. 1 – 6
[17] Fitzpatrick, J, " An E-Model based adaptation algorithm for AMR voice calls," Wireless Days, 2011
[18] 3GPP, "Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions," 3GPP, Tech. Rep. 26.090, Mar. 2011.
[19] Fastl H, Zwicker E. " Psychoacoustics Facts and models," Berlin Heidelberg New York: Springer, 2007.
[20] B.C.J. Moore and B.R. Glasberg, " Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," Journal of the Acoustical Society of America 74: 750-753, 1983.
[21] ITU-T Recommendation P.800. " Methods for Subjective Determination of Transmission Quality," August 1996

Weiwei Zhang, is currently a Ph.D. candidate with Wireless Theory and Technology laboratory (WT&T) of Beijing University of Posts and Telecommunications (BUPT), China. Her research interests include wireless networking, VoIP networking, Quality of experience (QoE), Speech coding. Email:zweiwei506@gmail.com

Yongyu Chang, is a professor with WT&T of Beijing University of Posts and Telecommunications. As director or a key member, she has made remarkable achievements by leading or participating in many research programs, some of them are supported by government such as China National 973 Program and China National 863 Program, some are company-sponsored projects. She has published more than 100 papers in major journals in China or in high level international conferences, as well as other 9 books. She also applied for 5 patents. Her current research interest is focused on mobile communications.

Yitong Liu, is a lecturer with Wireless Theory and Technology laboratory, School of Information and Communication Engineering, of Beijing University of Posts and Telecommunications, Her current research interests include wireless communication and Quality of experience (QoE).

Leilei Xiao, received his B.S. degree from Beijing University of Posts and Telecommunications (BUPT), China, in 2011. He is currently studying for a M.S. degree in the School of Information and Communication Engineering, BUPT. His research interests lie in wireless communication and Quality of experience (QoE).

A Combined DWT and DCT Watermarking Scheme Optimized Using Genetic Algorithm

Cheng Mingzhi
College of Information & Mechanical Engineering, Beijing Institute of Graphic Communication, Beijing, China
Email: chengmz@bigc.edu.cn

Li Yan*, Zhou Yajian and Lei Min
School of Computing, Beijing University of Posts and Telecommunications, Beijing, China
Email: liyanphoto@yahoo.com.cn, yajian@bupt.edu.cn, leimin@bupt.edu.cn

***Abstract*—To protect the copyright of digital image, this paper proposed a combined Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) based watermarking scheme. To embed the watermark, the cover image was decomposed by a 2-level DWT, and the HL2 sub-band coefficient was divided into 4x4 blocks, then the DCT was performed on each of these blocks. The watermark bit was embedded by predefined pattern_0 or pattern_1 on the middle band coefficients of DCT. After watermark insertion, inverse DCT was applied to each of the 4x4 blocks of HL2 sub-band coefficient, and inverse DWT was applied to obtain the watermarked image. For watermark extraction, the watermarked image, which may be attacked by various image attacks, was decomposed with 2-level DWT and DCT similarly as watermark embedding process, then correlation between middle band coefficients of block DCT and the predefined pattern (pattern_0 and pattern_1) was calculated to decide whether a bit 0 or a bit 1 was embedded. Genetic algorithm was used for embedding and extraction parameters optimization. Optimization is to maximize PSNR of the watermarked image and NCC of the extracted watermark. Experiment results show that the proposed scheme in this paper is robust against many image attacks, and improvement can be observed when compared to other existing schemes.**

***Index Terms*—watermark, DWT, DCT, genetic algorithm, image**

I. Introduction

The development of Internet and computer has changed our society. Digital multimedia can be copied and transmit easily and conveniently. Images, audio and video can be edited and modified illegally by some advanced processing software. The protection of copyright of digital multimedia becomes an important issue. Researchers are aware of this problem and some solutions have been proposed.

Digital watermarking techniques can provide detection and protection of copyrighted digital multimedia. Watermarking technique embeds specific information or copyright codes called watermark into the original media invisibly. In the case of dispute over the ownership of the media, embedded watermark can be extracted and it can be used to identify the ownership.

Digital watermarking algorithms can be classified into different types according to their embedding domain and characteristics.

Based on the resistance to attacks, it can be divided into robust watermarking [1-5], fragile watermarking [6-8] and semi-fragile watermarking algorithms [9-11]. Robust watermarking is designed to resist intentional and unintentional destroy to the watermark. Conversely, fragile watermark is designed to be easily destroyed by any kind of manipulations on the protected media. It is used for strict authentication. On the other hand, semi-fragile watermarking combines characteristics of fragile and robust watermarking techniques. In semi-fragile watermarking scheme, a watermark is embedded such that it can undergo some specific image processing operations while it is still possible to detect malevolent alterations to the media.

