

Workshop: 2002 International Computer Symposium (ICS2002)
Workshop on Multimedia Technology

Title of paper: Block-Based Motion Field Segmentation for Video Coding

Abstract:

In the past few years, motion compensation has been widely used in the coding of image sequences. Most of motion estimation and compensation are based on the Block-Based framework. The framework simplifies the complexity of motion estimation, but it gives over constraints to the motion field, which results in inaccuracies on the boundary of moving objects.

This paper presents a novel technique for raising motion field accuracy. The method improves motion compensation along boundaries of moving objects by using several pattern types to segment the motion field of previous frames of a sequence. The segmentation is based on the MAP framework that uses iterative method to obtain the solution. In addition, we also develop a predictive scheme to predict the location of motion field discontinuities in the current frame, which further reduce the side information of segmentation.

Keyword : MAP- Maximum a Posteriori Probability

Author : Chung-Ming Kuo, kuocm@isu.edu.tw ,
Yong Ren Huang, d890302d@stmail.isu.edu.tw ,
Feng-Chung Haung, b651218@ms13.hinet.net ,
Chaur-Heh Hsieh, hsieh@isu.edu.tw
* Dept. of Information Engineering, I-Shou University Tahsu , Kaohsiung,
840, Taiwan
Tel. 886-7-6577711-6501, Fax. 886-7-6578930

Contact author: Chung-Ming Kuo, kuocm@isu.edu.tw

Block-Based Motion Field Segmentation for Video Coding

Chung-Ming Kuo, Yong Ren Huang,
Feng-Chung Haung and Chaur-Heh Hsieh
* Dept. of Information Engineering, I-Shou University
Tahsu, Kaohsiung, 840, Taiwan
Tel. 886-7-6577711-6501, Fax. 886-7-6578930
NSC-90-2213-E-214-033
E-mail: {[kuocm](mailto:kuocm@isu.edu.tw), [hsieh](mailto:hsieh@isu.edu.tw)}@isu.edu.tw

Abstract

In the past few years, motion compensation has been widely used in the coding of image sequences. Most of motion estimation and compensation are based on the Block-Based framework. The framework simplifies the complexity of motion estimation, but it gives over constraints to the motion field, which results in inaccuracies on the boundary of moving objects.

This paper presents a novel technique for raising motion field accuracy. The method improves motion compensation along boundaries of moving objects by using several pattern types to segment the motion field of previous frames of a sequence. The segmentation is based on the MAP framework that uses iterative method to obtain the solution. In addition, we also develop a predictive scheme to predict the location of motion field discontinuities in the current frame, which further reduce the side information of segmentation.

Keyword : MAP- Maximum a Posteriori Probability

Introduction

Motion compensation plays an important role in video coding since it can reduce the temporal redundancy significantly [1]~[7]. The block-matching algorithm (BMA) is the most popular motion estimation/compensation technique and was adopted in the current video coding standards such as H.261/3, MPEG-1/2. To limit the transmission bit rate of MVs, the BMA assumes that the constant motion field over entire block. Therefore, the assumption results the inaccurate estimate of moving objects boundary, and a significantly compensation error is introduced. For the low bit rate applications, the constraint of bit rate of prediction error results the block artifacts along the boundary of moving objects.

To solve this problem, the idea of motion field segmentation has been proposed [5][8]. The methods improve the performance of BMA by relaxing the constraints of single translation motion in a block. Since the blocks may contain boundary of moving object, segmenting the motion field in these blocks and applying different motion vector to each region in a block achieve the performance improvement. In [8], the motion segmentation technique can reconstruct discontinuity of motion field at pixel resolution with much better performance than conventional BMA.

In this paper, we will propose a block-based motion field segmentation technique. Instead of pixel resolution, the new methods using block edge pattern to represent the object boundary in a block. Since the block pattern types are much simpler than the pixel based segmentation, and the condition of smoothness and connectivity between neighboring blocks can be easily defined, therefore, the new technique not only reduces computation complexity significantly but also achieve smooth motion field. On the other hand, the segmentation of each block only contains edge types, thus the extra side bits are much less than arbitrary shapes that obtained by conventional techniques. Furthermore, reducing the block size, the new technique can progressively refine the segmentation precision from coarse (16×16) to fine (2×2) that is the same resolution as conventional methods. Finally, when considering the application of motion estimation/compensation in video coding, the predictive motion field segmentation scheme will also be developed with same bit rate and higher performance than Orchard's method.

