JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION
Vol. 8, No. 1, March, pp. 83–96, 1997
ARTICLE NO. VC970343

A New Motion Estimation Method Using Frequency Components

Yung-Ming Chou and Hsueh-Ming Hang

Department of Electronics Engineering, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan, Republic of China

August 26, 1996; accepted November 25, 1996

Motion estimation techniques are widely used in today's video processing systems. The most frequently used techniques are the block matching method and the differential method. In this paper, we have studied this topic from a viewpoint different from the above to explore the fundamental limits and tradeoffs in image motion estimation. The underlying principles behind two conflict requirements in motion estimation, accuracy and ambiguity, become clear when they are analyzed using this tool—frequency component analysis. This analysis also suggests new motion estimation algorithms and ways to improve the existing algorithms. The so-called frequency component motion estimation algorithm is thus proposed. Compared to the conventional block matching and phase correlation algorithms, this approach provides more reliable displacement estimates particularly for the noisy pictures.

1. INTRODUCTION

Motion estimation techniques have been explored by many researchers in the past 20 years [1]. They are useful in many applications, such as computer vision, target tracking, and industrial monitoring. In interframe video coding, for example, motion estimation and compensation can reduce the bit rate significantly. Many motion estimation schemes have been developed. They can be classified, roughly, into three groups: (i) the block matching method, (ii) the differential (gradient) method, and (iii) the Fourier method [2].

Block matching is a very popular method. It finds the best match between the current image block and certain selected candidates in the previous frame under the assumption that the motion of pixels within the same block is uniform. When the block size is small enough, most movements in a scene can be approximated by piecewise translation. Jain and Jain [3] first proposed this method and applied it to interframe coding. Since then many papers have been published describing various fast search algorithms that reduce the computational load [4–7]. On the other hand, the differential approach is developed based on the assumption that the image intensity can be viewed as an analytic function in spatial and temporal domains. It was first proposed by Cafforio and Rocca [8], and later Netravali and Robbins developed an iterative algorithm, the so-called pel-recursive method [9]. The optical flow method [10] in computer vision is much like the pel-recursive scheme even though they were derived from different bases. There are many refined versions of the pel-recursive however, and optical flow methods [11–18]. The Fourier method used for motion estimation, however, is not as popular as these two approaches.

Phase correlation [19] is the most well-known method in this class that utilizes phase information of frequency components in estimating the motion vectors. Thomas [20, 21] did a rather extensive study on phase correlation and also suggested a two-stage process and a weighting function to improve this method. In this paper, we first analyze the fundamental limitation in motion estimation from the viewpoint of frequency components. Then a new scheme using frequency components to estimate motion vectors is proposed.

In the study of motion estimation techniques there exist two fundamental issues: (i) accuracy problem—inaccurate motion vector estimates due to noise (including object deformation) and/or due to the low spatial resolution of the motion vector field—and (ii) the ambiguity problem—incorrect estimates due to similar objects appearing at different locations in a picture. The desired performance, high accuracy and low ambiguity, leads to conflicts in the selection of motion estimation parameters. For example, in the block matching method, large-sized image blocks reduce content ambiguity but increase inaccuracy in estimating motion vectors because a single large block may contain several objects moving in different directions. On the other hand, small image blocks increase ambiguity because image blocks look similar when their sizes are small. The above simple qualitative analysis hardly goes beyond the intuitive level. In this paper, we try to give a more quantitative treatment to the above problems based on the frequency domain information. Frequency domain analysis provides insights on the performance limits of

E-mail: hmhang@cc.nctu.edu.tw

1047-3203/97 $25.00

motion estimation, and also helps us to construct new estimation algorithms and suggests ways of improving the conventional schemes.

In Section 2, we analyze motion estimation in the frequency domain under a noise-free environment. Error analysis of frequency components due to noise is discussed in Section 3. Section 4 compares the frequency component approach to the block matching approach and the phase correlation approach. A new motion estimation algorithm based on frequency components is developed in Section 5. Simulations in Section 6 demonstrate the performance of various schemes including block matching, phase correlation, and our method. Brief summary and discussions in Section 7 conclude this paper.

2. NOISE-FREE ANALYSIS

Our goal is to estimate the motion vectors (or displacement vectors) of image blocks that move from the previous frame to the current frame. In image coding applications, the current image block is usually used as the fixed reference. We then search for the best matched image block in the previous frame. We assume that this displacement is purely translational and, in this part of analysis, we further assume that there is no noise involved in the movement. In other words, the displaced image block in the previous frame is identical to the image block in the current frame. The only difference between them is their locations. Mathematically, the above assumption can be written as

$$s_c(n_1, n_2) = s_p(n_1 + d_1, n_2 + d_2), \quad \text{for } (n_1, n_2) \in B_s(0, 0), \tag{1}$$

where $s_c(\cdot, \cdot)$ and $s_p(\cdot, \cdot)$ represent the image signals in the current and in the previous picture frames, respectively; $B_s(x, y) = \{(n_1, n_2) | x \le n_1 \le (x + M_1 - 1), y \le n_2 \le (y + M_2 - 1)\}$ is the 2D block in the spatial domain; and M_1 and M_2 are the horizontal and the vertical block sizes as shown in Fig. 1.

If $s_p(n_1, n_2)$ and $s_c(n_1, n_2)$ are represented by their discrete Fourier transform (DFT) components in $B_s(d_1, d_2)$ and $B_s(0, 0)$, respectively, we have

$$s_p(n_1, n_2) = \sum_{(k_1,k_2)\in B_f} A_p(k_1, k_2) e^{j((2\pi/M_1)k_1 n_1 + (2\pi/M_2)k_2 n_2 + \phi_p(k_1,k_2))}, \quad (n_1, n_2) \in B_s(d_1, d_2),$$

and

$$s_c(n_1, n_2) = \sum_{(k_1,k_2)\in B_f} A_c(k_1, k_2) e^{j((2\pi/M_1)k_1 n_1 + (2\pi/M_2)k_2 n_2 + \phi_c(k_1,k_2))}, \quad (n_1, n_2) \in B_s(0, 0), \tag{2}$$

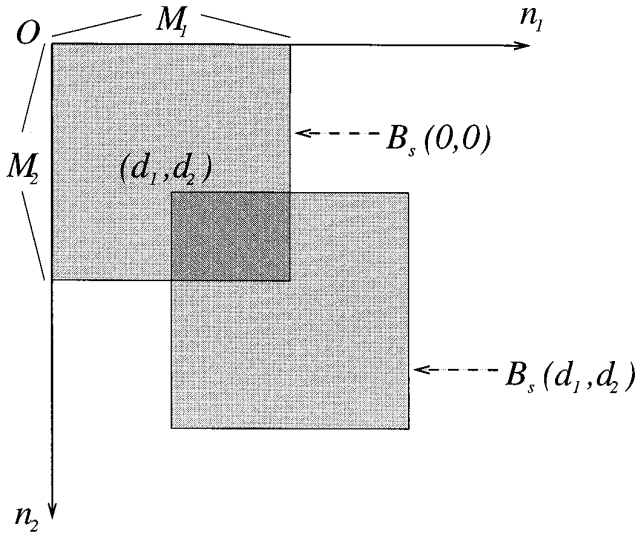


FIG. 1. Two-dimensional block and its shift.

where $B_f = \{(k_1, k_2) | -M_1/2+1 \le k_1 \le M_1/2, -M_2/2 + 1 \le k_2 \le M_2/2\}$ is the corresponding 2D window in the frequency domain, and (d_1, d_2) is the displacement vector of this image block. Note that we use the term "block" in the spatial domain and "window" in the frequency domain.

Under the assumption in Eq. (1), it is well-known in signal analysis that a shift in the spatial domain corresponds to a linear phase term in the frequency domain. That is, for any $(k_1, k_2) \in B_f$,

$$A_c(k_1, k_2) = A_p(k_1, k_2), \tag{3}$$

and

$$\phi_c(k_1, k_2) = \phi_p(k_1, k_2) + \frac{2\pi}{M_1} k_1 d_1 + \frac{2\pi}{M_2} k_2 d_2. \tag{4}$$

Equation (4) can be rewritten in matrix form as

$$\left(\frac{2\pi}{M_1} k_1 \; \frac{2\pi}{M_2} k_2\right) \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \phi_c(k_1, k_2) - \phi_p(k_1, k_2) \equiv \Delta\phi(k_1, k_2). \tag{5}$$

Ideally, in order to compute (d_1, d_2), we only need to measure $\Delta\phi(\cdot, \cdot)$ at two (independent) frequency components and then solve the above simultaneous equations using these two measurements. In practice, to reduce the noise effect, measurements are made at multiple frequency points $(k_1, k_2) = (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots$ Therefore,

$$\begin{pmatrix} \frac{2\pi}{M_1}X_1 & \frac{2\pi}{M_2}Y_1 \\ \frac{2\pi}{M_1}X_2 & \frac{2\pi}{M_2}Y_2 \\ \frac{2\pi}{M_1}X_3 & \frac{2\pi}{M_2}Y_3 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} \Delta\phi(X_1, Y_1) \\ \Delta\phi(X_2, Y_2) \\ \Delta\phi(X_3, Y_3) \\ \vdots \end{pmatrix}, \tag{6}$$

or

$$\mathbf{W}\,\mathbf{d} = \mathbf{\Delta\Phi}, \tag{7}$$

where

$$\mathbf{W} = \begin{pmatrix} \frac{2\pi}{M_1}X_1 & \frac{2\pi}{M_2}Y_1 \\ \frac{2\pi}{M_1}X_2 & \frac{2\pi}{M_2}Y_2 \\ \frac{2\pi}{M_1}X_3 & \frac{2\pi}{M_2}Y_3 \\ \vdots & \vdots \end{pmatrix},$$

$$\mathbf{d} = (d_1\, d_2)^T,$$

and

$$\mathbf{\Delta\Phi} = \begin{pmatrix} \Delta\phi(X_1, Y_1) \\ \Delta\phi(X_2, Y_2) \\ \Delta\phi(X_3, Y_3) \\ \vdots \end{pmatrix}.$$

