

A Performance Comparison of Block-Based Matching Cost Evaluation Models for FRUC Techniques

Jin-soo Kim, Jae-Gon Kim, *Member, KIMICS*

Abstract— DVC (Distributed Video Coding) and FRUC (Frame Rate Up Conversion) techniques need to have an efficient motion compensated frame interpolation algorithms. Conventional works of these applications have mainly focused on the performance improvement of overall system. But, in some applications, it is necessary to evaluate how well the MCI (Motion Compensated Interpolation) frame matches the original frame. For this aim, this paper deals with the modeling methods for evaluating the block-based matching cost. First, several matching criteria, which have already been dealt with the motion compensated frame interpolation, are introduced and then combined to make estimate models for the size of MSE (Mean Square Error) noise of the MCI frame to original one. Through computer simulations, it is shown that the block-based matching criteria are evaluated and the proposed model can be effectively used for estimating the MSE noise.

Index Terms— Motion Compensated Interpolation, Block-based Evaluation, Matching Cost, FRUC

I. INTRODUCTION

VIDEO frame rate up-conversion (FRUC) technique has been a technique of great interest due to its diversified consumer applications such as HDTV and multimedia environments [1], [2]. Besides these scanning format applications, this technique has also been considered in low bandwidth video coding. In low bandwidth applications, some frames are skipped in the encoding stage and missing frames are interpolated during the decoding process. Similarly to these applications, distributed video coding (DVC) schemes need to have good side information (SI), MCI (Motion Compensated Interpolation) frame which plays great role in determining the performance of overall system and it is necessary to be efficiently generated by using key frames at decoder [3], [4]. However, although theoretical R-D performance bounds of distributed video coding have been studied based on the information theory, the performances of DVC algorithms are still inferior to those of traditional

video coders, such as H.264, which use a hybrid of motion compensated prediction and transform coding. As one method to improve the performance of DVC systems, several attempts have been made to generate higher quality MCI frame.

Many MCI algorithms have been developed, which are divided broadly into two approaches. The first approach exploits the motion vector, derived from the reference frames, to interpolate the block with same spatial location for the interpolated frame. Ascenso et al [5] proposed a spatial motion smoothing algorithm, which is useful for interpolating frames with symmetric and linear motions. Peixoto et al [6] proposed a linear motion mode to produce SI associated with, and a way to deal with object-based motion search. However, this method introduces the overlapped and hole regions. This approach interpolates new block information at the half value of motion vector. Specially, this method has been considered in effectively interpolating the frames with non-linear motion.

Since the conventional DVC schemes are still inferior to those of traditional video coders, new attempts have been made to improve the overall system performances. Park et al [7] proposed a new DVC scheme with channel division scheme, which classifies the blocks in the SI into reliable ones and unreliable ones. In this scheme, parity bits are transmitted only for unreliable blocks, and thus achieving the better coding gain. Kim et al [8] proposed a novel DVC scheme suitable for some application areas where source statistics such as motion information can be provided to the encoder side from the decoder side. That is, a block-based selective encoding scheme was proposed to improve the coding efficiency by using the feedback of motion information derived in SI generation at the decoder side. But, in order to improve the performance in these applications, it is necessary to evaluate how well the MCI frame matches the original frame. So, this paper deals with the modeling methods for evaluating the block-based matching cost. First, several matching criteria are discussed and then combined to make estimate models for the MSE (Mean Square Error) noise of the MCI frame to original one.

This paper is organized as follows. In section II, the basic concept of conventional FRUC schemes are reviewed and symmetric motion estimation is discussed [9]. In section III, the block-based matching cost evaluation criteria are analyzed for the MCI frame. In section IV, the estimate models for the MSE noise of the MCI frame are presented and then, the experimental

Manuscript received October 17, 2011; revised November 25, 2011; accepted November 30, 2011.

