Motion Restoration — A Method for Object and Global Motion Estimation

Jih-Shi Su†, Hsueh-Ming Hang†, and David W. Lin†‡

†Dept. of Electronics Engineering and Center for Telecommunications Research
National Chiao Tung University
Hsinchu, Taiwan 30050, ROC

‡Bellcore
445 South Street
Morristown, NJ 07960-6438, USA

Abstract

A new technique called motion restoration method (MRM) for estimating the global motion due to zoom and pan of the camera is proposed. It is composed of three steps: (a) block-matching motion estimation, (b) object assignment, and (c) global motion restoration. In this method, each image is first divided into a number of blocks. Step (a) may employ any suitable block-matching motion estimation algorithm to produce a set of motion vectors which capture the compound effect of zoom, pan, and object movement. Step (b) groups the blocks which share common global motion characteristics into one object. Step (c) then extracts the global motion parameters (zoom and pan) corresponding to each object from the compound motion vectors of its constituent blocks. The extraction of global motion parameters is accomplished via singular value decomposition (SVD). Experimental results show that this new technique is efficient in reducing the entropy of the block motion vectors for both zooming and panning motions and may also be used for image segmentation.

Keywords: motion restoration, object assignment, central projection, singular value decomposition

1 Introduction

Motion estimation plays an important role in video data compression which exploits the high temporal redundancy between successive frames of a video sequence to achieve high compression ratio. It is also used in segmentation of images for computer vision applications. The most common technique of motion estimation employed in video coding is block matching[1]–[3]. In this technique, a single motion vector is estimated for each image block by comparing the current-frame image block to the blocks in the previous frame that correspond to different displacement vectors. And the displacement vector that minimizes a predetermined error criterion is chosen. The assumption underlying

0-8194-1638-X/94/$6.00

block-matching motion estimation is that all the pixels inside a block are undergoing the same translational motion. As a result, this approach may generate a significant proportion of motion vectors that do not correspond to true motion. This imprecise estimation will increase the prediction error and reduce the compression ratio. Therefore, methods that can cope with more general forms of motion (including translation, zoom, pan, and deformation) have been the focus of a great deal of research in recent years [4]–[7].

We propose a method called *motion restoration* to estimate local as well as global motion. This method consists of three steps: (a) block-matching motion estimation, (b) object assignment, and (c) global motion restoration. The first step estimates translational motion in a block-by-block fashion and it may employ any appropriate block-matching algorithm (BMA). The two remaining steps then extract the zooming and panning components from the block motion vectors obtained in the first step. The entropy in the motion vectors is thereby reduced. As a result, we can reduce the amount of data to be transmitted. Or we may use a smaller block size so that the amount of data is not reduced but the BMA could yield a more accurate estimate to start with, alleviating the inaccuracy problem associated with traditional block-based motion estimation.

This paper is organized as follows. In Section 2, we give a mathematical description for general global and object motion. In Section 3, the proposed motion restoration method (MRM) is derived. Section 4 is devoted to the presentation and discussion of experimental results. Section 5 is the conclusion.

2 Mathematical Description of Global Motion

To match the mechanism of ordinary video cameras, we use central projection to model the motion traces on the recorded images caused by object or camera movement (i.e., zoom, pan, etc.). Figure 1 illustrates our model. P is a point of interest on an object. Let

(x, y, z) = object-space coordinates of the point P,

(X, Y) = image-plane coordinates of the image point P', and

F = z-coordinate of the image-plane in object-space.

Based on similarity between the triangles $\triangle OPR$ and $\triangle OP'S$, we have

$$\frac{F}{z} = \frac{\overline{OS}}{\overline{OR}} = \frac{Y}{y}. \tag{1}$$

Therefore,

$$Y = F\frac{y}{z}. \tag{2}$$

Similarly, we also have

$$X = F\frac{x}{z}. \tag{3}$$

A general movement consisting of zoom, pan, and object motion is depicted in Figure 2. Let V_o be the displacement vector of point P and let V_{ox}, V_{oy}, and V_{oz} be its x-directional, y-directional and z-directional components, respectively. Geometry then gives

$$\begin{cases} X_2 = \frac{F_2}{z'}x' = \frac{F_2}{z+V_{oz}}(x + V_{ox} - V_{px}) \\ Y_2 = \frac{F_2}{z'}y' = \frac{F_2}{z+V_{oz}}(y + V_{oy} - V_{py}) \end{cases}, \tag{4}$$