The least-squared estimate of the motion vector $\mathbf{d}$ is simply

$$\hat{\mathbf{d}} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\,\mathbf{\Delta\Phi}. \tag{8}$$

However, there is a major problem associated with Eq. (6). Our measurement of $\Delta\phi(k_1, k_2)$ is between $-\pi$ and π, but its true value could be this principal value added by multiples of 2π. In other words, when $\Delta\psi(k_1, k_2)$, the measurement of $\Delta\phi(k_1, k_2)$, is used, the $\Delta\phi(k_1, k_2)$ term in Eq. (6) should be replaced by

$$\Delta\psi(k_1, k_2) + i(k_1, k_2)2\pi,$$

where $i(k_1, k_2)$ is a specific integer that satisfies the entire set of simultaneous equations in Eq. (6). More explicitly, we need to find both (d_1, d_2) and $\mathbf{I} = (\mathrm{i}(X_1, Y_1), i(X_2, Y_2), \ldots)^{\mathrm{T}}$ which simultaneously satisfy the vector equation

$$\begin{pmatrix} \frac{2\pi}{M_1}X_1 & \frac{2\pi}{M_2}Y_1 \\ \frac{2\pi}{M_1}X_2 & \frac{2\pi}{M_2}Y_2 \\ \frac{2\pi}{M_1}X_3 & \frac{2\pi}{M_2}Y_3 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} \Delta\psi(X_1, Y_1) + i(X_1, Y_1)2\pi \\ \Delta\psi(X_2, Y_2) + i(X_2, Y_2)2\pi \\ \Delta\psi(X_3, Y_3) + i(X_3, Y_3)2\pi \\ \vdots \end{pmatrix}, \tag{9}$$

or

$$\mathbf{W}\,\mathbf{d} = \mathbf{\Delta\Psi} + \mathbf{I}2\pi. \tag{10}$$

Note that the displacement information is contained in the phase portion of the frequency components. The magnitude portion does not provide information for estimating $\mathbf{d}$.

3. ERROR ANALYSIS DUE TO NOISE

The term "noise" in this paper is used in a rather broad sense. In addition to the measurement of noise, we treat object deformation, nontranslational movement, uncovered background, and incorrect block location as noise. The last item in the above statement—incorrect block location—may need a brief explanation. The analysis in Section 2 assumes that the block location in the previous picture is $B_s(d_1, d_2)$, the exact displaced location. However, in a real situation, (d_1, d_2) is unknown. It is the target that is to be estimated. In the estimation process, we start with an initial guess of (d_1, d_2), say (d_1^0, d_2^0), which is often different from the true (d_1, d_2). Even all of the other types of noise are not present; $s_p(n_1, n_2)$, $(n_1, n_2) \in B_s(d_1^0, d_2^0)$ is not merely a linear shift of $s_c(n_1, n_2)$, $(n_1, n_2) \in B_s(0, 0)$. As illustrated in Fig. 2 (a simple case in 1D), a portion of s_c in $B_s(0, 0)$ is outside $B_s(d_1^0, d_2^0)$ and a new segment of signals, which is not a part of s_c in $B_s(0, 0)$, is included in $B_s(d_1^0, d_2^0)$. These signals located outside the overlapped area are treated as noise in the following analysis.

If $v(n_1, n_2)$ represents the noise described in the above, we have

$$s_c(n_1, n_2) = s_p(n_1 + d_1, n_2 + d_2) + v(n_1, n_2), \quad \text{for } (n_1, n_2) \in B_s(0, 0). \tag{11}$$

Since $v(\cdot, \cdot)$ is a mix of several noise sources, it is very difficult to have an accurate model of such a noise. For analytical purposes, we assume that $v(\cdot, \cdot)$ is a well-behaved

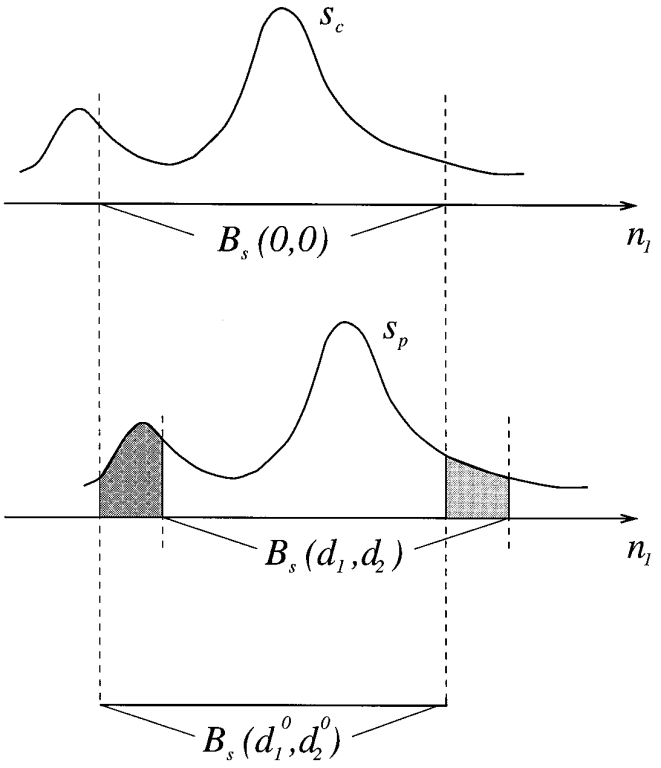


FIG. 2. Nonoverlapped area due to incorrect d_1 estimate ($d_2 = 0$).

stationary random process. It can then be represented by its frequency components; that is,

$$v(n_1,n_2) = \sum_{(k_1,k_2)\in B_f} A_v(k_1,k_2)e^{j((2\pi/M_1)k_1n_1+(2\pi/M_2)k_2n_2+\phi_v(k_1,k_2))}, \quad (n_1,n_2)\in B_s(0,0). \tag{12}$$

Consequently, for the (k_1, k_2)th component, we have the following relationship:

$$\begin{aligned} A_c(k_1,k_2)e^{j((2\pi/M_1)k_1n_1+(2\pi/M_2)k_2n_2+\phi_c(k_1,k_2))} &= A_p(k_1,k_2)e^{j((2\pi/M_1)k_1n_1+(2\pi/M_2)k_2n_2+\phi_p(k_1,k_2))} \\ &+ A_v(k_1,k_2)e^{j((2\pi/M_1)k_1n_1+(2\pi/M_2)k_2n_2+\phi_v(k_1,k_2))}, \\ (n_1,n_2)\in B_s(0,0). \end{aligned} \tag{13}$$

The phasor diagram of Eq. (13) is drawn in Fig. 3. If we skip the index (k_1, k_2) in the following equations (since it does not change in the calculations), it is clear that

$$A_c = [(A_p + A_v\cos(\phi_v-\phi_p))^2 + (A_v\sin(\phi_v-\phi_p))^2]^{1/2}, \tag{14}$$

and

$$\phi_c = \phi_p + \left(\frac{2\pi}{M_1}k_1d_1 + \frac{2\pi}{M_2}k_2d_2\right) + \arctan\left(\frac{A_v\sin(\phi_v-\phi_p)}{A_p + A_v\cos(\phi_v-\phi_p)}\right). \tag{15}$$

In order to go one step further, we need some additional assumptions. If we assume that $v(\cdot,\cdot)$ is white, then the probability distribution of $\phi_v(\cdot,\cdot)$ is uniform between $-\pi$ and π [22]. Consequently, $(\phi_v - \phi_p)$ can be replaced by ϕ_v statistically since $\text{prob}(\phi_v - \phi_p) = \text{prob}(\phi_v) = 1/(2\pi)$ for $\phi_v \in (-\pi,\pi]$. The subsequent analysis is very similar to the noise analysis in the phase or the frequency modulation systems [23]. If $A_p \gg A_v$, Eqs. (14) and (15) can be simplified to

$$A_c \approx A_p + A_v\cos(\phi_v), \tag{16}$$

and

$$\phi_c \approx \phi_p + \left(\frac{2\pi}{M_1}k_1d_1 + \frac{2\pi}{M_2}k_2d_2\right) + \arctan\left(\frac{A_v\sin(\phi_v)}{A_p}\right). \tag{17}$$

Because $E[(A_v\cos(\phi_v))^2] = \frac{1}{2}E[A_v^2]$, the noise disturbance to the phase information is less than its effect on the original signal. Therefore, the displacement estimate using phase information is more robust than that using the original signal provided that the signal magnitude is much higher than the noise magnitude. This is the well-known noise-reduction property of the phase and frequency modulation techniques in communications. Furthermore, our desired information, (d_1, d_2), is scaled by (k_1, k_2) in phase. For example, if $k_1 = 0$, then