Jin-soo Kim is with the School of Information Communication and Computer Engineering, Hanbat National University, Daejeon, 305-719, Korea (Email: jskim67@hanbat.ac.kr)

Jin-Gon Kim is with the School of Elec. Telecom. and Computer Engineering, Korea Aerospace Univ., Goyang-city, Gyeonggi-do, 412-791, Korea (Email: jgkim@kau.ac.kr)

results are shown. Finally, section V concludes this paper.

II. FRAME RATE UP CONVERSION AND SYMMETRIC MOTION ESTIMATION

Fig. 1 shows the overview of motion-compensated frame interpolation between two adjacent frames [9]. Conventional motion compensated interpolation schemes produce overlapped pixels and holes in the interpolated frame. This is mainly caused by two reasons. First, although the true motion trajectory is available for a moving object, the object is usually not under rigid translational motion. This often occurs in sequences with camera motion such as interframe zooming. Second, even if the object is under rigid motion, the estimated motion vector field may not be the same within the same object due to poor motion estimation. In either case, the motion trajectory is not a one-to-one mapping from moving object in the previous frame to that in the current frame. Therefore, the interpolated object tends to contain some overlapped pixels and holes. Specifically, for these areas, it is necessary to measure the matching characteristics based on the generated MCI frame. Thus, this paper evaluates each block of MCI frame based on temporal and spatial information as well as the size of motion information, which can be effectively measured in the MCI frame.

Usually, a motion vector is found using the concept of the sum of absolute difference (SAD), which represents the sum of the absolute difference for pixels between the candidate block in the previous frame I_{n-1} and the reference block in the current frame I_{n+1} as follows.

$$SAD(dx, dy) = \sum_{x \in S_x} \sum_{y \in S_y} |I_{n+1}(p) - I_{n-1}(p+v)|$$

$$v = \arg \min_{(dx, dy)} \{SAD(dx, dy)\} \quad (1)$$

where $v=(d_x, d_y)$ stands for the motion vector candidate and (S_x, S_y) denotes the horizontal and vertical search range. v denotes the selected motion vector, which locates the position of the block with the minimum SAD.

In this work, the symmetric motion estimation is considered to find motion vectors, as already described in [5] and [7]. By modifying the bilateral motion estimation, the block-centered symmetric motion estimation is used [8] [9]. By introducing the basic concept of [8], as shown in Fig. 2, in order that the motion vector of the co-located block in Y_n can reflect regular and irregular motions, both forward and backward motion vectors are designed to point to the opposite directions and also to be symmetric on the center of the block. Thus, the problem is to find the motion vector v^* such that minimizes the sum

of the forward sum of absolute difference (SAD) and the backward SAD and the block-centered bi-directional SAD, as expressed by

$$v^* = \arg \min_v [SAD_f(v) + SAD_b(v) + SAD_m(v)]$$

$$SAD_f(v) = \sum_{p \in M} |I_{n-1}(p) - I_{n+1}(p+v)| \quad (2)$$

$$SAD_b(v) = \sum_{p \in M} |I_{n+1}(p) - I_{n-1}(p-v)|$$

$$SAD_m(v) = \sum_{p \in M} |I_{n+1}(p+v/2) - I_{n-1}(p-v/2)|$$

where p denotes a pixel coordinate in the given block M . Based on the assumption that the current block experiences a constant motion between I_{n-1} and I_{n+1} , the side information is generated by

$$Y_n(p) = \frac{I_{n-1}(p-v/2) + I_{n+1}(p+v/2)}{2} \quad (3)$$

for $p \in M$.

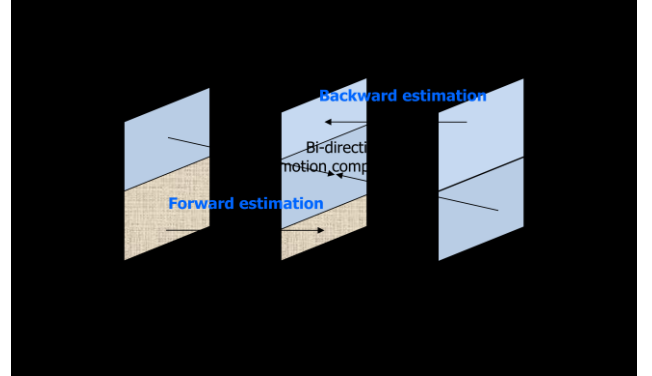


Fig. 1. Overview of motion-compensated frame interpolation between two adjacent frames [9].