2. Problem Formulation

In this Section, some notations and definition for motion estimation are introduced. The pixels in a frame are defined on a rectangular lattices S , and the elements $s \in S$ are located on the coordinates $(x, y)^T$ where x and y are the horizontal and vertical directions, respectively. Let $I(s, n)$ denote the intensity at pixel s of frame n , and $\hat{I}(s, n)$ denotes the intensity of the corresponding decoded frame. The set of motion vectors $\{v_s(n)\}_{s \in S}$ represent the motion field of frame n . Therefore, the displaced frame difference $DFD(s, v, n)$ caused by motion vector v at

pixel s of frame n is defined as

$$DFD(s, v, n) = I(s, n) - \hat{I}(s - v, n - 1). \quad (1)$$

Block-based motion estimation assumes that the frame S is partitioned into a rectangular blocks $b(i, j)$ each with $N \times N$, and each block $b(i, j)$ is assigned a single motion vector $v_{b(i, j)}(n)$ as

$$v_s(n) = v_{b(i, j)}(n) \text{ if } s \in b(i, j) \quad (2)$$

Under such constraint, although the bit rates required for transmitting motion vectors can be reduced significantly, the motion fields are also restricted. The most straightforward way to obtain the motion vectors is matching all the blocks in a pre-defined region at previous frame as

$$v_{b(i, j)}(n) = \arg \min_{v \in W} \sum_{s \in b(i, j)} (DFD(s, v, n))^2 \quad (3)$$

where W is the pre-defined region which motion vectors considered.

There are two main drawbacks with block-based motion estimation. One is that the blocks located at boundaries of moving objects are not compensated accurately. Since the blocks on the boundaries contain at least two different regions, therefore, only one motion vector that applied to entire block will cause incorrect compensation and thus degraded the performance of motion compensation. Because the blocks motion vectors are obtained independently, the other drawback is that for a regions or objects, which have same motion characteristics, the distributions of motion fields are not smooth enough.

In this Section, we will describe some necessary assumptions and then develop the framework of motion field segmentation. For block-based motion field segmentation, some assumptions are given beforehand.

(1). Each block small enough, so contains at most 2 regions and motion field in the block is smooth, thus only one boundary in a block. See Fig. 1.

(2). High correlation exist adjacent blocks, so the motion vectors are selected from neighborhood system. See Fig. 2.

(3). The motion field between blocks are connected and smoothed, so a probability model can be defined.

According to assumption 1, we can specific the edge types of block to simplify the motion field distributions in a block. According to assumption 2, the motion vectors in a block can select from the neighboring blocks. Finally, using assumption 3, a MAP framework of motion field segmentation can be constructed.

Let X is two-dimensional (2-D) random field that represents the segmentation of motion filed in a block. Since at most two motion vectors is defined in a block, so we design five feature patterns: smooth and four types of edge patterns, as shown in Fig. 3. We denote the patterns as a set of $P = \{m, h, v, l, r\}$, where m represents the smooth pattern, and h, v, l, r represent horizontal, vertical, left-diagonal, and right-diagonal edge patterns, respectively. We aim to classify every block into one of the five designed patterns by using MAP criterion. That is, we would like to obtain the pattern-label mapping of each block. The pattern-label mapping of all blocks is a random field $X = \{X_b = x_b \mid b \in I(s, n), x_b \in \{m, h, v, l, r\}\}$. Let two frames $I(s, n)$ and $\hat{I}(s, n-1)$ are current frame and decoded previous frame, respectively. Then the *a posteriori* probability density function of segmentation can be expressed as $P(X = x \mid I(s, n), \hat{I}(s, n-1), v_{b_{i,j}(n)}^1, v_{b_{i,j}(n)}^2)$, where $v_{b_{i,j}(n)}^1$ and $v_{b_{i,j}(n)}^2$ are the two possible motion vectors for block $b(i, j)$ in frame n . For simplicity, hereafter we use $I_n, \hat{I}_{n-1}, v_b^1$ and v_b^2 to instead of original symbol. By using Bayesian formulation, the motion field segmentation \hat{x} can compute by solving equation

$$\begin{aligned} \hat{x}_b &= \arg \min_{x \in P} \{P(X = x \mid I_n, \hat{I}_{n-1}, v_b^1, v_b^2)\} \\ &= \operatorname{argmin}_{x \in P} \{P(I_n \mid X = x, \hat{I}_{n-1}, v_b^1, v_b^2) \cdot P(X = x)\} \end{aligned} \quad (4)$$

where P is the pattern set. Now the problem is how to calculate the conditional probability (the first term of the sum in Eq.(4)) and the *a priori* probability (the second term).

