GLOBAL MOTION ESTIMATION AND ROBUST REGRESSION FOR VIDEO CODING

Kui Zhang and Josef Kittler

Centre for Vision, Speech and Signal Processing,
School of EEIT&M,
University of Surrey,
Guildford GU2 5XH, United Kingdom
e-mail: K.Zhang@surrey.ac.uk, J.Kittler@surrey.ac.uk

ABSTRACT

In the H.263 Version 2 (H.263+) coding standard, the global motion compensation can be introduced by using Reference Picture Resampling (Annex. P) syntax. Such an application requires that the global motion parameters be estimated automatically. In this paper, we propose a global motion estimation algorithm based on the Taylor Expansion Equation and robust regression technique using probabilistic thresholding. The experimental results confirm that the proposed algorithm can improve both coding efficiency and the quality of motion compensation on sequences involving camera movement.

1. INTRODUCTION

Motion estimation and motion compensation techniques are widely used in the video compression algorithms. However, the motion caused by camera movement has not been taken into separate consideration[3, 5] until recent development of MPEG-4 video and H.263+ coding standards[4]. This development leads to the introduction of global motion estimation and global motion compensation. The global motion presented in the image sequence normally consists of complex motion pattern and often "contaminated" by the local motion caused by independent moving objects in the scene. Such a situation requires that the global motion algorithm be capable of dealing with complex motion model as well as local motion contamination.

To address the requirement for global motion estimation, some researchers have proposed approaches based on a so called pan-zoom motion model[2, 1] in which a zoom factor and two translational parameters are used to described the global motion field. However, the assumptions for pan-zoom motion model cannot hold well in video conference applications where the distance between the camera and shooting targets is limited. Furthermore, becasue of the use of differential coding of motion vectors in H.263, any global translational motion field can be encoded very efficiently. This is due to the fact that a zero motion field (compensated by the global motion compensation) and a constant motion field (not compensated by the global motion compensation) will make little difference in coding efficiency for motion vectors. The extra overhead in sending global motion parameters may even reduce the coding efficiency on such kind sequences. Thus, the simple pan-zoom motion model may only gains when there is a zooming field present. A more general approach is to use 6-parameter affine motion model which can described as:

$$\begin{aligned} u(x,y) &= a_0 + a_1 x + a_2 y \\ v(x,y) &= a_3 + a_4 x + a_5 y \end{aligned} \qquad (1)$$

for the global motion field[7] which can deal with majority of motion type encountered in video sequence coding.

2. GLOBAL MOTION ESTIMATION

The difficulty of using affine motion model is due to its high computational demand and the influence of local motion vectors. In our earlier approach[7], the affine motion parameters were obtained by using a Robust Hough Transform (RHT) based method which can separate the global motion field from local motion field during the estimation process. However, the computational load increases in geometric order as the number of parameters increases in the RHT algorithm which requires a search in the parameter space. Another method is to use Taylor expansion equations to derive the affine motion parameters from a set of local motion estimates as described in [6] which suggests to use randomly selected 3 block matching results for median based robust regression. The performance of the robust regression may vary due to the random selection process or large amount of iterations are needed to achieve a stable results. Here, we extend the idea in [6] and develop an analytic based global motion estimation algorithm and robust regression technique which maintain the low computational load of Taylor Expansion based affine motion estimation while obtain the ability to reduce the influence of local motion to certain extend in a limited number of iterations.

To reduce the computational load, a set of sparsely sampled points in the image frame are used as the block centre to calculate the best matching displacement (p, q) and its surrounding 3×3 positions' compensation error. The number of sampled points depends on the image size and normally in the range of 49 to 500. The sampled points should be evenly distributed inside the image frame.

Using the Taylor expansion equation, the local compensation error function at a block position (i, j) can be expressed by

$$\begin{aligned} E_{i,j}(u,v) = \bar{E_{i,j}}(p_{i,j},q_{i,j}) + (u-p_{i,j})\frac{\partial E_{i,j}}{\partial u} + \\ (v-q_{i,j})\frac{\partial E_{i,j}}{\partial v} + \frac{1}{2}(u-p_{i,j})^2\frac{\partial^2 E_{i,j}}{\partial^2 u} + \\ \frac{1}{2}(v-q_{i,j})^2\frac{\partial^2 E_{i,j}}{\partial^2 v} + (u-p_{i,j})(v-q_{i,j})\frac{\partial^2 E_{i,j}}{\partial u \partial v} \end{aligned} \quad (2)$$