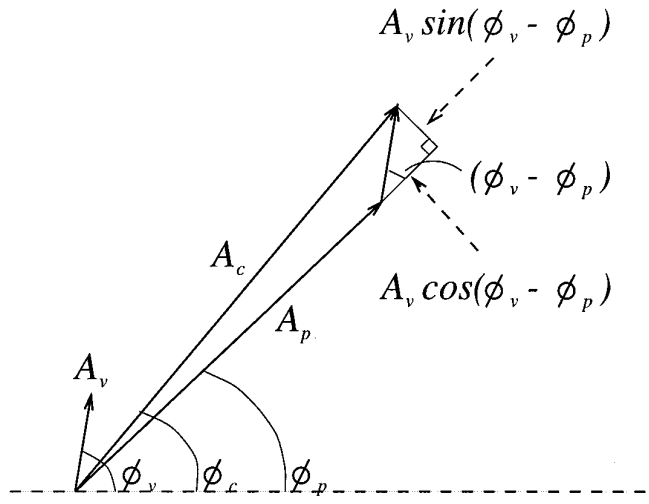


FIG. 3. Phasor diagram of A_c when $A_p \gg A_v$.

$$d_2 = \frac{(M_2/2\pi)\Delta\phi}{k_2} - \frac{(M_2/2\pi)\arctan((A_v \sin(\phi_v)/A_p)}{k_2}.$$

Since the second term—the noise term—is divided by k_2, given the same amount of noise the higher frequency components are more accurate in estimating motion vectors as compared to the lower frequency components.

However, the above noise-reduction situation is reversed when the noise magnitude is close to or higher than the signal magnitude at the same frequency. In this case, the phase information suffers more distortion than the original signal. This is the well-known "threshold effect" in continuous phase modulation [23]. Therefore, we should avoid using the phase information at the frequencies where the noise power is comparable to or even higher than the signal power.

4. BLOCK MATCHING, PHASE CORRELATION, AND FREQUENCY COMPONENTS

The frequency component analysis discussed in Sections 2 and 3 can be applied to the conventional motion estimation schemes and it helps us to gain insights into the underlying principles of these schemes. For example, the popular block matching method searches for the optimum (d_1, d_2) that minimizes

$$\mathscr{E} = \sum_{(n_1,n_2)\in B_s(0,0)} [s_p(n_1 + d_1, n_2 + d_2) - s_c(n_1, n_2)]^2. \quad (18)$$

In order to obtain more tractable results, the mean-squared error criterion is adopted in Eq. (18). According to Parseval's power theorem, Eq. (18) is equal to

$$\begin{aligned}\mathscr{E} &= \sum_{(k_1,k_2)\in B_f} |A_p(k_1, k_2)e^{j\phi_p(k_1,k_2)} - A_c(k_1, k_2)e^{j\phi_c(k_1,k_2)}|^2 \\ &\equiv \sum_{(k_1,k_2)\in B_f} e_f^2(k_1, k_2),\end{aligned} \quad (19)$$

where $e_f(\cdot, \cdot)$ denotes the frequency domain absolute error. For the (k_1, k_2)th component, assuming $A_p(k_1, k_2) \approx A_c(k_1, k_2)$, Eq. (19) becomes

$$\begin{aligned}e_f^2(k_1, k_2) &= A_c^2(k_1, k_2)[e^{j\phi_c(k_1,k_2)} - e^{j\phi_p(k_1,k_2)}][e^{j\phi_c(k_1,k_2)} \\ &\quad - e^{j\phi_p(k_1,k_2)}]^* \\ &= A_c^2(k_1, k_2)[e^{j(\phi_c(k_1,k_2)-\phi_p(k_1,k_2))/2} - e^{-j(\phi_c(k_1,k_2)-\phi_p(k_1,k_2))/2}] \\ &\quad \times [e^{j(\phi_c(k_1,k_2)-\phi_p(k_1,k_2))/2} - e^{-j(\phi_c(k_1,k_2)-\phi_p(k_1,k_2))/2}]^* \\ &= A_c^2(k_1, k_2)[2\sin(\Delta\phi(k_1, k_2)/2)]^2 \\ &= 2A_c^2(k_1, k_2)[1 - \cos(\Delta\phi(k_1, k_2))].\end{aligned} \quad (20)$$

The contribution of the (k_1, k_2)th component to $\mathscr{E}$ is proportional to the product of $A_c^2(k_1, k_2)$ and $(1 - \cos(\Delta\phi(k_1, k_2)))$. Because of the nonlinearity of $(1 - \cos(\Delta\phi(k_1, k_2)))$, when the value of $\Delta\phi(k_1, k_2) = ((2\pi/M_1)k_1d_1 + (2\pi/M_2)k_2d_2)$ is in between $-\pi/4$ and $\pi/4$ or their 2π multiples, it does not contribute much to $\mathscr{E}$. That is, when (d_1, d_2) is small, low frequency components are less important. However, another critical factor is $A_c^2(k_1, k_2)$. The low power components do not make significant contributions to $\mathscr{E}$ in general.

The phase correlation method [19] is similar to the block matching technique in the sense that it looks for the best matching location. However, it modifies the original signals and relies only on the phase information in the matching process. The block diagram of phase correlation is given in Fig. 4. Assume that $A_p(k_1, k_2)\, e^{j\phi_p(k_1,k_2)}$ and $A_c(k_1, k_2)\, e^{j\phi_c(k_1,k_2)}$ are the components of the discrete Fourier transforms of $s_p(\cdot, \cdot)$ and $s_c(\cdot, \cdot)$, respectively. Let C_r be the inverse discrete Fourier transform (IDFT) of the phase information of these two signals; i.e.,

$$\begin{aligned}C_r(n_1, n_2) &= \text{IDFT}[e^{j\phi_p(k_1,k_2)} \cdot e^{-j\phi_c(k_1,k_2)}] \\ &= \text{IDFT}[e^{-j\Delta\phi(k_1,k_2)}].\end{aligned} \quad (21)$$

It is clear that if there is no noise in the displacement process, a single peak which indicates the motion vector can be obtained. That is,

$$\begin{aligned}C_r(n_1, n_2) &= \text{IDFT}[e^{-j((2\pi/M_1)k_1d_1+(2\pi/M_2)k_2d_2)}] \\ &= \delta(n_1 - d_1, n_2 - d_2).\end{aligned} \quad (22)$$

Since the multiplication operation in frequency domain is equivalent to the convolution operation in spatial domain, $C_r(n_1, n_2)$ can also be written as

$$C_r(n_1, n_2) = \text{IDFT}[e^{j\phi_p(k_1,k_2)}] \otimes \text{IDFT}[e^{-j\phi_c(k_1,k_2)}]. \quad (23)$$

Let $p_x(n_1, n_2)$ denote $\text{IDFT}[e^{j\phi_x(k_1,k_2)}]$, the phase portion of the original signal $s_x(n_1, n_2)$ with its magnitude being replaced by unity. Then, $C_r(n_1, n_2)$ represents the correlation between $p_c(n_1, n_2)$ and $p_p(n_1, n_2)$. Compared to the correlation of the original signals $s_c(\cdot, \cdot)$ and $s_p(\cdot, \cdot)$, $C_r(\cdot, \cdot)$ has the even more attractive property that its nonzero value should appear only at (d_1, d_2) if noise is not present.

Ideally, we first compute $\text{IDFT}[e^{-j\phi_p(k_1,k_2)}]$ in Eq. (23) at the various candidate locations. Then we perform the convolution in Eq. (23). Finally, the maximum value in the (n_1, n_2) plane is selected as the motion vector. This procedure is similar to the block matching method except that signals are modified to retain the phase information only. It should produce good results at the price of computation—the DFT and IDFT are computed at every candidate location.

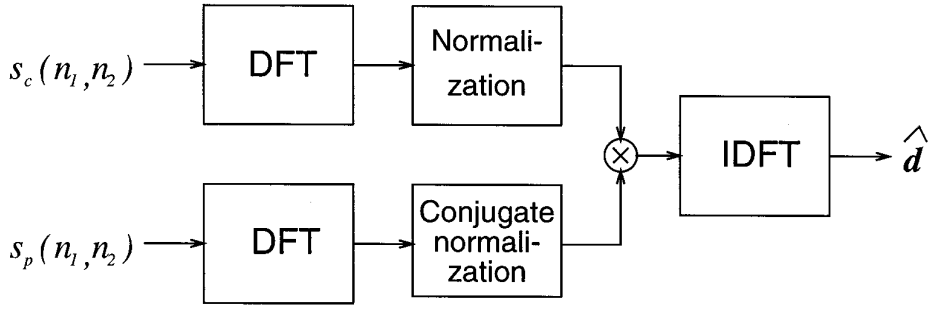


FIG. 4. Block diagram of the phase correlation method.

To reduce computation, the conventional approach uses the calculation of Eq. (21) instead. That is, the candidates at locations other than zero are cyclically shifted versions of the zero location candidate as illustrated by Fig. 5 (a 1D example). The cyclically shifted $s_p(n_1, n_2)$ for $(n_1, n_2) \in B_s(0, 0)$ is not the same as the truly displaced version (linear shift) of the reference signal $s_p(n_1, n_2)$, for $(n_1, n_2) \in B_s(d_1, d_2)$. Hence when (d_1, d_2) becomes large, the cyclically shifted version becomes incorrect. Signals in the nonoverlapped area may be viewed as noise. However, to avoid the noise threshold effect discussed in Section 3, this nonoverlapped area (shared region) has to be quite small compared to the block size. Therefore, large blocks must be used in the conventional phase correlation method.

