13—23

Mosaic Representations of Video Shots Based on Slice Image Analysis

Feng Shaolei, Lu Hangqing , Li Dalong
National Laboratory of Pattern Recognition, Institute of Auotomation,
Chinese Academy of Sciences,P.O.Box 2728, Beijing 100080, P.R. China
Phone:+86-010-62542971, Email: {slfeng,luhq,dlli}@nlpr.ia.ac.cn

Abstract

In this paper, we present a novel mosaic method, which is based on the analysis of video slice image, to get the initial values of the optimal estimation model so that we can avoid the local minim problem. Simultaneously, we can remove the accumulated distortions from the final mosaic that are common in the previous methods. Comparing with other methods, our method can greatly reduce the calculation and is more robust.

1. Introduction

As an explicit and compact scene-based representation, the mosaic representation has a wide application in video representation.
Generally, there are two categories of strategies for mosaic. One relies on the special feature correspondence [1], the other is direct methods based on optical flow from pixel correlation [2]. For their simplicity and reliability, the latter are used more widely. They are based on the direct minimization of an image-based misregistration measure, and solve a least-squares estimation problem in the unknown structure and motion parameters, which leads to statistically optimal estimate [3].

Using an optimization method to solve the least-squares estimation can only guarantee convergence to a local minimum, especially when large image motion exists between two consecutive frames or there are too many outliers. That means giving good initial values to the unknown structure and motion parameters is very important. In the case of the motion between successive frames is large, image pyramid [4] is often used, which first register subsampled smaller image to get parameter solutions as the initial values of the next level image registration. In another case of too many outliers existing, a robust M-estimate method has been developed by [5].

In this paper, we present a novel method, based on the video slice image analysis, to get the initial values. By cutting the video volume, we can get the slice image of the video sequence, in which the trajectory of the main motion of the image is left. Through the trajectory we can get the initial values of the motion parameters.

Comparing with other methods, our techniques have advantages as following:

- The initial values gotten by the analysis of slice images is more accurate, especially when large image motion exists between two consecutive frames.
- It avoids the distortion caused by the accumulated warps and distortions between every pairs of consecutive image. This is enabled by directly calculating the image motion of every frame to the reference frame.
- It is robust even large moving objects exist in the video sequence.
- It reduces the calculations greatly through at the cost of doing a little work in advance .

Finally, the experiment results are shown in this paper.

2. Defining and Making Slice Image

Video can be represented in a 3-D spatio-temporal coordinate system, three axises are x (horizontal), y (vertical) and t (time). Keeping x at a fixed value, cut the video stream along the t axis .we can get an image in y-t plane, this image is defined as the vertical slice of the video stream. By the same method, we can get a horizontal slice. Fig 1. Shows the method of making slice image.

In a single slice, the video scene and moving object trace out patterns that are slanted and curved according to the camera motion (main motion) and individual motions. That is, we can get some motion information from the slice image.

Thinking a most simple case: there is only the horizontal translation of the camera in a video sequence. Then every curve in the XT slice will accurately reflect the camera motion. Another simple case is that there is only the vertical translation of the camera, then the YT slice can do it. However, in most cases, the camera motion is complicated, which can be described by six parameters: the translation $T = (T_X, T_Y, T_Z)$, the

rotation $\Omega = (\Omega_X, \Omega_Y, \Omega_Z)$. Generally, It is impossible for a single slice to give the full information of the camera motion. But as for our work in this paper, we need only initial values of the model parameters. Further more, we use some knowledge to determine which curve in the slice is most useful to estimate the initial values. If needed, we can use more than one slice. Thus the validity of the initial values from slice images are guaranteed.

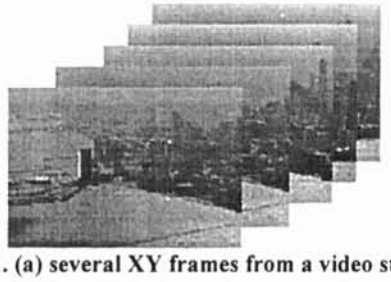
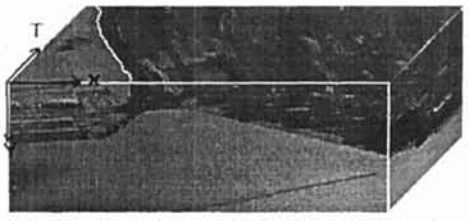


Fig 1. (a) several XY frames from a video stream



(b)XYT cube of the video stream, sliced at the middle height of the images. And the horizontal slice image (XT image plane) is gotten.