Let $\mathbf{S}$ be the 3×3 values of motion compensation error surrounding the best matched displacement (p, q) as

$$\begin{aligned} \mathbf{S} = (e(p-1,q-1), e(p-1,q), e(p-1,q+1), \\ e(p,q-1), e(p,q), e(p,q+1), \\ e(p+1,q-1), e(p+1,q), e(p+1,q+1))^t \end{aligned} \quad (3)$$

where $e(x, y)$ denotes the value of motion compensation error at position (x, y). Using differential operators, the partial derivative items in Equation (2) can be expressed as:

$$\begin{aligned} \bar{E}(p,q) &= \tfrac{1}{9}(-1,2,-1,2,5,2,-1,2,-1)\mathbf{S} \\ \tfrac{\partial E(p,q)}{\partial u} &= \tfrac{1}{6}(-1,0,1,-1,0,1,-1,0,1)\mathbf{S} \\ \tfrac{\partial E(p,q)}{\partial v} &= \tfrac{1}{6}(-1,-1,-1,0,0,0,1,1,1)\mathbf{S} \\ \tfrac{\partial^2 E(p,q)}{\partial^2 u} &= \tfrac{1}{3}(1,-2,1,1,-2,1,1,-2,1)\mathbf{S} \\ \tfrac{\partial^2 E(p,q)}{\partial^2 v} &= \tfrac{1}{3}(1,1,1,-2,-2,-2,1,1,1)\mathbf{S} \\ \tfrac{\partial^2 E(p,q)}{\partial u \partial v} &= \tfrac{1}{4}(1,0,-1,0,0,0,-1,0,1)\mathbf{S} \end{aligned} \quad (4)$$

To facilitate discussion, we rewrite Equation (1) in a matrix form as:

$$\left(\begin{array}{c} u \\ v \end{array} \right) = \mathbf{P}(\mathbf{x}_{ij})\mathbf{a} \quad (5)$$

where

$$\mathbf{P}(\mathbf{x}_{ij}) = \left\{ \begin{array}{cccccc} 1 & x_{ij} & y_{ij} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_{ij} & y_{ij} \end{array} \right\}$$

$$\mathbf{a} = (a_0, a_1, a_2, a_3, a_4, a_5)^t$$

If we substitute (u, v) in Equation (2) by Equation (5), the global motion estimation can be obtained by minimising

$$\sum_{i,j} E_{i,j}(\mathbf{P}(\mathbf{x}_{ij})\mathbf{a}) \quad (6)$$

where $\mathbf{x}_{ij}$ denotes the centre position of the block (i, j). It leads to Eular equation as

$$\frac{\partial \sum_{i,j} E_{i,j}(\mathbf{P}(\mathbf{x}_{ij})\mathbf{a})}{\partial \mathbf{a}} = 0 \quad (7)$$

and therefore

$$\mathbf{A}\mathbf{a} - \mathbf{b} = 0 \quad (8)$$

where

$$\mathbf{A} = \sum_{i,j} \left\{ \begin{array}{cc} \frac{\partial^2 E_{i,j}}{\partial^2 u}\mathbf{U}(\mathbf{x_{ij}}) & \frac{\partial^2 E_{i,j}}{\partial u \partial v}\mathbf{U}(\mathbf{x_{ij}}) \\ \frac{\partial^2 E_{i,j}}{\partial u \partial v}\mathbf{U}(\mathbf{x_{ij}}) & \frac{\partial^2 E_{i,j}}{\partial^2 u}\mathbf{U}(\mathbf{x_{ij}}) \end{array} \right\}$$

$$\mathbf{U}(\mathbf{x_{ij}}) = \left\{ \begin{array}{ccc} 1 & x & y \\ x & x^2 & xy \\ y & xy & y^2 \end{array} \right\}$$

$$\mathbf{b} = \sum_{i,j} \mathbf{P}(\mathbf{x}_{ij})^t \left(\begin{array}{c} -\frac{\partial E_{i,j}}{\partial u} + \frac{\partial^2 E_{i,j}}{\partial^2 u} p_{i,j} + \frac{\partial^2 E_{i,j}}{\partial u \partial v} q_{i,j} \\ -\frac{\partial E_{i,j}}{\partial v} + \frac{\partial^2 E_{i,j}}{\partial^2 v} q_{i,j} + \frac{\partial^2 E_{i,j}}{\partial u \partial v} p_{i,j} \end{array} \right)$$

