Estimating the Number of People in Crowded Scenes Based on Energy in Frequency Domain*

Guo-qing Yang and Rong-yi Cui**

Intelligent Information Processing Lab.
Department of Computer Science and Technology
Yanbian University
Yanji, Jilin, China
y.p.yu@sohu.com, cuirongyi@ybu.edu.cn

Abstract. A method of estimating the number of people in crowded scenes was proposed based on energy of the foreground image in the frequency domain. Firstly, the foreground image including the target area was obtained by background subtraction. And furthermore, with the second order wavelet decomposition of the foreground image, the energy of high frequency sub-band was computed. Finally, by analyzing the relation between population size and high frequency sub-band energy, the mathematical model was established and the model was solved by the least square method. The experimental results show that the proposed method is highly effective for estimating the number of moving people in crowded scenes.

Keywords: estimation of number of people, wavelet decomposition, energy in frequency domain, the least square method.

1 Introduction

With the development of human civilization and the advancement of society, crowd density estimation has become an important topic in the field of public safety. Recently, much attention has been paid to crowd monitoring and pre-warning in automatic surveillance systems. The traditional artificial way to safeguard crowd safety is using Closed Circuit Television (CCTV) by which specific objects and their behaviour could be monitored for a long period of time. However, a human observer might lose some information because monitoring crowds through CCTV is very laborious and cannot be performed for all the cameras simultaneously, and furthermore, may cause irreparable consequences.

In recent years, various methods for intelligent monitoring have emerged. Early work on crowd monitoring using image processing was reviewed by Davies et al. in 1995[1] by mainly analysing the crowd size and crowd motion estimation. Counting the number of pixels in foreground area and in edge of foreground, the "optimal"

* This work is supported by Sci. & Tech. Department of Jilin Prov. Grant#20050703-1.
** Corresponding author.

D. Zeng (Ed.): ICAIC 2011, Part II, CCIS 225, pp. 546–554, 2011.
© Springer-Verlag Berlin Heidelberg 2011

estimation of crowd density was achieved using linear regression training model and Kalman filter. Ref. [2] proposed Bayesian-Poisson regression for crowd counting and Ref. [3] adopted the foreground pixel area and edge pixels of foreground, and verified that there exists a linear relation between them, by which the crowd density was estimated through the least squares method. Nowadays, the most methods for density estimation are mainly researched in space domain, while some frequency-domain-dependent methods just simply accumulate the values of amplitude [4].

In this paper, under complicated background, background subtraction was adopted to obtain the moving target groups, and furthermore, the energy of the high frequency sub-band details in the wavelet transform was extracted as the feature. In our work it was found that there exists a linear relation between the number of people and the energy of the high frequency sub-band details. Finally, the least square was employed to fit the linear equation and estimate the number of crowd. Experimental results show that the proposed method is highly effective for estimating the number of moving people in crowded scenes.

2 Moving Object Extraction

Moving object detection and extraction is an important field of computer vision research, and is the basic part of the later tasks such as moving target tracking, identification, target behaviour analysis and understanding. Moving object detection methods are mainly divided to three types [5,6]: frame difference, optical flow and background subtraction. In view of the moving objects detected by background difference is more complete, we use background subtraction to extract the moving object. Extraction process is as shown below:

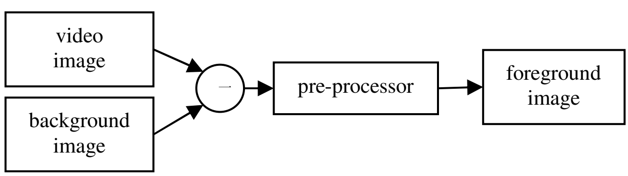

Fig. 1. The process of moving object detection

In the process of background reconstructing, available background models are median background, single Gaussian model and Gaussian mixture models, and the simplest model is the time average image. In our work, we adopted the Kalman-filter-based background updating algorithm (see Fig. 2).

As shown in Fig.3, the moving object detecting is always interfered by shadow. Using color information may improve the accuracy of shadow detection. Instead of RGB space, HSV color space is closer to human vision system in which the

