FAST PROGRESSIVE MODEL REFINEMENT GLOBAL MOTION ESTIMATION ALGORITHM WITH PREDICTION

Haifeng Wang, Jia Wang, Qingshan Liu, Hanqing Lu

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences

ABSTRACT

Global Motion Estimation (GME) is an important part in the object-based applications. In this paper, a fast progressive model refinement (FPMR) GME algorithm is proposed. It can select the appropriate motion model according to the complexity of the camera motion. Two techniques are used to accelerate the procedure of FPMR. The first is an outlier prediction based feature point selection method. It can predict outliers from that of the last frame and therefore can effectively remove the influence of outliers on parameter calculation. The second is an intermediate-level model prediction method, which is used to fast the model selection and the parameter calculation procedure. Experiments show that the proposed algorithm is above two times faster than that of the Feature-based Fast and Robust GME technique.

1. INTRODUCTION

The MPEG-4 standard permits operations on arbitrarily shaped video object planes (VOPs), such as transmitting, retrieving, downloading, storing and reusing [1]. The first step of all these operations is to segment video objects from the scene, which further needs to perform global motion compensation beforehand. GME plays an important role in these object-based applications.

The commonly used GME algorithms can be grouped into two classes: motion vector-based GME algorithms [2] [3] and pixel-based GME algorithms [4] [5]. The first class needs to get motion vectors (MVs), and then use these MVs to calculate the model parameters. It is suitable for compressed video analysis, because the motion vectors can be directly obtained from the compressed bit-streams. Pixel-based GME methods calculate the parameters by using gradient descent algorithm to minimize the sum of squared intensity errors between the current frame and the reference frame. The GME algorithm adopted in MPEG-4 is Feature-based Fast and Robust GME technique (FFRGMET), which is pixel-based [5].

Although FFRGMET is about 7 times faster than the original GME algorithm in MPEG-4 verification model (VM), it is still not fast enough to enable a real-time application [6]. Many improvements have been made to accelerate the speed of the three-level GME procedure.

These improvements mainly focused on the selection of feature points (FPs) and the computation complexity reduction of Levenberg-Marquadet algorithm. Keller used a Pixel Subset Selection method to reduce the number of FPs and an interpolation-free formulation of GM to decrease the computational complexity [7]. Chan introduced the motion vector prediction and early termination method which have been applied successfully in local ME to accelerate the GME algorithm [8]. Qi advised to use the object information to reject outliers [9].

All these GME algorithms use one motion model, usually affine model to describe the global motion. There are two disadvantages for this method. If the camera only undergoes a translational motion, the affine model may seem too complex when compared with the translational model. On the other hand, the affine model can not describe all kinds of camera motions. If the camera motion is too complex and is beyond the scope of affine model, the affine model may not describe the camera motion precisely. The perspective model is preferred in this situation.

In this paper, we proposed a Fast Progressive Model Refinement (FPMR) algorithm to select the appropriate motion model to describe different camera motions. An outlier prediction based FP selection (OPBFPS) method is used to reduce the influence of outliers. Based on the correlation of motion model and model parameters between neighbor frames, an intermediate-level model prediction (ILMP) method is proposed. If this method is used, the calculation can start from the intermediate level. The down-sampling, the translational estimation and the parameter calculation at the top level are omitted.

The following of the paper is arranged as follows: Section 2 describes the outlier prediction based feature point selection (FPS) method and the intermediate-level model prediction method. Section 3 presents the FPMR algorithm. Experiment results and analysis are shown in Section 4. Section 5 concludes the paper.

2. OUTLIER AND MODEL PREDICTION

2.1. Outlier Prediction based FPS Method

A pixel subset selection method was proposed in [7]. This method divides the whole image into 100 sub-regions and

selects the top 10% as FPs to avoid numerical instabilities caused by the concentration of FPs. We take this method to a further step and divide the image into 4 regions as shown in figure 1. For the first inter-predicted frame, no information can be used to predict the FPs, the number of FPs remained in each region is adjusted according to the common cases. That is, more FPs are remained in region 1 and 3, while fewer FPs are remained in region 2. FPs in region 4 are located at the boundaries of the image. These FPs are removed because their corresponding pixels in the reference frame are most likely out of the range of the frame.
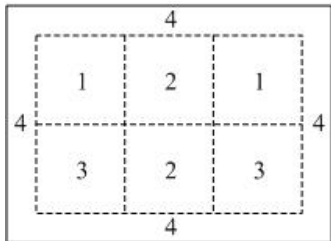

Figure1. Region Partition

As for the next inter-predicted frames, the outlier blocks of the last frame detected by the residual block-based outlier detection method proposed in [5] are used to predict the location of outliers in the current frame. In this case, region 1, 2 and 3 are treated in the same way and the top 10% FPs are remained. But FPs in region 4 are still removed no matter whether the corresponding block is an outlier block or not. Table 1 shows the relationship of the number of FPs needed by FFRGMET and OPBFPS. We can see that the number of FPs needed by OPBFPS is about one quarter of that of FFRGMET. Experiment results in Section 4 show that the PSNR performance when applying OPBFPS only drops by about 0.1 when compared with that of FFRGMET.