Based on embedding domain, digital watermarking algorithms can be classified into spatial domain [12, 13] and transform domain [14-21].Embedding watermarks in spatial domain are conceptually simple and have very low computational complexities, while algorithms based on the transform domain are more robust. Some of the transform domain methods available in the literature are listed as follow: methods based on discrete cosine transform (DCT) [14, 15], discrete wavelet transform (DWT) [16, 17], discrete Hadamard transform [18, 19], singular value decomposition [20] and discrete Fourier transform [21].

In [14], Zhang et al. proposed an improved watermarking algorithm based on DCT. First, both rows and ranks of the watermark were extended by using the proposed method before the embedding stage. After expansion, Sine chaotic system was employed in encrypting the watermark. In the embedding stage, an effective and adaptive embedding method was proposed to embed the watermark into the blocked DCT coefficients. Experimental results demonstrated that the proposed algorithm works well in resisting both geometry attack and noise attack. It also worked well in recovering

Corresponding author: Li Yan.

doi:10.4304/jmm.8.3.299-305

the watermark after image suffered from JPEG compression.

In [22], Wang et al. proposed a wavelet-tree-based blind watermarking scheme for copyright protection. The wavelet coefficients of the host image were grouped into so-called super trees. The watermark was embedded by quantizing super trees. The trees were so quantized that they exhibit a large enough statistical difference, which will later be used for watermark extraction. Each watermark bit was embedded in perceptually important frequency bands, which renders the mark more resistant to frequency based attacks. Also, the watermark was spread throughout large spatial regions. This yields more robustness against time domain geometric attacks.

In this paper, we propose a robust watermarking scheme based on combined DWT and DCT. In the embedding scheme, firstly, the original image is decomposed using 2-level DWT, then a 4x4 block DCT decomposing is performed on the HL2 coefficients of DWT. After that, the watermark is embedded in the middle band coefficients of DCT. There are two predefined watermark pattern, according to the bit of the watermark image is 0 or 1, one of the watermark patterns is scaled by the gain factor and added to the middle band coefficients of DCT. Then the inverse DCT and DWT is applied to obtain the watermarked image. In the extraction scheme, the DWT and DCT transform are applied similarly as in the embedding scheme, then the correlation between the middle band coefficients of DCT and the two predefined watermark pattern is calculated to decide whether a bit 0 or bit 1 is embedded in the image. Moreover, we use genetic algorithm to find the optimal embedding parameters including the gain factor and frequency bands for watermark embedding into the DCT coefficients, which can improve robustness and image quality of the watermarked image.

The rest of the paper is organized as follows. The genetic algorithm principle is described in Section 2. The proposed method is presented in Section 3. Then, experimental results are presented in Section 4. Finally, we conclude this paper in Section 5.

Your goal is to simulate the usual appearance of papers in a Journal of the Academy Publisher. We are requesting that you follow these guidelines as closely as possible.

II. Genetic Algorithm Principles

Genetic algorithm (GA) is one of the most widely used artificial intelligent techniques for optimization. GA was first developed by John Holland [23]. GA is stochastic searching algorithm based on the mechanisms of natural selection and genetics, and is very efficient in searching for global optimum solutions.

GA process can be described based on three functional units. They are selection, crossover and mutation operations. The implementation of GA is briefly summarized as follows. Firstly, a set of number strings called population, which is composed of a group of chromosomes, are randomly generated to represent the solutions of the optimization problem. The elements in the strings, called “genes”, are adjusted to minimize or maximize the fitness value generated by the fitness function. The chromosome with good fitness value has greater chance to survive during the evolution process. The objective function is problem specific; it should be carefully defined by the designers. Then, the selection operator performs a selection function on the chromosomes. The chromosomes are copied from one set to the next according to their fitness values. Next, the crossover operator chooses pairs of chromosomes at random and produces two new offspring. A crossover point is selected between the two parents’ chromosomes. Two new offspring are produced by exchange the fractions of two parents’ chromosome after the crossover point. After that, the mutation operator randomly changes the value of bits in a chromosome. It can keep GA from converging too fast. The mutation rate should be low. Similarly, the fitness value of the offspring is calculated as their parents in order to replace the chromosomes in the current generation. The GA cycle is repeated until maximum number of generations is reached or the objective value is below a threshold. With the fundamental concepts in GA, we are able to design an optimized DWT and DCT based watermarking scheme with the aid of GA.

The genetic algorithm toolbox developed by the University of Sheffield is one of the most popular used genetic algorithm toolbox, and it is used in this paper.

