

Extended Analysis of Motion-Compensated Frame Difference for Block-Based Motion Prediction Error

Ko-Cheung Hui and Wan-Chi Siu, *Senior Member, IEEE*

Abstract—In the past, most design and optimization work on hybrid video codecs relied mainly on experimental evidence. A proper theoretical model is always desirable, since this allows us to explain the phenomena of existing codecs and to design better ones. In this paper, we make use of the first-order Markov model to derive an approximated separable autocorrelation model for the block-based motion compensation frame difference (MCFD) signal. A major assumption of our derivation is that the net deformation of pixels is directional, in general, rather than a uniform error distribution in a block. We have also shown that the imperfect block-based motion compensation is significant to the theoretical study and the behavior of motion-compensated codecs. Results of our experimental work show that the derived model can describe the statistical characteristics of the MCFD signals accurately. The model also shows that the imperfectly formulated block-based motion compensation can result in an incorrect MCFD autocorrelation function while, conversely, it can form a better block-based motion compensation scheme.

Index Terms—Autocorrelation model, compound covariance, hybrid video coding, motion compensation, motion model.

I. INTRODUCTION

THE transform coding method with motion-compensated prediction, abbreviated as hybrid interframe coding, is one of the most essential methods for modern video coding standards. In this method, block-based motion compensation frame difference signals are transform coded by the discrete cosine transform (DCT). The pixel errors between a block in a current frame and a motion compensated block in a reference frame are called the block-based motion compensation frame difference (MCFD) signals. The motion compensated block is a result of any effective motion estimation algorithm. To limit the required amount of bits, a quantizer carries out quantization to discard some part of the transform coefficients.

Most of the work for the design and optimization of the hybrid video codecs is carried out experimentally. A proper theoretical treatment of motion-compensated video coding is valuable for the design of state-of-the-art video codecs, even though it requires many assumptions and simplifications for the analysis of a complicated system processing real-world signals. Furthermore, even an approximate theory can provide useful insights into the underlying mechanisms of the video codecs. In 1987, the first comprehensive rate-distortion analysis of motion-compensated prediction (MCP) was presented [1]. This theoretical framework leads motion-compensated video coding away from

heuristics and towards an engineering science. After this initial analysis, a number of researchers investigated this subject in depth and developed many different techniques for efficiency improvement [2]–[12]. These techniques include motion-compensation with fractional-pixel accuracy [3], overlapped block motion compensation [4]–[7], and loop filtering technique [12].

To design an optimal coding algorithm, a signal-source model that is sufficiently accurate to reflect the practical signal characteristics is required. It is well known that the first-order Markov process is a successful model for still image analysis. This model is accurate for the correlation relationship for smooth image, and it works fairly well for the first few steps in active images. By using the model, the correlation coefficient ρ for natural image is often suggested as 0.95. Hence, we consider that the DCT is a very close approximation of the optimum KLT [13] and widely employed for video coding. However, for motion-compensated errors signal, the situation is very different. It has been observed that the MCP errors at block boundaries tend to be larger than those at block centers [5]. It means that they are space-dependent and the assumption of wide-sense stationary (WSS) is not valid. As a result, it is inaccurate to employ the simple Markov model for the MCP errors.

In [8], Chen and Pang proposed theoretically a compound covariance model (the CP model) and demonstrated that the DCT performs nearly optimally in intraframe coding. Nevertheless, this investigation still assumed that the prediction errors are WSS across a block. In order to reduce and to equalize the MCP errors across a block, overlapped block motion compensation (OBMC) was developed. In [9], Niehsen and Brüning confirmed that the statistical means and standard deviations of the errors may change significantly from block to block. Hence, they proposed another compound covariance model (the NB model) empirically, which takes the OBMC into account. According to their experimental results, they claimed that their model closely fitted the characteristic of practical signals. The major disadvantage of this model, however, is that it lacks a theoretical basis, and, thus, its use for other analytical purposes is limited. To resolve the problem, we assume that a net deformation of pixels in a block along a certain direction is a more general situation. In addition, we consider that the prediction errors are the result of a motion estimation algorithm. We improve the CP model to propose a covariance model analytically by making use of this assumption. Our proposed model reflects the characteristics of practical prediction errors well and is comparable to the empirical model proposed in [9]. Moreover, our model explains the deviation of the compound model [8].

The outline of this paper is as follow. Section II presents the derivation of a mathematical model for autocorrelation of block-based motion prediction error. The origin of our derivation is similar to that of [11], and it is easy to show that the compound model [8] can be obtained analytically from our model.

Manuscript received March 8, 2006; revised December 15, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bruno Carpentieri.

The authors are with the Centre for Multimedia Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: enwcsiu@polyu.edu.hk).

Digital Object Identifier 10.1109/TIP.2007.894263

This work is also given in Section II. Section III shows our simulation results, and compares our model with other models. The experimental results given in [8], [9], and [11] provide evidence which verifies the accuracy of our model. Section IV summarizes and concludes this study.

II. MODELING OF THE AUTOCORRELATION OF BLOCK-BASED MOTION PREDICTION ERROR

In [8] and [11], the CP model was proposed theoretically

$$C(I) = A \cdot \rho^{|I|} + (1 - A)\delta(I) = C_1(I) + C_2(I) \quad (1)$$

where $C(I)$ is the compound covariance of the prediction errors of the CP model [8], I is the pixel separation in the x dimension, and $\delta(\cdot)$ is the Kronecker delta function with $\delta(0) = 1$ and $\delta(I) = 0$ for $I \neq 0$; with $A = 0.5$ and $\rho = 0.95$ for motion-compensated frame difference. This model can be separated into two parts. The first component, $C_1(I)$, in (1) represents the autocorrelation of a first-order autoregressive process, AR(1), with $\rho = 0.95$. The second component represents the white noise with a flat power spectrum [8]. However, this model deviates significantly from experimental results.

In [9], Niehsen and Brünig, proposed a different compound covariance model (the NB model) empirically, which takes the OBMC into account. That is

$$C_e(I) = c\rho_0^{|I|} + (1 - c)\rho_1^{|I|^2} \quad (2)$$

where $C_e(I)$ is the compound covariance of the prediction error of the NB model, and c , ρ_0 , and ρ_1 are model parameters. The model parameters, $c = 0.17$, $\rho_0 = 0.91$, and $\rho_1 = 0.38$, were chosen to fit their empirical covariance in the l_1 -norm¹ sense.

The simplicity and analytical tractability make the first-order autoregressive [AR(1)] model a popular one in still image and image sequence processing. Our model is also based on the AR(1) one and with image correlation coefficient equal to ρ .

For a block of pixels $f_t(i, j)$ in a frame at time t , the block-based motion compensation uses a matched block $f_{t-1}(i+u, j+v)$ in a reference frame at time $t-1$ for prediction. Hence, the motion prediction error is given by

$$e(i, j) = f_t(i, j) - f_{t-1}(i+u, j+v) \quad (3)$$

where (u, v) represent the motion vector of the block.

The autocorrelation function, $C_e(I, J)$ of the prediction error is then given by

$$\begin{aligned} C_e(I, J) &= E\{f_t(i, j) - f_{t-1}(i+u, j+v)\} \\ &\quad \times \{f_t(i+I, j+J) \\ &\quad - f_{t-1}(i+u+I, j+v+J)\} \\ &= 2C_f(I, J) - C_{f,t}(I-u, J-v) \\ &\quad - C_{f,t}(I+u, J+v) \end{aligned} \quad (4)$$

¹ l_1 -norm is defined as the sum of absolute values. For example, the l_1 -norm of error between an interested model and the empirical data is defined as: $|\text{Err}|_{L_1} = \sum_{I=0}^I |C_{\text{mod}}(I) - C_{\text{exp}}(I)|$, where $|\text{Err}|_{L_1}$ is the l_1 -norm of error, $C_{\text{mod}}(I)$ represents the test autocorrelation model (e.g., the NB model) and $C_{\text{exp}}(I)$ represents the empirical data in pixel distance I .

where

$C_f(\cdot, \cdot)$ autocorrelation function with correlation coefficient ρ ;

$C_{f,t}(\cdot, \cdot)$ cross-correlation function between a frame at time t and the reference frame at time $t-1$, and I and J are pixel separations in the x and y dimensions, respectively.

We assume that the matched block $f_{t-1}(i+u, j+v)$ can be approximated to the current block $f_t(i, j)$ with reasonable deformation. That is, some physical models can approximate the deformations. For example, the affine transform is used in MPEG-4 for the block-matching algorithm in sprite construction. Thus

$$f_t(i+m_x, j+n_y) \approx f_{t-1}(i+u, j+v). \quad (5)$$

The deformation vector (m_x, n_y) represents the deformation of each pixel in the current block. The values of m_x and n_y rely on a number of factors, such as the motion activities, light variation, inaccuracy of motion compensation, quantization error, and noise. However, these two factors are generally independent in our consideration. Furthermore, in this modeling, the magnitudes of m_x and n_y are not related to that of the u and v directly. They depend only upon the matching or correlation between the matched block, $f_{t-1}(i+u, j+v)$ and the current block, $f_t(i, j)$. Substituting the L.H.S. of (5) into the R.H.S. of (4) results in

$$\begin{aligned} C_{f,t}(I-u, J-v) \\ \approx E[f_t(i+m_x, j+n_y)f_t(i+I, j+J)] \\ = C_f(I-m_x, J-n_y). \end{aligned} \quad (6)$$