where (X_2, Y_2) is the projection of $P(x', y', z')$ on the image plane and (V_{px}, V_{py}) is the panning vector of camera (or, equivalently, image coordinates). Now note that (from Equations (2) and (3))

$$x = \frac{z}{F_1}X_1, \qquad y = \frac{z}{F_1}Y_1. \tag{5}$$

Inserting Equation (5) into Equation (4), we obtain

$$\begin{cases} X_2 = \frac{F_2}{z+V_{oz}}(\frac{z}{F_1}X_1 + V_{ox} - V_{px}) \\ Y_2 = \frac{F_2}{z+V_{oz}}(\frac{z}{F_1}Y_1 + V_{oy} - V_{py}) \end{cases}. \tag{6}$$

Therefore, the corresponding vector (V_x, V_y) on the image plane due to the combination of object and camera movement is

$$\begin{cases} V_x = X_2 - X_1 = (\frac{F_2}{F_1}\frac{z}{z+V_{oz}} - 1)X_1 - \frac{F_2}{z+V_{oz}}V_{px} + \frac{F_2}{z+V_{oz}}V_{ox} \\ V_y = Y_2 - Y_1 = (\frac{F_2}{F_1}\frac{z}{z+V_{oz}} - 1)Y_1 - \frac{F_2}{z+V_{oz}}V_{py} + \frac{F_2}{z+V_{oz}}V_{oy} \end{cases}. \tag{7}$$

For simplicity, we rewrite Equation (7) as

$$\begin{cases} V_x = \mathcal{Z}X_1 + \mathcal{P}V_{px} + \mathcal{V}_{ox} \\ V_y = \mathcal{Z}Y_1 + \mathcal{P}V_{py} + \mathcal{V}_{oy} \end{cases}, \tag{8}$$

where

$$\begin{cases} \mathcal{Z} = (\frac{F_2}{F_1}\frac{z}{z+V_{oz}} - 1) \\ \mathcal{P} = -\frac{F_2}{z+V_{oz}} \\ \mathcal{V}_{ox} = \frac{F_2}{z+V_{oz}}V_{ox} \\ \mathcal{V}_{oy} = \frac{F_2}{z+V_{oz}}V_{oy} \end{cases}.$$

The first term in the righthand side of Equation (7) is due to camera zoom. The second term is caused by pan. And the third term is the projection of the object's movement on the image plane. In the next section, we derive a method to restore the motion components, i.e., the pan, zoom, and object motion parameters.

3 Motion Restoration Method (MRM)

The architecture of the motion restoration method is shown in Figure 3. It consists of three steps:

1. *Motion Estimation:*

   In this step, the motion vector of each block of an image is obtained using a suitable BMA (e.g., full search, 3-step search, or others). In our simulation, we use the full-search BMA to estimate these motion vectors. The resultant motion vector field forms the basis of the following two steps.

2. *Object Assignment:*

   We assign the the blocks that share certain common global motion characteristics to the same object. The assignment criteria are summarized in two *object-assignment theorems* to be described later.

3. *Motion Restoration:*

The motion components due to camera zooming and panning are extracted in this step. Hence, the object movement is separated from the camera motion.

3.1 Object Assignment

For simplicity, images are divided into blocks and each block is viewed as a single computational unit. Thus, let (X_1, Y_1) in Equation (8) refer to the center of a block. Let A and B be two image blocks. According to Equation (8), we have

$$\left\{ \begin{array}{lcl} V_{xA} & = & \mathcal{Z}_A X_A + \mathcal{P}_A V_{px} + \mathcal{V}_{oxA} \\ V_{yA} & = & \mathcal{Z}_A Y_A + \mathcal{P}_A V_{py} + \mathcal{V}_{oyA} \end{array} \right. \tag{9}$$

and

$$\left\{ \begin{array}{lcl} V_{xB} & = & \mathcal{Z}_B X_B + \mathcal{P}_B V_{px} + \mathcal{V}_{oxB} \\ V_{yB} & = & \mathcal{Z}_B Y_B + \mathcal{P}_B V_{py} + \mathcal{V}_{oyB} \end{array} \right. . \tag{10}$$

If these two blocks belong to the same object, then the corresponding motion parameters $(\mathcal{Z}, \mathcal{P}, \mathcal{V}_o)$ will be equal. Assuming this is true, we subtract Equation (10) from Equation (9) and obtain

$$\frac{V_{xA} - V_{xB}}{V_{yA} - V_{yB}} = \frac{X_A - X_B}{Y_A - Y_B}. \tag{11}$$

