MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-3

Fast Global Motion Estimation via Modified ILSE Method

Jia Wang, Hanqing Lu, Qingshan Liu
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences (CAS), Beijing 100080, P.R. China
Emails:{wangjia, luhq, qsliu}@nlpr.ia.ac.cn

Abstract

This paper presents a Modified ILSE technique for global motion estimation where Gradient Thresholding (GT) is used to analyze and preprocess the image blocks before motion estimation. By means of GT, treacherous blocks that are more likely to produce inaccurate motion estimations are identified at the beginning and discarded from the following process. So, the computational cost is reduced to a great extent comparing to the traditional method. Furthermore, a method for the selection of threshold is also presented which adapt the threshold to different types of images. The presented method has been tested on a variety of image sequences, and experimental results illustrate its promising performance.

1 Instruction

Motion estimation is one of the research topics having attracted many research activities in the video image processing community [1]. In video, motion is primarily due to the movement of a camera (pan, zoom), movement of objects in the scene, or movement of both. The former is often referred as *global* motion and the latter as *local* motion. Separating these two classes of motion is significant for video coding, video indexing, video object segmentation, and many other applications.

The global motion estimation procedure depends on parametric models of camera motion and the way that the model parameters are estimated. Various techniques and schemes have been proposed to deal with these two factors-[1-5], and each of them has its particular features and applications.

Iterative Least Square Estimation (ILSE) technique is a commonly used method for global motion estimation. Recently, Rath and Makur [2] proposed a four-parameter model to calculate global motion parameters using ILSE. In their motion model, only the camera pan and zoom parameters are considered, because they figure that, generally, camera rotation is comparatively much less frequent than zooming and panning. Rath and Makur's method consists of two steps: First, an initial motion field is calculated using Block Matching Algorithm (BMA) [6,7] considering all of the blocks in a frame. Note that the calculated motion field can hardly be accurate everywhere in the frame, which means some of the calculated motion vectors will be wrong for some of blocks, and these blocks with inaccurate motion vectors may introduce a bias into the final estimated global motion parameters. In the second step, ILSE technique is used to gradually eliminate the influence of these blocks and finally extract accurate global motion parameters. Although the results of ILSE are fairly accurate, many computational work and time are already wasted in BMA and ILSE for those blocks with inaccurate motion estimations. From this point of view, if these blocks could be identified and eliminated before processing by BMA and ILSE, the computational cost and execution time would be saved to a great extent.

To implement this strategy, we propose a Modified ILSE method in this paper. In our method, Gradient Thresholding (GT) is used to analyze and preprocess all of the blocks before motion estimation with BMA and ILSE. By means of GT, the treacherous blocks that are more likely to produce inaccurate motion estimations are identified at the beginning and then discarded from the following process. As the result, the computational cost is reduced to a great extent comparing with the traditional method [2]. Furthermore, we also propose a method to automatically select the threshold for GT based on gradient histogram. So the selection of threshold becomes self-adaptive to different types of images.

2 Global Motion Estimation

The traditional structure of global motion estimation using ILSE technique [2] involves two steps. First, the frame image is segmented into several $n \times n$ blocks, and block-matching algorithm (BMA) is performed to estimate the motion vector for each block. Second, ILSE technique is used to compute global motion parameters from the estimated motion field constructed by the blocks and their motion vectors. As discussed earlier, Rath and Makur's method will consider all the blocks in the frame.

Let there be N blocks in a video frame, and assume that the motion vector of a block is the motion vector of the central pixel of that block. Let $(v_x(k), v_y(k))$ be the measured motion vector, gained by BMA, of the block k, k=0,1,…,N-1, whose central pixel's coordinates are $(s_x(k), s_y(k))$ with respect to the center of the frame. In this regard, the global motion estimation model represented in [2] for camera zoom and pan is as:

$$\begin{bmatrix} v_x(k) \\ v_y(k) \end{bmatrix} = \begin{bmatrix} a_1 \cdot s_x(k) \\ a_3 \cdot s_y(k) \end{bmatrix} + \begin{bmatrix} a_2 \\ a_4 \end{bmatrix} \quad (1)$$

Where

$$\begin{aligned} a_1 &= z_x \quad and \quad a_2 = f_1(p_x, z_x) \\ a_3 &= z_y \quad and \quad a_4 = f_2(p_y, z_y) \end{aligned} \quad (2)$$

In the above definition, z_x and z_y are the zoom factors along the *x*-axis and *y*-axis respectively, (p_x, p_y) is the pan vector.