In Fig 1(a), we found that the flank of the building (marked as white) in the scene keep vertical in this sequence. So in the XT slice image, its trace (marked as white in the Fig 1) can reflect the image's horizontal motion very well. For the same reason, if we can find horizontal edge in the frame image, its relevant trace in the YT slice can reflect the image's vertical motion.

3. Model of Motion

The instantaneous image motion of a general 3D scene can be expressed as:

$$\begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \begin{bmatrix} -(\frac{T_X}{Z} + \Omega_Y) + x\frac{T_Z}{Z} + y\Omega_Z - x^2\Omega_Y + xy\Omega_X \\ -(\frac{T_Y}{Z} - \Omega_X) - x\Omega_Z + y\frac{T_Z}{Z} - xy\Omega_Y + y^2\Omega_X \end{bmatrix} \quad (1)$$

If we assume that all the scene points (X,Y,Z) are in a 3D plane, the points will satisfy a plane equation $Z = A + BX + CY$. We also have the projective relation between a scene point and its corresponding image point $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} xZ \\ yZ \end{bmatrix}$. Substituting these two equations in Equation (1) yields the 2D quadratic transformation:

$$\begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \begin{bmatrix} a + bx + cy + gx^2 + hxy \\ d + ex + fy + gxy + hy^2 \end{bmatrix} \quad (2)$$

This equation describes the 2D parametric image motion of a 3D planar scene by eight parameters (a,b,c,d,e,f,g,h).

It should be noted that all the equations above are established in a normalized coordination.

The sum of the squared intensity difference (SSD) measure integrated over two image is used as a match measure:

$$E(\vec{\alpha}) = \sum (I(x,y,t) - I(x - u(x,y;\vec{\alpha}), y - v(x,y;\vec{\alpha}), t-1))^2 \quad (3)$$

We use the Levenberg-Marquardt iterative non-linear minimization algorithm to perform the minimization of the SSD error measure. Its advantage over straightforward gradient decent is that its convergence speed is faster.

But the Levenberg-Marquardt algorithm can not guarantee to converge to the minim of the whole solution space; it is liable to get into a local minim. So only when given the better initial values can we get the correct solution. Usually we make the initial values as zero, this means that the motion between two image cannot be too large.

To solve this problem, hierarchical matching based on the image pyramid is used by traditional methods. But even this method can not dispose larger motion very well, and what is more, they cost much more calculation. Another problem is that the traditional methods first calculate the motion between two successive frame images in turn, then take one frame as the reference frame and warp other frames into the reference frame. As a result, large distortion may be generated in the final mosaic because the errors between every two successive frames are accumulated even though image pyramid method has made them very small.

In order to overcome these limitations, we have tried to develop a novel method based on the slice image analysis. In our method, we first take one frame as the reference frame, then calculate the initial values of the motion parameters of every frame through the selected

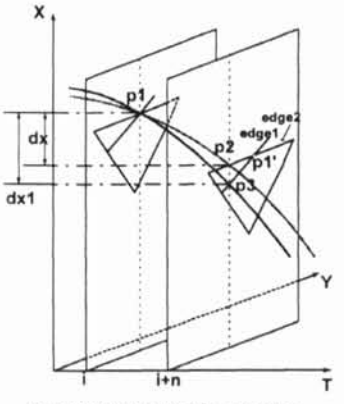


Fig.2. The illustration of calculating initial value from the slice curve.

curve in the slice image. Using these initial values, we directly search for the motion parameters of every frame to the reference frame and warp it into the final mosaic so that we avoid the accumulated distortion.