The first term is a likelihood function, we assume that all blocks are independent each other, given the block label configuration. The second term, we consider *a priori* probability $P(X)$, which is used to impose the spatial smoothness on the block patterns. The underlying pattern of the block random field is assumed to be a Gibbs random field (GRF), which is a

special class of GRF called multilevel logistic (MLL) Gibbs distribution [6]. Then

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{c \in C} -V_c(x)\right), \quad (5)$$

where Z is a normalizing constant, $V_c(x)$ is a clique potential for a clique c , and C is the set of all cliques associated with the specific neighborhood system.

4. Block-Based Motion Field Segmentation (BMFS)

4.1 MAP Estimation of Block Segmentations

In this Section, we will describe the segmentation algorithm in detail. In this work, the first order neighborhood system η^1 is adopted because it is simple and sufficient to define smoothness and connectivity. η^1 consists of the closest four neighbors of a block, hence it is well known as the nearest-neighbor model. The neighborhood system and the corresponding clique type that contains single-block and pair-block clique potentials are shown in Fig. 4. In the algorithm, only pair-cliques are considered. According Eq. (4), the motion field segmentation can be expressed as

$$\begin{aligned} \hat{x}_b &= \underset{x \in P}{\operatorname{argmin}} \left\{ P(I_n | X = x, \hat{I}_{n-1}, v_b^1, v_b^2) \cdot P(X = x) \right\} \\ &\propto \underset{x \in P}{\operatorname{argmin}} \left\{ -\ln P(I_n | X = x, \hat{I}_{n-1}, v_b^1, v_b^2) - \ln P(X = x) \right\} \end{aligned} \quad (6)$$

The first term is the compensation error, which assume white Gaussian distribution with $N \sim (0, \sigma^2)$. So

$$\left\{ -\ln P(I_n | X = x, \hat{I}_{n-1}, v_b^1, v_b^2) \right\} = \frac{1}{2\sigma^2} \sum_{x \in P} DFD((s, v(D, x_b = p, f), n))^2 \quad (7)$$

The $v(D, x_b = p, f)$ is the motion vector corresponding to block segmentation. Where D is the selected two motion vectors corresponding to block b with segmentation $x_b=p$ and motion vectors indicator f . Therefore, the calculation of Eq. (7) contains two parts: 1. Find the two motion vectors in a block, 2. using the two motion vectors to find the optimal segmentation x_b and the motion vectors location. For the first part, we define the neighborhood system (see Fig.4) of block $b(i,j)$ as

$$N_{b,i,j} = \{b(i-1, j), b(i, j-1), b(i+1, j), b(i, j+1)\}$$

The motion vector $v_{b(i,j)}^1$ and $v_{b(i,j)}^2$ for block $b(i,j)$ are selected from the neighborhood system.

For each block, we start with the motion vector selection, and then estimate the segmentation and the motion vectors location accordingly. The motion vectors selection is as:

$$(v_b^1, v_b^2) = \arg \min_D \sum_{s \in b(i,j)} \min(DFD(s, v(D), n)^2) \quad (8)$$

and then

$$\hat{x}_b = \arg \min_{(P,F)} \sum_{s \in b(i,j)} \min(DFD(s, v(x=p, f), n)^2) \quad (9)$$

where $D \in \{(v_b^1, v_b^2); v_b^1, v_b^2 \in (v_{b(i,j)} \cup v_{N_{b(i,j)}})\}$ is the two motion vectors of a block selected from neighborhood system, $p \in P$ is the block edge pattern types and $f \in F = \{0,1\}$ is the indicator to identify the motion vectors corresponding to the region of segmentation.

The second term of Eq. (5) represents *a priori* information provided by the smoothness constraints between the segmentation fields of adjacent blocks. To impose this smoothness constraint, a Gibbs random fields (GRF) model is used as the *a priori* distribution $P(X=x)$ of the segmentation field.

4.2 GRF Model for the segmentation Field

Now we consider *a priori* probability $P(X=x)$, which is used to impose the spatial smoothness on the block patterns. Then

$$P(X=x) = \frac{1}{Z} \exp\left(\sum_{c \in C} -V_c(x)\right), \quad (10)$$

where Z is a normalizing constant, $V_c(x)$ is a clique potential for a clique c , and C is the set of all cliques associated with the specified neighborhood system. The detailed descriptions of GRF and the related topics can be found in [6].

In this work, the first order neighborhood system η^1 is adopted because it is simple and sufficient to define smoothness and connectivity. η^1 consists of the closest four neighbors of a block, hence it is well known as the nearest-neighbor model. The neighborhood system and the corresponding clique type that contains single-block and pair-block clique potentials are shown

in Fig. 4. As mentioned in Section 3, we only consider the pair-clique in our work.