III. Proposed Method

In the proposed method, both DWT and DCT are performed on the image. For DCT, there are three frequency sub-bands: low band frequency, middle band frequency and high band frequency. The low band frequency contains most of the signal energy and most important visual parts of the image. If watermark embeds in low band frequency, the imperceptibility will be poor although it will be robust to many image attacks. Also, the watermark cannot be embedded into high band frequency, because high band frequency may be removed by JPEG compression and noise attacks. Therefore, the watermark should be better embedded into the middle band frequency to trade-off between robustness and imperceptibility [24-27].

DWT decompose an image into four non-overlapping multi-resolution sub-bands LL1, HL1, LH1 and HH1. LL1 is the approximation coefficients, HL1, LH1 and HH1 are details coefficients (horizontal, vertical, and diagonal, respectively). To obtain 2-level DWT coefficients, the sub-band LL1 is further decomposed. And as mentioned in [28], HL2 is selected to embed watermark in this paper.

A. Watermark Embedding

The watermark embedding procedure is represented in Figure 1, followed by a detailed explanation.

Step 1: Perform DWT on the cover image to decompose it into four non-overlapping sub-bands: LL1, HL1, LH1, and HH1.

Step 2: Perform DWT again on the HL1 sub-band to get four smaller sub-bands: LL2, HL2, LH2, and HH2 as show in Figure 2.

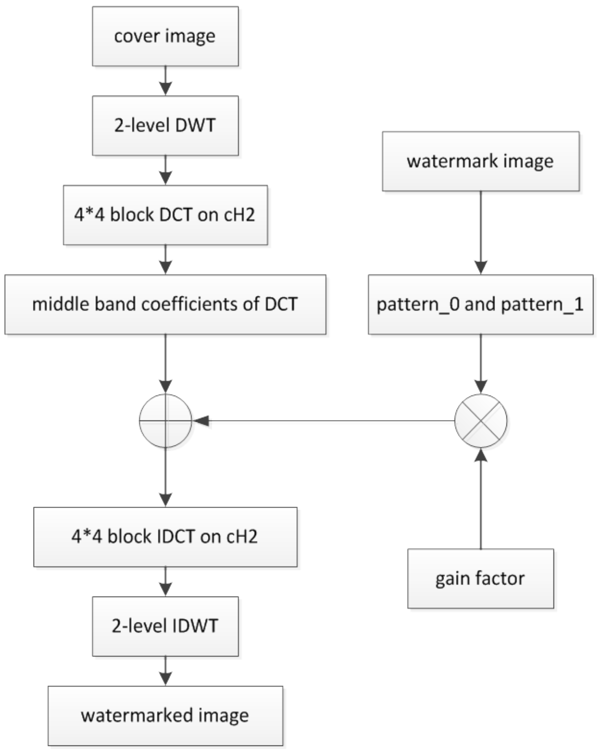


Figure 1 Embedding scheme

Step 3: Divide the sub-band HL2 into 4x4 blocks.

Step 4: Perform DCT to each block in the HL2 sub-band.

Step 5: Generate a pseudorandom sequence or called "pattern_0" by a key, and inverse each bit of pattern_0 to get pattern_1. The length of pattern_0 and pattern_1 is equal to the number of selected middle band coefficients of DCT. Where pattern_0 is used to embed the watermark bit 0 and pattern_1 is used to embed the watermark bit 1.

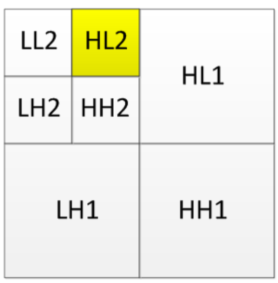


Figure 2 2-level DWT sub-bands

Step 6: Embed pattern_0 and pattern_1 with a gain factor in the middle band coefficients of DCT as follow:

$$X' = \begin{cases} X + \alpha * pattern\_0 & if\ watermark\ bit = 0 \\ X + \alpha * pattern\_1 & if\ watermark\ bit = 1 \end{cases} \quad (1)$$

Where, α is the gain factor, X is the selected middle band coefficients of DCT before embedding, and X' is middle band coefficients of DCT after embedding.

Step 7: Perform the inverse DCT on each 4x4 block after pattern_0 and pattern_1 are added to the selected middle band coefficients.

Step 8: Perform the inverse DWT on the image to produce the watermarked image.

B. Watermark Extraction

The watermark extraction scheme is show in Figure 3, and the details are described as follows.