The reasons for the block deformation include failure of the block-based motion model for moving parts, light variation, inaccuracy of motion compensation, quantization error, and noise. We regard the deformation vector as a pair of independent random variables: $(-\infty \leq m_x \leq \infty, -\infty \leq n_y \leq \infty)$, and the expected value of the autocorrelation function, $C_e(I, J)$ can subsequently be represented by

$$E[C_e(I, J)] = 2C_f(I, J) - 2E[C_f(I-m_x, J-n_y)]. \quad (7)$$

By applying the separable 2-D AR(1) model, we have

$$C_f(I, J) = \sigma_f^2 \rho^{|I|} \rho^{|J|} \quad (8)$$

where σ_f^2 is the variance of the pixel values in the AR(1) model, and

$$E[C_f(I-m_x, J-n_y)] = \sigma_f^2 E[\rho^{|I-m_x|} \rho^{|J-n_y|}]. \quad (9)$$

For the sake of simplicity, a separable autocorrelation model is our objective. Hence, m_x and n_y are generally independent in our consideration. Equation (9) along the x axis can then be expressed as

$$\begin{aligned} E[C_f(I-m_x, J-n_y)] &= \sigma_f^2 E[\rho^{|I-m_x|}] E[\rho^{|J-n_y|}] \\ &I \geq 0, \quad J = 0. \end{aligned} \quad (10)$$

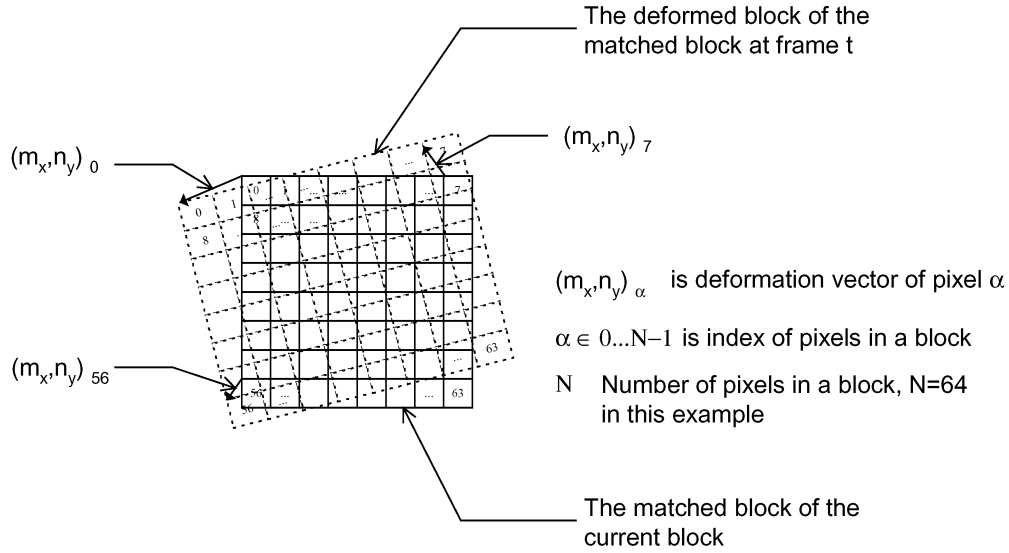


Fig. 1. Illustration of the assumption that pixels in a deformed block tend to deform along a definite direction.

The term $E[\rho^{I-m_x}]$ can be computed as

$$E[\rho^{I-m_x}] = \int p(m_x) \rho^{I-m_x} dm_x \quad (11)$$

where $p(m_x)$ is the probability density function (pdf) of m_x .

At this stage, let us assume that pixels in a deformed block tend to deform along a definite direction rather than deforming randomly. In other words, the mean vector of the deformation vectors (m_x, n_y) is not regarded as zero according to this assumption. This assumption is based on the nature of 1) part of an object being translated, 2) partial rotation of a moving part, 3) zooming, and 4) inaccuracy of motion compensation.

Fig. 1 illustrates our assumption on the deformation using a simplified version of the affine transform as an example. The relationship between the coefficients of this deformation model can be written as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}_\alpha = M \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}_\alpha \quad (12)$$

where

$(x, y)_\alpha$ coordinate of pixel α in the current block;

$(x', y')_\alpha$ coordinate of the corresponding pixel α in the deformed block;

and

M 3×3 affine transform matrix, which represents rotation and scaling in this example.

Let us define the deformation vector of pixel α as, $(m_x, n_y)_\alpha = (x', y')_\alpha - (x, y)_\alpha$, and the mean deformation vector of $(m_x, n_y)_\alpha$ in a block as (μ_x, μ_y) , where

$$(\mu_x, \mu_y) = \left(\sum_{\alpha=0}^{N-1} \frac{(m_x)_\alpha}{N}, \sum_{\alpha=0}^{N-1} \frac{(n_y)_\alpha}{N} \right) \quad (13)$$

for N is the number of pixels in a block and is equal to 64 in this example.

Let also σ_{mvx} and σ_{mvy} be the standard deviations of the x - and y -components of the deformation vectors $(m_x, n_y)_\alpha$ in the block. Fig. 1 demonstrates that part of a current block is not predicted accurately enough, because not all pixels in the block are translated in the same direction and their moving distances are not identical. However, they still present some motion tendency. Thus, a finite mean vector (μ_x, μ_y) can be defined for the deformation vectors of (m_x, n_y) in a block. Hence, we refine (11) with the above consideration. It then becomes

$$E[\rho^{I-m_x}] = \int p(m_x, \mu_x) \rho^{I-m_x} dm_x \quad (14)$$

where $p(m_x, \mu_x)$ is the probability density function (pdf) of m_x with mean deformation μ_x , where μ_x is the x component of the mean deformation vector in each block.

Equation (9) represents the variance-normalized cross-correlation function between the matched block and the current frame in terms of the image correlation coefficient, ρ . It is a matching or correlation measure of the matched block in the current frame expressed conceptually. The assumed block deformation makes the cross-correlation function, $E[C_{f,t}(I - m_x, J - n_y)]$ and the error autocorrelation function depend on the direction of the mean deformation vector. However, an error autocorrelation function must be an even function with respect to I . To remedy this directional dependence, we only consider the absolute value of μ_x . In addition, the physical meaning can be interpreted graphically in Fig. 2, which shows the difference between the autocorrelation function of a frame at time t , and the cross-correlation function between the frame at time t and the reference frame at time $t - 1$. To consider the absolute value of μ_x , let us consider only the prediction errors within the domain of a compensated block. Then, we further assume that μ_x is randomly distributed. It gives

$$E[\rho^{I-m_x}] = \iint p(m_x || \mu_x) p(\mu_x) \rho^{I-m_x} dm_x d\mu_x \quad (15)$$

where $p(\mu_x)$ is the probability density function of μ_x .

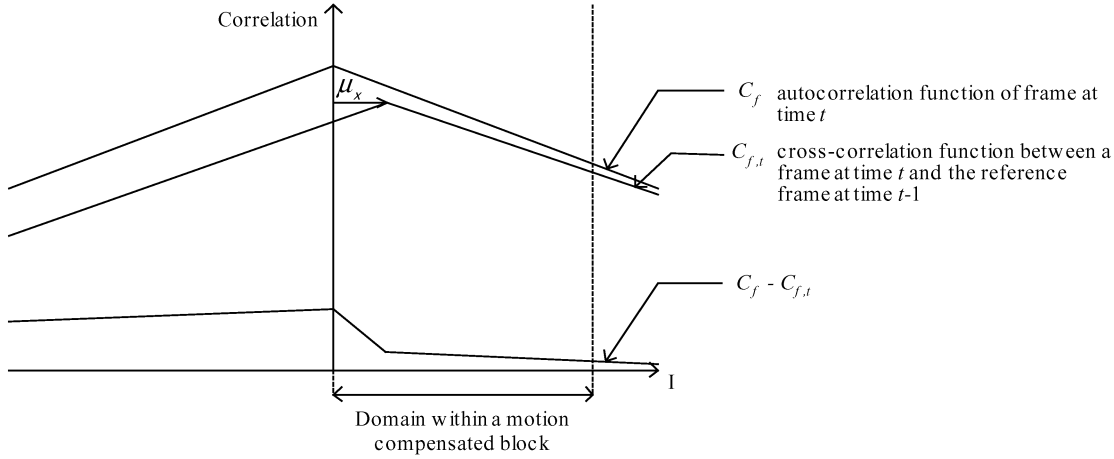


Fig. 2. Difference between the autocorrelation function of a frame at time t and the cross-correlation function between the frame at time t and the reference frame at time $t - 1$ in the 1-D case.

For the sake of simplicity and because of its separable property, we use Gaussian distributions to represent the conditional distribution function, $p(m_x|\mu_x)$ and the pdf, $p(\mu_x)$. The $E[\rho^{|I-m_x|}]$ is then given by

$$E[\rho^{|I-m_x|}] = \frac{2}{\sigma_\mu\sqrt{2\pi}} \frac{1}{\sigma_{mv}\sqrt{2\pi}} \int_0^\infty e^{-\frac{\mu_x^2}{2\sigma_\mu^2}} \times \int_{-\infty}^\infty e^{-\frac{(m_x-|\mu_x|)^2}{2\sigma_{mv}^2}} \rho^{|I-m_x|} dm_x d\mu_x \quad (16)$$

where

- σ_μ standard deviation of the mean deformation vector;
- σ_{mv} standard deviation of the deformation vectors in a single block.

Assume that the block deformation in the x and y dimensions can be modeled with the same standard deviations σ_μ and σ_{mv} . Making use of (8) and (10), the error autocorrelation function along x direction in (7) can then be expressed as

$$E[C_e(I, J)] = 2\sigma_f^2 \left\{ \rho^{|I|}\rho^{|J|} - E[\rho^{|I-m_x|}]E[\rho^{|J-n_y|}] \right\} \quad I \geq 0, \quad J = 0 \quad (17)$$

and the variance-normalized autocorrelation function of the block prediction error is given by

$$\frac{E[C_e(I, 0)]}{E[C_e(0, 0)]} = \frac{\rho^{|I|}\rho^{|0|} - E[\rho^{|I-m_x|}]E[\rho^{|-n_y|}]}{1 - E[\rho^{|m_x|}]E[\rho^{|n_y|}]} \quad I \geq 0, \quad J = 0. \quad (18)$$

The result of (18) can be obtained by numerical calculation. However, this form is not convenient for analytical purposes. We will describe the derivation of its approximated form below.