As pointed out in [6], the pair-block clique potentials $V_c(x)$ in Eq. (10) represent the continuity and smoothness with the neighboring block configurations. Therefore, the $V_c(x)$ is defined as

$$V_c(x) = \begin{cases} k & \text{if } c \in \text{full - connected (f.c.)} \\ 0 & \text{if } c \in \text{half - connected (h.c.)} \\ -k & \text{if } c \in \text{non - connected (n.c.)} \end{cases} \quad (11)$$

where f.c., h.c. and n.c. are illustrated in Fig. 5. It is a simple idea, when neighboring blocks are connected with the same motion vector, which satisfies continuity and smoothness, so should be given a reward. Otherwise, when neighboring blocks are connected with the different motion vector, should be given a penalty.

Now taking Eq. (8), (9), combining the result with Eq. (11), and discarding the constant terms, the objective function of Eq. (5), which is to be minimized, become

$$\hat{x}_b = \arg \min_{(P,F)} \left\{ \sum_{s \in b_{i,j}} \min(DFD(s, v(x = p, f), n)^2) + \sum_{c \in C} V_c(x = p) \right\} \quad (12)$$

Finding the optimum block segmentation \hat{x}_b that satisfies Eq. (12) for all possible configurations of P is computational prohibitive. In this work, we used a deterministic relaxation scheme in [6] to update block pattern type iteratively. That is, we determine the segmentation of each block in a frame one-by-one, and then update the segmentation in an iterative manner. As in [6], we employ the criterion $\arg \min_{x \in P} \left\{ -\ln P(I_n | X = x, \hat{I}_{n-1}, v_b^1, v_b^2) \right\}$ to determine the pattern types of blocks. The results are used as initial conditions for the subsequent iteration process. Then, traveling every block in a frame, we update the block type x_b depending on its neighboring block types. By doing the procedure iteratively until there is no more change of the block type or the allowed maximum number of iterations is reached, we can obtain the final block segmentation.

4.3 Refinement of BMFS

The BMFS start with block size 16×16 , so the segmentation of motion field is coarse.

Therefore, we can further refine with small block size and using the coarse segmentation as an initial guess. When we partition a 16×16 block into four 8×8 sub-blocks, then these 8×8 sub-blocks have specified segmentation pattern types depending on the 16×16 segmentation pattern types. For example, a 16×16 block has a horizontal segmentation type, and then the four 8×8 sub-blocks has uniform segmentation types, as shown in Fig. 6. According to the initial segmentation, we can apply the MAP estimation for the 8×8 sub-blocks in the frame and new motion field segmentation with 8×8 resolution can be obtained. Repeating this procedure, we can refine the block motion field segmentation to 4×4 and 2×2 and then more and more detailed motion field segmentation can be achieved.

4.4 Predictive BMFS (P-BMFS)

In this Section, we assume that block segmentation computed by BMFS is a portion of an object, which moves through time with purely translation motion. And then we can backward along the motion trajectory and identify the motion field segmentation \hat{x}_b relating to the corresponding portion of the same object. Fig. 7 demonstrates such idea in the new approach. The object moves with velocity v_{object}^n against the stationary background. Using BMFS the block $b_{i,j}$ is segmented into two regions, one is assigned the motion v_{object}^n and one is assigned the motion of background. The predictive BMFS identifies the block $b_{i,j}^*$ in the frame $n-1$, in which the block $b_{i,j}$ is offset by motion vector $-v_{object}^n$. So the segmentation of the motion fields of block $b_{i,j}^*$ in frame $n-1$ approximates the segmentation $b_{i,j}$ in frame n . Since the segmentation of $b_{i,j}^*$ is using motion vector v_{n-1} and $DFD(s,v,n-1)$. Thus the segmentation can be calculated at the decoder directly.

Under such assumption, the motion vector to identify the b^* for each block b in current frame is finding first. And then we select the motion vectors from the b^* neighborhood system. The selection is

$$(v_b^1, v_b^2) = \arg \min_{D^*} \sum_{s \in b_{i,j}^*} \min(DFD(s, v, n-1)^2) \quad (13)$$

where $D^* \in \{(v_b^1, v_b^2); v_b^1, v_b^2 \in (v_{b^*(i,j)} \cup v_{N_{b^*(i,j)}})\}$.

When the motion vectors are selected, then we can apply BMFS to segment the motion field of b^* . The resulting motion field segmentation can be used as block $b_{i,j}$ segmentation, and the procedures can be computed in the decoder directly. Considering the computational complexity and bit rate saving, we restrict one of the motion vectors of b^* to be motion vector of block $b_{i,j}$, and the other one is selected from neighborhood system. Therefore, we need 2 bits to represent the motion vector of the selection, 1 bit to represent the block motion vector of block $b_{i,j}$, and 1 bit to represent the motion distribution. So 4 bits is sufficiently to represent the block segmentation.