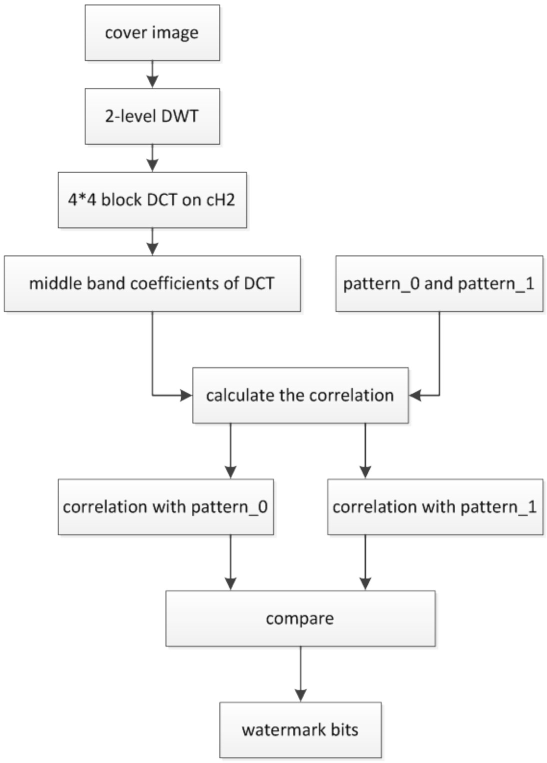


Figure 3 Extraction scheme

Step 1: Perform DWT on the cover image to decompose it into four non-overlapping sub-bands: LL1, HL1, LH1, and HH1.

Step 2: Perform DWT again on the HL1 sub-band to get four smaller sub-bands: LL2, HL2, LH2, and HH2 as show in Figure 2.

Step 3: Divide the sub-band HL2 into 4x4 blocks.

Step 4: Perform DCT to each block in the HL2 sub-band.

Step 5: Regenerate the two pseudorandom sequences using the same key which used in the watermark embedding scheme.

Step 6: For each block in the sub band HL2 calculate the correlation between the middle band coefficients and the two generated pseudorandom sequences. If the correlation with pattern_0 is higher than the correlation with pattern_1, then the extracted watermark bit is considered to be 0, otherwise the extracted watermark is considered to be 1.

Step 7: The watermark image is reconstructed according to the watermark bits.

C. Performance Improvement using GA

Watermarking problem can be viewed as an optimization problem. There are three conflicting requirements: robustness, imperceptibility and capacity. NCC indicates the amount of similarity between the original watermark image and the extracted watermark image. It reflects the robustness of a watermark scheme. Imperceptibility of a watermark scheme can be described by PSNR. PSNR is inversely proportional to the amount of distortion introduced to the host image. Capacity describes the amount of bits can be embedded into the cover image. So, in a good watermarking scheme, all the three elements, PSNR, NCC and capacity should be as large as possible. At the other hand, the three elements are related with each other. Maximization of PSNR decreases the value of other two elements, and so does NCC and capacity. So, GA is used to find the optimum value of parameters used in the watermarking scheme to obtain the trade-off between the three conflicting elements. However, capacity is often application specific. In this paper, capacity is 1024 bits, PSNR is targeted to 40 dB and NCC is targeted to 1.0. So the fitness function abs ((40-PSNR) + (1-NCC)) is defined for the GA optimize procedure.

PSNR can be defined as follows:

$$\mathrm{PSNR} = 10 * log_{10} * \frac{255*255}{1/(M*N)\sum_{x=1}^{M}\sum_{y=1}^{N}[f(i,j)-g(i,j)]^2} dB \quad (2)$$

There, M and N are the height and width of the image, respectively. $f(i,j)$ and $g(i,j)$ are the pixel value of cover image and the attacked image, respectively. The NCC is computed after the watermark is extracted from the attacked image. NCC can be defined as follow:

$$\mathrm{NCC} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[w(i,j)-w_{mean}][v(i,j)-v_{mean}]}{\sqrt{(\sum_{i=1}^{m}\sum_{j=1}^{n}[w(i,j)-w_{mean}]^2)(\sum_{i=1}^{m}\sum_{j=1}^{n}[v(i,j)-v_{mean}]^2)}} \quad (3)$$

There, m and n are the height and width of the watermark, respectively. $w(i,j)$ and $v(i,j)$ are the pixel value of the original watermark and the extracted watermark respectively.