Then according to the ILSE algorithm, the optimal values for camera parameters (a_1, a_2, a_3, a_4) are obtained by using the following criteria:

$$\min_{a_1,a_2} \sum_{k=0}^{N-1} (v_x(k) - a_1 \cdot s_x(k) - a_2)^2 \quad (3)$$

$$\min_{a_3,a_4} \sum_{k=0}^{N-1} (v_y(k) - a_3 \cdot s_y(k) - a_4)^2 \quad (4)$$

By differentiating with respect to the parameters, and setting the derivatives to zero, the following solution is obtained as:

$$a_1 = \frac{N\sum_{k=0}^{N-1} v_x(k)s_x(k) - \left[\sum_{k=0}^{N-1} v_x(k)\right]\left[\sum_{k=0}^{N-1} s_x(k)\right]}{N\sum_{k=0}^{N-1} s_x^2(k) - \left[\sum_{k=0}^{N-1} s_x(k)\right]^2} \quad (5)$$

$$a_2 = \frac{\left[\sum_{k=0}^{N-1} v_x(k)\right]\left[\sum_{k=0}^{N-1} s_x^2(k)\right] - \left[\sum_{k=0}^{N-1} v_x(k)s_x(k)\right]\left[\sum_{k=0}^{N-1} s_x(k)\right]}{N\sum_{k=0}^{N-1} s_x^2(k) - \left[\sum_{k=0}^{N-1} s_x(k)\right]^2} \quad (6)$$

$$a_3 = \frac{N\sum_{k=0}^{N-1} v_y(k)s_y(k) - \left[\sum_{k=0}^{N-1} v_y(k)\right]\left[\sum_{k=0}^{N-1} s_y(k)\right]}{N\sum_{k=0}^{N-1} s_x^2(k) - \left[\sum_{k=0}^{N-1} s_x(k)\right]^2} \quad (7)$$

$$a_4 = \frac{\left[\sum_{k=0}^{N-1} v_y(k)\right]\left[\sum_{k=0}^{N-1} s_y^2(k)\right] - \left[\sum_{k=0}^{N-1} v_y(k)s_y(k)\right]\left[\sum_{k=0}^{N-1} s_y(k)\right]}{N\sum_{k=0}^{N-1} s_y^2(k) - \left[\sum_{k=0}^{N-1} s_y(k)\right]^2} \quad (8)$$

As shown by Rath and Makur in [2], to eliminate the influence of the blocks with inaccurately estimated motion, the above procedure is evaluated iteratively, and each iteration eliminates blocks whose motion vectors (estimated by BMA) do not match with the current global motion fields. Matching means that a motion vector lies within a threshold distance from the corresponding global motion field. So, in Rath and Makur's method, the influence of those blocks with inaccurate motion estimations will be gradually removed, and after several iterations, the estimated parameters will converge to the final results.

3 Gradient Thresholding

As described in section 2, Rath and Makur's method considers all the rows and columns of macroblocks in a frame to estimate the global motion parameters. Yet in our opinion, it's not necessary to involve all the blocks into the task, because many blocks will finally be discarded for their inaccurately estimated motion vectors. So, in this paper, Gradient Thresholding (GT) method is proposed to preprocess the blocks before the motion estimation. The basic idea comes from the analysis of BMA.

BMA [6,7] has been a popular motion estimation technique because of its simplicity, robustness, and implementary advantages. In this technique, a frame is segmented into square blocks of pixels. Motion of each block is estimated as a displacement vector by finding its best match in a search area in the previous frame. The set of displacement vectors is called displacement vector field or motion field. The matching process in BMA could be density-based, color-based or texture-based, etc. Whichever feature is used, there is a fact that, when there is no enough texture or gradient information, the estimated motion field is always treacherous. The reason is: some blocks are so similar to their neighbors that BMA can hardly find the right matches and produce the accurate motion vectors for them. Therefore, we should find these blocks first and throw them away before all of the other process. Since gradient is the critical information, we propose the Gradient Thresholding (GT) method to identify those treacherous blocks that are more likely to produce inaccurate motion estimations.

3.1 Block Analysis

To identify the blocks that are more likely to produce inaccurate motion vectors, the current frame should first be processed to make the Gradient Map. Many methods can be used for this task, such as the *Sobel Operator* [8], *Roberts Operator* [9], etc.

Having the gradient map, we will check the gradient information of each block.

For a block B with $n \times n$ pixels, let $p_{i,j}$ be the pixel with coordinates (i, j) in the block. We can find the gradient value $G(p_{i,j})$ of pixel $p_{i,j}$ in the gradient map. Then, the gradient value of block B can be defined as

$$G_B = \sum_{i=1}^{n}\sum_{j=1}^{n} [G(p_{i,j}) / n^2] \quad (9)$$

which is the mean gradient of all the pixels in B. Since in gradient map, there is $G(p_{i,j}) \in [0,255]$, there will be $G_B \in [0,255]$.

So for all the blocks in the frame, we define a threshold $\theta \in [0,255]$ to classify them as:

A block B will be treated as a treacherous block if, its gradient value $G_B < \theta$.