We use an expected value of the mean deformation vector in (15), instead of using the pdf, $p(\mu_x)$ to approximate the autocorrelation model. The expected value of the mean deformation vector is

$$\bar{\mu}_x(\sigma_\mu) = \frac{2}{\sigma_\mu\sqrt{2\pi}} \int_0^\infty \mu_x e^{-\frac{\mu_x^2}{2\sigma_\mu^2}} d\mu_x = \sigma_\mu \sqrt{\frac{2}{\pi}}. \quad (19)$$

Making use of the expected value of μ_x , $\bar{\mu}_x$, (16) can be simplified, approximately, in the form of a single integral

$$E[\rho^{|I-m_x|}] \approx \int p(m_x|\bar{\mu}_x) \rho^{|I-m_x|} dm_x = \frac{1}{\sigma_{mv}\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{(m_x-|\bar{\mu}_x|)^2}{2\sigma_{mv}^2} + |I-m_x| \ln \rho} dm_x = \frac{1}{\sigma_{mv}\sqrt{2\pi}} \left[\int_{-\infty}^{|I|} e^{-\frac{(m_x-|\bar{\mu}_x|)^2}{2\sigma_{mv}^2} + (|I-m_x|) \ln \rho} dm_x + \int_{|I|}^\infty e^{-\frac{(m_x-|\bar{\mu}_x|)^2}{2\sigma_{mv}^2} + (m_x-|I|) \ln \rho} dm_x \right]. \quad (20)$$

These two integrations are expressed in terms of the error function, $\text{erf}(z)$. Finally, we have

$$E[\rho^{|I-m_x|}] \approx \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ \rho^{(|I-|\bar{\mu}_x||)} \left[1 + \text{erf} \left(\frac{|I| - |\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] + \rho^{-(|I-|\bar{\mu}_x||)} \left[1 - \text{erf} \left(\frac{|I| - |\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] \right\} = \rho^{|I|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ \rho^{|\bar{\mu}_x|} \left[1 + \text{erf} \left(\frac{|I| - |\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] + \rho^{-(2|I-|\bar{\mu}_x||)} \left[1 - \text{erf} \left(\frac{|I| - |\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] \right\} = \rho^{|I|} \tilde{R}(I, \bar{\mu}_x, \sigma_{mv}, \rho). \quad (21)$$

Using (21), we can approximate the variance-normalized autocorrelation function of the block prediction error which is given by (18) as

$$\frac{E[C_e(I, 0)]}{E[C_e(0, 0)]} \approx \tilde{C}_{e, \bar{\mu}_x, \sigma_{mv}}(I, 0) = \rho^{|I|} \rho^{|0|} \frac{1 - \tilde{R}(I, \bar{\mu}_x, \sigma_{mv}, \rho) \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho)}{1 - \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho) \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho)} = \rho^{|I|} \rho^{|J|} \Re(I, J, \bar{\mu}_x, \sigma_{mv}, \rho)|_{J=0}. \quad (22)$$

Finally, we express a separable 2-D variance-normalized autocorrelation function as

$$\frac{E[C_e(I, 0)]E[C_e(0, J)]}{E[C_e(0, 0)]^2} = \rho_x^{|I|} \rho_y^{|J|} \mathfrak{R}(I, 0, \bar{\mu}_x, \sigma_{mvx}, \rho_x) \mathfrak{R}(0, J, \bar{\mu}_y, \sigma_{mvy}, \rho_y). \quad (23)$$

Details of the deviation of (21) from (20) are shown in Appendix I.

In fact, the function $\mathfrak{R}(I, J, \bar{\mu}_x, \sigma_{mv}, \rho)|_{J=0}$ is equivalent to the role of A(a) in [11] in the 1-D case, which is shown by (1).

Moreover, we can show that the compound covariance model, (1), proposed in [8] can be obtained from (22). The model of (1) is a 1-D case that is derived based on a translational motion model and a composite motion-estimation-error probability density function (pdf) consisting of a uniform pdf for the granular estimation errors and an impulse pdf for the background regions with zero estimation errors.

To derive (1), we assume that a video sequence with very low motion activity is under simulation. For this reason, the block-based motion model can properly compensate the motion of each block between successive frames, and, thus, $\bar{\mu}_x$ and σ_{mv} are set to zero and 0.5, respectively. Using the properties of $\text{erf}(z)$

$$\text{erf}(z) = \begin{cases} \frac{2}{\sqrt{\pi}} e^{-z^2} z \left[1 + \frac{2z^2}{1.3} + \frac{(2z^2)^2}{1.3 \cdot 5} + \dots \right], & \text{if } z \ll 1 \\ 1, & \text{if } z \gg 1 \end{cases} \quad (24)$$

an approximation of (21) can be obtained. For $|I|$ equal to zero, we express the $\text{erf}(\cdot)$ in (21) as the first term of its expanded series, and for $|I|$ greater or equal to one, the $\text{erf}(\cdot)$ is expressed as unity. Hence, (21) is approximated as

$$\begin{aligned} E[\rho^{|I-m_x|}] &\approx \rho^{|I|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ \left[1 + \frac{2}{\sqrt{\pi}} e^{-\frac{(\sigma_{mv} \ln \rho)^2}{2}} \right. \right. \\ &\quad \times \left. \left. \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \delta(|I|) + (1 - \delta(|I|)) \right] \right. \\ &\quad \left. + \left[1 - \frac{2}{\sqrt{\pi}} e^{-\frac{(\sigma_{mv} \ln \rho)^2}{2}} \cdot \frac{-\sigma_{mv} \ln \rho}{\sqrt{2}} - (1 - \delta(|I|)) \right] \right\} \\ &= \rho^{|I|} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left[1 + \sqrt{\frac{2}{\pi}} e^{-\frac{(\sigma_{mv} \ln \rho)^2}{2}} \sigma_{mv} \ln \rho \cdot \delta(|I|) \right]. \end{aligned} \quad (25)$$

Let $\beta = \sqrt{(2/\pi)} e^{-((\sigma_{mv} \ln \rho / \sqrt{2})^2)} \sigma_{mv} \ln \rho$, and substitute it into (25). We have

$$\begin{aligned} \tilde{C}_{e, \bar{\mu}_x, \sigma_{mv}}(I, 0) &\approx \rho^{|I|} \frac{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta \delta(|I|)] \cdot [1 + \beta]}{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]^2} \\ &= \rho^{|I|} [A + (1 - A) \delta(|I|)] \end{aligned} \quad (26)$$

where

$$A = \frac{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]}{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]^2}. \quad (27)$$

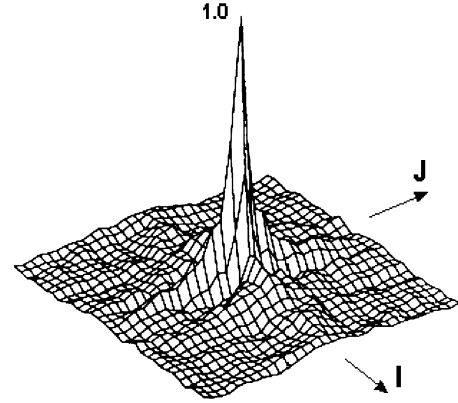


Fig. 3. Autocorrelation function of MCFD signals generated from Trevor sequence.

The numerical value of A is equal to 0.497 with $\rho = 0.95$ and $\sigma_{mv} = 0.5$. Hence, (1) is derived. Details of the deviation of (26) are shown in Appendix II.

III. SIMULATION RESULTS

Let us consider an analytical treatment of the motion-compensated frame differences (MCFD). A covariance model that can represent the empirical covariance result accurately enough is required. Hence, it is necessary to verify a covariance model by comparing its fitness to empirical results. We have made use of the experimental results from [8], [9], and [11] to compare the fitness of our model with that of the CP model and the NB model. The empirical result of the ‘‘Trevor’’ sequence extracted from [11] and the result of the sequence ‘‘Miss America’’ in [8] are used to evaluate the accuracy of our model and compared it with the CP model in the 2-D situation. Because some experimental parameters, such as the coordinates of the blocks of interest, the motion estimation algorithm used and the resulting motion vectors, were not indicated in [8] and [11], we had to choose some typical values to fit these parameters in our realization. We have also implemented additional experiments, in which the required parameters have been obtained experimentally, in order to compare the performance of the three models. We compared the accuracy of our model with that of the NB model in a 1-D case by using the statistical result of a number of standard sequences in [9].