Therefore, we have the following *object-assignment theorem.*

Theorem 1 If zoom motion exists ($\mathcal{Z}$ is nonzero) and two blocks 1 and 2 belong to the same object, then

$$\frac{V_{x1} - V_{x2}}{V_{y1} - V_{y2}} = \frac{X_1 - X_2}{Y_1 - Y_2},$$

where

$$\begin{array}{lcl} (X_i, Y_i) & = & \textit{the central coordinates of block } i, \\ (V_{xi}, V_{yi}) & = & \textit{the observed motion vector in the image plane.} \end{array}$$

Theorem 1 is valid under rather general conditions; that is, when both zoom and pan exist. Assume that only panning exists, then Equations (9) and (10) become

$$\left\{ \begin{array}{lcl} V_{xA} & = & \mathcal{P}_A V_{px} + \mathcal{V}_{oxA} \\ V_{yA} & = & \mathcal{P}_A V_{py} + \mathcal{V}_{oyA} \end{array} \right. \tag{12}$$

and

$$\left\{ \begin{array}{lcl} V_{xB} & = & \mathcal{P}_B V_{px} + \mathcal{V}_{oxB} \\ V_{yB} & = & \mathcal{P}_B V_{py} + \mathcal{V}_{oyB} \end{array} \right. . \tag{13}$$

If the two blocks belong to the same object, then

$$\left\{ \begin{array}{lcl} V_{xA} & = & V_{xB} \\ V_{yA} & = & V_{yB} \end{array} \right. . \tag{14}$$

Thus we obtain the *second object-assignment theorem.*

Theorem 2 If zoom motion does not exist and the two blocks A and B belong to the same object, then

$$\begin{cases} V_{xA} = V_{xB} \\ V_{yA} = V_{yB} \end{cases}.$$

Based on the above object-assignment theorems, the blocks in the whole image can be grouped into a number of objects. The blocks that belong to the same object have the same global and object motion vectors. Therefore, for an object containing p blocks, we have

$$\begin{cases} V_{x1} = \mathcal{Z}X_1 + \mathcal{P}V_{px} + \mathcal{V}_{ox} \\ V_{y1} = \mathcal{Z}Y_1 + \mathcal{P}V_{py} + \mathcal{V}_{oy} \\ \vdots \\ V_{xp} = \mathcal{Z}X_p + \mathcal{P}V_{px} + \mathcal{V}_{ox} \\ V_{yp} = \mathcal{Z}Y_p + \mathcal{P}V_{py} + \mathcal{V}_{oy} \end{cases}, \tag{15}$$

where $\mathcal{Z}$ and $\mathcal{P}$ are identical for all the p blocks. This set of linear equation can be abbreviated as $AW = b$.

In object assignment, we first index each block in an image in ascending numerical order as shown in Figure 4. The blocks are denoted B_i, $i = 1, \ldots, N$. We then invoke the following procedure.

Step 0: Set j=1. Let all blocks be unmarked.

Step 1: Among all unmarked blocks, choose the one with the smallest index as the reference block and denote it B_{ref}. Mark this block and assign it to object j.

Step 2: For each remaining unmarked block, test it against B_{ref} for Equality (11) or (14). If equality holds, then mark it and assign it to object j.

Step 3: If all blocks are marked, then stop. Otherwise let j=j+1 and go to Step 1.

In the next subsection, we discuss how motion restoration is performed on the "objectized" image to compute the global motion parameters for each object.

3.2 Cascaded Motion Restoration

One way to implement the motion restoration block shown in Figure 3 is to decompose it into two cascaded sub-steps for separate zoom and pan estimation as depicted in Figure 5. This figure shows that the motion vectors of an arbitrary object are first processed for zooming estimation which extracts the zoom vector V_z from the motion vector V. The difference vector $V_r = V - V_z$ is then processed for panning estimation and is separated into a pan vector V_p and an object motion vector $V_{obj} = V_r - V_p$. The two estimation sub-steps may be reversed to yield a pan-plus-zoom (P+Z) architecture instead of the depicted zoom-plus-pan (Z+P) architecture. The overall organization of the complete motion restoration process can therefore have a number of variants. The four that we considered are denoted as schemes A, A', B, and B', respectively, in Figure 6. Schemes A and A' employ *Theorem 1* in object assignment, while schemes B and B' employ *Theorem 2*.