As soon as a block is decided to treacherous, it will not be considered in the following process by BMA and ILSE.

An example of gradient thresholding is shown in Figure 1. The original frame is selected from the *Flower Garden* sequence. Figure 1 gives the treacherous blocks identified by GT method, which are marked by crosshairs. The threshold used for GT is $\theta = 70$.

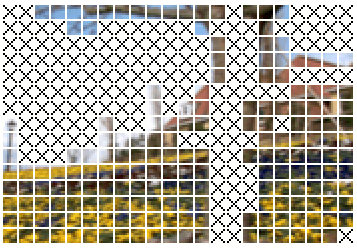
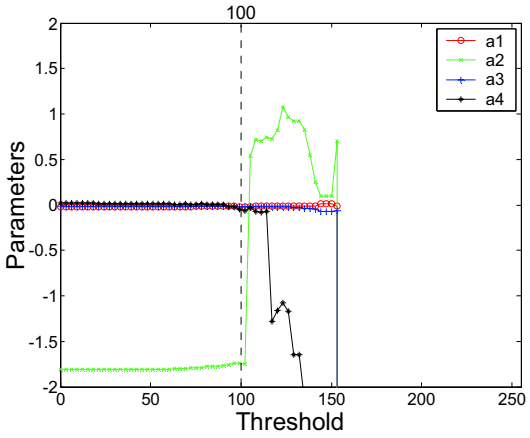


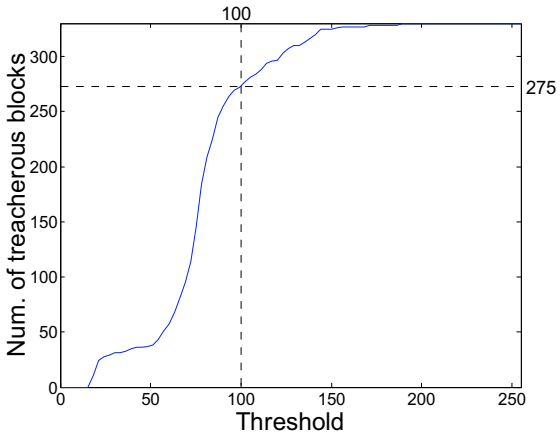
Figure 1. Gradient thresholding

The original frame is separated into 330 blocks with block-size 16×16. After gradient thresholding with $\theta = 70$, 145 blocks (about 44% to 330) is marked as treacherous and will not be processed by the following operation. So the execution time will be saved to a great extent. In the following subsection, we will analyze the performance of GT method quantitatively.

3.2 Performance Analysis



(a)Estimated parameters



(b) Num. of remaining blocks after GT

Figure 2. Performance analysis

In this subsection, the performance of GT method using different thresholds will be quantitatively analyzed.

Figure 2 shows the statistical information according to different thresholds derived from experiments on *Flower Garden* frame. Image-size for the sequence is 352×240, and block-size is 16×16. Figure 2(a) shows the curves of estimated parameters (a_1, a_2, a_3, a_4) due to different thresholds. It can be seen that the estimated parameters are fairly steady over a large range of thresholds. However, the threshold cannot be too high, otherwise no block could pass through the GT process. On the other hand, if there were too few blocks remaining after GT by a high threshold, the estimated parameters would be wrong, either. So, to balance the execution time and estimation accuracy, there should be rules to rationally choose the threshold.

3.3 Self-adaptive Threshold Selection

We have studied the relationship between number of remaining blocks and estimation accuracy, and found that, in many cases, if there were less than 30% blocks remaining, the estimated parameters would lose their accuracy.

For instance, the testing image for figure 2 has totally 330 blocks. As shown in figure 2(b), when the threshold for GT is higher than 100, the number of treacherous blocks will be more than 275 (83.3% to 330). Then in figure 2(a), the curves of estimated parameters become fluctuant. In this case, the proportion of remaining blocks should be larger than 17%.

Based on the above analysis, we recognize that the remaining blocks after GT process should be no less than 30%. Considering the diversity of different image sequences, in our work, we choose such threshold that no less than 50% of the blocks can pass through the GT process.

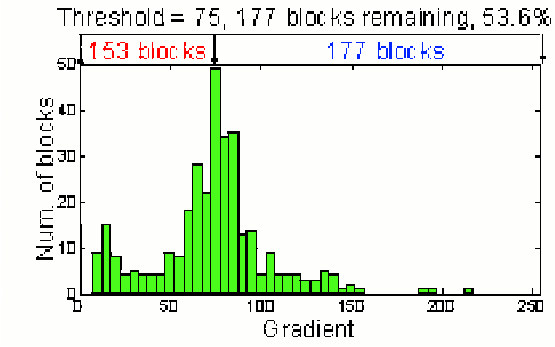


Figure 3. Gradient histogram

By means of gradient histogram, the threshold can be selected automatically. Figure 3 shows the gradient histogram and automatically selected threshold for the testing *Flower Garden* frame.