5. Simulation Results

The simulations are performed on the test sequences ‘‘Salesman’’ and ‘‘Table-tennis’’, which are CIF format with 352×288 luminance pixels, $2 \times 176 \times 144$ chrominance pixels and the frame rate of 30 Hz. In our simulations, we only consider the luminance sequences. For comparison, the full search algorithm (FSA), the Orchard’s method C and method D are also performed using the same test sequences. For all BMA, the block size is 16×16 , and search ranges are ± 7 . We compare the experimental results of different methods in terms of average PSNR, the necessary bits to represent the block segmentation, the computation cost in time, visual quality and the distribution of motion field.

The PSNR (dB) is defined as

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}, \quad (14)$$

MSE is the mean square error between the original image frame and predicted frame. Table 1 list the FSA, Orchard’s method C, and BMFS, respectively. We can find that no matter how Orchard’s or proposed methods improves the PSNR significantly. The method C improves PSNR from 0.5dB to 1.1dB on average. The BMFS for block size 16×16 is slight lower than method C, but superior to that FSA about 0.3dB to 0.8 dB on average. While BMFS starts to refine using smaller sub-blocks, the performance increases immediately. We can find that the

PSNR is closing to method C with 8×8 sub-blocks, and higher than method C with 4×4 sub-blocks. Therefore, the block-based motion field segmentation with specified block pattern types achieves very good performance.

Compare P-BMFS with Orchard's method D, we will see new method also achieves better performance with lower computational cost. So we can conclude that the block-based motion field segmentation with specific segmentation pattern not only have better performance but also have lower computational cost and smaller overhead.

Now we consider the visual quality. The first example is ping-pong paddle; see Fig. 8. Compare the method D and P-BMFS, and you will see method D have some compensation errors especially in the edge of the paddle. Finally, the same situation occurs in the Salesman. Please see the triangle object in the circle of the Figure 9.

Fig. 10 is the motion field of the Table tennis; we can see that in BMFS the motion field is more and more detailed from coarse to fine, also smooth and uniform compared with method C. We can see that the motion field of ping-pong ball and paddle is very consistent with the real object and better than method C. Also in the P-BMFS, not only motion field is consistent with the real object but also smoother than Orchard's method. The same results can be seen in the Fig. 11, which is the motion field of Salesman.

Observing the performances evaluation, we can conclude that the new techniques have the advantages as: higher PSNR and visual quality, lower computational cost, overhead and smooth motion field distribution. It is also adjustable depending on our emphasis on motion compensation or object extraction.

6. Conclusion

In this paper, a novel block-based motion field segmentation with specified block segmentation pattern types were proposed. There are three algorithms developed: 1. BMFS, 2. P-BMFS. The new methods are suitable for both motion compensation and object extraction. The BMFS efficiently achieve higher performance in both PSNR and visual quality. If the transmission rate is limited, P-BMFS can reduce bit rate significantly with reasonable performance. On the contrary, if the application of video object extraction is considered, the

proposed method is also a very good approach. The proposed methods provide smooth and uniform motion field especially at the object boundary. In the future, the object extraction using motion field segmentation will be further investigated.

References

1. A.N. Netravali and J.D. Robbins, "Motion compensated television coding—Part I," *Bell Syst. Tech. J.*, vol. 58, pp.631-670, Mar. 1979.
2. J.R. Jain and A.K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. 29(12), pp.1799-1808,1981
3. T. Koga, K. Inuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," In *Proc. NTC81*, pp. C9.6.1-9.6.5 New Orleans, LA, 1981.
4. R. Srinivasan and K. Rao, "Predictive coding based on efficient motion estimation," *IEEE Trans. Commun.*, vol. 33(8), pp.888-896,1985
5. Michael M. Chang, A.M. Tekalp, M. Ibrahim Sezsn, "Simultaneous Motion Estimation and Segmentation," *IEEE Trans. Image Processing* vol. 6, no.9, pp.1326-1333,SEP 1997.
6. C.M. Kuo, C.H. Hsieh, Y.R. Huang and S. L. Zen, "A New Mesh-Based Temporal-Spatial Segmentation for Image Sequence," *IEEE COMPSAC-2000*, pp. 395-400, Oct. 2000.
7. C.H Hsieh, P.C. Lu, J.S. Shyu and E.H., "A motion estimation algorithm using interblock correlation," *Electronic Letters*, vol. 26(5), pp.276-277 ,1990
8. Orchard, M.T., "Predictive Motion-Field Segmentation for Image Sequence Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3(1) , pp. 54 –70, Feb. 1993