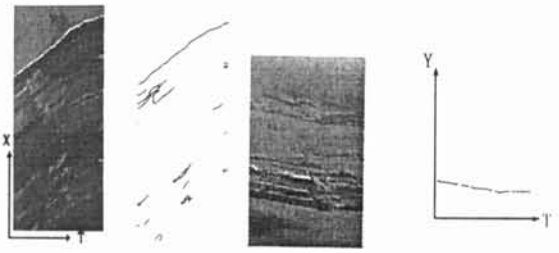
4. Analyzing Slice to Get the Initial Values

From the eight parameters motion model in the section 3, we can find that the parameters a, d can be explained as the homogeneous translation of the whole image, since they are irrelevant to the position of every pixel in the image. In our method, we first calculate the initial values of these two parameters using the trace curve gotten from slice images. Fig. 2 can illustrate it.

Legible trace curves in the slice image is left by motions of edges in the source frame images, which requires that we should making slice by cutting through a certain edge of the frame images. If there are many trace curves in the slice image, which is select to calculate the initial values? In the XT-slice, we select the trace left by edges paralleling to Y-axis in frame image; In the YT-slice, we select the trace left by edges paralleling to X-axis in frame image. This could be explained in Fig. 2, in which two traces (the red curve and the blue curve) in the XT-slice are left by the edge1 and edge2 respectively, and edge2 is paralleling to the Y-axis but edge1 is not. All the points in the edge2 have the same x coordinate if it keep paralleling to Y-axis in the all frames, this is maybe a strong assumption and we will discuss it. later. So we still can use the p1 and p2 to calculate the dx even though they are not the same points after the image motion. The corresponding point p1' of p1 in the (i+n) frame has not same x coordinate to p3, so dx1 can not reflect the x-motion of the image accurately. Based on the reason above, we select the red trace, which is left by the edge2, to calculate the initial value of the image x-motion.

In the above analysis, we need search for the edges that keep paralleling to the x-axis or y-axis in the video sequence. From the equation (1), we can find that to keep the slope of a straight edge approximately constant, the following need be satisfied: (i) The depth change of the points in the 3D straight line corresponding to the image straight edge should be small relative to the overall distance of the 3D straight line from the camera.($\Delta Z << Z$), (ii) Camera rotation should be small enough so that the terms contain the $\Omega_X, \Omega_Y, \Omega_Z$ are relative small to other terms in equation (1) for the motion of the points in the straight edges. All the striction above is cast on the straight edges but not the whole image.

Fig. 3 (a) shows a XT slice at the middle height of the frame image from a video sequence (city). The trace marked as white is generated by the flank of the building that parallels to the y-axis. So this trace can reflect the x-motion of images. Applying a Canny edge operator to this slice image, we can transfer it into a binary edge map and then erase the useless edges but just reverse the needed curve.



(a) (b) (c) (d)

Fig.3. (a) The XT slice of the video sequence at the middle height of the image. (b) The binary edge map after the canny edge operator applied to (a). (c)The vertical (YT) slice cut from the left of the frame images. (d) The final YT slice curve used to calculate the y-motion of the images after erasing the useless curves.

Now the points in the curve can be represented as $x(t; y)$, where y is a constant for this is a horizontal (XT) slice image. Then once we have select one frame as the reference frame, the horizontal translation of every other frame can be calculated as follows:

$$dx(t) = x(t; y) - x(t_r; y)$$

where $x(t; y), x(t_r; y)$ are the x coordinates of the points in the curve correspond to the frame at the time of t and the reference frame respectively.

Using the same method on the vertical slice image, we can get the vertical translation of every frame to the reference frame: $dy(t) = y(t; x) - y(t_r; x)$.

Now we can use the $dx(t)$ and $dy(t)$ as the initial values of the parameters a and d:

$a = dx(t)$, $d = dy(t)$

to estimate the motion between the frame at the time of t and the reference frame.

5. Experiment results

We estimated the motion parameters of every frame to the reference frame, using the Levenberg-Marquardt optimal algorithms. Then we warped every frame to one whole panorama. Fig 4 shows the results using the image pyramids and using our method based on the slice image respectively.