Fig. 3 depicts the 3-D plot of an autocorrelation function of an MCFD signal generated from the ‘‘Trevor’’ images [11]. The spatial autocorrelation function of the ‘‘Trevor’’ image has a high value with $\rho = 0.99$ measured in [11]. Authors of [11] stated that the MCFD ρ was generated from frames 59 to 60 but did not state definitively the coordinates of their target block. Frames 59 to 60 show multiple people video conferencing. The motion activities in the x and y directions are presented, including activities such as the swinging of papers, closing a book, etc. We regard these activities as having a medium amount of movement. The whole conferencing background is a slowly moving curtain. Hence, we considered the presence of deformation in the

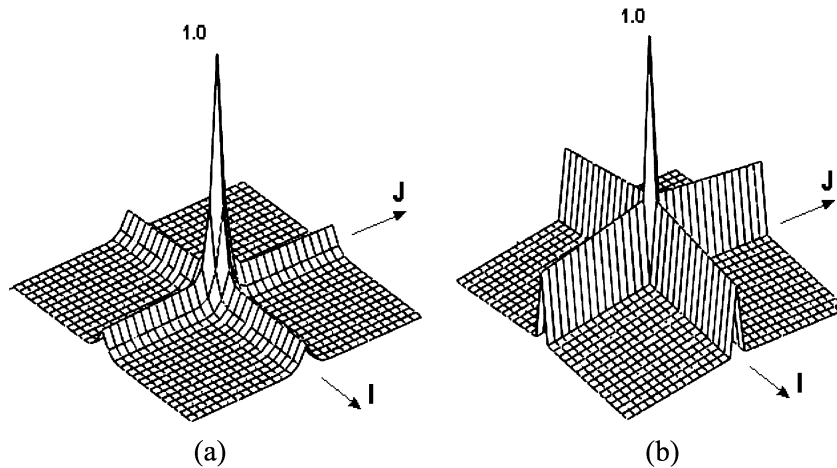


Fig. 4. Representation of the autocorrelation function of Trevor's MCFD signals by using separable models: (a) our improved CP model and (b) original CP model.

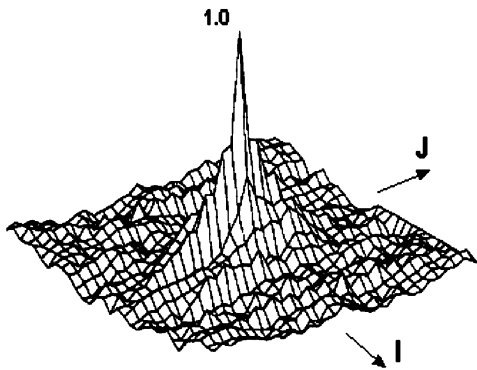


Fig. 5. Autocorrelation function of the MCFD of sequence Miss America.

x and y directions only. We set $\mu_x = \mu_y$ for simplicity, and we do not know the position of the target block. In H.261, integer pixel accuracy is applied. We considered that the medium motion activities have a deformation larger than the mean motion estimation error, 0.5. We selected $\mu_x = \mu_y = 0.75$. Similarly, we set σ_x equal to σ_y . The value $\sigma_x = \sigma_y = 1.15$ is arbitrarily selected to fit the shape of Fig. 3.

The result of our improved model is plotted in Fig. 4(a) by substituting $\rho_x = \rho_y = 0.99$, $\sigma_x = \sigma_y = 1.15$, and $\mu_x = \mu_y = 0.75$ into (21). We plotted the result of the CP model with $\rho_x = \rho_y = 0.99$ in Fig. 4(b) for comparison. According to the prediction from the CP model, the autocorrelation function is decreasing slowly for $|I|, |J| \geq 1$. It deviates significantly from the empirical result. However, the improved model represents the autocorrelation function more accurately than the original CP model. The improved model correctly represents the rapidly decreasing autocorrelation for $|I|, |J| \geq 1$.

Let us use another experimental result, the autocorrelation function of the MCFD of the sequence "Miss America" in [8], to evaluate the accuracy of our model and that of the original CP model. The pixel correlation coefficients are 0.88 and 0.80 in the vertical direction (indicated by arrow J) and horizontal direction (indicated by arrow I) as stated in [8]. The corresponding auto-

correlation function of the MCFD is shown in Fig. 5. Authors of [8] chose the motion compensated blocks (32×32) around the head area in the 15th frame of the sequence to generate the result in Fig. 5. The motion of the head area in the 15th frame is slow but the lips are opening with medium motion activity. We assumed the symmetrical motion of the lips and the slow motion of the other region would generate a low degree of deformation, with $\mu_x = \mu_y$ approximated to zero. The value $\sigma_x = \sigma_y = 1.15$ was selected identical to the "Trevor" sequence because it is within the typical range of our experimental results.

Fig. 6(a) illustrates the representation of the autocorrelation function of "Miss America's" MCFD signals by our improved model with the following parameters $\rho_x = 0.80$, $\rho_y = 0.88$, $\sigma_x = \sigma_y = 1.15$, and $\mu_x = \mu_y = 0$ in (24). Instead, Fig. 6(b) gives the result of the original CP model with $\rho_x = 0.80$ and $\rho_y = 0.88$. Fig. 6 shows that both models compare favorably with the experimental results in Fig. 5. Because the original CP model is not concerned with the practical motion properties of a video sequence, it assumes that the block-based motion estimation can successfully compensate for the motion of each pixel in a block. This assumption is only valid in a sequence with very slow motion activities, such as the "Miss America." However, we can still find that, in the I direction, the CP model shows a sudden drop of autocorrelation value at $|I| = 1$, which is incorrect when compared to the empirical result in Fig. 5. The improved model can represent correctly the trend of the autocorrelation function in both directions.

To further justify the improved model, we measured the autocorrelation functions of the MCFD of the "Football" sequence and the "Salesman" sequence. The typical positions of two sample Macroblocks (MB) with a size of 16×16 pixels of the "Football" and "Salesman" are indicated in Fig. 7(a) and (b), respectively. Fig. 8(a) and (b) illustrates, respectively, the 3-D plots of the autocorrelation functions of the MCFD of the MBs.

We performed the simplest exhaustive search motion estimation between frame 2 and frame 1 of the "Football" sequence

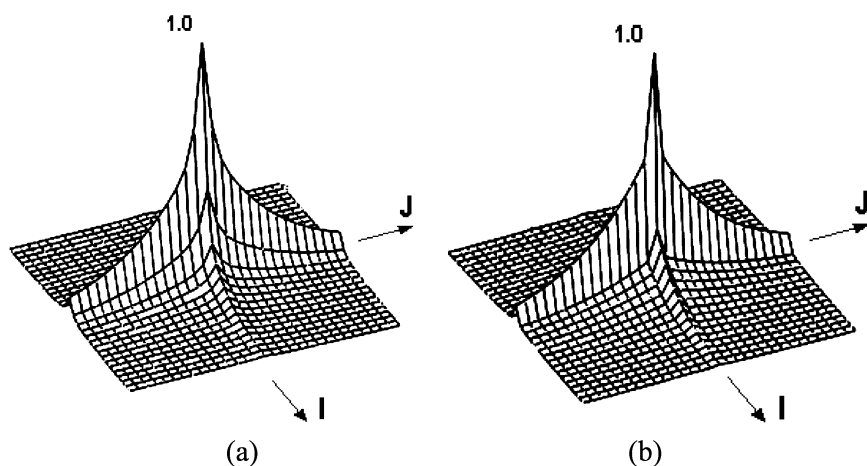


Fig. 6. Representation of the autocorrelation function of Miss America's MCFD signals by using separable models: (a) improved CP model and (b) original CP model.

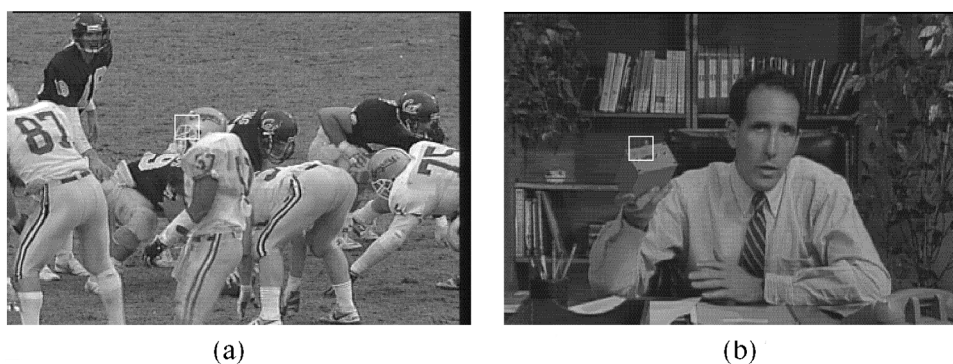


Fig. 7. Position of the target MBs at the measured frames of the sequences: (a) Football and (b) Salesman.

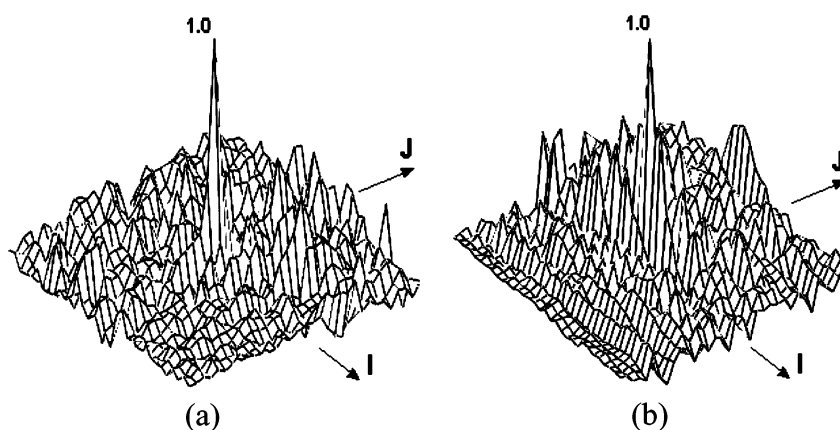


Fig. 8. Autocorrelation function of the MCFD of sequences: (a) Football and (b) Salesman.

(Frame 1 is identical to Frame 0 in the "Football" sequence). The motion estimation for the Salesman sequence was performed between frame 1 and frame 0. We performed the motion estimation with a search range of -16 to 16 pixel distance with integer pixel accuracy to obtain motion vectors of each block. The deformation vectors of each pixel were measured by dividing a block into a number of 2×2 pixels sub-blocks and then the motion estimation with a search range of -2 to 2 pixels and with

half pixel accuracy was carried out. Hence, a mean deformation vector and the standard deviations of the deformation vectors in a block can be estimated. Moreover, the averaged pixel correlation coefficients of the target MBs were measured along the vertical and horizontal directions within a window of 33×33 pixels centered with the target MBs, respectively.