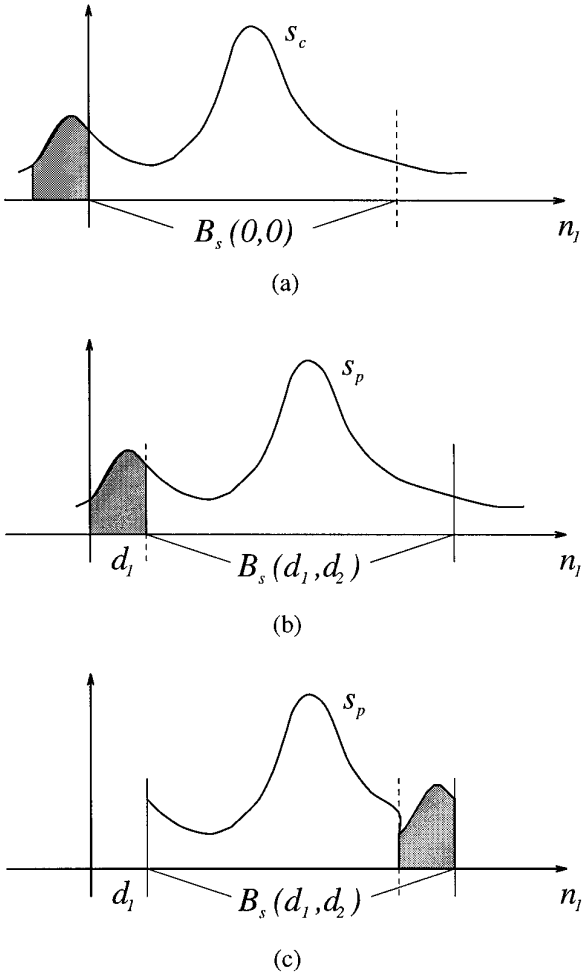


FIG. 5. Cyclic shift effect in phase correlation ($d_2 = 0$): (a) reference signal in the current frame, (b) correctly shifted version, and (c) cyclically shifted version in the previous frame.

Since the normalization procedure of phase correlation is similar to the equalization in communication, an alternate approach to bringing the block matching method closer to the phase correlation method is to filter the pictures with a high-pass filter that produces a flat output spectrum. This is consistent with the earlier discussion in Section 3 that the high frequency components could offer more accurate estimates, provided that the high frequency noise is small. In practice, a carefully designed bandpass filter may offer the best compromise between estimation accuracy and noise immunity.

5. FREQUENCY COMPONENT ALGORITHM

One way of using frequency components to estimate motion vectors is to solve Eq. (9). However, we encounter two problems: (i) incorrect block location and (ii) unknown values of $\{i(\cdot, \cdot)\}$. In Section 3, we modeled all signals in the nonoverlapped area as "noise." However, due to the threshold effect, this noise level must be kept low; otherwise, the result would be invalid. This problem becomes even more serious in 2D cases, because a 20% shift in each direction would result in a 36% nonoverlapped 2D area. To overcome this problem, we propose an iterative algorithm. The basic idea is "move-on" and "follow-up." The initial block location is the zero displacement. If the estimated displacement is greater than a certain value, we anticipate encountering large estimation errors. Hence, we shift the block to the latest estimated location, assuming that our last estimate would lead us to a location closer to the true

displacement. The frequency domain motion estimation procedure is then repeated again at the new location. The above iterative process can help to relieve some of the nonoverlapped block constraint although it is not always successful in handling very large displacements (greater than 1/3 of the block length).

Then, we look into the unknown $\{i(\cdot, \cdot)$ problem. Assume that the image block is square; that is, $M_1 = M_2 = M$. Our goal is to find the $\mathbf{d} = (d_1, d_2)$ that minimizes the following mean-squared error criterion:

$$E\left\{\left[\frac{1}{2\pi}\Delta\phi(k_1, k_2) - \left(\frac{k_1}{M}d_1 + \frac{k_2}{M}d_2\right)\right]^2\right\}.$$

As discussed in Section 2, we can only measure $\Delta\psi(\cdot)$, the principal value (between $-\pi$ and π) of $\Delta\phi(\cdot)$. In other words, we need to estimate $\mathbf{d}$ and $\{i(\cdot, \cdot)\}$ simultaneously. The difficulty is that $i(k_1, k_2)$ varies as a function of (k_1, k_2). Here, we adopt a modified version of the least-mean-squared (LMS) algorithm [24]. Define the error signal to be

$$\varepsilon(k_1, k_2, \mathbf{d}) = \frac{1}{2\pi}\Delta\phi(k_1, k_2) - \mathbf{d}^{\mathrm{T}}\mathbf{m}(k_1, k_2), \quad (24)$$

where the coordinate vector $\mathbf{m}(k_1, k_2) = (k_1/M\ k_2/M)^{\mathrm{T}}$. Using the steepest descent algorithm, we can update the motion vector point-by-point in the frequency domain; that is,

$$\begin{aligned}\hat{\mathbf{d}}^{(l+1)} &= \hat{\mathbf{d}}^{(l)} - \mu\varepsilon(k_1, k_2, \hat{\mathbf{d}}^{(l)})\nabla_{\mathbf{d}}\varepsilon(k_1, k_2, \hat{\mathbf{d}}^{(l)}) \\ &= \hat{\mathbf{d}}^{(l)} + \mu\varepsilon(k_1, k_2, \hat{\mathbf{d}}^{(l)})\mathbf{m}(k_1, k_2),\end{aligned} \quad (25)$$

where $\nabla_{\mathbf{d}}$ denotes the 2D gradient operator with respect to $\hat{\mathbf{d}}^{(l)}$ and μ is the updating constant controlling the speed of convergence. The initial value $\hat{\mathbf{d}}^{(0)}$ is set to zero.

As stated earlier, the measured $\Delta\psi(k_1, k_2)$ is the principal value of $\Delta\phi(k_1, k_2)$; that is,

$$-\pi < \Delta\psi(k_1, k_2) = (\Delta\phi(k_1, k_2) - i(k_1, k_2)2\pi) \le \pi. \quad (26)$$

If the true motion vector $\mathbf{d}$ is known, $\Delta\phi(k_1, k_2)$ can be replaced by $\mathbf{d}^{\mathrm{T}}\mathbf{m}(k_1, k_2)2\pi$ as Eq. (5) implies. Then, dividing Eq. (26) by 2π, we obtain

$$-\frac{1}{2} < \mathbf{d}^{\mathrm{T}}\mathbf{m}(k_1, k_2) - i(k_1, k_2) \le \frac{1}{2}. \quad (27)$$

Let $c(k_1, k_2) = \mathbf{d}^{\mathrm{T}}\mathbf{m}(k_1, k_2) - \frac{1}{2}$. Then $i(k_1, k_2)$ is an integer between $c(k_1, k_2)$ and $c(k_1, k_2) + 1$. We can thus determine $i(k_1, k_2)$ by

$$i(k_1, k_2) = \lceil c(k_1, k_2) \rceil, \quad (28)$$

where $\lceil c \rceil$ denotes the smallest integer greater than or equal to c. Given the best displacement estimate $\hat{\mathbf{d}}^{(l)}$ at step l (Eq. 25), the phase ambiguity factor $i(\cdot)$ can be estimated by

$$\hat{i}(k_1, k_2, \hat{\mathbf{d}}^{(l)}) = \lceil (\hat{\mathbf{d}}^{(l)})^{\mathrm{T}}\mathbf{m}(k_1, k_2) - \tfrac{1}{2} \rceil. \quad (29)$$

Therefore, the error signal in Eq. (24) can be rewritten as

$$\begin{aligned}\varepsilon(k_1, k_2, \hat{\mathbf{d}}^{(l)}) = \frac{1}{2\pi}\Delta\psi(k_1, k_2) + \hat{i}(k_1, k_2) \\ - (\hat{\mathbf{d}}^{(l)})^{\mathrm{T}}\mathbf{m}(k_1, k_2).\end{aligned} \quad (30)$$

Now, we arrive at a recursive procedure as shown in Fig. 6. At each frequency (k_1, k_2) we perform a two-step calculation: (1) estimate $i(k_1, k_2)$ (Eq. (29)) based on the most recent $\hat{\mathbf{d}}$, and (2) update $\hat{\mathbf{d}}$ (Eq. (25)) using the estimated $\hat{i}(k_1, k_2)$ and the previous $\hat{\mathbf{d}}$ (Eq. (30)). We run the above calculation through the selected frequency components recursively until the result converges.

Selecting proper frequency components for this recursion is critical to obtaining reliable motion vector estimates. As discussed earlier, the high frequency components usually have small signal power and larger phase-ambiguity values. Therefore, a frequency constraint

$$|k_1| + |k_2| < \frac{M}{2}$$

is adopted to exclude high-frequency components. Figure 7 shows the locations of admissible frequency components in the 2D Fourier plane for block size $M = 16$ (the size used in our simulations in Section 6). Since image data are real numbers, their frequency components are conjugate symmetric with respect to the origin. Consequently, only 56 distinct components are included in the admissible set. Note that the DC component is excluded because it contains no motion information. In addition, the components of small magnitude in the admissible set should not be used to avoid the threshold effect in computing displacement vectors. Furthermore, data points at different regions in the frequency domain often lead to less correlated noise and hence can reduce estimation errors. Therefore, for a block size around 16×16, we select the 10 largest data points in the admissible set: three from the first quadrant, three from the fourth quadrant, two from the x axis, and the last two from the y axis.