In more general video sequences there are moving objects. Obviously, the existing of

moving objects as outliers has influence on the estimation of the dominant motion, especially when moving objects are large comparing with the background scene. In our method, we use the slice image to get initial values instead of the M-estimation. In fact, knowledge has been applied to determine which trace curve of the slice could be used in our method, so that the initial values is reliable since the selected slice curve is generated by the background scene of the video.

Here, besides the city sequence above, we analyzed another video sequence with a moving object in it. We use the horizontal and vertical edges of platform in the right of the frames to get the YT and XT slice images. Because the platform has moved out of the frames at the back section of this sequence, we utilized other feathers such as the vertical stain in the surface of the left wall. Furthermore, the moving object as the outliers in this case has almost completely been deleted by a temporal median filter since their location are not highly correlated over time. A real scenario mosaic is shown in the Fig.5.

6. Conclusion

We have developed a mosaic technique based on the slice analysis in this paper. And we have tested it on two video sequences. The experiment results based on this technique show that it is robust to outliers and avoid the distortion in the final mosaic.

7. References

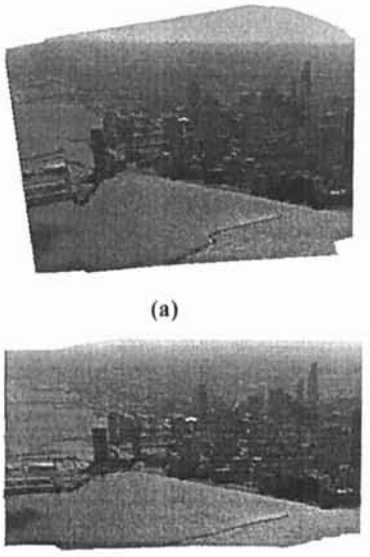
[1] Huang, T.S. and Netravali, A.. Motion and structure from feature correspondences: a review. 1994. Proc. IEEE,vol.82, pp.252-269.

[2] Hanna, K.. 1991. Direct multi-resolution estimation of ego-motion and structure from motion. Proceedings of IEEE Workshop on Visual Motion, pp. 156-162.

[3] Szeliski, R. and Kang, S.B.. 1995. Direct Methods for Visual Scene Reconstruction. IEEE Workshop on Representation of Visual Scene.,Cambridge,MA

[4] Rosenfeld, A., 1984. editor. Multi-resolution Image Processing and Analysis. Springer-Verlag

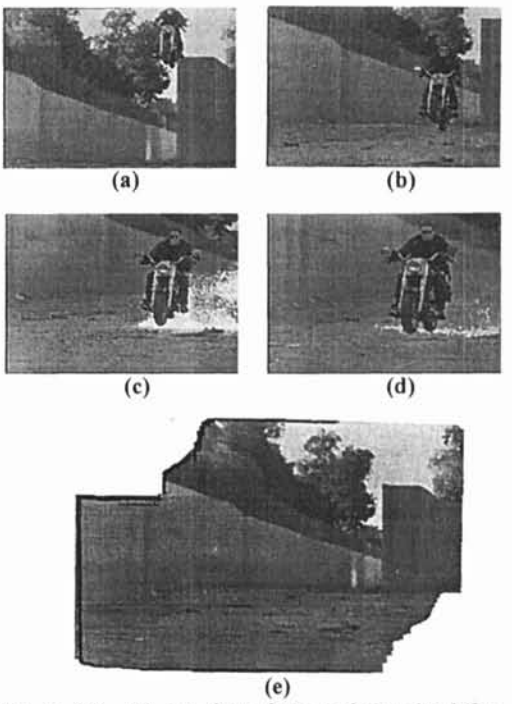
[5] Sawhney, H., Ayer, S.. 1996. Compact representations of videos through dominant and multiple motion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.18, No.8, pp.814-830.



(a)

(b)

Fig 4. (a) The left panorama used the image pyramids to get the initial value of the parameter, and (b) the right panorama used the slice image to get the initial value of the parameter. We can see that the distortion in the left panorama has been greatly improved in the right.



(a) (b) (c) (d) (e)

Fig. 5. (a)---(d) are four frames from the bike sequence, which is the 20th, 80th, 100th , 140th frame respectively.(e) shows the final whole mosaic.