Fig. 8(a) gives the autocorrelation function of Football's MCFD signals. Our improved model with the following pa-

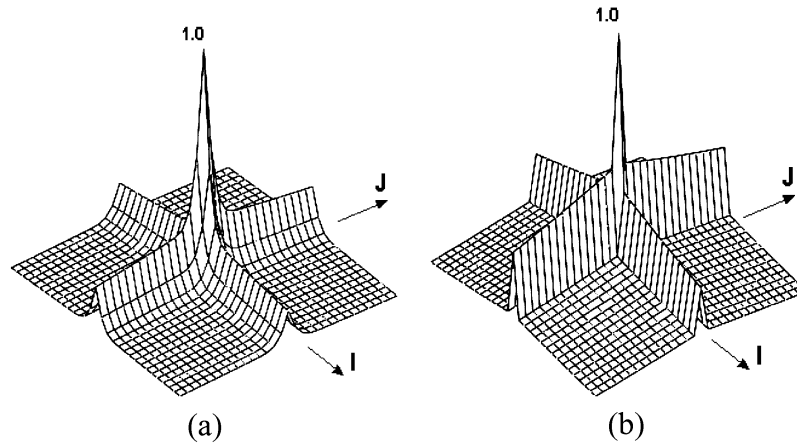


Fig. 9. Representation of the autocorrelation function of Football's MCFD signals by using separable models: (a) improved CP model and (b) original CP model.

rameters: $\rho_x = 0.979$, $\rho_y = 0.982$, $\sigma_x = 1.29$, $\sigma_y = 1.35$, and $\mu_x = 0.453$, $\mu_y = 0.539$ is plotted in Fig. 9(a) using (21). Meanwhile, Fig. 9(b) gives the result of the original CP model with $\rho_x = 0.979$ and $\rho_y = 0.982$. Fig. 9 shows that both models can predict that, on average, the autocorrelation in the J direction is larger than that in the I direction for $|I|, |J| \geq 1$. However, in terms of numerical values, both of these two models are not as good as the NB model when comparing the results as shown in Fig. 10. This figure shows that the results of the 1-D normalized autocorrelation of the Football's MCFD in the I and J directions. Fig. 10(a) depicts that the real autocorrelation signal decreases extremely fast from $I = 0$ to $|I| = 2$ and tends to zero in the I direction. In the J direction, the real autocorrelation signal decreases relatively slowly and tends to about 0.05. The original CP model seriously overestimates the autocorrelation values. It predicts that the autocorrelation function tends to about 0.41 and 0.48 in the I and J directions, respectively. The estimation from the improved model is more accurate, but it still suffers from serious overestimation. It predicts that the autocorrelation function approaches to about 0.22 and 0.27 in the I and J directions, respectively. On the other hand, the NB model indicates that the autocorrelation function decreases rapidly from $I = 0$ to $|I| = 2$ and then tends to about 0.1 in both I and J directions on average. The "Football" is a video sequence with fast motion activities. Moreover, the helmet part in the target MB has moderate light variation. The improved CP model can predict successfully that the autocorrelation in the J direction tends to have larger values than that in the I direction.

Fig. 8(b) shows a 3-D plot of the autocorrelation function of the Salesman's MCFD signal. From our measurement, parameters of the sequence "Salesman" are: $\rho_x = 0.970$, $\rho_y = 0.960$, $\sigma_x = 1.327$, $\sigma_y = 1.140$, and $\mu_x = 0.469$, $\mu_y = 0.789$. Our improved model is shown in Fig. 11(a), while the original CP model is shown in Fig. 11(b) with $\rho_x = 0.970$ and $\rho_y = 0.960$. Our improved model shows an extremely different autocorrelation function between the I and J directions. From Fig. 12, our improved model predicts that the autocorrelation

decreases slowly and then tends to 0.3 in the I direction. While in the J direction, the autocorrelation is predicted decreasing rapidly when $|J| \leq 2$ and tends to zero. The CP model predicts that the autocorrelation decreases rapidly from $I = 0$ to $|I| = 1$ and then tends to about 0.5 and 0.3 in the I and J directions, respectively, as shown in the Fig. 12. Our improved model provides an extremely good prediction in the I direction, for $|I| < 4$, and the best prediction in the J direction when compared with the other two models on real signals. In the I direction at $|I| < 4$, the autocorrelation function of the Salesman's MCFD signals decreases slowly to about 0.29. The improved model decreases gradually to about 0.36. However, the CP model and the NB model decrease too rapidly and tend to 0.52 and 0.13, respectively. In the J direction, only our improved model can predict correctly that the signal tends to zero.

In [9], the authors used four MPEG test sequences to verify and proposed their 1-D NB model, which empirically fitted in the sense of the l_1 -norm. Their experimental results are shown in Fig. 13. This figure gives the maximal, minimal and mean of the normalized autocorrelation functions of the MCFD at different pixel distances. Let us use these results to evaluate the fitness of the original 1-D CP model, the NB model and the improved CP model. These models are given in (1), (2), and (22), respectively. In this simulation, we set $\rho_x = 0.95$, $\sigma_x = 1$, and $\mu_x = 1$ in (22). We had to choose the values for σ_x and μ_x in order to fit close enough to the experimental results. The value of $\rho_x = 0.95$ is chosen, making it identical to the original CP model.

In Fig. 13, we find that the original CP model does not fit accurately with the empirical autocorrelation of the MCFD. The CP model decreases slowly with pixel distances, for $I \geq 1$, which differs from the empirical results significantly. Tables I and II summarize the performance of the NB-model, CP-model, and our improved CP-model for football and salesman video sequences. The NB model fits the experimental results closely, because the required parameters of the NB model are chosen to fit the experimental result in the l_1 -norm sense. The fitness of each of the three test models is compared in terms of their statistical results, the l_1 -norm errors. The NB model gives the

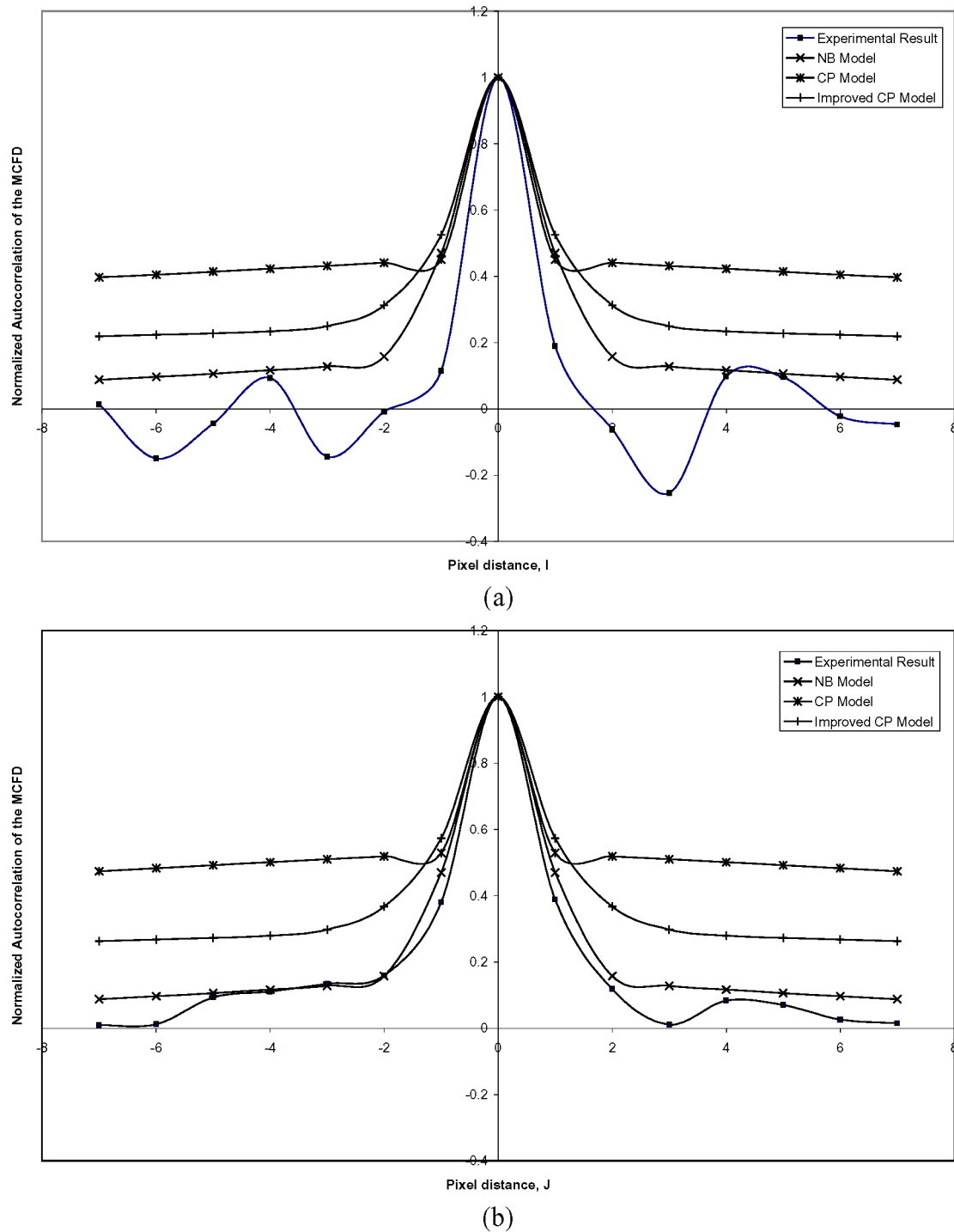


Fig. 10. One-dimensional normalized autocorrelation of the Football's MCFD in (a) I direction and (b) J direction; (a) 1-D normalized autocorrelation of the MCFD in I direction; (b) 1-D normalized autocorrelation of the MCFD in J direction.