By solving the Equation (8), we can obtain the affine coefficients as

$$\mathbf{a} = \mathbf{A}^{-1}\mathbf{b} \quad (9)$$

3. ROBUST REGRESSION

Because of the presence of local motion, the above procedure cannot guarantee an accurate estimation of the global motion parameters without eliminating the influence of local motion. Here, we propose a robust regression method based on the statistic distribution of the compensation error and use a pair of probability thresholds to classify inliers and outliers. The basic idea of this method is to assume that in each iteration, only a small portion of the compensation error is much greater than the average compensation error. Those points have a compensation error much greater than the average compensation error should be identified as outliers. This classification is based on the assumption that the motion parameters are computed only using inliers. Apparently, when some of the points in the inlier set are classified as outliers, the original assumption is no longer valid. Furthermore, as the previous classification of inliers and outliers is not completely validated, we need to check the outlier set to see if there is any mis-classification in the previous iteration. We assume that compensation error of a mis-classified point is smaller than a threshold close to the mean of compensation errors for inliers. When this process repeated certain number of times, we shall be able to separate points under influence of local motion from those points contributing to global motion only. The whole robust regression procedure can be described as follows:

1. For all M points, where M is the total number of sampled points used for global motion estimation, using block matching algorithm calculate the best match (p, q) for each point and its surrounding 3×3 values of motion compensation error. Using Equation (3) and (4) to calculate derivative items.

2. Mark all M points as inliers and set the iteration counter to N.

3. Estimate global motion parameters using only inliers;

4. Calculate the motion compensation error for each central point using Equation (2), (4) and (5).

5. Calculate the mean and standard deviation of values of the motion compensation error for all inliers using:

$$\bar{E} = \frac{1}{K} \sum_{\mathbf{x_{i,j}} \in Inliers} E_{i,j}(\mathbf{P}(\mathbf{x}_{ij})\mathbf{a}) \quad (10)$$

$$\sigma = \sqrt{\frac{1}{K-1} \sum_{\mathbf{x_{i,j}} \in Inliers} (E_{i,j}(\mathbf{P}(\mathbf{x}_{ij})\mathbf{a}) - \bar{E})^2} \quad (11)$$

   where K is the number of inliers.

6. Calculate the upper threshold and lower threshold using

$$\begin{aligned} T_{upper} &= \bar{E} + C_{upper}\sigma \\ T_{lower} &= \bar{E} + C_{lower}\sigma \end{aligned} \quad (12)$$

   where $\bar{E}$ and σ are the mean and the standard deviation of the compensation error as given in (10) and (11), C_{upper} and C_{lower} are constants. Both constants correspond to probabilities in Normal distribution. Given a pair of predetermined probability P_{upper} and P_{lower}, the constants C_{upper} and C_{lower} can be obtained by solving:

$$\begin{aligned} P_{upper} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{C_upper} e^{\frac{x^2}{2}} dx \\ P_{lower} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{C_lower} e^{\frac{x^2}{2}} dx \end{aligned} \quad (13)$$

7. For all inliers, if an inlier's motion compensation error is greater than the upper threshold, it is marked as an outlier.

8. For all outliers, if an outlier's motion compensation error is less than the lower threshold, it is marked as an inlier.

9. Decrease iteration counter by one; if the iteration counter is greater than zero, then goto step 3; otherwise return the estimated parameters.

The robust regression procedure effectively improves the accuracy of global motion estimation when local motion is present. However, the computational load is much higher due to the iteration procedure. Because of the way we classify outliers and inliers, the minimum number of iterations can be estimated by the following equation:

$$I_{min} = \frac{log(P_{upper})}{log(P_{lower})} \quad (14)$$

where P_{upper} and P_{lower} are the probabilities corresponding to C_{upper} and C_{lower} respectively, as defined in Equation (13).

4. EXPERIMENTAL RESULTS

The experiments were carried out using several CIF image sequences with global motion. We use evenly sampled 81 points in the image as the centre for block matching with a window size of 15×15. The search windows are of the size of 64×64. The P_{upper} and P_{lower} are fixed at 0.975 and 0.64 respectively which correspond to C_{upper} and C_{lower} being 1.96 and 0.36 respectively. The minimum required iteration number using the given P_{upper} and P_{lower} is 17.627. In our experiment, 20 iterations are used for the robust regression.