Fig. 2. The background image

Fig. 3. Result after removing background

chromaticity and intensity of pixel are distinguishable more easily than in RGB space, so that shadow and foreground targets may be separated conveniently in HSV space. Hence, in our work the shadow was removed in HSV space. Shadow model [7] is described as a shadow mask *SP* in (1), where $I_k^V(x,y)$, $I_k^S(x,y)$ and $I_k^H(x,y)$ describe the *V*, *S* and *H* components in HSV space respectively for the pixel of foreground image at coordinate (*x*, *y*) in the *k*-th frame. Same as above, $B_k^V(x,y)$, $B_k^S(x,y)$ and $B_k^H(x,y)$ denote corresponding components for background image at the same position in the same frame. In (1), the constants α, β are less than 1, and α depends on light intensity, the higher light intensity, the less the value of α; variable β could avoid misidentification of shadow caused by background pixels which are interfered by noise. The image after removing shadow is shown in Fig. 4. When the color of moving object is close to that of background, empty areas may appear in the region of moving object, and this may be solved by morphological processing.

$$SP(x,y)=\begin{cases}1 & for\ \alpha \le I_k^V(x,y)/B_k^V(x,y)\le \beta\ and \\ & I_k^S(x,y)-B_k^S(x,y)\le T_S\ and \\ & I_k^H(x,y)-B_k^H(x,y)\le T_H \\ 0 & otherwise\end{cases}. \tag{1}$$

Fig. 4. Result after removing shadow

3 Wavelet Decomposition

A. Wavelet Decomposition of Image

With wavelet decomposition of image we can obtain the high frequency sub-graph and the low frequency sub-graph. The high frequency sub-graph retains the edge information of original image, and the low frequency sub-graph retains the original image details [8].

From the perspective of image processing, wavelet decomposition has some advantages as follows:

1) Wavelet decomposition can cover the entire frequency domain;

2) By selecting the appropriate filter, Wavelet transform can greatly reduce or remove the extracted correlation between different features;

3) The wavelet transform has a “zoom” feature, high frequency resolution and low temporal resolution may be used in low frequency band, low-frequency resolution and high temporal resolution may be used in high frequency band.

As a tool for video processing, wavelet transform has the following advantages:

1) Wavelet transform maintains the spatial characteristics of the original image, which allows us to extract the high frequency information of the image;

2) Wavelet can remove correlation and related energy;

3) Wavelet components have orientation selectivity, and can be divided into horizontal, vertical and diagonal direction;

4) Proper wave filter can improve the bad image quality and gain the better video reconstruction.

Thus, selecting high frequency sub-graph has two advantages: firstly, we can get the marginal information accurately; secondly, dimensionality reduction of image can reduce the computing amount for feature extracting. The *db2* of Daubechies wavelet radical is employed for decomposition in this paper.

B. Energy

We can get low frequency image and high frequency image after wavelet decomposition. High frequency image includes horizontal detail sub-band, vertical detail sub-band and diagonal detail sub-band. Because the high frequency sub-graph retains the marginal information of original image, and the marginal information of image represents the basic feature of image, which is invariant for changing illumination [9]. So we used high frequency sub-graph to extract marginal information. In the wavelet transform domain, the amplitude of factor represents change rate of original image gray scale under this resolution, so the energy can represent the definite feature of original image as possible as it can.

The energy of high frequency sub-graph in this paper adopts the definition given in Ref. [10]:

$$E = \sum_{i=1}^{M} \sum_{j=1}^{N} \left| f(i,j) \right| . \tag{2}$$

where M and N are the size of high frequency sub-graph, $f(i,j)$ is the image of high frequency sub-graph. Because there are three images of high frequency sub-graph,

Table 1. Number of People And The energy of High Frequency Sub-Band

Actual number of people	Energy of high frequency sub-band	Actual number of people	Energy of high frequency sub-band
5	149.3141	13	367.6555
7	163.4963	17	449.5771
9	216.3634	24	846.8471
11	279.2682	29	907.0659

so the energy in this paper is the sum of three detail sub-bands calculated by Eq.(2). Tab.1 presents the number of crowd and corresponding energy of high frequency sub-band. It can be seen from the table that the energy of high frequency sub-band increases with the increase of number of people.

4 Curve Fitting

A. Principle of Least Square

The least square method—a very popular technique—is used to compute the estimations of parameters to fit given data. As a mathematical optimization technique, it aims to find the best matching function for given data by minimizing the error sum squares. The unknown data can be easily calculated by the least square method, which produces minimum squared error between actual data and calculated data [11].

Curve fitting can be roughly divided into the following steps:

1) Draw the scatter diagram and choose suitable curve type.

2) Variable transformation

$$\begin{cases} x' = g_1(x) \\ y' = g_2(y) \end{cases} . \tag{3}$$

The method of variables replacement is adopted in (3) so that there exists linear relation between the two new variables.

3) The experimental data are used to seek for undetermined coefficient according to the least square principle.

4) The linear formula is transformed to the function about original variables x, y.