We next describe in more detail how each sub-step in the cascaded motion restoration can be performed, assuming a Z+P architecture. Equations for the P+Z architecture can be similarly derived.

a. *Zooming estimation*

Assuming $\mathcal{P}V_p + \mathcal{V}_o = V_r$, we can rewrite Equation (15) as

$$\left\{ \begin{array}{lcl} V_{x1} & = & \mathcal{Z}X_1 + V_{rx} \\ V_{y1} & = & \mathcal{Z}Y_1 + V_{ry} \\ & \vdots & \\ V_{xp} & = & \mathcal{Z}X_p + V_{rx} \\ V_{yp} & = & \mathcal{Z}Y_p + V_{ry} \end{array} \right. . \tag{16}$$

The above equation can be expressed in matrix notations as

$$A_z W_z = b_z,$$

where

$$\begin{array}{rcl} A_z & = & \begin{bmatrix} X_1 & 1 & 0 \\ \vdots & & \\ X_p & 1 & 0 \\ Y_1 & 0 & 1 \\ \vdots & & \\ Y_p & 0 & 1 \end{bmatrix}, \\ W_z & = & [\mathcal{Z}, V_{rx}, V_{ry}]^T, and \\ b_z & = & \left[\underbrace{V_{x1}, \ldots, V_{xp}}_{p}, \underbrace{V_{y1}, \ldots, V_{y1}}_{p} \right]^T . \end{array}$$

Using the *singular value decomposition* (SVD) technique[8], we can obtain the solution as

$$W_z = A_z^\dagger b_z.$$

After removing the zooming factor $\mathcal{Z}$, $V_r = (V_{rx}, V_{ry})$ is passed to the next sub-step.

b. *Panning estimation*

From the result of zooming estimation, we remove the $\mathcal{Z}$ component in Equation (15) and obtain

$$\left\{ \begin{array}{lcl} V_{rx} & = & \mathcal{P}V_{px} + \mathcal{V}_{ox} \\ V_{ry} & = & \mathcal{P}V_{py} + \mathcal{V}_{oy} \\ & \cdots & \\ V_{rx} & = & \mathcal{P}V_{px} + \mathcal{V}_{ox} \\ V_{ry} & = & \mathcal{P}V_{py} + \mathcal{V}_{oy} \end{array} \right. . \tag{17}$$

This can be rewritten as

$$A_p W_p = b_p,$$

where

$$
\begin{aligned}
A_p &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & & & \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \vdots & & & \\ 0 & 0 & 1 & 1 \end{bmatrix}, \\
W_p &= [\mathcal{P}V_{px}, \mathcal{V}_{ox}, \mathcal{P}V_{py}, \mathcal{V}_{oy}]^T, and \\
b_p &= \Bigg[\underbrace{V_x, \ldots, V_x}_{p}, \underbrace{V_y, \ldots, V_y}_{p}\Bigg]^T .
\end{aligned}
$$

Applying SVD again, we can obtain the panning vector from

$$W_p = A_p^{\dagger} b_p.$$

4 Experimental Results

The proposed algorithm is tested on a variety of image sequences. We present the results from using the flower garden and the table tennis sequences. Each sequence contains 30 pictures at a resolution of 720×480 per picture. The flower garden sequence contains panning activity only, whereas the table tennis sequence has individual object movement as well. Besides the four MRM schemes outlined previously, we also consider a *zero-forcing* (ZF) MRM in which the object motion vectors are set to zero. This approach is based on the assumption that object motion vectors do not affect significantly the estimated zoom and pan vectors and hence can be neglected in their estimation.

The numerical results are summarized in Figures 7–10, in which we compare the entropy of the block motion vectors as well as the PSNR before and after motion restoration. The entropy values are computed frame-by-frame using the statistics of each frame separately. In addition, since the block motion vectors after the extraction of global motion components may not be integers, they are quantized prior to entropy computation. The block size is 16×16 in all experiments. The figures show that the MRM can reduce the entropy of the block motion vectors and increase the PSNR. Interestingly, the ZF MRM is found to significantly outperform other MRM schemes in both entropy reduction and PSNR gain in some cases.

For the flower garden sequence, Figure 8 shows that schemes A' and B' yield a higher compression ratio than schemes A and B. This is intuitively reasonable since the sequence contains pan motion only, and schemes A' and B' conduct panning estimation first while schemes A and B do zooming estimation first. As a result, schemes A and B may produce incorrect zooming vectors and thereby result in a higher distortion in the subsequent panning estimation. In the case of the table tennis sequence, there is no significant global motion before the 23rd frame, at which camera zoom commences. This causes the MRM to produce an increase in the entropy of the motion vectors for the first 22 frames. This undesirable anomaly of the MRM can be avoided by developing an improved method or by turning off the MRM in adverse conditions.