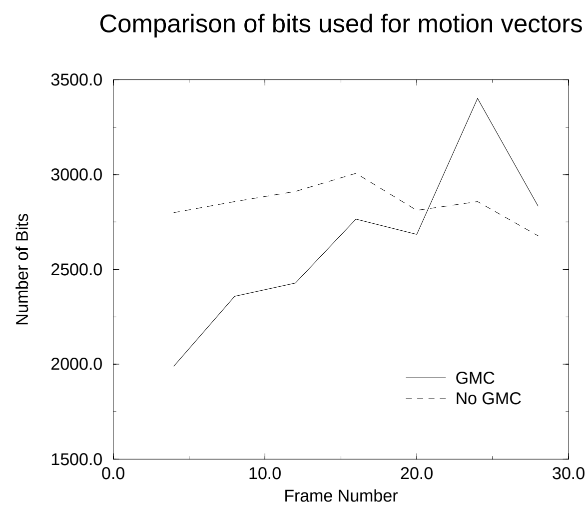

Figure 1: Bits spending on motion vector coding

Figure 1 gives the comparison results on the number of bits used for motion vector coding for "MIT" sequence. It can be seen that the use of global motion estimation and global motion compensation effectively reduce the number of bits required to encode the motion vectors. The curve shows that the gain from the use of global motion compensation varies from frame to frame.

Sequence	INTRA Blocks		INTER Blocks	
	NoGMC	GMC	NoGMC	GMC
Bus	550	445	1025	1134
Mall	512	411	1031	1119
MIT	53	6	1503	1539

Table 1: Number of INTRA and INTER coded blocks

Further investigation shows that the use of global motion compensation can significantly reduce the number of blocks being encoded in INTRA mode (and increase the number of blocks being encoded in INTER mode). Table 1 gives the comparison results on coding mode for the test sequences. It is clear that the use of global motion compensation on sequences with global motion can effectively improve the quality of motion compensation which is indicated by the reduction of the number of INTRA coded blocks (where motion compensation fails). Because of the increase in the number of blocks being coded in INTER mode, the number of motion vectors also increases. This may explain why the number of bits used for encoding motion vectors increases sometimes for GMC enabled algorithm.

The coding performance comparison is given in Table 2. The results show that significant gain comes from sequences with complex motion pattern like “MIT”. In spite of only panning is involved in sequence “Bus” and “Mall”, the global motion compensation still gains due to the fact that the close range panning of camera will result in a perspective motion field rather than the pure translational motion field. As we have mentioned in the introduction, if the global motion is a pure translational motion, the use of global motion compensation may not gain any thing. Therefore, the use of affine motion model is justified even for sequences only involving pan and tilt movement of the camera.

Sequence	PSNR (dB)		Bitrate (kbits/sec.)	
	NoGMC	GMC	NoGMC	GMC
Bus	30.11	29.97	514.43	505.27
Mall	33.67	33.51	198.57	195.34
MIT	29.80	29.67	356.54	318.08

Table 2: Performance comparison - PSNR and Bitrate

5. CONCLUSION

In this paper, we extended work on global motion estimation algorithms based on block matching and the Taylor Expansion Equation. The advantage of using the Taylor Expansion Equation is to avoid computationally expensive high-dimensional search for the affine parameters. To reduce the influence of local motion, a robust regression technique using probabilistic thresholding is proposed. Given the required probability threshold, the minimum number of iterations can easily be determined so that any unnecessary iterations can be avoided. The experimental results show that the use of global motion compensation can improve the coding efficiency on sequences involving complex global motion field. It also indicates that the use of global motion compensation can improve the quality of motion compensation so that the number of INTRA coded blocks is reduced. The gain on sequences only with close range camera panning show that the existence of complex motion in such kind of sequences in which the simple pan-zoom motion model may not work very well.

6. REFERENCES

[1] Marco Barbieri and Rosa Lancini. Effective motion field description based on affine models and global motion information. In *EUSIPCO-96,Trieste, Italy*, 10-13 September, 1996.

[2] H.Jozawa, K.Kamikura, K.Yanaka, and H.Watanabe. Video coding using adaptive global mc and local affine mc. In *EUSIPCO-96,Trieste, Italy*, 10-13 September, 1996.

[3] ITU-T. *Recommendation H.261: Video Codec for Audiovisual Services at p×64 kbits*. International Telecommunication Union, 1994.

[4] ITU-T. *Draft ITU-T Recommendation H.263: Video Coding for Low Bitrate Communication*. International Telecommunication Union, 26 September, 1997.

[5] ITU-T. *Draft ITU-T Recommendation H.263: Video Coding for Low Bitrate Communication*. International Telecommunication Union, 5 July, 1995.

[6] ISO/IEC JTC1/SC29/WG11 N1385. *Description of core experiments on coding efficiency in MPEG-4 video*. International Organisation for Standardisation, September, 1996.

[7] K. Zhang, M. Bober, and J. Kittler. Video coding using affine motion compensated prediction. In *ICASSP-96,Atlanta, USA*, volume 4, pages 1978–1981, 7-10 May, 1996.