smallest error value of 0.05. The improved CP model follows the NB model, with an error value of 0.10 whilst the original CP model results in the largest error value of 1.69 as shown in Fig. 13. However, the NB model is purely empirical and without any theoretical foundation. The usage of this model for analytical purposes is very limited. On the other hand, the improved CP model successfully represents the autocorrelation function of the MCFD. Moreover, the improved model is derived from

the simple first-order Markov model and takes into consideration the net deformation of pixels in a block due to imperfect block-based motion compensation. Hence, the improved CP model is suitable for analytical design and the investigation of signal decomposition algorithms in motion compensation. For instance, using $\sigma_x = 1$ and $\mu_x = 1$ indirectly suggests that the motion compensation errors in a block are not distributed uniformly. It is a property of the motion compensation error that

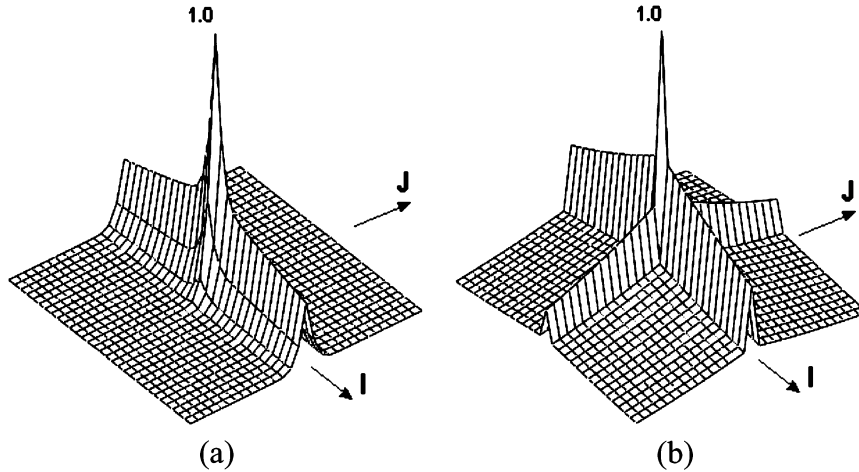


Fig. 11. Representation of the autocorrelation function of Salesman's MCFD signals by using separable models: (a) improved CP model and (b) original CP model.

induces a wide study of the OBMC algorithms in video coding. In these studies of the OBMC, experimental results [6] showed that the motion compensation error magnitude is larger at the boundary pixels than that of the pixels around the center of a block. In our derivation, this error distribution is assumed to be caused by the deformation in a block that cannot be compensated for by block-based motion compensation. Hence, our model can be used to analyze the error distribution of different motion activities. Moreover, it also explains the deviation of the CP model from the empirical results, which is due to the lack of concern relating to the practical situation of the block-based motion compensation in an encoder.

IV. CONCLUSION

A proper theoretical treatment of motion-compensated video coding is valuable for the design of most advanced video coding systems, even though it requires a number of assumptions and simplifications for the analysis of real-world signals. We have shown that the first-order Markov model can be used to derive an approximate separable autocorrelation signal model for the block-based motion compensation difference signal. In the derivation, we have assumed that a net deformation of pixels is directional in general rather than having a uniform error distribution in a block. We have improved the original CP model by proposing a covariance model analytically making use of this assumption. Simulation results show that the improved CP model can describe the characteristics of the MCFD signals accurately. We have also found that the concern of imperfect block-based motion compensation is one important step in the study of the motion-compensated coder; otherwise, the autocorrelation function of the MCFD signals cannot be expressed correctly. We can make use of this improved model to provide some useful insights into the analytical design and investigation of video signal decomposition algorithms.

Even though we borrowed some concepts from the affine transform which is a basis for sprite coding in the MPEG-4 stan-

dard, our model is general and the theory behind is fundamental. We have set forth a new analytical tool to model the deformation of pixels for block-based motion-compensated video coding. This is a useful tool which allows a good insight into the underlying principle of video codecs, which in turn facilitates the future design of new coding algorithms. For example, the H.264 supports video coding with variable block sizes. Besides the typical 16×16 block size, we can partition a block into 16×8 , 8×16 , and 8×8 sub-blocks. If the sample size is 8×8 , the H.264 allows the further partitioning of it into 8×4 , 4×8 , or 4×4 sub-blocks. Hence, the H.264 encoder involves the mode decision of different sub-block sizes, motion estimation algorithms for different sub-blocks, and bitrate control for the possible modes. These require an in depth study of the tradeoff between encoding workload, the resulting bitrate and the quality of the encoded video. Our proposed model can be used as an analytical tool to study the effect of different sub-block sizes in an encoder, because the use of various sub-blocks reflects some kinds of deformation within the target macroblock. This enables us to enhance our understanding of mode selection and eventually facilitates the design of an efficient video encoder. This is a fruitful direction for further research.

APPENDIX

I. DETAILED DERIVATION OF THE PROPOSED COMPOUND COVARIANCE MODEL FOR MOTION PREDICTION ERROR

From (20)

$$\begin{aligned}
 E \left[\rho^{|I-m_x|} \right] &\approx \frac{1}{\sigma_{mv} \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(m_x - |I_x|)^2}{2\sigma_{mv}^2} + |I-m_x| \ln \rho} dm_x \\
 &= \frac{1}{\sigma_{mv} \sqrt{2\pi}} \left[\int_{-\infty}^{|I|} e^{\frac{-(m_x - |I_x|)^2}{2\sigma_{mv}^2} + (|I-m_x|) \ln \rho} dm_x \right. \\
 &\quad \left. + \int_{|I|}^{\infty} e^{\frac{-(m_x - |I_x|)^2}{2\sigma_{mv}^2} + (m - |I|) \ln \rho} dm_x \right]. \quad (20)
 \end{aligned}$$

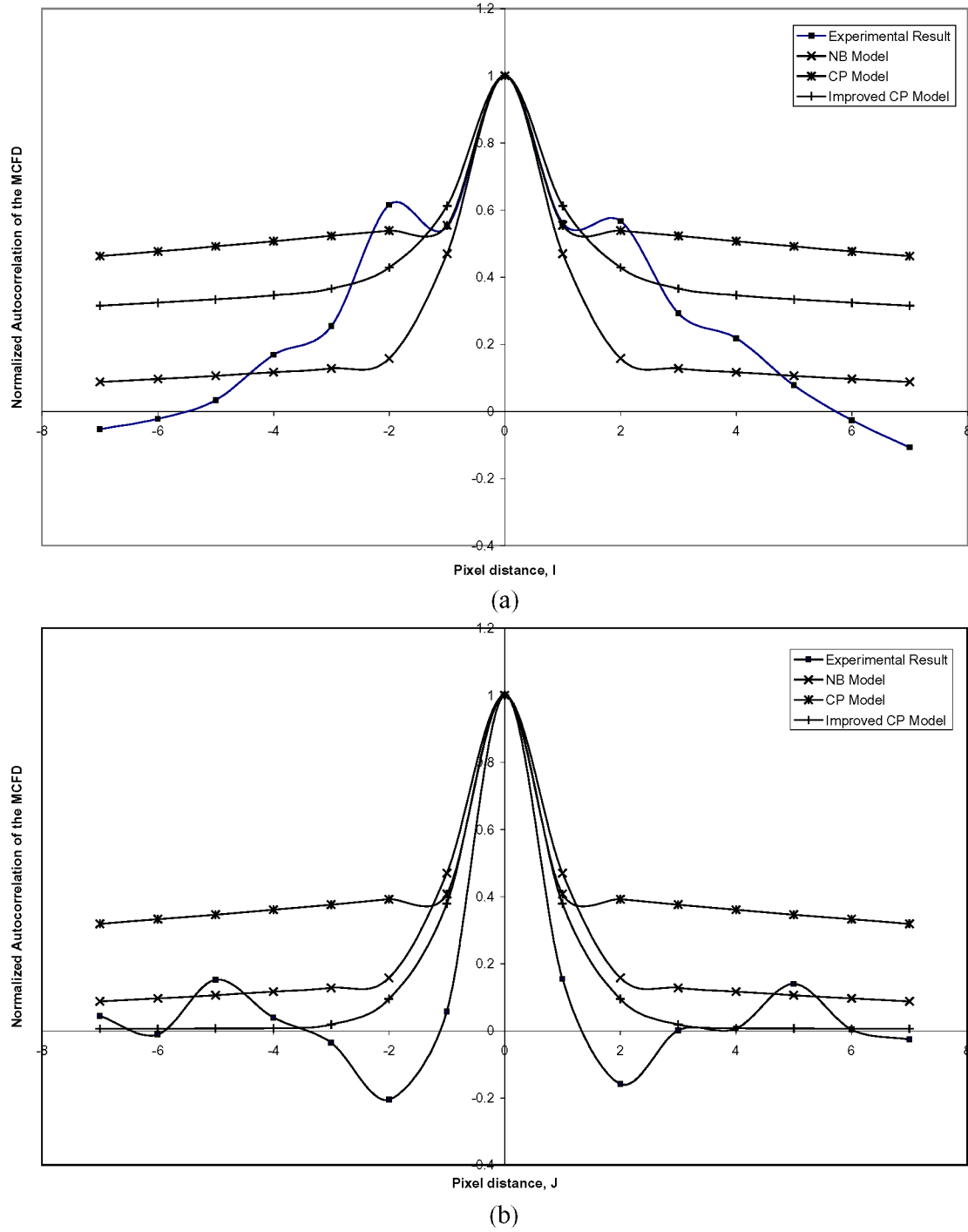


Fig. 12. One-dimensional normalized autocorrelation of the Salesman's MCFD in (a) I-direction and (b) J-direction. (a) 1-D Normalized Autocorrelation of the Salesman's MCFD in I-direction. (b) 1-D Normalized Autocorrelation of the Salesman's MCFD in J-direction.