The same set of frequency components (data points) are repetitively used in the recursive process until it converges. Assume that L data points are selected ($L = 10$ in our simulations); we sum up the absolute errors for every L calculations (i.e., processing the entire data set once) to check the convergence. For convenience, let l denote the

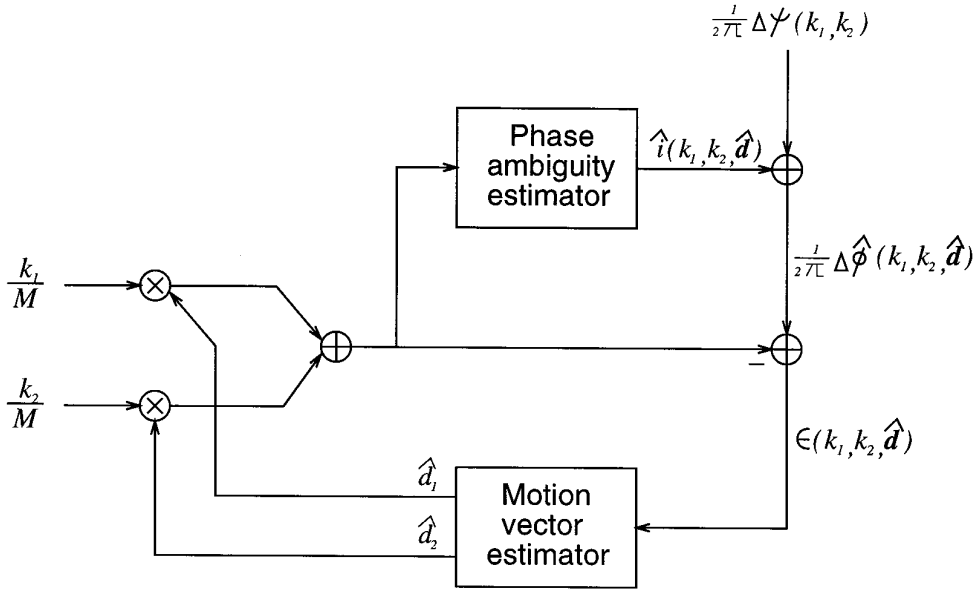


FIG. 6. Block diagram of the recursive process in the frequency component algorithm.

accumulated calculation (recursion) number. Then, when $l = nL$, $n = 2,3,4, \ldots$, we compute the accumulated absolute error by

$$\sigma(n) \equiv \sum_{j=(n-1)L+1}^{nL} |\varepsilon(k_1, k_2, \hat{d}^{(j)})|. \tag{31}$$

Given a proper threshold T_0, if the ratio

$$r(n) \equiv \frac{\sigma(n)}{\sigma(n-1)} > T_0, \tag{32}$$

then the process terminates. For convenience, the *recursive procedure* refers to the process that computes Eqs. (29) and (25) using the selected data points. When the recursive procedure converges, a motion vector estimate is obtained and one *iteration* is done (Fig. 8).

Incorrect block location is a major cause of errors in estimating motion vectors using frequency components. After obtaining the displacement vector in the last iteration we check the estimated motion vector value. If its value is significant, we shift the data block location to the estimated displacement and then start a new iteration by performing FFT and the recursive process (Fig. 6). Because FFT is computed on the integer sampled grid, the displacement between the old and new locations is approximated by the nearest integer of the estimated motion vector. Let $\bar{\mathbf{d}}^{(l)} = (\bar{d}_1^{(l)}\bar{d}_2^{(l)})^{\mathrm{T}}$ be the rounded-off value of $\hat{\mathbf{d}}^{(l)}$. Assume that $l = nL$, where n satisfies the convergence criterion (Eq. (32)). If $\bar{\mathbf{d}}^{(l)} \neq \mathbf{0}$, then the data block location is shifted to $B_s(\bar{d}_1^{(l)}, \bar{d}_2^{(l)})$ and the recursive computation process starts again; otherwise, the iterative procedure is ended. The final estimate $\hat{\mathbf{d}}$ is $\hat{\mathbf{d}}^{(l)}$ plus the accumulated block shifts. The flow chart of the entire frequency component algorithm is shown in Fig. 8.

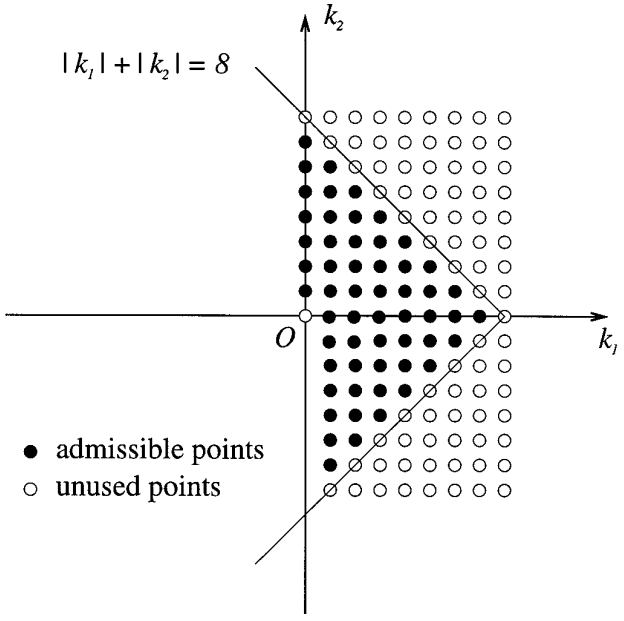


FIG. 7. Admissible set in the 2D Fourier plane ($M = 16$).

6. SIMULATION RESULTS

Several simulations have been conducted to compare the performance of different motion estimation algorithms.

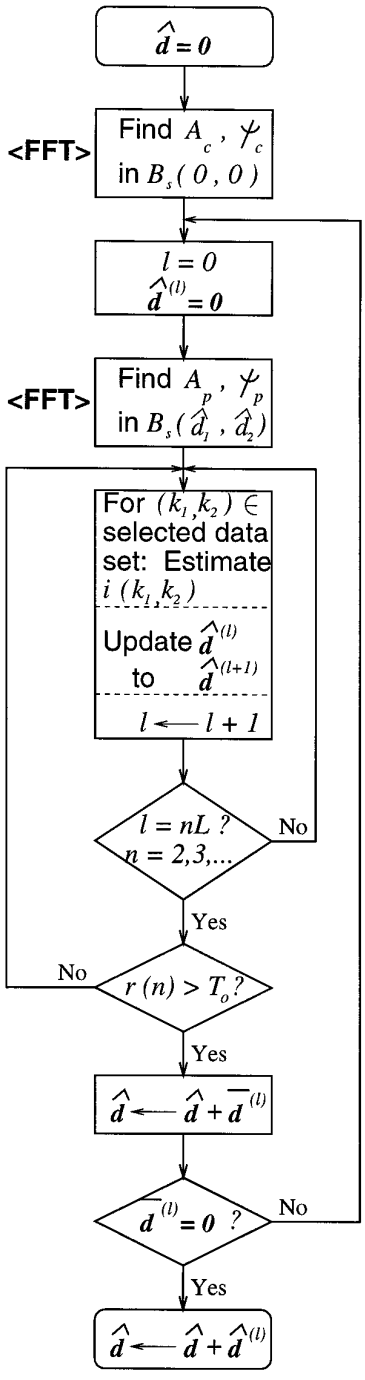


FIG. 8. Flow chart of the proposed frequency component algorithm.

Two test picture sequences shown in Fig. 9 are used: Mobile & Calendar (352 pixels by 240 lines) and Miss America (352 pixels by 288 lines). In the Mobile & Calendar sequence, a toy train is pushing a ball slowly from right to left. The background is a cartoon painting and a calendar sliding vertically on the right side. The Miss America sequence is a typical head-and-shoulder video with a simple background. The major motion is the head, shoulder, eyes, and lips. This sequence contains a fairly significant amount of noise. Thus it can be used to illustrate the effect of noise on displacement estimation. Although the original sequences are in color, only the luminance (brightness) component is used to estimate the motion vectors.

Three schemes are compared: (i) the full-search block matching algorithm (BMA), (ii) the phase correlation algorithm (PCA), and (iii) the frequency component algorithm (FCA). The image block size in all of these schemes is 16×16. The search range of BMA is ± 7 pixels along both the horizontal and vertical directions, which is sufficient for these two sequences. In FCA, we select the updating constant $\mu = 4$ and the convergence threshold $T_0 = 0.99$.