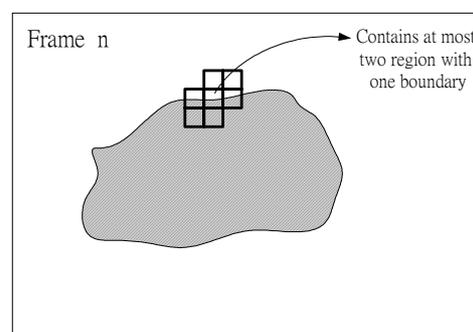


Fig. 1. The block small enough contains at most two regions and one boundary

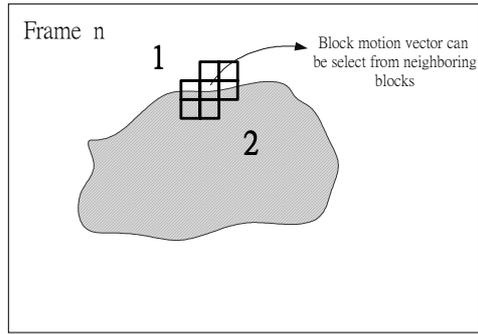


Fig. 2. The motion vector in a block can select from the neighborhood

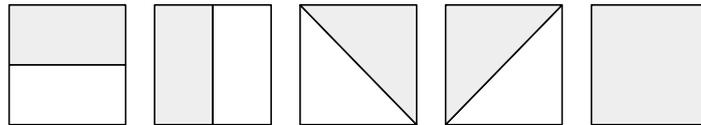


Fig. 3. Five specified feature pattern include horizontal, vertical, right-diagonal, left-diagonal and smooth for the block motion field segmentation

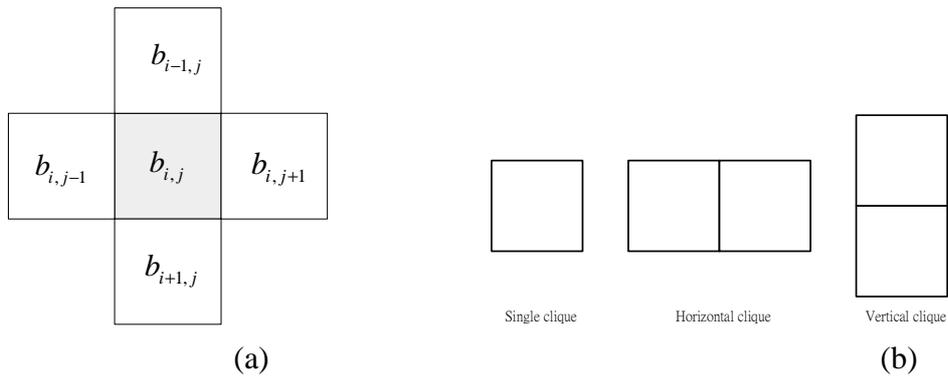


Fig. 4. (a) The first order neighborhood system, (b) The corresponding clique types.

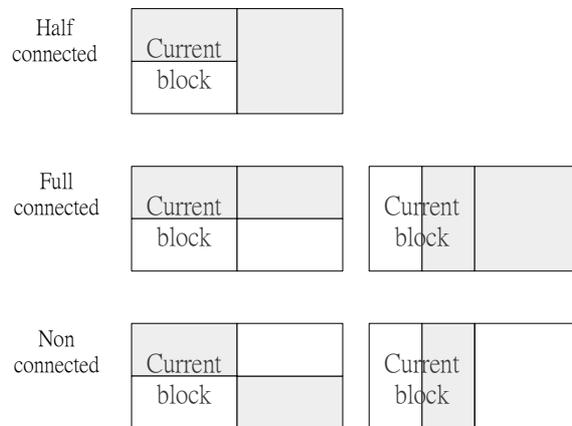


Fig. 5. The conditions of h.c., f.c. and n.c.

