

# An Adaptive Occluded Region Detection and Interpolation for Robust Frame Rate Up-Conversion

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

**Abstract**— FRUC (Frame Rate Up-Conversion) technique needs an effective frame interpolation algorithm using motion information between adjacent neighboring frames. In order to have good visual qualities in the interpolated frames, it is necessary to develop an effective detection and interpolation algorithms for occluded regions. For this aim, this paper proposes an effective occluded region detection algorithm through the adaptive forward and backward motion searches and also by introducing the minimum value of normalized cross-correlation coefficient (NCCC). That is, the proposed scheme looks for the location with the minimum sum of absolute differences (SAD) and this value is compared to that of the location with the maximum value of NCCC based on the statistics of those relations. And, these results are compared with the size of motion vector and then the proposed algorithm decides whether the given block is the occluded region or not. Furthermore, once the occluded regions are classified, then this paper proposes an adaptive interpolation algorithm for occluded regions, which still exist in the merged frame, by using the neighboring pixel information and the available data in the occluded block. Computer simulations show that the proposed algorithm can effectively classify the occluded region, compared to the conventional SAD-based method and the performance of the proposed interpolation algorithm has better PSNR than the conventional algorithms.

**Index Terms**— Frame Rate Up-Conversion (FRUC), Normalized Cross-Correlation Coefficient (NCCC), Occluded Region.

## I. INTRODUCTION

VIDEO frame rate up-conversion (FRUC) 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, FRUC 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 [3], [4]. Similarly to these applications, distributed video coding

(DVC) schemes need to have good side information 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 [5]-[7].

Many FRUC algorithms have been developed, which are divided broadly into two categories. The first approach exploits the motion vector, derived from the reference frames, to interpolate the block with same spatial location for the interpolated frame [7]. This is simple and useful for interpolating frames with symmetric and linear motions. However, this algorithm tends to have severe visual degradation in some regions with non-linear motions. On the other hand, the second 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. However, this method introduces the overlapped (multi-passing of motion trajectories) and hole (no motion trajectory is passing) regions. To handle the occluded regions, first of all, we need to find an effective detection algorithm for these regions. Furthermore, this algorithm needs to have an efficient interpolation scheme for these occluded regions and overlapped areas.

In order to achieve the performance improvement in interpolated frames, in this paper, by developing the adaptive forward and backward motion searches and also by introducing new measurement parameters, a new occluded region detection algorithm is proposed. That is, when bi-directional motion estimation is performed with adjacent reference frames to construct motion vector field for the frame to be interpolated, the normalized cross correlation coefficient (NCCC) as well as the sum of absolute difference (SAD) is measured, simultaneously and compared in terms of SAD and the size of motion vector. Then the proposed algorithm decides whether the given block is the occluded region or not. Once the occluded regions are classified, an adaptive interpolation algorithm for occluded regions is designed by using the neighboring pixel information and the available data in the occluded block.

This paper is organized as follows. In section II, the basic concepts of conventional motion-compensated frame rate up conversion algorithm and its problems are discussed. In section III, bi-directional motion estimation and measurements are presented. In section IV and V, the occluded region detection algorithm and the adaptive interpolation algorithm are proposed, respectively and then experimental results are given. Finally, section VI concludes this paper.

Manuscript received March 7, 2011; revised March 30, 2011; accepted April 12, 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)

## II. MOTION-COMPENSATED FRAME INTERPOLATION AND OCCLUDED REGIONS

Fig.1 shows the conceptual diagram of motion-compensated frame interpolation. In this technique, a segmentation mask is created between two successive frames. This mask is divided into stationary background, moving objects, and covered and uncovered regions. Then, based on the segmentation information, an appropriate interpolation method is determined and the pixel value of the interpolated frame is predicted. Moving objects can be reconstructed by bi-directional motion-compensated interpolation and static background is easily generated by linear interpolation. Also, covered background and uncovered background can be reconstructed by forward prediction and backward prediction, respectively. But, the motion compensated interpolation requires accurate motion estimation for the true motion vector field. Although the motion vector reflects the true object motion more closely, if object motion is nonlinear, the block cannot be perfectly reconstructed.

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 have good performances. Thus, this paper focuses on the FRUC scheme which generates new block information at the half value of motion vector. This method has been considered in effectively interpolating the frames with non-linear motion. However, this method introduces the overlapped and hole (no motion trajectory is passing) regions.