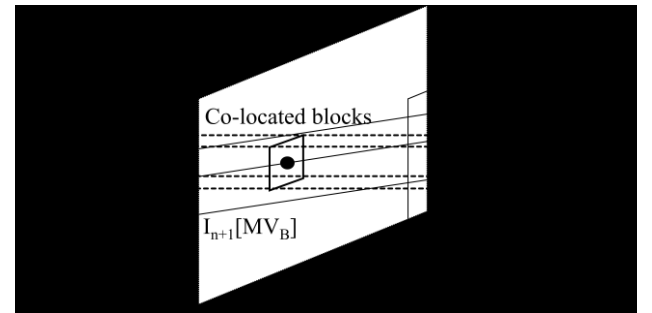


Fig. 2. Motion-compensated frame interpolation between two adjacent frames [8].

III. BLOCK-BASED MATCHING COST EVALUATION CRITERIA

In order to improve the performance of MCI frame, as shown in Fig.2, it is necessary to effectively estimate the distortion of each block in the side information $Y_n(p)$. In this paper, four criteria are tested for measuring the distortion of each block. First, the temporal matching criterion is defined by calculating the absolute difference between the forward predictive block and the backward predictive block as follows.

$$C_t = \frac{1}{N_M} \sum_{p \in M} (I_{n+1}(p - v/2) - I_{n-1}(p + v/2))^2 \quad (4)$$

where N_M is the number of pixels in the given block M . v is the motion vector, which is defined and used in (2) and (3), respectively. This criterion measures the temporal matching cost between the forward predictive block and the backward predictive block, assuming that the given block experiences a linear motion between I_{n-1} and I_{n+1} .

Second, the spatial matching criterion is defined by calculating the square difference of neighbor pixels. That is, the pixels of the interpolated block should be compared to those of the neighbor blocks and need to be smoothed. Therefore, the spatial matching criterion is defined as follows.

$$C_s = \frac{1}{N_s} \sum_{p \in M_s} \{Y_n(p) - Y_n(n(p))\}^2 \quad (5)$$

where M_s is the set of boundary pixels in the given block M and N_s is the number of the boundary pixels. $n(p)$ denotes the neighboring pixel of p in the adjacent block. This criterion reflects the smoothness between neighboring blocks.

Third, if the current motion vector is similar to the motion vectors of neighboring blocks, the matching cost of the current block is small. So, the minimal difference of the current motion vector to the nearest neighboring motion vectors checks the reliability. The motion consistency is measured as follows.

$$C_v = \min \{ \|v - w\|^2 : w \in W \} \quad (6)$$

where W is the set of motion vectors in the nearest neighboring blocks.

Fourth, if the current block contains high-frequency contents, it may be difficult to find an exact matching block. In this case, the high frequency components may result in the mismatch of the given block. So, in order to measure how much high frequency contains, a new measure of the high frequency contents is introduced as follows.

$$C_f = \frac{N_M^2 - 1}{N_M^2} \left(\sum_{i=0}^{N_M-1} \sum_{j=0}^{N_M-1} \sqrt{(i+1)(j+1)} (|f(i,j)| - f(0,0)) \right) \quad (7)$$

where N_M is the block size and $f(i,j)$ is the DCT coefficient at the (i, j) -th location. (7) means that the higher the DCT coefficient is, the larger the weighting is imposed.