Table1. Comparison between FP Selection Methods

Frame No.	115	120	125	130	135
FFRGMET	15902	17584	18602	18298	18965
OPBFPS	6193	4095	7412	4244	4344

2.2. Intermediate-level Model Prediction

The three-level pyramid structure provides great flexibility for model prediction. It can be started from the top level, from the intermediate level, or from the bottom level. Experiments show that prediction from the intermediate level performs better than the other two cases when considering the prediction precision and the execution time. From figure 2, we can see that the PSNR performance of ILMP is a little better, and the average time needed is 120ms as compared with 133ms of the top level and 116ms of the bottom level. Another reason for starting from the intermediate level is that there is a model refinement step between the intermediate level and the bottom level. We can increase the order of model selected according to the prediction precision of the intermediate level.

In ILMP, the model information (model type and model parameters) at the intermediate level of the current frame is predicted from that of the corresponding level in the last frame. The top level down-sampling and the 3-step block matching step are omitted. The GME starts directly from the intermediate level.
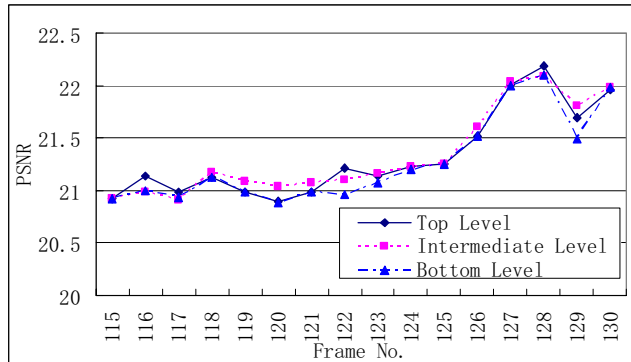

Figure2. PSNR Comparison among Different Level Prediction (basketball)

3. FAST PROGRESSIVE MODEL REFINEMENT

Using one motion model for all kinds of sequences has its disadvantages. Su proposed an adaptive model decision (AMD) method, in which lower order model is used to fit the input MV fields first, and then increase the model order to see if better matching can be obtained [10]. This method has to perform an extra estimation with higher order model before determining the final model. The GME algorithms adopted by MPEG-4 are three-level based, which facilitates the model selection.

In FPMR, Model refinement (MR) takes place between levels. Three motion models, that is, translational model, affine model and perspective model are used in our algorithm. For the first inter-prediction frame, the top level always utilizes the translational model. Whether the motion models of the next two levels remain the same or upgrade to higher order models depends on the GMC results. The progressive here means MR will follow an order from the translational model to the affine model and to the perspective model. The motion model won't jump from the translational model directly to the perspective model. For the following inter-prediction frames, the top level calculation is omitted. The model information of the intermediate level is predicted from that of the last frame. MR step only happens between the intermediate level and the bottom level. The flowchart of FPMR is shown in figure 3.

For the first inter-prediction frame, all the three switches connect to the vertical line. This is the basic structure of PMR. After the top level calculation, the parameters obtained are projected to the intermediate level. The global motion compensated result is used by MR step to determine whether to increase the model order or not. The compensated result, which we called residual, is partitioned

into blocks. We count the number of blocks whose sum of residual exceeds a predefined threshold. If the number is larger than a percent of the total block number, higher order motion mode will be used in the next level. It should be noted that both the two MR steps use the image data after one down-sample step. This is because the image at the intermediate level contains more information than that of the top level and is much simpler than that of the bottom level.
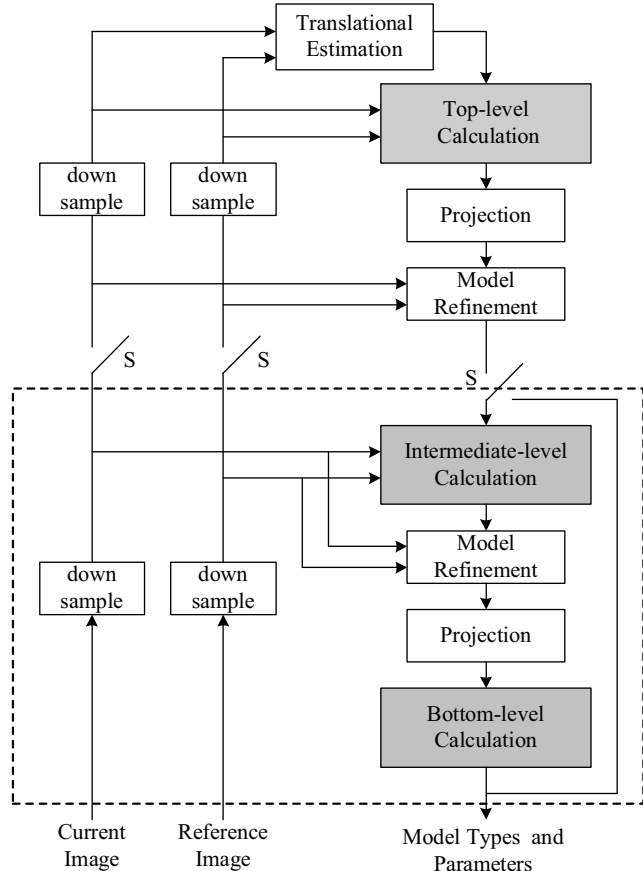