From coding experiments on different video material with a CCITT H.261-type coder, we note that the amount of motion information can vary from 10% to over 20% of the total compressed video data. Therefore, depending on the video material, the bit-rate saving from the above entropy reduction can be quite significant.

5 Conclusion

We gave a mathematical model describing global motions in an image sequence. Based on this model, the motion restoration method (MRM) was derived which can restore the zoom, pan, and object motion vectors in an image. Four variants were considered, plus one which forces the object motion vectors to zero (the zero-forcing MRM). Simulation results show that, for images containing both panning and zooming, the proposed method can achieve roughly 30% to 40% of entropy reduction in the block motion vectors. And the zero-forcing MRM can be quite advantageous compared to the four more elaborate alternatives.

Due to the object-assignment step, the method is inherently hierarchical. The proposed object-assignment technique can also be used for image segmentation in various applications such as computer vision and pattern recognition.

References

[1] H. G. Musmann, P. Pirsch, and H. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, no. 4, pp. 523–548, Apr. 1985.

[2] A. Zaccarin and B. Liu, "Fast algorithm for block motion estimation," *Proc. IEEE ICASSP'92*, pp. III449–III452.

[3] S. Iu, "Comparison of motion compensation using different degrees of sub-pixel accuracy for interfield/interframe hybrid coding of HDTV image sequences," *Proc. IEEE ICASSP'92*, pp. III465–III468.

[4] V. Seferidis, "Three dimensional block matching motion estimation," *Electron. Lett.*, vol. 28, pp. 1770–1772, August 1992.

[5] J. Konrad and E. Dubois, "Estimation of image motion field: Bayesian formulation and stochastic solution," in *Proc. IEEE ICASSP'88*, pp. 1072–1075, Apr. 1988.

[6] S. F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale and translation," *Signal Process.: Image Commun.*, vol. 2, pp. 69–80, 1990.

[7] M. Hoetter, "Differential estimation of the global motion parameters zoom and pan," *Signal Process.*, vol. 16, pp. 249–265, 1989.

[8] G. H. Golub and C. F. van Loan, *Matrix Computations.* The Johns Hopkins University Press, 1989.