IV. ESTIMATE MODELS AND EXPERIMENTAL RESULTS

A. Normalized Cross Correlation

The matching criteria defined in Section III need to be compared to the real MSE noise of MCI frame. The normalized cross correlation (NCC) between x and y with mean ζ_x and ζ_y is used.

$$NCC_{xy} = \frac{\sum_i (x_i - \zeta_x)(y_i - \zeta_y)}{\left[\sum_i (x_i - \zeta_x)^2 \right]^{1/2} \left[\sum_i (y_i - \zeta_y)^2 \right]^{1/2}} \quad (8)$$

Fig. 3 shows the experimental results of the NCC between the real MSE and the criteria defined in Section III. The block size (N_M) was 8x8 pixels. The real MSE of the MCI frame is measured and the four criteria defined in Section are compared. In these results, the temporal matching criterion (mse_{C_t}), defined in (4), shows the best performance of the four criteria and is dominant, while the spatial matching criterion (mse_{C_s}), defined in (5), is independent of the MSE characteristics of the given sequences. In some conventional studies [7][8], the spatial criterion was used as one of the cost evaluation measures in the motion compensated interpolation frame. But, these results show that the spatial matching criterion is not efficient. The motion consistency (mse_{C_m}) and the high frequency characteristics (mse_{C_f}) are also correlated with the real MSE. In this paper, in order to find a proper model, which describes well the characteristics of the MSE for the motion compensated interpolation block, three estimate models are designed by using the criteria, that is, C_t , C_m and C_f .

B. Estimate Models

Based on the experimental results denoted in Fig. 3, three different estimate models are designed. Simply, since the temporal matching criterion is dominant, first, one-parameter model is designed as

$$\hat{\sigma}^2 = \alpha \cdot C_t + \varepsilon \quad (9)$$

where $\hat{\sigma}^2$ denotes the estimating value for the mean square error of the MCI frame and α is the proportional

factor between the MSE and the temporal matching cost and ε is the white Gaussian noise with zero mean. In this case, the proportional factor is obtained by

$$\alpha = \frac{\sum_i \sigma_i^2 \cdot C_{t,i}}{\sum_i \sigma_i^4} \quad (10)$$

Second, since the motion consistency plays important role on the low-motion sequences like Salesman, so a two-parameter model by introducing the C_t and C_m as follows.

$$\hat{\sigma}^2 = \alpha \cdot C_t + \beta \cdot C_m + \varepsilon \quad (11)$$

The proportional factors α and β are obtained by

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_i C_{t,i}^2 & \sum_i C_{m,i} C_{t,i} \\ \sum_i C_{t,i} C_{m,i} & \sum_i C_{m,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i \sigma_i^2 C_{t,i} \\ \sum_i \sigma_i^2 C_{m,i} \end{bmatrix} \quad (12)$$

Third, a three-parameter model considering the high frequency coefficients may be defined as

$$\hat{\sigma}^2 = \alpha \cdot C_t + \beta \cdot C_m + \gamma \cdot C_f + \varepsilon \quad (13)$$

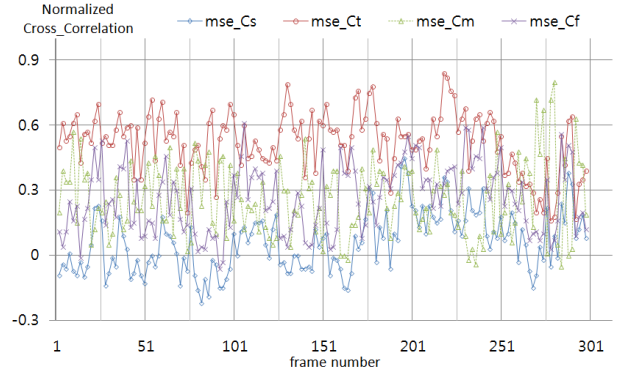
The model parameters α , β and γ are determined by