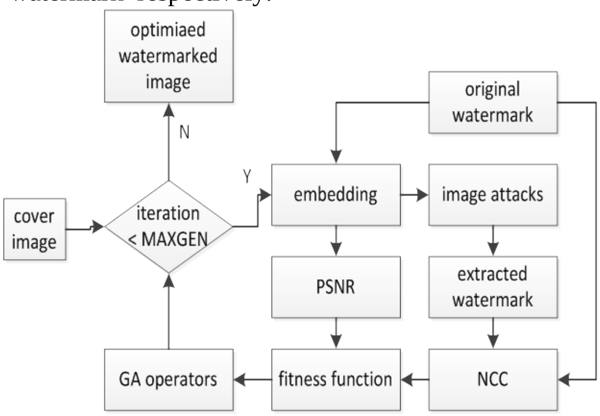


Figure 4 GA-based watermark embedding and extraction

Figure 4 shows the flow chart for GA-based watermark embedding and extraction. If the number of iteration is less than the defined maximal generation of GA, embed the original watermark into the cover image. After embedding, PSNR of the watermarked image is calculated, and the watermarked image may go through one or more image attacks, such as JPEG compression, Gaussian filter, cropping, rotation and so on. Then, extracted the watermark from the attacked image. Next, NCC between the original watermark and the extracted watermark is calculated. With PSNR and NCC, fitness value is obtained by the fitness function, and then GA operators (include selection, crossover, and mutation) yield the next generation parameters. This loop goes until the iteration of GA is equal to the defined maximum generation.

IV. EXPERIMENTAL RESULTS

Lena image of size 512x512 is used as cover image and is shown in Figure 5. The watermark image of size 32x32 is shown in Figure 6. The watermarked image is shown in Figure 7. GA is used to find the optimum values of embedding gain factor and which four middle band coefficients of DCT to embed. To test the performance of the proposed watermarking scheme, MATLAB 7.12 and the genetic algorithm toolbox developed by the University of Sheffield is used. The GA parameters used in this paper are listed as follow: population size is 20; crossover rate is 0.8, mutation rate is 0.0056. Various attacks used to test the robustness of the watermarking scheme proposed in this paper are JPEG compression (with different quality factors), median filter, Gaussian filter, average filter, image sharpening, 50% and 200% scaling, 20%, 25% and 40% cropping, Gaussian noise, salt & pepper noise, row & column copying, row & column blanking, and rotation with different degrees. Moreover, the search space for GA optimize are: 12 to 50 for embedding gain factor and 1 to 15 for the DCT coefficients to embed.

With GA, gain factor used in the embedding scheme are 21 and the four DCT middle band coefficients are 3, 4, 5 and 8. Table 1 shows extracted watermark from attacked image and NCC values. PSNR value of the watermarked image without attack is 41.5213dB. Table 2 shows the performance comparison with existing schemes [22, 29, 30]. The proposed scheme in this paper is better than most of the schemes shown in the table.

Figure 5 Cover image

Figure 6 Watermark

Figure 7 Watermarked image, PSNR = 41.5213 dB

TABLE I. EXTRACTED WATERMARK FROM ATTACKED IMAGE AND NCC VALUES

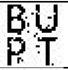
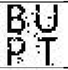
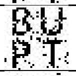
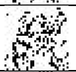
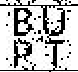
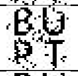
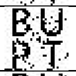
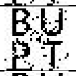
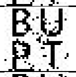
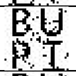
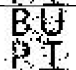
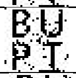
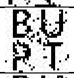
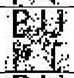
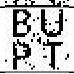
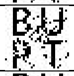
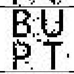
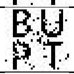
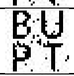
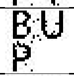
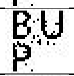
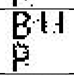
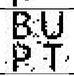
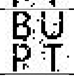
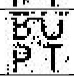
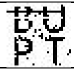
Attacks		Extracted watermark	NCC
JPEG	QF=90		0.9783
	QF=70		0.9721
	QF=50		0.9342
	QF=30		0.8943
	QF=20		0.8095
	QF=10		0.3675
rotation	10°		0.7817
	30°		0.7580
	0.25°		0.8235
	0.75°		0.7995
	1°		0.8390
	-0.25°		0.8242
	-0.75°		0.7986
	-1°		0.8394
Median filter	3x3		0.9032
	4x4		0.6387
Gaussian filter (3x3)			0.9117
Average filter (3x3)			0.8432
Image sharpening			0.9520
Scaling	50%		0.8667
	200%		0.9245
Cropping	25%		0.9364
	20%		0.9311
	40%		0.8731
Gaussian noise (0.001)			0.8895
Salt & pepper noise(0.001			0.8990
Row & Column copying (10-30, 40-70, 100-120)			0.7962
Row & column Blanking (30, 70, 120)			0.7920

TABLE II. COMPARISON OF THE PROPOSED METHOD WITH THE METHODS OF LIEN ET AL. [22], LI ET AL. [29]AND WANG ET AL. [30]