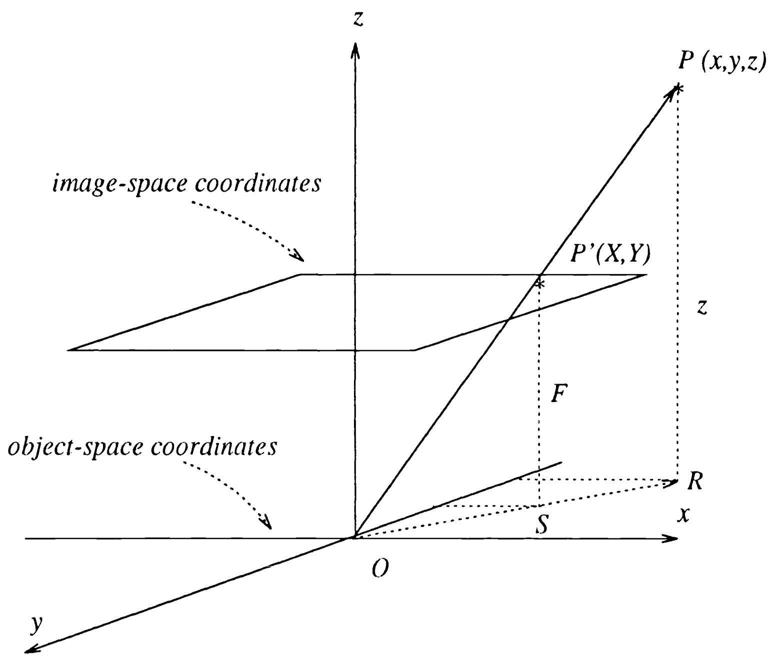

Figure 1: *Central projection.*

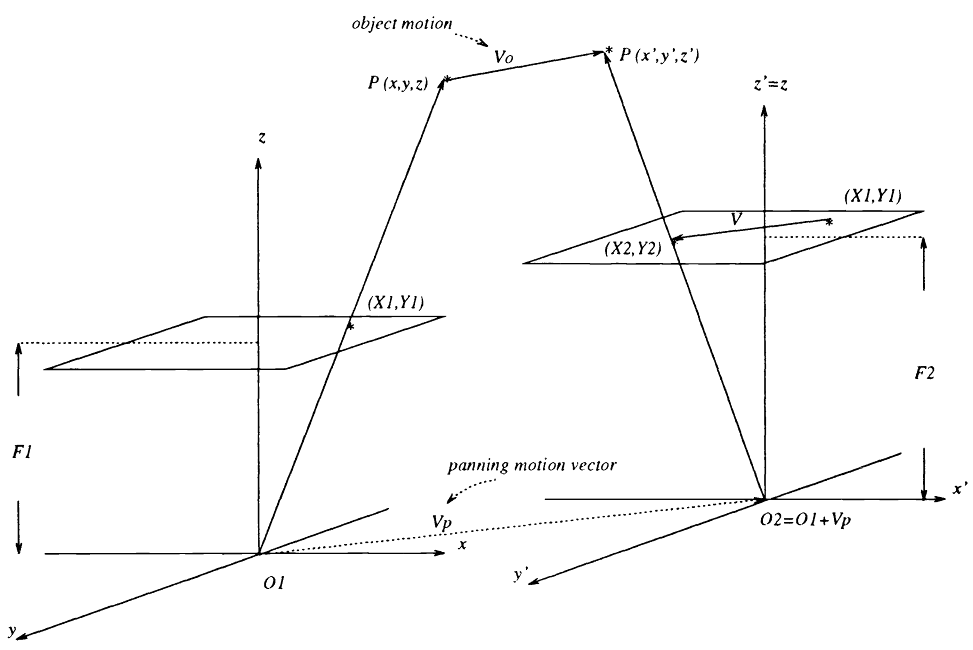

Figure 2: *General motion including zoom, pan, and object movement.*

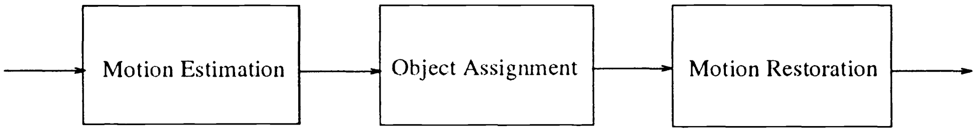

Figure 3: *Motion restoration method.*

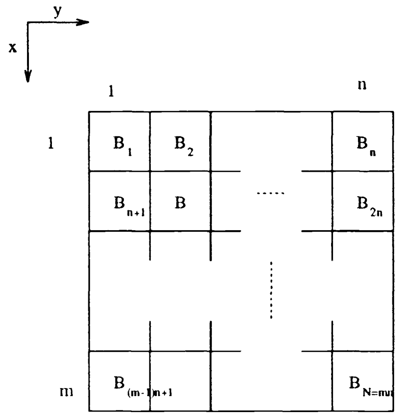

Figure 4: *Ordering of image blocks.*

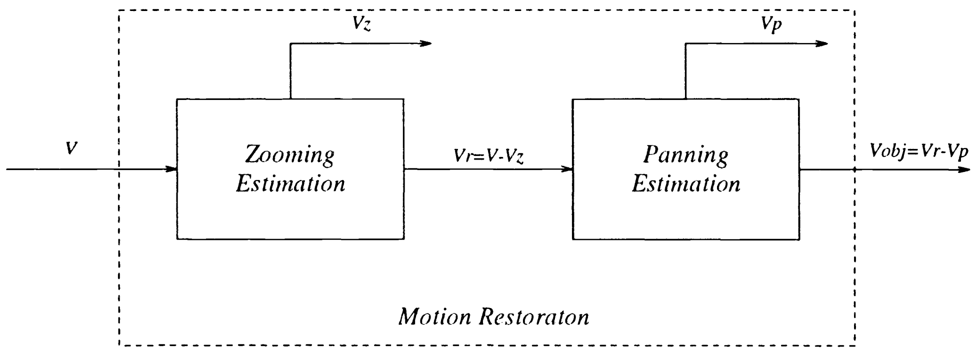