B. Correlation Coefficient

The correlation coefficients stand for degree of the pertinence between variables. Let $r \in [-1,1]$ stands for correlation coefficient of samples, so that larger value of $|r|$ corresponds to smaller error, which represents higher correlative linear degree; on the other hand, if the value $|r|$ is close to 0 then the error becomes large, the correlative linear degree is lower. The two variables may be considered highly correlative in the case of $|r|>0.75$. The equation for calculating correlation coefficient is as follow:

$$r = \left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right) / \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} . \tag{4}$$

where, x_i , y_i are sample variables and $\bar{x}$, $\bar{y}$ stand for average value of sample variables and n stands for sample size.

5 Experimental Results and Analysis

A. Experimental Procedures

We choose 40 frames of images (there are different number of people in different frame) randomly from video to form training image base. In order to reduce the error caused by the incompleteness of the foreground images, the training images are divided into 4 groups, each group has 10 frames of images, and the numbers of people in these groups are 5, 7, 9 and 11 respectively. Experimental training process is as follows:

Step1: Count the number of each frame manually and store the number into array y;

Step2: Execute background subtraction for training image base, and obtain moving object, i.e., foreground image;

Step3: Wavelet decompose the foreground image, and store the high frequency sub-band energy into array x;

Step4: Draw scatter diagram according to statistical data;

Step5: Calculate correlation coefficient of the data;

Step6: Ascertain curve type according to scatter diagram and correlation coefficient, compute undetermined coefficient and obtain curve equation.

The experimental process chart is as follows:

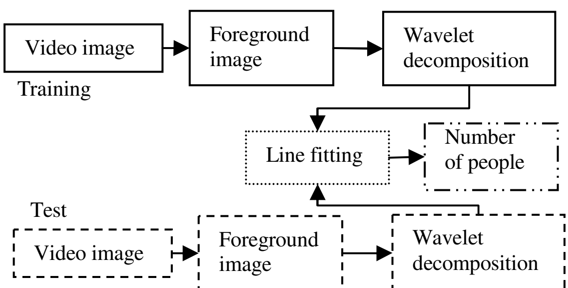

Fig. 5. Experiment process

B. Experimental Results and Analysis

After wavelet decomposition of all foreground images and calculation of the energy of their high-frequency sub-bands, curve can be fitted with all the energy and respective number of people. Part of energy for foreground images in the training base is indicated in Tab.II, and the experimental result is shown as in Fig.6.

Table 2. The energy of frame difference image

Foreground image	Energy	Number of people	Foreground image	Energy	Number of people
1	141.75	5	6	168.48	7
2	149.31	5	7	216.72	9
3	136.05	5	8	213.09	9
4	161.21	7	9	278.78	11
5	171.99	7	10	289.13	11

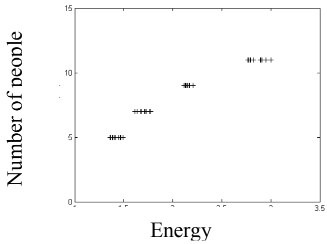

Fig. 6. Scatter of experimental data

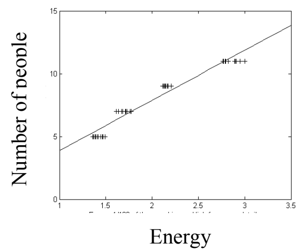

Fig. 7. Result of linear fitting

According to the scatter diagram, the data distributed intensively with the shap of strap. Conclusion can be drawn through experimental analysis that the correlation coefficient between the number of people and the energy of the high frequency sub-bands of the foreground images in the crowd scenes is 0.9759, which indicates that there exist a strong linear relation between the number of people and the energy. Hence, the method of linear fitting can be adopted, in other words, estimation of the number of people in the crowd scenes can be described in the form of linear equation $y = ax + b$, where a, b are undetermined coefficients, y stands for the number of people in the crowd scenes, and x stands for the energy of high frequency sub-bands of the foreground images. Fig.7 is the resunlt of linear fitted by the least square method.

We use the relative error to measure the results, namely:

$$error = \left|N_t - N_c\right| / N_t \times 100\% \,. \tag{5}$$

where N_t is the actual number, N_C is the test value.

Tab. 3 is the testing results, and Fig.8 is the result of comparison between test value and real value, where '*' indicates the actual number of people, '□' stands for the tested number of people. According to the experimental result, the maximum relative error of the test result is 12.50%, and the maximum difference number of people is 2. Tab. 4 and Fig. 9 are the results of comparisons between proposed methods and the method of Canny operator edge detection in Ref. [3]. As shown in the figure, our result has a higher preciseness rate than the method proposed in Ref. [3], especially when the number of people increases, its estimation becomes closer to the actual situation.