Type of attacks	Wang et al. [30] (PSNR=38.2 dB)	Li et al. [29] (PSNR = 40.6 dB)	Lien et al. [22] (PSNR = 41.54 dB)	This paper (PSNR = 41.52 dB)
Median filter(3x3)	0.51	0.35	0.79	0.90
Median filter(4x4)	0.23	0.26	0.51	0.64
JPEG, QF= 10	NA	0.15	0.17	0.37
JPEG, QF = 20	NA	0.34	0.61	0.81
JPEG, QF= 30	0.15	0.52	0.79	0.89
JPEG, QF= 50	0.28	0.52	0.89	0.93
JPEG, QF = 70	0.57	0.63	0.97	0.97
JPEG, QF = 90	1	0.78	1	0.98
Image Sharpening	0.46	0.38	0.88	0.95
Gaussian filter	0.64	0.70	0.84	0.91
Rotation 0.25°	0.37	0.46	0.53	0.82
Rotation 0.75°	0.26	0.36	0.16	0.80
Rotation 1°	0.24	0.33	0.07	0.84
Rotation - 0.25°	0.32	0.50	0.47	0.82
Rotation - 0.75°	0.24	0.29	0.10	0.80
Rotation - 1°	0.16	0.33	0.16	0.84
Cropping 25%	NA	0.61	0.92	0.94
Scaling 50%	NA	0.35	0.79	0.87

V. CONCLUSION

In this paper, a robust watermarking scheme based on DWT and DCT is presented, and is optimized with GA. HL2 sub-band of DWT coefficient is decomposed with 4x4 block DCT. In each block, the four middle band coefficients are used to embed the watermark. To embed a bit 0, the predefined pattern_0 is scaled by a gain factor and added to the middle band coefficient of the block DCT. To embed a bit 1, the predefined pattern_1 is done similarly as pattern_0 when embed a bit 0. To extract the watermark, correlation between the predefined pattern (pattern_0 and pattern_1) and the middle band coefficients of block DCT is calculated. If the correlation between pattern_0 and the middle band coefficients of block DCT is larger, then the extracted watermark bit is consider to be 0, otherwise, the extracted watermark bit is consider to be 1. GA is used to find the optimum value of gain factor and to find which four middle bands of coefficient of block DCT are used to embed the watermark. The PSNR and NCC of the proposed scheme are tested with Lena image and a binary watermark image of size 32x32. Compared with other methods [22, 29, 30], the proposed scheme is better than most of the schemes.

ACKNOWLEDGMENT

This work was supported in part by a grant from Beijing Natural Science Foundation under Grant No. 4122026,Scientific Research Common Program of Beijing Municipal Commission of Education under Grant No. KM201210015007, KM201210015006 and KZ201210015015, Institute Level Key Projects Funded by Beijing Institute of Graphic Communication under Grant No. E-a-2012-25 and E-a-2012-27, Beijing University of Posts and Telecommunications under Grant No. 2011RC0210 and 2012RC0217.

REFERENCES

[1] Lee Chin-Feng, Chen Kuo-Nan, Chang Chin-Chen. A hierarchical fragile watermarking with VQ index recovery. Journal of Multimedia, 2011, 6(3): 277-284.

[2] Chen Hongyuan, Zhu Yuesheng. A robust watermarking algorithm based on QR factorization and DCT using quantization index modulation technique. Journal of Zhejiang University-science C-computer & Electronics, 2012, 13(8): 573-584.

[3] Zhou Xinmin, Yu Jianping. Attack Model and Performance Evaluation of Text Digital Watermarking. Journal of Computers, 2010, 12(5): 1933-1941.

[4] Zhang, Q.; Li, Y.; Wei, X. An Improved Robust and Adaptive Watermarking Algorithm Based on DCT. Journal Of Applied Research And Technology, 2012, 10(3): 405-415.

[5] Chen Huang-Chi, Chang Yu-Wen, Hwang Rey-Chue. A watermarking technique based on the frequency domain. Journal of Multimedia, 2012, 7(1): 82-89.

[6] Di Martino, Ferdinando; Sessa, Salvatore. Fragile watermarking tamper detection with images compressed by fuzzy transform. Information Sciences, 2012, 195: 62-90.

[7] Xiao, Di; Shih, Frank Y. An improved hierarchical fragile watermarking scheme using chaotic sequence sorting and sub-block post-processing. Optics Communications, 2012, 285(10-11): 2596-2606.

[8] Huo, Yaoran; He, Hongjie; Chen, Fan. Alterable-capacity fragile watermarking scheme with restoration capability. Optics Communications, 2012, 285(7): 1759-1766.