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \sum_i C_{t,i}^2 & \sum_i C_{m,i} C_{t,i} & \sum_i C_{f,i} C_{t,i} \\ \sum_i C_{t,i} C_{m,i} & \sum_i C_{m,i}^2 & \sum_i C_{f,i} C_{m,i} \\ \sum_i C_{t,i} C_{f,i} & \sum_i C_{m,i} C_{f,i} & \sum_i C_{f,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i \sigma_i^2 \cdot C_{t,i} \\ \sum_i \sigma_i^2 \cdot C_{m,i} \\ \sum_i \sigma_i^2 \cdot C_{f,i} \end{bmatrix} \quad (14)$$

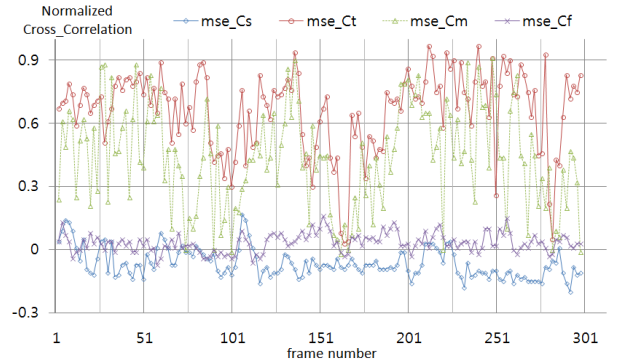
C. Experimental Results

The three models are tested through computer simulations by using two QCIF sequences (Foreman, Salesman). Simulation results are shown in Fig. 4. *real_MSE* denotes the real MSE value, which is obtained by comparing original frame with the MIC frame. *1-para model*, *2-para model* and *3-para model* represent the models corresponding to (9), (11), and (13), respectively. Foreman sequence is temporally active and so the MCI frames provide low visual qualities. In this case, 2-parameter model and 3-parameter model are effective, specially, in the highly active frames. On the contrary, Salesman sequence has temporally low details. So, it is easy to find a highly correlated information based on the motion compensated interpolation technique. In the relatively high MSE regions of this sequence, 2-parameter model and 3-parameter model show very similar performances. Their performances are better than that of 1-parameter model. Also, it is shown that in low active

frames, 1-parameter tends to overestimates the MSE values. From these experiments, it is shown that these linear models are effective in estimating the noise of the MCI frame.



(a) Foreman sequence



(b) Salesman sequence

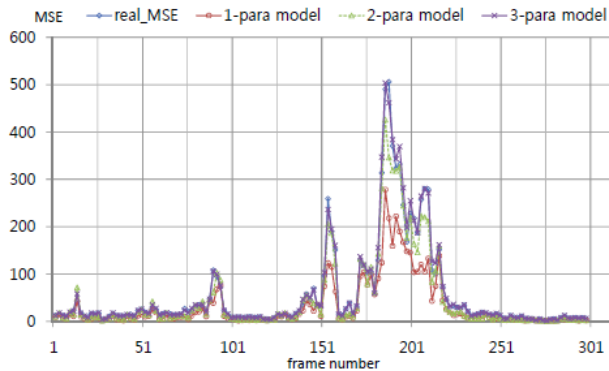
Fig. 3. Experimental results for cross-correlation.

V. CONCLUSIONS

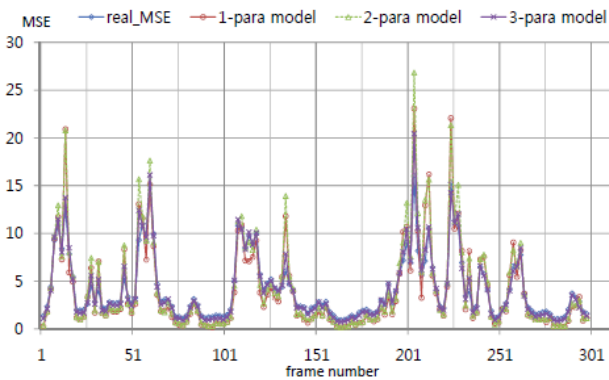
This paper dealt with the block-based matching cost evaluation models for the MCI frames. Several conventional matching criteria were introduced and evaluated through computer simulations. Contrary to the conventional works, it was shown that the spatial matching criterion was not useful for estimating the MSE noise of the MCI frame. But, it was observed that the temporal matching criterion and the motion consistency matching criterion were relatively effective. Furthermore, this paper presented three estimate models for estimating the MSE noise of the MCI frame and by experimental results, it was shown that the combining model with the temporal and the motion consistency criteria is very effective and can be effectively used for estimating the MSE noise of the MCI frame.