Table 3. Testing results

Test image	Test value	Actual value	Relative error(%)
1	18	20	10.00
2	19	20	5.00
3	30	30	0.00
4	7	8	12.50
5	14	16	12.50
6	30	28	7.14
7	27	28	3.57
8	13	13	0.00
9	4	4	0.00
10	15	15	0.00

Table 4. Results of comparison

Test image	Test value	Test value in Ref. [3]	Actual value
1	18	22	20
2	19	22	20
3	30	34	30
4	7	34	8
5	14	17	16
6	30	34	28
7	27	32	28
8	13	15	13
9	4	4	4
10	15	17	15

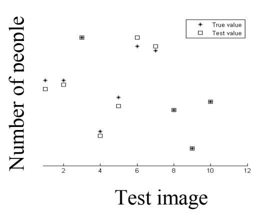

Fig. 8. Comparison between test and real value

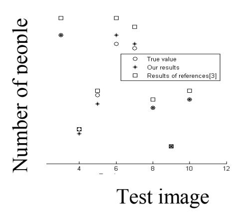

Fig. 9. Comparison with Ref.[3]

6 Conclusion

In this paper, moving objects were captured by background subtraction, and through wavelet decomposition of foreground images, the energy of high frequency sub-bands was extracted as the feature for estimation of number of people. And furthermore, with the least square method, the linear equation for number of people and energy was found. Finally, the estimation for number of people in the crowd scenes was conducted. The experimental results show that simple and fast method of linear equation proposed in this paper is very effective to estimate the number of people in the crowd scenes.

Sometimes moving objects are incomplete, which affects the subsequent work greatly. The incompleteness of moving objects is mainly due to the similar color between background and moving objects. So extracting the complete moving objects is a problem to be solved during the work of algorithm improvement in the future.

References

1. Davies, C., Yin, J.H., Velastin, S.A.: Crowd monitoring using image processing. IEE Electronic and Commnications Engineering Journal 7, 43–45 (1995)
2. Chan, B., Vasconcelos, N.: Bayesian poisson regression for crowd counting. In: IEEE Int'l Conf. on Computer Vision, Kyoto, Japan, pp. 1–7 (2009)
3. Du, J.: Crowd's density and motion estimation, pp. 21–25. Dalian University of Technology, Dalian (2006) (in Chinese with English abstract)
4. Liu, X.R., Zhou, J.L., Li, X.H.: Grey level dependent matrix based crowd density estimation in Fourier spectrum. Control & Automation 24, 310–311 (2008) (in Chinese with English abstract)
5. Wang, S.L., Kang, M.N., Deng, Z.H., et al.: Object separating algorithm based on background alternation and dynamic expanding filter. Computer Engineering and Applications 46, 246–247 (2010)(in Chinese with English abstract)
6. Shi, Z.L., Chen, Y.M., Xu, J.: Enhanced human motion detection algorithm and its implementation based on video surveillance. Computer Engineering and Applications 44, 206–207 (2008)(in Chinese with English abstract)
7. Cucchiara, R., Grana, C., Piccardi, M., et al.: Improving shadow suppression in moving object detection with HSV color information. In: Proc. IEEE Intel. Transp. Syst. Conf., Oakland, USA, pp. 335–336 (2001)
8. Xue, L.Y., Pan, J.J.: Edge Detection Combining Wavelet Transform and canny operator based on fusion rules. Proceedings of the 2009 International Conf. on Wavelet Analysis and Pattern Recognition, Baoding, China, pp. 324–325 (2009)
9. He, Y., Wang, L.: Image stitching algorithm based on feature block and wavelet transform. Computer Engineering and Design 31, 1959–1960 (2010) (in Chinese with English abstract)
10. Yang, D.Q., Xiao, S.P., Jiang, J.F.: Texture image recognition based on differentia evolution probabilistic neural network. Computer Engineering and Applications 44, 180–181 (2008) (in Chinese with English abstract)
11. Abdi, H.: Least squares. In: Lewis-Beck, M., Bryman, A., Futing, T. (eds.) Encyclopaedia for Research Methods for the Social Sciences, pp. 559–561. Sage, Thousand Oaks (2003)

返回检索结果

Web of Knowledge
第 **1** 页 (记录 **1 -- 1**)
[1]

打印此页

第 **1** 条，共 **1** 条

标题: Estimating the Number of People in Crowded Scenes Based on Energy in Frequency Domain

作者: Yang, GQ (Yang, Guo-qing); Cui, RY (Cui, Rong-yi)