Figure3. The Flowchart of FPMR

After the first inter-prediction frame GME calculation, the intermediate-level model type and parameters are used as the initial values of the next frame. In this case, the left two switches in figure 3 are disconnected from the vertical line and the right one is connected to the horizontal line. By using ILMP, one step down-sampling, 3-step translational estimation, top-level parameter calculation, projection and MR are omitted, which accelerate the GME procedure.

## 4. EXPERIMENT RESULTS AND ANALYSIS

In order to demonstrate the performance of FPMR algorithm proposed here, we built a test environment using C++ language and realized it in Visual C++. The CPU of the computer used is Pentium 4 2.4GHz, and the memory is 512MB. Two CIF-format sequences: basketball and coastguard are tested.
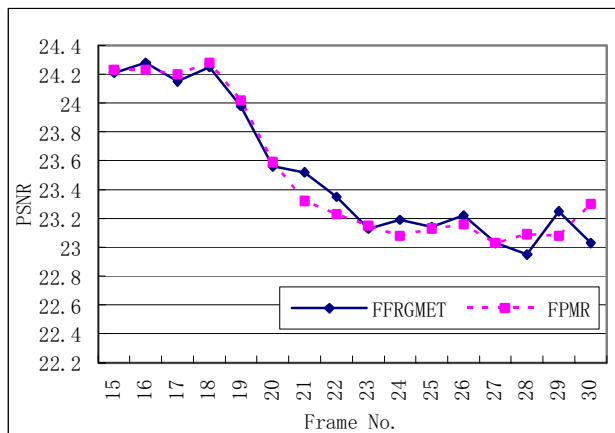
(a) FPMR-Translational vs. FFRGMET
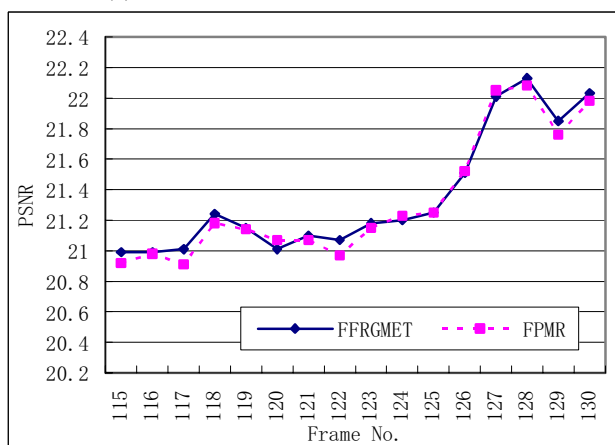
(b) FPMR-Affine vs. FFRGMET
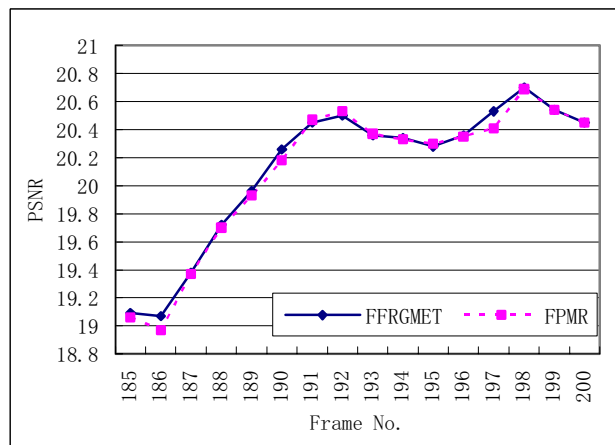
(c) FPMR-Perspective vs. FFRGMET
Figure4. PSNR Comparison
(a) coastguard, (b) basketball, and (c) basketball

FFRGMET algorithm is also realized to compare its performance with that of FPMR. Here, we use the peak signal-noise ratio (PSNR) between the current frame and the frame reconstructed using the estimated global motion parameters as an objective measure for the GME precision.