By completing the square method, the first term of LHS of (20), we have we have the equation shown at the bottom of the next page. Let

$$\zeta = \frac{m_x}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \quad \text{and} \quad = \frac{1}{\sqrt{\pi}} e^{(|I| - |\bar{\mu}_x|) \ln \rho + \frac{(\sigma_{mv} \ln \rho)^2}{2}}$$

$$I' = \frac{|I|}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \quad \times \left(\int_{-\infty}^0 e^{-(\zeta)^2} d\zeta + \int_0^{I'} e^{-(\zeta)^2} d\zeta \right)$$

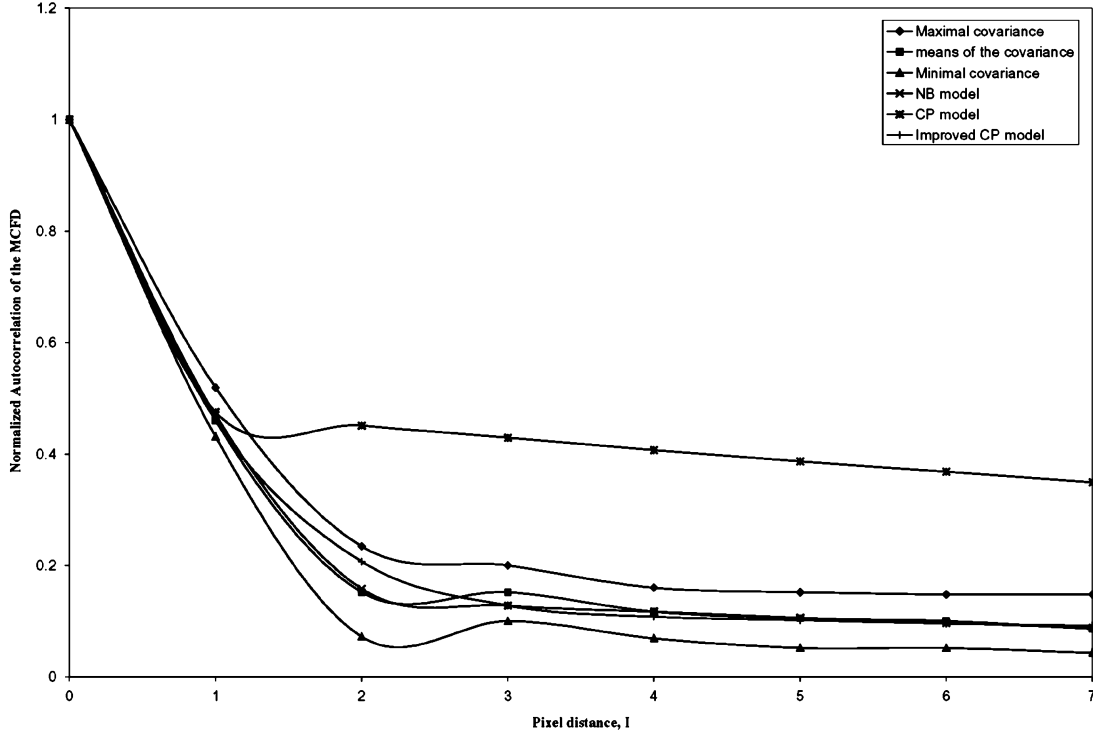


Fig. 13. Comparison of the predicted autocorrelation function using the tested models with experimental results in [9].

$$\begin{aligned} \because \int_{-\infty}^0 e^{-(\zeta)^2} d\zeta &= \frac{\sqrt{\pi}}{2} \quad \text{and} \\ \text{erf}(z) &= \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx \\ &= \frac{1}{2} e^{(|I| - |\bar{\mu}_x|) \ln \rho + \frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left(1 + \text{erf} \left(\frac{|I|}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \right) \right). \end{aligned}$$

Similarly, the second term of LHS of (18) becomes

$$\begin{aligned} &\frac{1}{\sigma_{mv}\sqrt{2\pi}} \left(\int_{|I|}^{\infty} e^{-\frac{(m_x - |\bar{\mu}_x|)^2}{2\sigma_{mv}^2} + (|I| - m_x) \ln \rho} dm_x \right) \\ &= \frac{1}{2} e^{(|\bar{\mu}_x| - |I|) \ln \rho + \frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left[1 - \text{erf} \left(\frac{|I|}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \right) \right]. \end{aligned}$$

Then, we can have

$$\begin{aligned} E \left[\rho^{|I - m_x|} \right] &\approx \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left\{ \rho^{(|I| - |\bar{\mu}_x|)} \left[1 + \text{erf} \left(\frac{|I| - |\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho}{\sigma_{mv}\sqrt{2}} \right) \right] \right. \\ &\quad \left. + \rho^{-(|I| - |\bar{\mu}_x|)} \left[1 - \text{erf} \left(\frac{|I| - |\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho}{\sigma_{mv}\sqrt{2}} \right) \right] \right\} \\ &= \rho^{|I|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left\{ \rho^{|\bar{\mu}_x|} \left[1 + \text{erf} \left(\frac{|I| - |\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho}{\sigma_{mv}\sqrt{2}} \right) \right] \right. \\ &\quad \left. + \rho^{-(2|I| - |\bar{\mu}_x|)} \left[1 - \text{erf} \left(\frac{|I| - |\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho}{\sigma_{mv}\sqrt{2}} \right) \right] \right\} \\ &= \rho^{|I|} \tilde{R}(I, \bar{\mu}_x, \sigma_{mv}, \rho) \quad (21) \end{aligned}$$

$$\begin{aligned} &\frac{1}{\sigma_{mv}\sqrt{2\pi}} \left(\int_{-\infty}^{|I|} e^{-\frac{(m_x - |\bar{\mu}_x|)^2}{2\sigma_{mv}^2} + (|I| - m_x) \ln \rho} dm_x \right) \\ &= \frac{1}{\sigma_{mv}\sqrt{2\pi}} \left(\int_{-\infty}^{|I|} e^{-\left(\frac{m_x}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \right)^2 + \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)^2}{2\sigma_{mv}^2} + \frac{-|\bar{\mu}_x|^2 + 2\sigma_{mv}^2 |I| \ln \rho}{2\sigma_{mv}^2}} dm_x \right) \\ &= \frac{1}{\sigma_{mv}\sqrt{2\pi}} e^{(|I| - |\bar{\mu}_x|) \ln \rho + \frac{(\sigma_{mv} \ln \rho)^2}{2}} \left(\int_{-\infty}^{|I|} e^{-\left(\frac{m_x}{\sqrt{2}\sigma_{mv}} - \frac{(|\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho)}{\sqrt{2}\sigma_{mv}} \right)^2} dm_x \right) \end{aligned}$$

TABLE I
I₁-NORM OF ERRORS OF THE INTERESTED MODELS OF THE FOOTBALL SEQUENCE

	NB-Model	CP-Model	Improved CP-Model
I-direction	0.16	0.40	0.27
J-direction	0.05	0.36	0.20

TABLE II
I₁-NORM OF ERRORS OF THE INTERESTED MODELS OF THE SALESMAN SEQUENCE

	NB-Model	CP-Model	Improved CP-Model
I-direction	0.14	0.28	0.20
J-direction	0.15	0.33	0.10

where

$$\begin{aligned} \tilde{R}(\mathbf{I}, \bar{\mu}_x, \sigma_{mv}, \rho) &= \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ \rho^{|\bar{\mu}_x|} \left[1 + \operatorname{erf} \left(\frac{|\mathbf{I}| - |\bar{\mu}_x| + \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] \right. \\ &\quad \left. + \rho^{-(2|\mathbf{I}| - |\bar{\mu}_x|)} \left[1 - \operatorname{erf} \left(\frac{|\mathbf{I}| - |\bar{\mu}_x| - \sigma_{mv}^2 \ln \rho}{\sigma_{mv} \sqrt{2}} \right) \right] \right\}. \end{aligned}$$

II. TO DEDUCE THE COMPOUND COVARIANCE MODEL, CP MODEL: (1) FROM (22)

To derive (1), we assume that $\bar{\mu}_x$ and σ_{mv} are equal to 0 and 0.5, respectively. Using the properties of $\operatorname{erf}(z)$, (21) is approximated as

$$\begin{aligned} E \left[\rho^{|\mathbf{I} - m_x|} \right] &\approx \rho^{|\mathbf{I}|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left\{ \left[1 + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \delta(|\mathbf{I}|) + (1 - \delta(|\mathbf{I}|)) \right] \right. \\ &\quad \left. + \rho^{-2|\mathbf{I}|} \left[1 - \frac{2}{\sqrt{\pi}} e^{-\left(\frac{-\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \right. \right. \\ &\quad \left. \left. \cdot \frac{-\sigma_{mv} \ln \rho}{\sqrt{2}} \delta(|\mathbf{I}|) - (1 - \delta(|\mathbf{I}|)) \right] \right\}. \end{aligned}$$