[9] Hu Yanjun, Wang Guanjun, Cao Xinde, Yang Lei. A robust public-key watermarking algorithm based on contourlet ransform and its application. Journal of Software, 2011, 6(11): 2247-2254.

[10] Zhang Xiaoqiang, Zhu Guiliang, Wang Weiping , Wang Mengmeng. New Public-Key Cryptosystem Based on Two-Dimension DLP. Journal of Computers, 2012, 7(1): 169-178..

[11] Qi, Xiaojun; Xin, Xing. A quantization-based semi-fragile watermarking scheme for image content authentication. Journal Of Visual Communication And Image Representation, 2011, 22(2): 187-200.

[12] D. Lin and G. Liao. Swarm intelligence based fuzzy c-means clustering for motion vector selection in video watermarking. International Journal of Fuzzy Systems, 2008, 10(3): 185–194.

[13] P. Chan, M. Lyu, and R. Chin. A novel scheme for hybrid digital video watermarking: approach, evaluation and experimentation.*Circuits and Systems for Video Technology, IEEE Transactions*, 2005, 15(12): 1638–1649.

[14] Zhang Q. Li Y.; Wei, X. An Improved Robust and Adaptive Watermarking Algorithm Based on DCT. Journal Of Applied Research And Technology, 2012, 10(3): 405-415.

[15] Bi Hongbo, Zhang Yubo, Li Xueming. Video watermarking robust against spatio-temporal attacks. Journal of Networks, 2011, 6(6): 932-936.

[16] Yang Yu, Lei Min, Liu Huaqun, Zhou Yaiian, Luo Qun. A novel robust zero-watermarking scheme based on discrete wavelet transform. Journal of Multimedia, 2012, 7(4): 303-308.

[17] Hu Yuping, Wang Zhijian, Liu Hui, Guo Guangjun. A geometric distortion resilient image watermark algorithm based on DWT-DFT. Journal of Software, 2011, 6(9): 1805-1812.

[18] Aung Aye, Ng Boon Poh, Rahardja Susanto. A Robust Watermarking Scheme Using Sequency-ordered Complex Hadamard Transform. Journal Of Signal Processing Systems For Signal Image And Video Technology, 2011, 64(3): 319-333.

[19] Maity Santi P., Kundu Malay K. DHT domain digital watermarking with low loss in image information. Aeu-International Journal Of Electronics And Communications, 2010, 64(3): 243-257.

[20] Tsai Hung-Hsu, Jhuang Yu-Jie, Lai Yen-Shou. An SVD-based image watermarking in wavelet domain using SVR and PSO. Applied Soft Computing, 2012, 12(8): 2442-2453.

[21] Liu Yan, Zhao Jiying. A new video watermarking algorithm based on 3D DFT and Radon transform. Signal Processing, 2010, 90(2): 626-639.

[22] Wang S.H., Lin Y.P. Wavelet tree quantization for copyright protection watermarking. *IEEE Trans. Image Process*, 2004, 13(2): 154-165.

[23] Holland, J.H. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. Cambridge, MA: MIT Press: 1992.

[24] Chu W. DCT-Based Image Watermarking Using Subsampling . *IEEE Trans. Multimedia*, 2003, 5(1): 34-38.

[25] Lin, S. and C. Chin. A Robust DCT-based Watermarking for Copyright Protection. *IEEE Trans. Consumer Electronics*, 2000, 46(3): 415-421.

[26] Deng F., B. Wang. A novel technique for robust image watermarking in the DCT domain. *In Proc. of the IEEE 2003 Int. Conf. on Neural Networks and Signal Processing*. Nanjing: 2003, 2: 1525-1528.

[27] Wu C., W. Hsieh. Digital watermarking using zero tree of DCT. *IEEE Trans. Consumer Electronics*, 2000, 46(1): 87-94.

[28] Ehsan Vahedi, Reza Aghaeizadeh Zoroofi, Mohsen Shiva. Toward a new wavelet-based watermarking approach for color images using bio-inspired optimization principles. Digital Signal Processing, 2012, 22: 153–162.

[29] Li, E., Liang, H., Niu, X. An integer wavelet based multiple logo-watermarking scheme. Computer and Computational Sciences, *2006. IMSCCS '06. First International Multi-Symposiums on. Hanzhou*, Zhejiang: 2006, 10256-10260.

[30] Zhang Weiwei, Zhang Ru, Liu Xianyi, Wu Chunhua, Niu Xinxin. A video watermarking algorithm of H.264/AVC for content authentication. Journal of Networks. 2012, 7(8): 1150-1154.