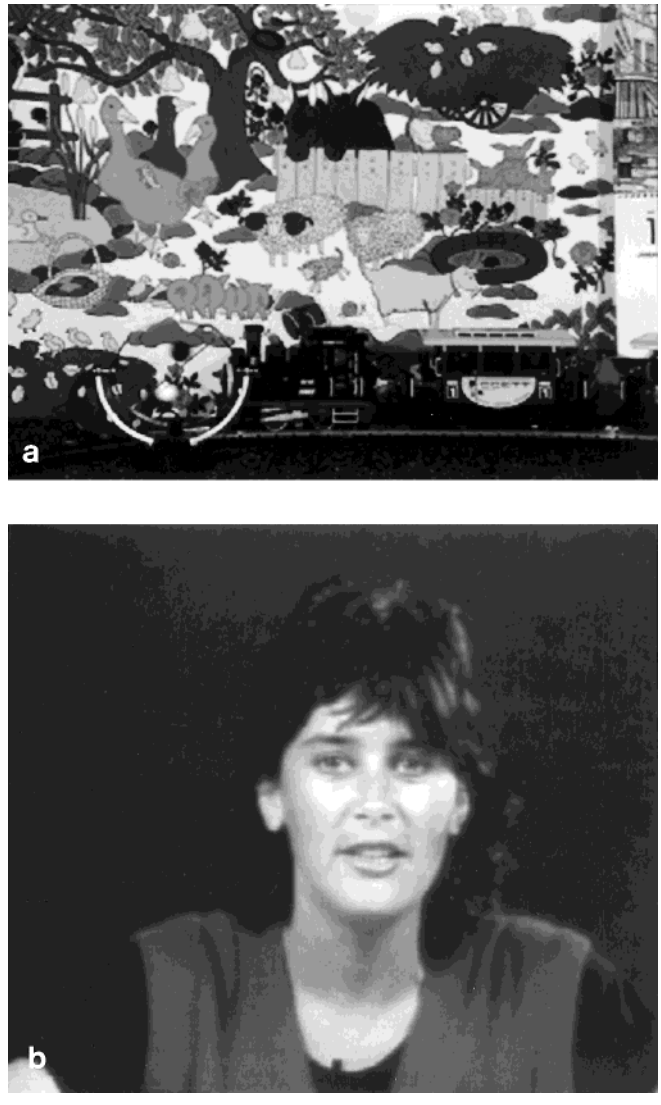


FIG. 9. Test sequences: (a) Mobile & Calendar and (b) Miss America.

Since the motion vector derived from FCA is no longer an integer, we use bilinear interpolation to generate the predicted pixel values. Let $(\hat{d}_1, \hat{d}_2)$ be the estimated motion vector of a block using FCA; the predicted value of pixel (n_1, n_2) is

$$\begin{aligned}\hat{s}_c(n_1, n_2) = {} & (1-\alpha)(1-\beta)s_p(\lfloor n_1' \rfloor, \lfloor n_2' \rfloor) \\ & + (1-\alpha)\beta s_p(\lfloor n_1' \rfloor + 1, \lfloor n_2' \rfloor) \\ & + \alpha(1-\beta)s_p(\lfloor n_1' \rfloor, \lfloor n_2' \rfloor + 1) \\ & + \alpha\beta s_p(\lfloor n_1' \rfloor + 1, \lfloor n_2' \rfloor + 1),\end{aligned} \tag{33}$$

where $n_1' = n_1 + \hat{d}_1$ and $n_2' = n_2 + \hat{d}_2$, and

$$\alpha = n_1' - \lfloor n_1' \rfloor,$$

$$\beta = n_2' - \lfloor n_2' \rfloor,$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Figure 10 shows the estimated motion vector fields for the Mobile & Calendar sequence using the three aforementioned motion estimation methods. Note that for a fair comparison we use BMA with both integer and half-pel accuracy that are widely adopted by international video coding standards such as H.263 and MPEG. In order to see motion vectors clearly, their magnitudes are enlarged four times in Fig. 10. (To save space only the half-pel BMA is shown.) For this particular sequence, FCA provides the most consistent and reliable motion vector field. Both BMA and PCA fail to detect the relatively slow motion of the toy train and the background calendar. Similar results are shown in Fig. 11 for the Miss America sequence. The motion vectors in Fig. 11 are also magnified by a factor of 4. In this case, due to the noise in the still background area, both BMA and PCA produce abrupt motion vector fields. Although these abrupt motion vectors may lead to lower numerical mean-squared errors, they are incorrect motion vectors. Because of the noise-resistant property of FCA, it produces more reliable estimates. In terms of our terminologies, some incorrect estimates (on the background) caused by phase ambiguity due to noise are removed.

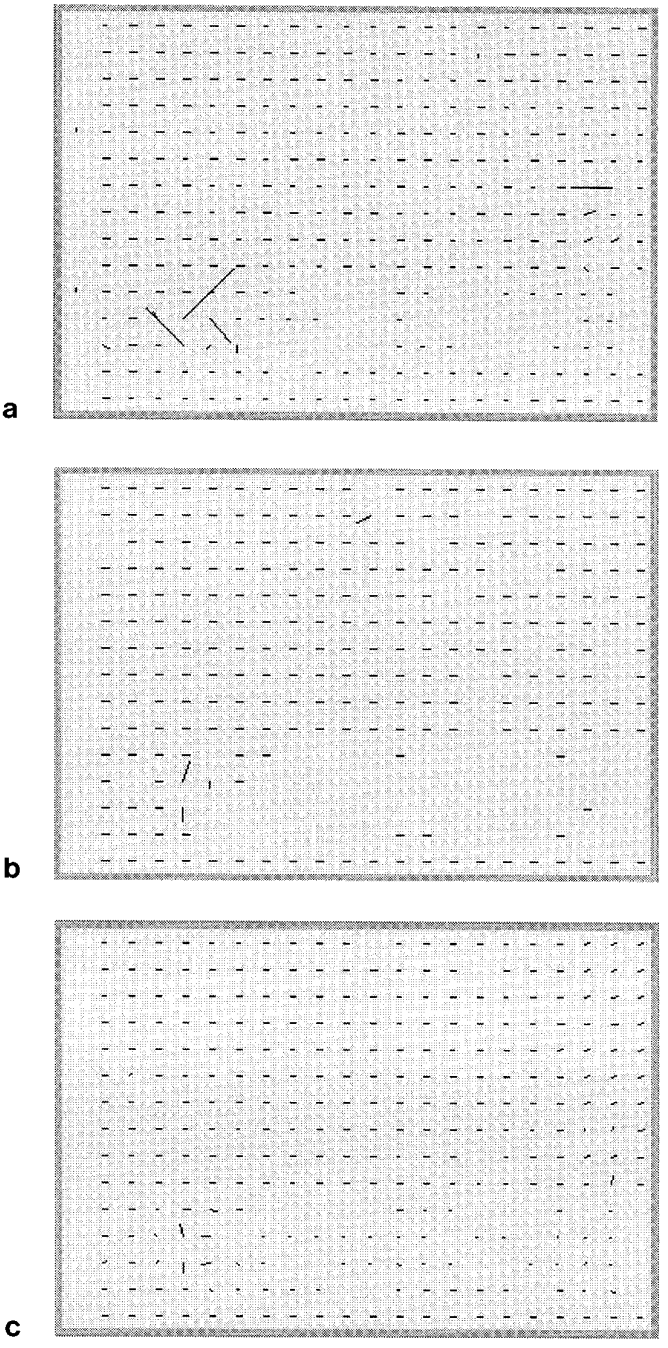


FIG. 10. Motion vector fields of Mobile & Calendar sequence using (a) block matching with half-pel accuracy, (b) phase correlation, and (c) frequency component methods (4× magnification).

To see more clearly the correctness of motion estimation, we use Mobile & Calendar as an example. The motion-compensated pictures using BMA and FCA are shown in Fig. 12. Portions of these two pictures are enlarged in Fig. 13 to show the differences. We find more predominant artifacts on the BMA-compensated image, such as the loss of the nail spot on the fence (the upper left in Fig. 13) and the discontinuity near the rotational toy (the upper right in Fig. 13). On the other hand, the FCA-compensated image appears to be a little bit more blurring due to bilinear interpolation. Overall, our frequency component approach typically offers better visual quality (motion-compensated) images than the other methods.

Mean-squared error (MSE) between the original and the motion-compensated pictures is a popular performance measure for motion estimation. However, it is not necessar-

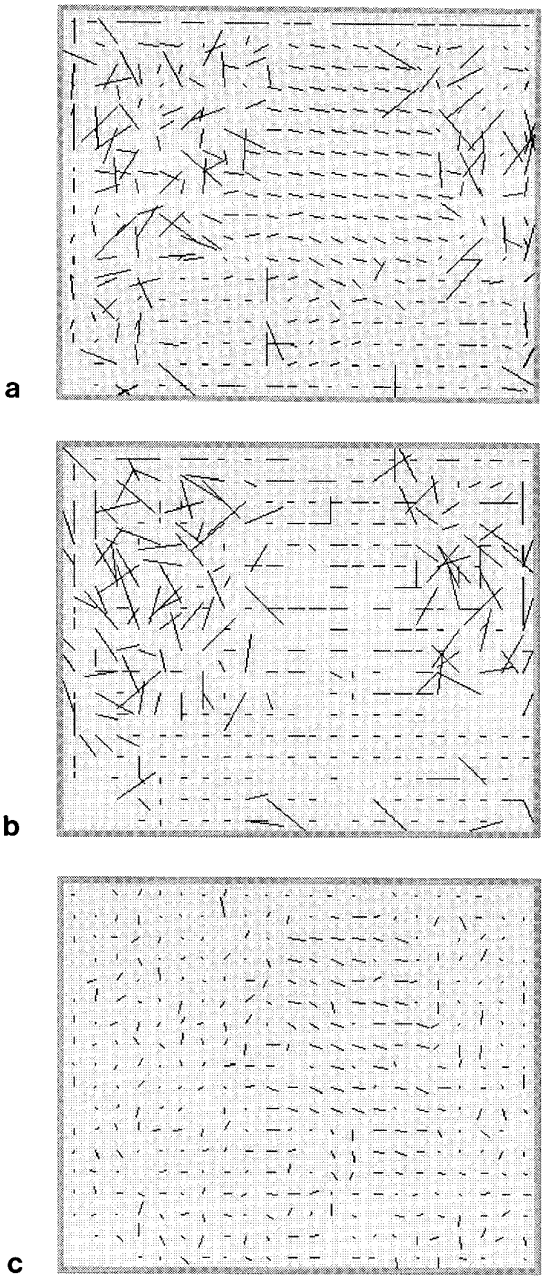


FIG. 11. Motion vector fields of Miss America sequence using (a) block matching with half-pel accuracy, (b) phase correlation, and (c) frequency component methods (4× magnification).