编者: Zeng D

来源出版物: APPLIED INFORMATICS AND COMMUNICATION, PT 2 **丛书:** Communications in Computer and Information Science **卷:** 225 **页:** 546-554 **子辑:** Part ii **出版年:** 2011

在 **Web of Science** 中的被引频次**:** 0

被引频次合计: 0

引用的参考文献数: 11

摘要: A method of estimating the number of people in crowded scenes was proposed based on energy of the foreground image in the frequency domain. Firstly, the foreground image including the target area was obtained by background subtraction. And furthermore, with the second order wavelet decomposition of the foreground image, the energy of high frequency sub-band was computed. Finally, by analyzing the relation between population size and high frequency sub-band energy, the mathematical model was established and the model was solved by the least square method. The experimental results show that the proposed method is highly effective for estimating the number of moving people in crowded scenes.

入藏号: WOS:000307481100069

语种: English

文献类型: Proceedings Paper

会议名称: International Conference on Applied Informatics and Communication (ICAIC 2011)

会议日期: AUG 20-21, 2011

会议地点: Xian, PEOPLES R CHINA

作者关键词: estimation of number of people; wavelet decomposition; energy in frequency domain; the least square method

地址: [Yang, Guo-qing; Cui, Rong-yi] Yanbian Univ, Dept Comp Sci & Technol, Intelligent Informat Proc Lab, Yanji, Jilin, Peoples R China.

通讯作者地址: Cui, RY (通讯作者),Yanbian Univ, Dept Comp Sci & Technol, Intelligent Informat Proc Lab, Yanji, Jilin, Peoples R China.

电子邮件地址: y.p.yu@sohu.com; cuirongyi@ybu.edu.cn

出版商: SPRINGER-VERLAG BERLIN

出版商地址: HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY

Web of Science 类别: Computer Science, Artificial Intelligence; Computer Science, Information Systems

研究方向: Computer Science

IDS 号: BBN36

ISSN: 1865-0929

ISBN: 978-3-642-23219-0

29 字符的来源出版物名称缩写: COMM COM INF SC

来源出版物页码计数: 9

返回检索结果

Web of Knowledge
第 **1** 页 (记录 **1 -- 1**)
[1]

打印此页

© 2013 Thomson Reuters | 使用条款 | 隐私策略 | *请提供使用 Web of Knowledge 的反馈。*

Engineering Village

Print Close

Accession number: 20113614307529

Title: **Estimating the number of people in crowded scenes based on energy in frequency domain**

Authors: Yang, Guo-Qing[1] ; Cui, Rong-Yi[1]

Author affiliation: 1 Intelligent Information Processing Lab., Department of Computer Science and Technology, Yanbian University, Yanji, Jilin, China

Corresponding author: Cui, R.-Y. (cuirongyi@ybu.edu.cn)

Source title: Communications in Computer and Information Science

Abbreviated source title: Commun. Comput. Info. Sci.

Volume: 225 CCIS

Issue: PART 2

Monograph title: Applied Informatics and Communication - International Conference, ICAIC 2011, Proceedings

Issue date: 2011

Publication year: 2011

Pages: 546-554

Language: English

ISSN: 18650929

ISBN-13: 9783642232190

Document type: Conference article (CA)

Conference name: 2011 International Conference on Applied Informatics and Communication, ICAIC 2011

Conference date: August 20, 2011 - August 21, 2011

Conference location: Xi'an, China

Conference code: 86306

Publisher: Springer Verlag, Tiergartenstrasse 17, Heidelberg, D-69121, Germany

Abstract: A method of estimating the number of people in crowded scenes was proposed based on energy of the foreground image in the frequency domain. Firstly, the foreground image including the target area was obtained by background subtraction. And furthermore, with the second order wavelet decomposition of the foreground image, the energy of high frequency sub-band was computed. Finally, by analyzing the relation between population size and high frequency sub-band energy, the mathematical model was established and the model was solved by the least square method. The experimental results show that the proposed method is highly effective for estimating the number of moving people in crowded scenes. © 2011 Springer-Verlag.

Number of references: 11

Main heading: Least squares approximations

Controlled terms: Estimation - Frequency domain analysis - Information science - Mathematical models - Population statistics - Wavelet decomposition

Uncontrolled terms: Background subtraction - Foreground images - Frequency domains - High frequency - Least square methods - Moving peoples - Population sizes - Second orders - Sub-bands

Classification code: 903 Information Science - 921 Mathematics - 922.2 Mathematical Statistics

DOI: 10.1007/978-3-642-23220-6_69

Database: Compendex

Compilation and indexing terms, © 2011 Elsevier Inc.

© 2011 Elsevier Inc. All rights reserved.