Cheng Mingzhi was born 1974 in Hubei China. He received the B.S. degree in Communication Engineering from Chongqing University of Posts and Telecommunications in 1996 and the M.S. degree in 2002 and the Ph.D. degree in Cryptography from Beijing University of Posts and Telecommunications in 2010.

From 2010, he worked in Beijing Institute of Graphic Communication as a lecturer.

Dr. Cheng is working in packaging anti-counterfeiting and Digital Content Security.

Li Yan was born 1982 in Beijing China. He received the B.S. degree in information engineering from The PLA Information Engineering University in 2004, the M.S. degree in military communication from The PLA Information Engineering University in 2007.

Dr. Li is working on information security as a Ph.D candidate at Beijing University of Posts and Telecommunications.

Zhou Yajian was born 1971 in Shan Xi China. He received the B.S. degree in materials science and engineering from Beijing University of Aeronautics and Astronautics in 1993, the M.S. degree in materials science and engineering from Beijing University of Aeronautics and Astronautics in 1996 and Ph.D degree in communication engineering from Xidian University.

Dr. Zhou is working on information security as an associate professor in Beijing University of Posts and Telecommunications.

Lei Min was born 1979 in Jiangxi China. He received the B.S. degree of Science in Computer Science from Nanchang University in 1999, the M.S. degree in Computer software and theory from Beijing University of Posts and Telecommunications in 2002 and Ph.D degree in Cryptography from Beijing University of Posts and Telecommunications in 2011.

Dr. Lei is working on information security as a lecturer in Beijing University of Posts and Telecommunications.

Call for Papers and Special Issue Proposals

Aims and Scope.

Journal of Multimedia (JMM, ISSN 1796-2048) is a scholarly peer-reviewed international scientific journal published bimonthly, focusing on theories, methods, algorithms, and applications in multimedia. It provides a high profile, leading edge forum for academic researchers, industrial professionals, engineers, consultants, managers, educators and policy makers working in the field to contribute and disseminate innovative new work on multimedia.

The Journal of Multimedia covers the breadth of research in multimedia technology and applications. JMM invites original, previously unpublished, research, survey and tutorial papers, plus case studies and short research notes, on both applied and theoretical aspects of multimedia. These areas include, but are not limited to, the following topics:

- Multimedia Signal Processing
- Multimedia Content Understanding
- Multimedia Interface and Interaction
- Multimedia Databases and File Systems
- Multimedia Communication and Networking
- Multimedia Systems and Devices
- Multimedia Applications

JMM EDICS (Editors Information Classification Scheme) can be found at http://www.academypublisher.com/jmm/jmmedics.html.

Special Issue Guidelines

Special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call for Papers or by invitation, edited by guest editor(s). We encourage you to submit proposals for creating special issues in areas that are of interest to the Journal. Preference will be given to proposals that cover some unique aspect of the technology and ones that include subjects that are timely and useful to the readers of the Journal. A Special Issue is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

The following information should be included as part of the proposal:

- Proposed title for the Special Issue
- Description of the topic area to be focused upon and justification
- Review process for the selection and rejection of papers.
- Name, contact, position, affiliation, and biography of the Guest Editor(s)
- List of potential reviewers
- Potential authors to the issue
- Tentative time-table for the call for papers and reviews

If a proposal is accepted, the guest editor will be responsible for:

- Preparing the “Call for Papers” to be included on the Journal’s Web site.
- Distribution of the Call for Papers broadly to various mailing lists and sites.
- Getting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Instructions for Authors.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

Special Issue for a Conference/Workshop

A special issue for a Conference/Workshop is usually released in association with the committee members of the Conference/Workshop like general chairs and/or program chairs who are appointed as the Guest Editors of the Special Issue. Special Issue for a Conference/Workshop is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

Guest Editors are involved in the following steps in guest-editing a Special Issue based on a Conference/Workshop:

- Selecting a Title for the Special Issue, e.g. “Special Issue: Selected Best Papers of XYZ Conference”.
- Sending us a formal “Letter of Intent” for the Special Issue.
- Creating a “Call for Papers” for the Special Issue, posting it on the conference web site, and publicizing it to the conference attendees. Information about the Journal and Academy Publisher can be included in the Call for Papers.
- Establishing criteria for paper selection/rejections. The papers can be nominated based on multiple criteria, e.g. rank in review process plus the evaluation from the Session Chairs and the feedback from the Conference attendees.
- Selecting and inviting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Author Instructions. Usually, the Proceedings manuscripts should be expanded and enhanced.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

More information is available on the web site at http://www.academypublisher.com/jmm/.

(Contents Continued from Back Cover)