ily an accurate measure for justifying the motion field correctness, particularly for the noisy pictures. The MSEs of these three motion estimation algorithms are shown in Fig. 14. In the case of the Mobile & Calendar sequence (Fig. 14a), FCA has the least MSE partially due to its correct motion vector field and partially due to its fractional motion vector values. However, FCA has a slightly larger MSE than BMA for the Miss America sequence (Fig. 14b). In certain portions of this sequence, the displacement vectors are large and FCA fails to track large movement due to nonoverlapped block data as discussed in the previous sections. Table 1 lists the average mean-squared errors of both test sequences using these three methods. From Fig. 14 and Table 1, it is clear that FCA improves PCA significantly (about 40% MSE reduction) although both are frequency domain algorithms. Because the normalization (equalization) operation in PCA enhances the noise power at high frequencies, it produces incorrect displacement estimates on slow moving and noisy pictures. Also, FCA uses only large magnitude components to avoid the threshold effect.

Finally, we examine the convergence of our iterative algorithm. Figure 15 shows the average mean-squared errors of the two test image sequences under different calculation (recursion) numbers for the first iteration. Note that

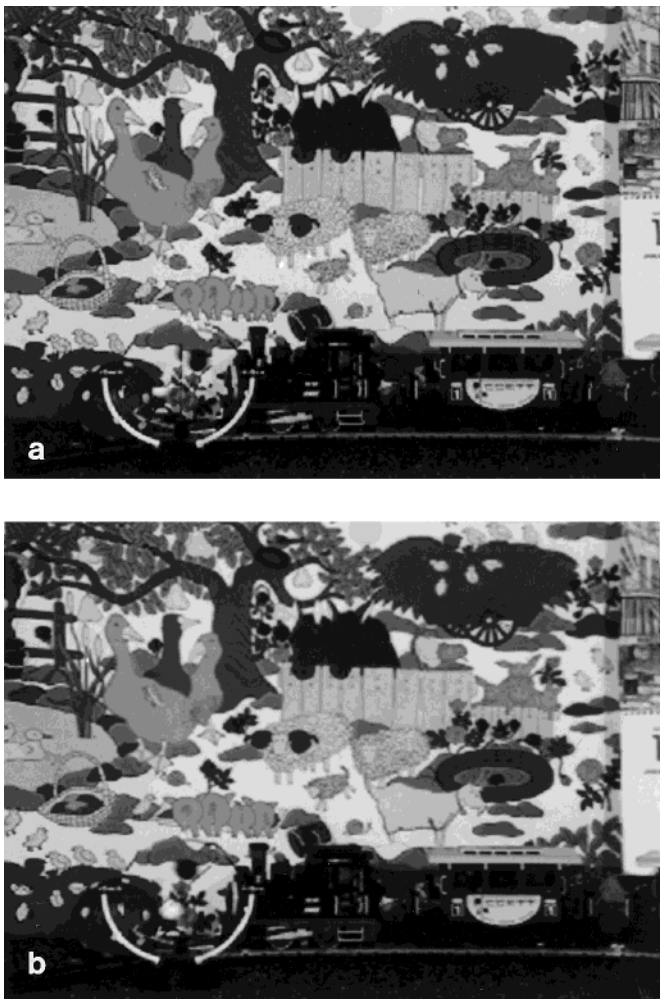


FIG. 12. Motion-compensated pictures of Mobile & Calendar sequence using (a) block matching with half-pel accuracy and (b) frequency component methods.

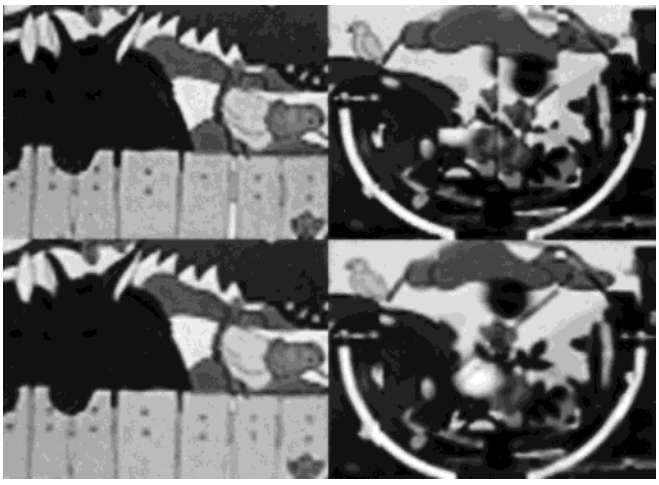


FIG. 13. Enlarged portions of the motion-compensated pictures of the Mobile & Calendar sequence using block matching with half-pel accuracy (top) and frequency component methods (bottom).

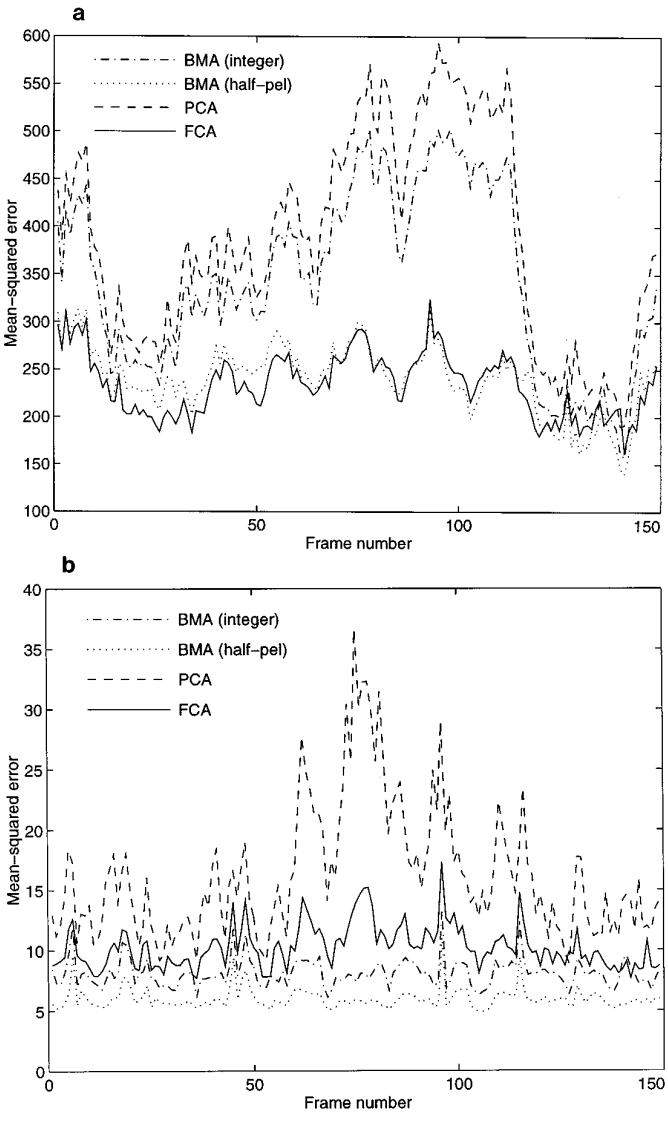


FIG. 14. Mean-squared errors for (a) Mobile & Calendar and (b) Miss America sequences.

the curves in Fig. 15 are derived without using a convergence threshold. As we expect, the MSE decreases as the number of recursions increases. These curves reach their minimum in approximately 100 or so recursions. Thus we can further limit our recursion number below 100 in our algorithm to reduce the computational load. In this case, the average numbers of iterations and numbers of recursions in one iteration for both sequences are shown in Table 2. On average, it requires less than one block shift to obtain the final result in our experiments. That is, typically FCA needs fewer than three FFT operations for a block. On the other hand, PCA uses exactly three FFT operations in estimating each motion vector. However, FCA needs an additional 120 or so recursions. Each recursive process shown in Fig. 6 requires seven additions and seven multiplications. The recursion is essentially a search procedure and its complexity can be reduced by using faster search algorithms.