The second middle bracket term in the braces

$$\begin{aligned} &\rho^{-2|\mathbf{I}|} \left[1 + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \cdot \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \delta(|\mathbf{I}|) - (1 - \delta(|\mathbf{I}|)) \right] \\ &= 0, \quad \text{if } \mathbf{I} \neq 0 \end{aligned}$$

if $\mathbf{I} = 0$

$$\begin{aligned} &E \left[\rho^{|\mathbf{I} - m_x|} \right] \\ &\approx \rho^{|\mathbf{I}|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ \left[1 + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \right] \right. \\ &\quad \left. + \left[1 + \frac{2}{\sqrt{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \cdot \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \right] \right\} \\ &= \rho^{|\mathbf{I}|} \frac{1}{2} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \left\{ 2 + \frac{4}{\sqrt{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \frac{\sigma_{mv} \ln \rho}{\sqrt{2}} \right\} \end{aligned}$$

Hence, for $\mathbf{I} \geq 0$

$$\begin{aligned} E \left[\rho^{|\mathbf{I} - m_x|} \right] &= \rho^{|\mathbf{I}|} e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} \\ &\times \left[1 + \sqrt{\frac{2}{\pi}} e^{-\left(\frac{\sigma_{mv} \ln \rho}{\sqrt{2}}\right)^2} \sigma_{mv} \ln \rho \cdot \delta(|\mathbf{I}|) \right]. \quad (25) \end{aligned}$$

Let $\beta = \sqrt{(2/\pi)} e^{-((\sigma_{mv} \ln \rho / \sqrt{2})^2)} \sigma_{mv} \ln \rho$, and substitute it into (25). Now, we have $\tilde{R}(\mathbf{I}, \bar{\mu}_x, \sigma_{mv}, \rho) = e^{((\sigma_{mv} \ln \rho)^2 / 2)} [1 + \beta \cdot \delta(|\mathbf{I}|)]$ and let us substitute it into

$$\begin{aligned} \frac{E[C_e(\mathbf{I}, 0)]}{E[C_e(0, 0)]} &= \tilde{C}_{e, \bar{\mu}_x, \sigma_{mv}}(\mathbf{I}, 0) = \rho^{|\mathbf{I}|} \rho^{|\mathbf{0}|} \frac{1 - \tilde{R}(\mathbf{I}, \bar{\mu}_x, \sigma_{mv}, \rho) \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho)}{1 - \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho) \tilde{R}(0, \bar{\mu}_x, \sigma_{mv}, \rho)} \\ \tilde{C}_{e, \bar{\mu}_x, \sigma_{mv}}(\mathbf{I}, 0) &= \rho^{|\mathbf{I}|} \frac{1 - \left[e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} [1 + \beta \cdot \delta(|\mathbf{I}|)] \right] \left[e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} [1 + \beta] \right]}{1 - \left[e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} [1 + \beta] \right] \left[e^{\frac{(\sigma_{mv} \ln \rho)^2}{2}} [1 + \beta] \right]} \\ &= \rho^{|\mathbf{I}|} \frac{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta \delta(|\mathbf{I}|)] \cdot [1 + \beta]}{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]^2} \\ &= \rho^{|\mathbf{I}|} \frac{1 - e^{(\sigma_{mv} \ln \rho)^2} (1 + \beta) - e^{(\sigma_{mv} \ln \rho)^2} \{\beta \delta(|\mathbf{I}|) + \beta^2 \delta(|\mathbf{I}|)\}}{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]^2} \\ &= \rho^{|\mathbf{I}|} \frac{1 - e^{(\sigma_{mv} \ln \rho)^2} (1 + \beta) + (1 - 1) \delta(|\mathbf{I}|) - e^{(\sigma_{mv} \ln \rho)^2} \{(1 + \beta)^2 - (1 + \beta)\} \delta(|\mathbf{I}|)}{1 - e^{(\sigma_{mv} \ln \rho)^2} [1 + \beta]^2} \end{aligned}$$

(22), we have the equation shown at the bottom of the previous page. As a result

$$\begin{aligned} \tilde{C}_{e, \bar{\mu}_x, \sigma_{mv}}(I, 0) &= \rho^{|I|} \left\{ \frac{1 - e^{(\sigma_{mv} \ln \rho)^2 (1 + \beta)}}{1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]^2}} \right. \\ &\quad + \frac{[1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]^2}] \delta(|I|)}{1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]^2}} \\ &\quad \left. - \frac{[1 - e^{(\sigma_{mv} \ln \rho)^2 (1 + \beta)}] \delta(|I|)}{1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]^2}} \right\} \\ &= \rho^{|I|} [A + (1 - A)\delta(|I|)] \end{aligned} \quad (28)$$

where

$$A = \frac{1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]}}{1 - e^{(\sigma_{mv} \ln \rho)^2 [1 + \beta]^2}}. \quad (29)$$

The numerical value of A is equal to 0.497 with $\rho = 0.95$ and $\sigma_{mv} = 0.5$. The CP Model is then obtained.

REFERENCES

- [1] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 7, pp. 1140–1154, Aug. 1987.
- [2] B. Girod, "Why B-pictures work: A theory of multi-hypothesis motion-compensated prediction," in *Proc. Int. Conf. Image Processing*, Oct. 1998, vol. 2, pp. 213–217.
- [3] B. Girod, "Motion-compensation prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 604–612, Apr. 1993.
- [4] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding," in *Proc. Int. Symp. Circuits and Systems*, May 1992, vol. 1, pp. 184–187.
- [5] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 693–699, May 1994.
- [6] B. Tao and M. T. Orchard, "A parametric solution optimal overlapped block motion compensation," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 341–350, Mar. 2001.
- [7] W. Zheng, Y. Shishikui, M. Naemura, Y. Kanatsugu, and S. Itoh, "Analysis of space-dependent characteristics of motion-compensated frame differences based on a statistical motion distribution model," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 377–386, Apr. 2002.
- [8] C. F. Chen and K. K. Pang, "The optimal transform of motion-compensated frame difference images in a hybrid coder," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 40, no. 6, pp. 393–397, Jun. 1993.
- [9] W. Niehsen and M. Brüning, "Covariance analysis of motion-compensated frame difference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 536–539, Jun. 1999.

- [10] B. Tao and M. T. Orchard, "Prediction of second-order statistics in motion-compensated video coding," in *Proc. Int. Conf. Image Processing*, Oct. 1998, vol. 3, pp. 910–914.
- [11] C. F. Chen and K. K. Pang, "Hybrid coders with motion compensation," *Multidimensional Syst. Signal Process.*, vol. 3, no. 3, pp. 241–266, 1992.
- [12] Y. L. Lee and H. W. Park, "Loop filtering and post-filtering for low-bit-rates moving picture coding," *Signal Process. Image Commun.*, vol. 16, pp. 871–890, 2001.
- [13] A. K. Jain, "A fast Karhunen–Loeve transform for a class of random processes," *IEEE Trans. Commun.*, vol. COM-24, no. 5, pp. 1023–1029, Sep. 1976.

Ko-Cheung Hui received the B.Sc. degree (with first class Honors) in engineering physics in 1997 from The Hong Kong Polytechnic University, the M.Sc. degree in physics from the Chinese University in 1999, and the Ph.D. degree in electronic and information engineering from The Hong Kong Polytechnic University in 2005.

He is currently a Senior Application Engineer with the NXP Semiconductor, Ltd. His research interests include video coding technology, imaging, and digital camera.

Wan-Chi Siu (S'77–M'77–SM'90) received the Associateship from The Hong Kong Polytechnic University in 1975, the M.Phil. degree from The Chinese University of Hong Kong in 1977, and the Ph.D. degree from Imperial College of Science, Technology, and Medicine, London, U.K., in October 1984.

He was with The Chinese University of Hong Kong as a Tutor and later as an Engineer between 1975 and 1980. He then joined The Hong Kong Polytechnic University as a Lecturer in 1980. He was promoted to Senior Lecturer, Principle Lecturer, and Reader in 1985, 1987, and 1990, respectively, and has been Chair Professor of the Department of Electronic and Information Engineering since 1992. He was Head of Department of Electronic and Information Engineering Department and subsequently Dean of Engineering Faculty between 1994 and 2002. He is currently the Director of the Centre for Multimedia Signal Processing of the same university. He has published 300 research papers, over 130 of which appeared in international journals, such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and he is an Editor of the recent book *Multimedia Information Retrieval and Management* (Springer, 2003). His research interests include digital signal processing, fast computational algorithms, transforms, wavelets, image and video coding, and computational aspects of pattern recognition.

Prof. Siu was a Guest Editor, Associate Editor, and Editorial Board Member of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: ANALOG AND DIGITAL SIGNAL PROCESSING, the IEEE TRANSACTIONS ON PATTERN RECOGNITION, the *Journal of VLSI Signal Processing Systems for Signal, Image, Video Technology*, the *EURASIP Journal on Applied Signal Processing*, in addition to other journals. He has been a keynote speaker for many international conferences, including IEEE PCM-2002 (Taiwan, R.O.C.) and the IEEE 2003 ICNNSP (Nanjing, China). He has held the position of General Chair or Technical Program Chair of many international conferences, including IEEE Society sponsored flagship conferences such as ISCAS 1997, ICASSP 2003, and ICIP 2010 (to be held in Hong Kong). Between 1991 and 1995, he was a member of the Physical Sciences and Engineering Panel of the Research Grants Council (RGC), Hong Kong Government, and, in 1994, he chaired the first Engineering and Information Technology Panel of the Research Assessment Exercise (RAE) to assess the research quality of 19 Cost Centers (departments) from all universities in Hong Kong. He has received many awards, including the Distinguished Presenter Award (1997), the IEEE Third Millennium Medal (2000), the Best Teacher Award (2003), the Outstanding Award in Research (2003), the Plaque for Exceptional Leadership from IEEE SPCB (2003), and the Honorable Mention Winner Award from *Pattern Recognition* (2004).