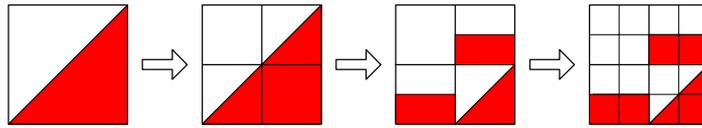


Fig. 6. BMFS refinement with sub-blocks

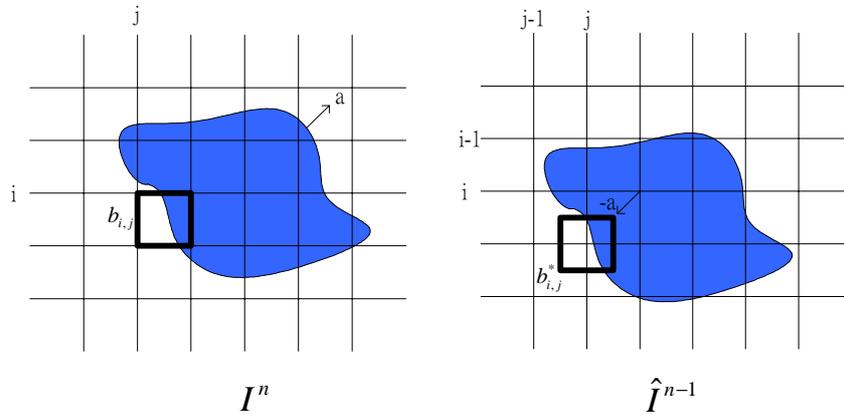


Fig. 7. The motion field segmentation of $b_{i,j}$ in frame n is approximated by $b_{i,j}^*$ in frame