4 Experimental results

The proposed method was tested on a variety of image sequences, some of which are shown below. In all cases, the method was independently tested between every pair of consecutive frames.

So far, the performance of ILSE with and without GT has been analyzed in which consecutive pairs of frames have been considered for global motion estimation. Table 1 shows the simulation results derived from several testing image sequences. It can be observed that the proposed method has a similar performance compared to the traditional ILSE

technique [2] while reducing computational cost almost 50% in camera motion estimation.

Experimental results have shown that global motion estimation does not require a consideration of all blocks of a frame. It is also shown that the presented Gradient Thresholding (GT) method provides a reasonable pretreatment to the blocks, so that the estimation accuracy is well guaranteed while the execution time is greatly reduced.

5 Conclusion

A fast global motion estimation method is presented in this paper. In this method, traditional ILSE method is modified by using Gradient Thresholding (GT) to analyze and preprocess all of the blocks in a frame before BMA and ILSE. By means of GT, the treacherous blocks who are more likely to produce inaccurate motion estimations will be discarded from the following process. As the result, the computational cost is reduced to a great extent (approximately 50%). We also propose a method to automatically select the threshold for GT based on gradient histogram. So the selection of threshold becomes self-adaptive to different types of images. The presented method has been tested on a variety of image sequences, and the experimental results illustrate its promising performance.

6 Acknowledgements

This research is supported by France Telecom R&D Division, and the National Natural Science Foundation of China (Grant No. 60135020 and 60121302).

References

[1] A. Tekalp, *Digital Video Processing*. Prentice Hall, Englewood Cliffs, NJ, 1995.

[2] G.B. Rath, A. Makur, "Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation", IEEE Trans. on Circuits & Systems for Video Technology, vol. 9, pp. 1075–1099, 1999.

[3] G. Sorwar, M. Murshed, and L. Dooley, "Fast Global Motion Estimation using Iterative Least-Square Technique", 4th International Conference on Information, Communications & Signal Processing and 4th IEEE Pacific-Rim Conference On Multimedia, Dec. 2003.

[4] C.T. Hsu, Y.C. Tsan, "Mosaics of video sequences with moving objects", Proceedings of International Conferences on Image Processing (ICIP'01), vol. 2, pp. 387-390, 2001.

[5] H. Jozawa, K. Kamikura, A. Sagata, H. Kotera, and H. Watanabe, "Two-stage motion compensation using adaptive global MC and local affine MC", IEEE Trans. on Circuits & Systems for Video Technology, vol.7, pp. 75-85, 1997.

[6] H. G. Musmann, P. Pirsh, and H. J. Grallert, "Advances in picture coding", Proc. IEEE, vol. 73, no. 4, pp. 523–548, 1985.

[7] J. R. Jain and A.K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. on Communications, vol. COM-29, pp. 1799–1808, 1981.

[8] L.S. Davis, "A survey of edge detection techniques", Computer Graphics and Image Processing, vol. 4, no. 3, pp. 248-270, 1975.

[9] J.K. Aggarwal, R.O. Duda and A. Rosenfeld, *Computer methods in image analysis*, IEEE Press, New York, 1977.

Table 1. Experimental results

Input frames	Selected Threshold	Statistical comparison			
		Method	**Remaining blocks**	**Time**	**Results** (a_1, a_2, a_3, a_4)
Table Tennis #32, #33 Frame size: 352×240 Block size: 16×16 Number of blocks: 330	85	Without GT	330	1422ms	(-0.02, -0.07, -0.02, -0.15)
		After GT	166 (50.3%)	740ms (52.0%)	(-0.02, -0.06, -0.02, -0.17)
Flower Garden #31, #32 Frame size: 352×240 Block size: 16×16 Number of blocks: 330	75	Without GT	330	1359ms	(0.00, -1.81, 0.00, +0.04)
		After GT	177 (53.6%)	797ms (58.6%)	(0.00, -1.90, 0.00, +0.03)
Skater #30, #31 Frame size: 352×240 Block size: 8×8 Number of blocks: 1320	39	Without GT	1320	5297ms	(0.00, +0.93, 0.00, -0.05)
		After GT	661 (50.1%)	2687ms (50.7%)	(0.00, +0.99, 0.00, -0.04)
Coast Guard #04, #05 Frame size: 176×144 Block size: 8×8 Number of blocks: 396	73	Without GT	396	1594ms	(0.00, +0.94, 0.00, +0.04)
		After GT	203 (51.3%)	828ms (51.9%)	(-0.01, +0.96, 0.00, +0.03)