Figure 5: *Cascaded motion restoration.*

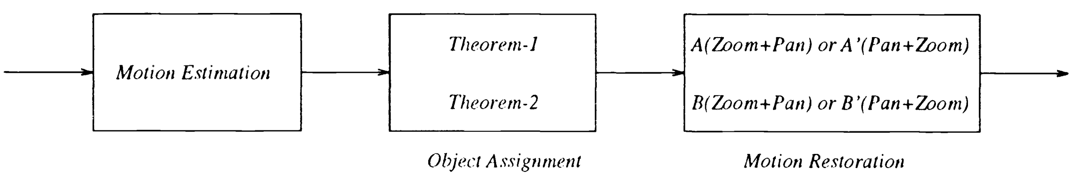

Figure 6: *Four MRM schemes.*

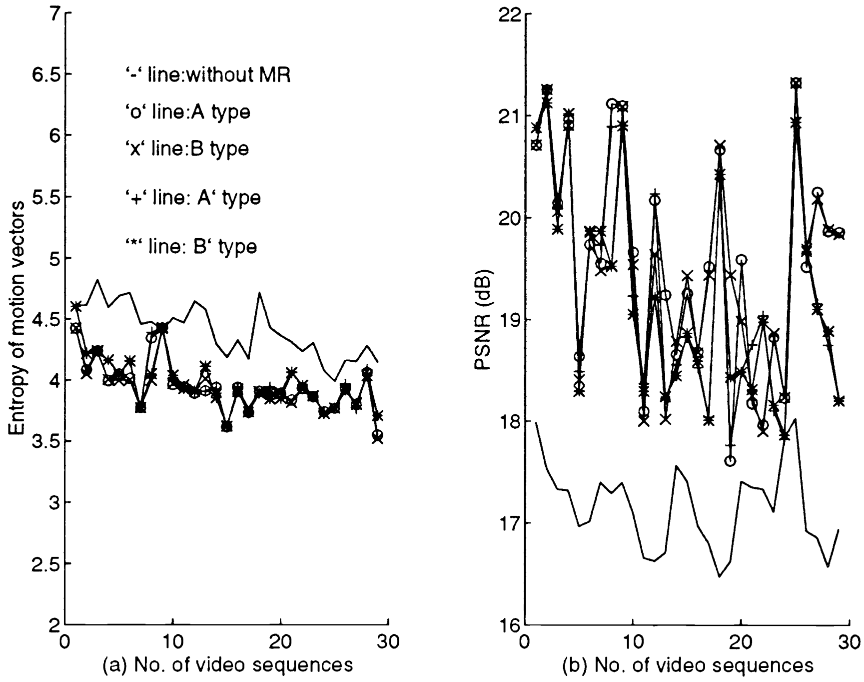

Figure 7: *Entropy reduction and PSNR gain for the flower garden sequence with motion restoration.*

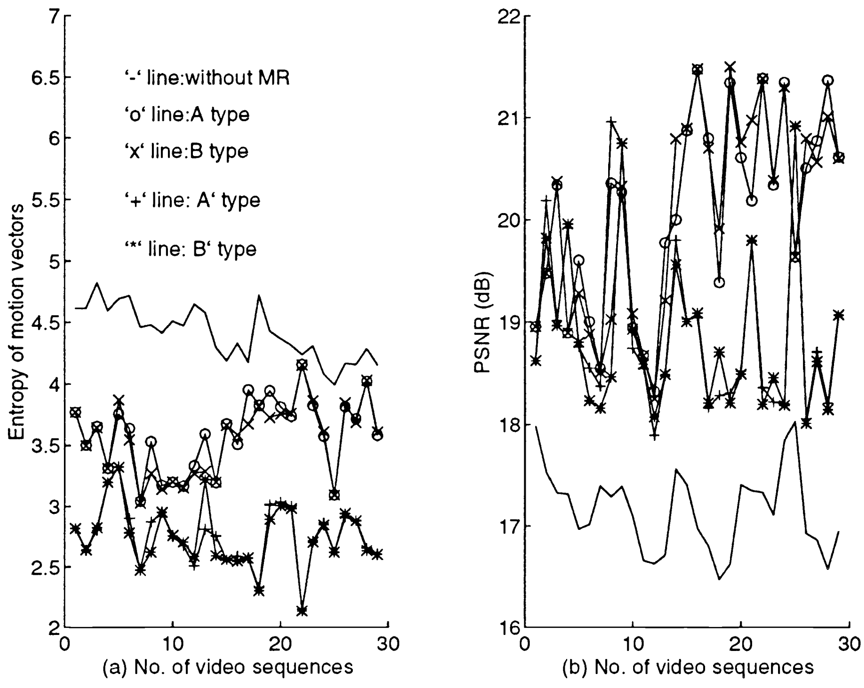

Figure 8: *Entropy reduction and PSNR gain for the flower garden sequence with zero-forcing motion restoration.*