It is expected that the proposed modeling can be efficiently modified and is applicable for FRUC technique and DVC schemes. For practical implementations, we are

analyzing the matching cost criteria and also we are finding how to implement the real model parameter.



(a) Foreman sequence



(b) Salesman sequence

Fig. 4. Experimental results of three different models for estimating the MSE of MCI frame.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2011-0026730, No. 2011-0023182)

REFERENCES

- [1] R. Castagno, P. Haavisto, and G. Ramponi, "A Method for Motion Adaptive Frame Rate Up Conversion," *IEEE Trans. Circuits Systems for Video Technology*, vol. 6, no. 5, pp. 436-446, Oct. 1996.
- [2] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate," *IEEE Trans. Broadcasting*, Vol.33, No. 1, pp.441-446, Mar. 2007
- [3] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. on Information Theory* 19, pp.

471-480, July 1973.

- [4] A. Wyner and J. Ziv, "The Rate-distortion Function for Source Coding with Side Information at the Decoder", *IEEE Trans. on Information Theory* 22, pp. 1-10, Jan. 1976.
- [5] J. Ascenso, C. Brites, and F. Pererira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," *Proc. EURASIP Conf. Speech and Image Processing*, July 2005. Pp. 311-316
- [6] E. Peixoto, R. L. Queiroz and D. Mukherjee, "On Side Information Generation for Wyner-Ziv Video Coding," *Proc. of Simposio Brasileiro de Tel., Rio de Janeiro, Brazil*, Sept. 2008.
- [7] Sang-Uk Park, Jin-Woo Choi, Chang-Su Kim, Sang-Uk Lee and Jung-Won Kang, "Efficient Distributed Video Coding Using Symmetric Motion Estimation and Channel Division," *PACRIM09*, Aug. 2009.
- [8] Jin-soo Kim, Jae-Gon Kim and Kwang-deok Seo, "A Selective Block Encoding Scheme Based on Motion Information Feedback in Distributed Video Coding," *IEICE Transactions on Communications*, Vol. E94-B, No.3, pp.860-862, March 2011
- [9] Jin-soo Kim, Jae-Gon Kim, "An Adaptive Occluded Region Detection and Interpolation for Robust Frame Rate Up-Conversion," *International Journal of KIMICS*, Vol.8, No. 2, pp.201-206, March 2011.



Jin-soo Kim received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1991, and the MS and Ph.D. degrees in electrical engineering from KAIST, Korea, in 1993 and 1998, respectively. From 1998 to 2000, he was with the Business Division of System LSI at Samsung Electronics, where he was involved in the development of MCU chipsets. Since March 2000, he has been a faculty member in the

School of Information Communication and Computer Engineering, Hanbat National University, Korea, where he is a professor. His research interests include distributed video coding (DVC), high efficiency video coding, networked video rate shaping and adaptation and media convergence.



Jae-Gon Kim received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1990, and the MS and PhD degrees in electrical engineering from KAIST, Korea, in 1992 and 2005, respectively. From 1992 to 2007, he was with Electronics and Telecommunications Research Institute (ETRI), where he was involved in the development of digital broadcasting media services, MPEG-4/7/21

standards and related applications, and convergence media technologies. From 2001 to 2002, he was a staff associate at the Department of Electrical Engineering, Columbia University, New York. He is currently an associate professor in School of Electronics, Telecommunications and Computer Engineering, Korea Aerospace University, Korea. His research interests include distributed video coding, video adaptation, media convergence, and multimedia applications.