7. CONCLUSIONS

This paper contains two major points, analyzing the motion estimation problem in the frequency domain and proposing a frequency-component-based robust motion estimation algorithm. The analysis in the previous sections reveals the contribution of various frequency components in the motion estimation process, particularly their roles in the accuracy and the ambiguity problems. When properly used, the low frequency components can reduce ambiguity and the high frequency components can increase accuracy. The displacement information is contained in the phase portion of the frequency components. Extraction of motion information directly from the phase components is disturbed by "noise." An interesting phenomenon is that the noise effect in this procedure is similar to that in frequency modulation (FM). That is, there is a noise-reduction effect when the noise is much smaller than the signal and the threshold effect appears when the noise strength is comparable to the signal strength. The preceding analysis can

TABLE 1
Average Mean-Squared Errors for the Test Sequences

Test Sequence	BMA (integer)	BMA (half-pel)	PCA	FCA
Mobile & Calendar	343.72	239.69	388.46	234.10
Miss America	8.09	6.00	16.24	10.26

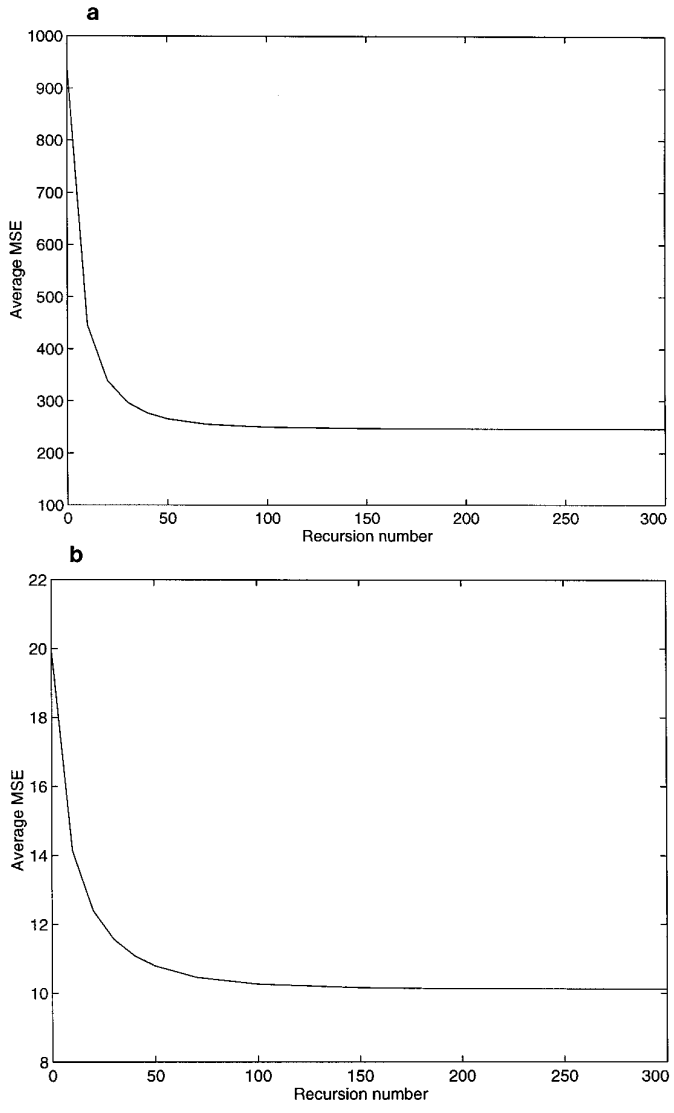


FIG. 15. Average mean-squared errors versus recursions of the first iteration in the frequency component method for (a) Mobile & Calendar and (b) Miss America sequences.

be used to improve the conventional motion estimation schemes. For example, since both the very low and the high frequency components do not help in motion estimation, a properly designed band-pass filter applied to the images should increase the reliability of the block matching algorithm.

Then, we propose a new frequency component motion estimation algorithm to resolve the dilemma of accuracy and ambiguity. A recursive procedure including a phase ambiguity estimator is designed to extract motion information from the phase components. The major limitation of this approach is the nonoverlapped block data. We increase the motion estimation range by shifting the evaluation block locations progressively. Our experiments indicate that for displacement of less than 1/3 of the block length, this scheme produces more reliable motion vector estimates than the conventional block matching and phase correlation algorithms, particularly for noisy images. To further increase the displacement estimation range, a hierarchical structure similar to that in the hierarchical block matching algorithm [25] may be used.

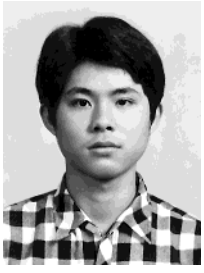
TABLE 2
Average Iterations and Recursions of the Frequency Component Algorithm

Test Sequence	Iterations	Recursions per iteration
Mobile & Calendar	1.61	80.62
Miss America	1.60	67.93

REFERENCES

1. H. G. Musmann, P. Pirsch, and H.-J. Grallert, Advances in picture coding, *Proc. IEEE* **73,** 1985, 523–548.
2. H.-M. Hang and Y.-M. Chou, Motion estimation for image sequence compression, in *Handbook of Visual Communications* (H.-M. Hang and J. W. Woods, Eds.), pp. 147–188, Academic Press, San Diego, CA, 1995.
3. J. R. Jain and A. K. Jain, Displacement measurement and its application in interframe image coding, *IEEE Trans. Commun.* **29,** 1981, 1799–1808.
4. T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, Motion-compensated interframe coding for video conferencing, in *Proceedings, National Telecommunication Conference, New Orleans, LA, 1981*, pp. G5:3.1–3.5.
5. S. Kappagantula and K. R. Rao, Motion compensated predictive interframe coding, *IEEE Trans. Commun.* **33,** 1985, 1011–1015.
6. R. Srinivasan and K. R. Rao, Predictive coding based on efficient motion estimation, *IEEE Trans. Commun.* **33,** 1985, 888–896.
7. A. Puri, H.-M. Hang, and D. L. Schilling, An efficient block-matching algorithm for motion-compensated coding, in *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, 1987*, pp. 25.4.1–25.4.4.
8. C. Cafforio and F. Rocca, Methods for measuring small displacements of television images, *IEEE Trans. Inform. Theory* **22,** 1976, 573–579.
9. A. N. Netravali and J. D. Robbins, Motion-compensated television coding; I, *Bell System Tech. J.* **58,** 1979, 631–670.
10. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Artif. Intell.* **17,** 1981, 185–204.
11. R. Paquin and E. Dubois, A spatio-temporal gradient method for estimating the displacement field in time-varying imagery, *Comput. Vision Graphics Image Process.* **21,** 1983, 205–221.
12. D. R. Walker and K. R. Rao, Improved pel-recursive motion estimation, *IEEE Trans. Commun.* **32,** 1984, 1128–1134.
13. J. Biemond, L. Looijenga, D. E. Boekee, and R. H. J. M. Plompen, A pel-recursive Wiener-based displacement estimation algorithm, *Signal Process.* **13,** 1987, 399–412.
14. H. Yamaguchi, Iterative method of movement estimation for television signals, *IEEE Trans. Commun.* **37,** 1989, 1350–1358.

15. R. Chellappa and A. A. Sawchuk (Eds.), *Digital Image Processing and Analysis, Vol. 2, Digital Image Analysis*, IEEE Press, New York, 1985.
16. J. K. Aggarwal and N. Nandhakumar, On the computation of motion from sequences of images—A review, *Proc. IEEE* **76,** 1988, 917–935.
17. W. N. Martin and J. K. Aggarwal (Eds.), *Motion Understanding: Robot and Human Vision*, Kluwer, Norwell, MA, 1988.
18. M. I. Sezan and R. L. Lagendijk (Eds.), *Motion Analysis and Image Sequence Processing*, Kluwer, Norwell, MA, 1993.
19. C. D. Kuglin and D. C. Hines, The phase correlation image alignment method, in *Proceedings, IEEE International Conference on Cybernetics and Society, San Francisco, 1975*, pp. 163–165.
20. G. A. Thomas, Television motion measurement for DATV and other applications, *British Broadcasting Corporation Research Department Tech. Rep. 1987/11*, 1987.
21. G. A. Thomas, HDTV bandwidth reduction by adaptive subsampling and motion-compensated DATV techniques, *Soc. Motion Picture Television Eng. J.* 1987, 460–465.
22. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice–Hall, Englewood Cliffs, NJ, 1986.
23. A. B. Carlson, *Communication Systems*, McGraw–Hill, New York, 1986.
24. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice–Hall, Englewood Cliffs, NJ, 1985.
25. M. Bierling, Displacement estimation by hierarchical blockmatching, in *Proceedings, SPIE Conference on Visual Communications and Image Processing, Cambridge, MA, 1988*, pp. 942–951.



YUNG-MING CHOU was born in Taipei, Taiwan, on November 30, 1966. He received his B.S. degree in control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1989. He is now pursuing the Ph.D. degree in electronics at National Chiao Tung University. His current interests are motion estimation, motion segmentation, video coding, and image processing.

HSUEH-MING HANG received his B.S. and M.S. degrees from National Chiao Tung University, Hsinchu, Taiwan, in 1978 and 1980, respectively, and his Ph.D. in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1984. From 1984 to 1991, he was with AT&T Bell Laboratories, Holmdel, NJ. He joined the Electronics Engineering Department of National Chiao Tung University, Hsinchu, Taiwan, In December 1991. Dr. Hang was a conference co-chair of the Symposium on Visual Communications and Image Processing (VCIP), 1993, and the Program Chair of the same conference in 1995. He guest coedited two *Optical Engineering* special issues on visual communications and image processing in July 1991 and July 1993. He was an associate editor of *IEEE Transactions on Image Processing* from 1992 to 1994 and a coeditor of the book *Handbook of Visual Communications* (Academic Press, San Diego, 1995). He is currently an editor of *Journal of Visual Communication and Image Representation.* He is a senior member of IEEE and a member of Sigma Xi.