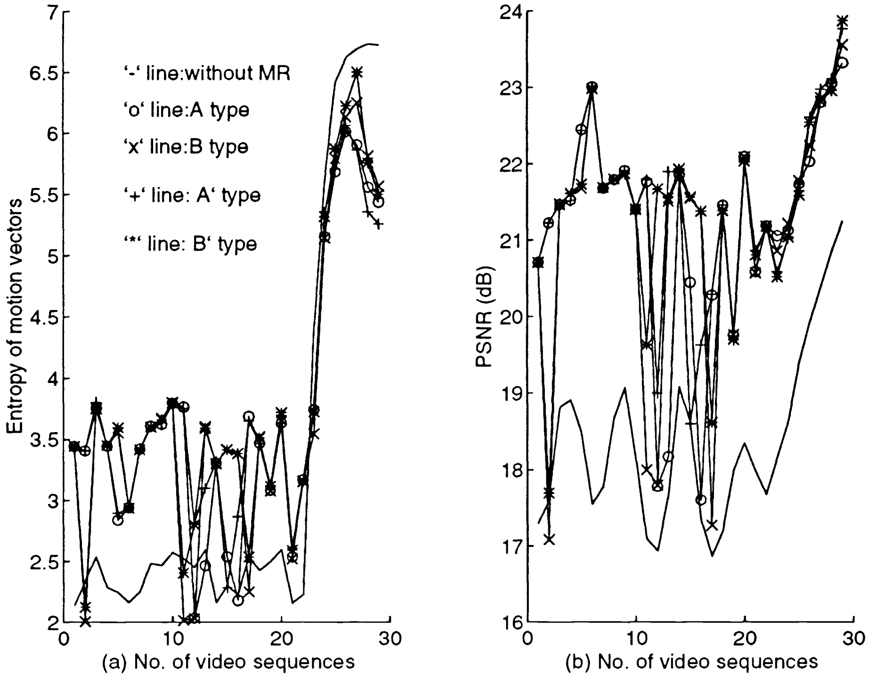

Figure 9: *Entropy reduction and PSNR gain for the table tennis sequence with motion restoration.*

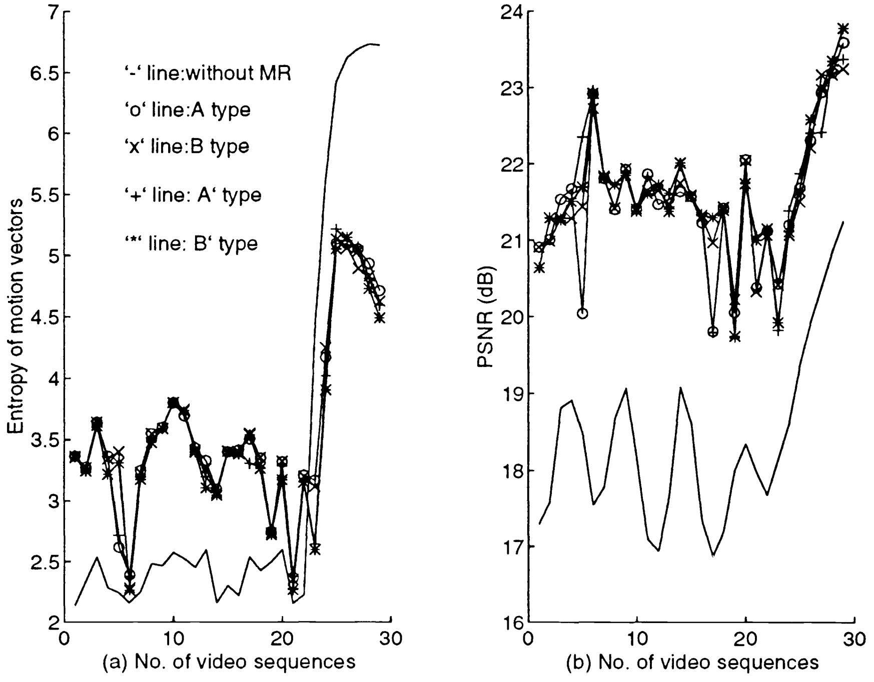

Figure 10: *Entropy reduction and PSNR gain for the table tennis sequence with zero-forcing motion restoration.*