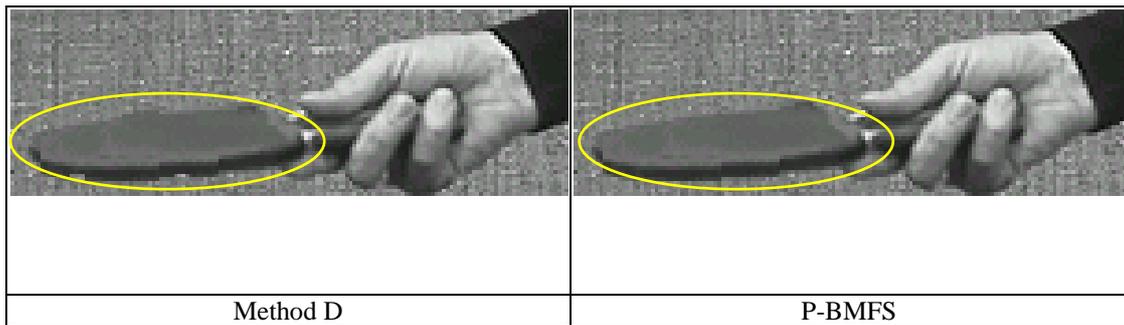


Fig. 8. The visual quality of paddle for predictive method

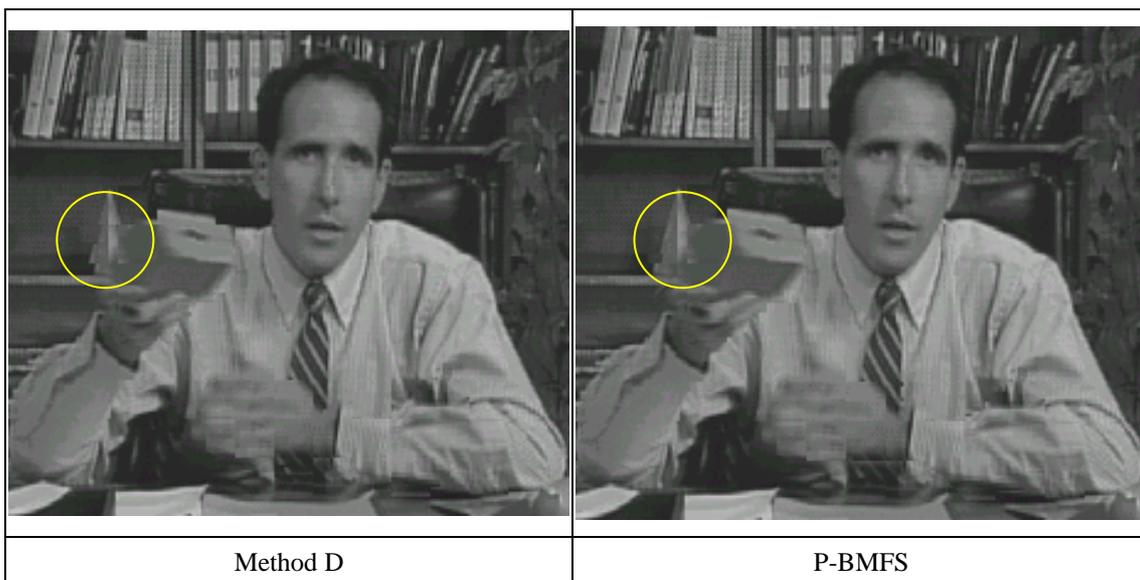


Fig. 9. The visual quality of Salesman for various methods

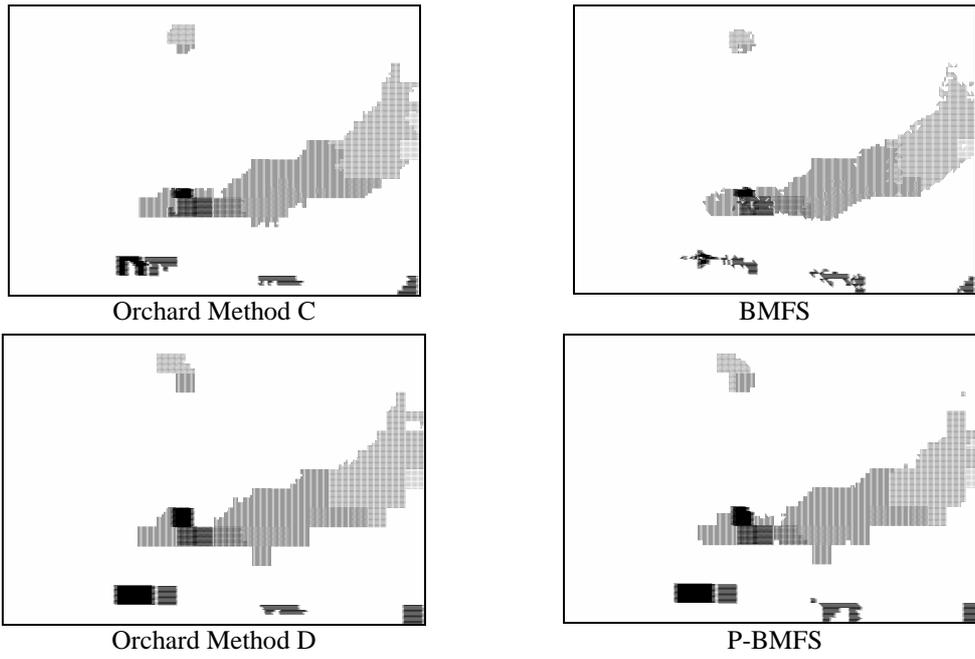


Fig. 10. The motion field of Table tennis for various methods

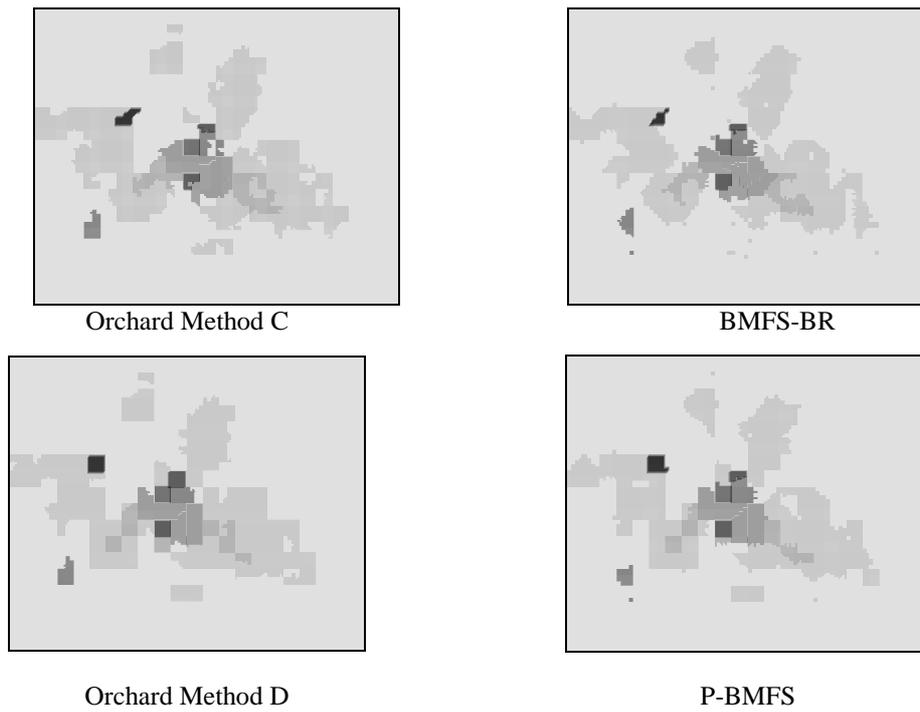


Fig. 11. The motion field of Salesman for various methods

Table 1. The performance indexes for various methods

Sequences	Method Indexes	FSA	Orchard's		BMFS			P-BMFS
			C	D	16×16	8×8	4×4	
Salesman	PSNR	34.80	35.30	34.95	35.17	35.38	35.51	35.00
Salesman	Overhead		20~30	4	4	16	64	4
Salesman	Time	54.61	177.06	333.15	82.37	98.65	115.1	187.89
Table-Tennis	PSNR	30.17	31.27	30.50	30.98	31.51	31.92	30.60
Table-Tennis	Overhead		20~30	4	4	16	64	4
Table-Tennis	Time	93.05	195.72	385.27	117.94	133.28	146.90	246.94