In order to demonstrate the performance of FPMR algorithm under different situations, different parts in basketball and coastguard test sequences are selected. The PSNR results are shown in Figure 4. For FPMR, the final model selected in figure 4(a), (b) and (c) is translational model, affine model and perspective model, respectively. We can see that the prediction precision of our FPMR algorithm is totally comparable with that of FFRGMET. The average PSNR reduced in the three cases is 0.008, 0.029 and 0.021, respectively.

The average time needed by FFRGMET and FPMR in the three cases in figure 4 is shown in table 2. Figure 5 gives a detailed comparison between the two algorithms when the affine model is the final model selected in FPMR. We can see that the speed of FPMR is about 2.5 times faster than that of FFRGMET.

Combine the PSNR and the speed results, we can see that FPMR is about 2.5 faster than FFRGMET with little precision loss.

Table2 Average Time Needed by FFRGMET and FPMR (second)

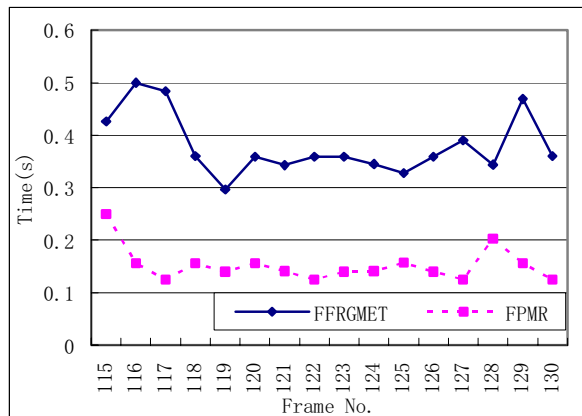| Final Model	Translational	Affine	Perspective
FFRGMET	0.329	0.380	0.404
FPMR	0.124	0.152	0.161

Figure5. Speed Comparison

5. CONCLUSION

In this paper, we proposed a fast progressive model refinement algorithm, which can select the appropriate model according to the complexity of the camera motion. In order to speed up the algorithm, we proposed an outlier prediction based feature point selection method and an intermediate-level model prediction method. These two methods can reduce the computation complexity and effectively reduce the influence of outliers on the parameter calculation. Experiments show that the speed of our algorithm is above two times faster than that of FFRGMET.

6. REFERENCES

[1] Jen-Chi Huang, and Wen-Shyong Hsieh, "Automatic Feature-based Global Motion Estimation in Video Sequences", IEEE Transactions on Consumer Electronics, pp. 911-915, Aug. 2004.

[2] Rath. G.B., and Makur. A., "Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation", Circuits and Systems for Video Technology, IEEE Transactions on, Volume 9, Issue 7, pp:1075-1099, Oct. 1999.

[3] J. Wang, H.F. Wang, Q.S. Liu, and H.Q. Lu, "Fast Global Motion Estimation via Iterative Least-Square Method", accepted by ACCV2006, India, Jan. 2006.

[4] Dufaux. F., and Konrad. J., "Efficient, robust, and fast global motion estimation for video coding", IEEE Transactions on Image Processing, Volume 9, Issue 3, pp:497-501, Mar. 2000.

[5] Y.W. He, W. Qi, S.Q. Yang, Y.Z. Zhong, Feature-based Fast and Robust Global Motion Estimation Technique for Sprite Coding, Meeting of ISO/IEC JTC1/SC29/WG11, MPEG00/M6226, Beijing, July 2000.

[6] MPEG, "Text of 14496-7 PDTR", ISO/IEC JTC1/SC29/WG11 N4344, Sydney, July 2001.

[7] Y. Keller, and A. Averbuch, "Fast Global Motion Estimation For MPEG-4 Video Compression" International Conference on Image Processing (ICIP) 2002, Rochester, USA, September 2002.

[8] W.C. Chan, O.C. Au, and M.F. Fu, "Improved Global Motion Estimation Using Prediction and Early Termination", International Conference on Image Processing, Volume 2, pp. II-285 - II-288, Sept. 2002.

[9] B. Qi, A. A, "Robust and Fast Global Motion Estimation Oriented to Video Object Segmentation", IEEE International Conference on Image Processing, pp. 153-156, Sept. 2005.

[10] Y.P. Su, M.T. Sun, and V. Hsu, "Global Motion Estimation from Coarsely Sampled Motion Vector Field and the Applications", IEEE Transactions on Circuits and Systems for Video Technology, pp.232-242, Feb. 2005.