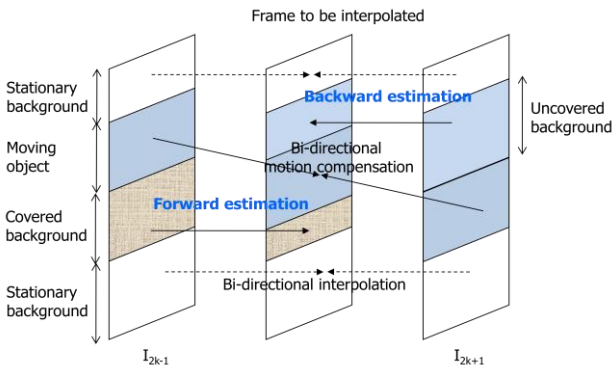


Fig. 1. Overview of motion-compensated frame interpolation between two adjacent frames.

## III. BI-DIRECTIONAL BLOCK SEARCH AND MEASUREMENTS

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 and the reference block in the current frame as follows.

$$SAD(dx, dy) = \sum_{x \in S_x} \sum_{y \in S_y} |I_{2k}(x - dx, y - dy) - I_{2k-1}(x, y)|$$

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

where  $(d_x, d_y)$  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.

By introducing the SAD, block-based motion search is performed bi-directionally and the motion-compensated interpolation is obtained at the half of the motion vector between adjacent frames as follows.

$$I_{2k,B}(m + \frac{1}{2}B_x, n + \frac{1}{2}B_y) = \frac{1}{2} \{I_{2k-1}(m + B_x, n + B_y) + I_{2k+1}(m, n)\}$$

$$I_{2k,F}(m + \frac{1}{2}F_x, n + \frac{1}{2}F_y) = \frac{1}{2} \{I_{2k-1}(m, n) + I_{2k+1}(m + F_x, n + F_y)\} \quad (2)$$

where  $(F_x, F_y)$  and  $(B_x, B_y)$  are the forward and backward motion vectors, respectively.  $I_{2k-1}$  and  $I_{2k+1}$  denote the previous and current frames, respectively.  $I_{2k,F}$  and  $I_{2k,B}$  contains occluded regions (overlapped pixels and holes). These regions mean the mismatch between the original frame and interpolated frame. So, we need to investigate how well the current block matches the original block, the temporal matching characteristics are measured as follows:

$$SAD_f = \min_{M \in S} \sum_{(m,n) \in M} |I_{2k-1}(m, n) - I_{2k+1}(m + F_x, n + F_y)|$$

$$SAD_b = \min_{M \in S} \sum_{(m,n) \in M} |I_{2k+1}(m, n) - I_{2k-1}(m + B_x, n + B_y)| \quad (3)$$

where  $SAD_f$  measures the mismatch between the current block ( $I_{2k+1}(m, n)$ ) and the reference block ( $I_{2k-1}(m + F_x, n + F_y)$ ). Similarly,  $SAD_b$  measures the mismatch between the current block ( $I_{2k+1}(m, n)$ ) and the

reference block  $(I_{2k-1}(m+B_x, n+B_y))$ . If the interpolated block has no or multiple motion trajectories, the value has small value. But, even if SAD is a high value, the interpolated block can have good visual quality. That is, (3) does not guarantee that the quality of the interpolated block can be determined only by (2). (3) can be used as one of the matching characteristics in temporal direction. Thus, in order to reflect the spatial matching characteristics, the cross correlation characteristics need to be calculated and compared. The cross correlation is measured by introducing NCCC (Normalized Cross Correlation Coefficient) as follows.

$$NCCC_f = \frac{\max_{M \in S} \sum_{(m,n) \in M} I_{2k-1}(m,n) I_{2k+1}(m+F_x, n+F_y)}{\left[ \sum_{(m,n) \in M} I_{2k-1}(m,n) \right]^{1/2} \left[ \sum_{(m,n) \in M} I_{2k+1}(m+F_x, n+F_y) \right]^{1/2}}$$

$$NCCC_b = \frac{\max_{M \in S} \sum_{(m,n) \in M} I_{2k-1}(m+B_x, n+B_y) I_{2k+1}(m,n)}{\left[ \sum_{(m,n) \in M} I_{2k-1}(m+B_x, n+B_y) \right]^{1/2} \left[ \sum_{(m,n) \in M} I_{2k+1}(m,n) \right]^{1/2}}$$

(4)

where  $(F_x, F_y)$  with  $NCCC_f$  means the most highly correlated location in the forward search range and  $NCCC_f$  indicates how well spatially is correlated.  $(B_x, B_y)$  with  $NCCC_b$  denotes the most highly correlated location in the backward search range and  $NCCC_b$  indicates how well spatially is correlated. Fig.3 shows an experimental result, which is obtained from Foreman sequences with QCIF, 30 frames (block size = 8x8). In this experiment, each block

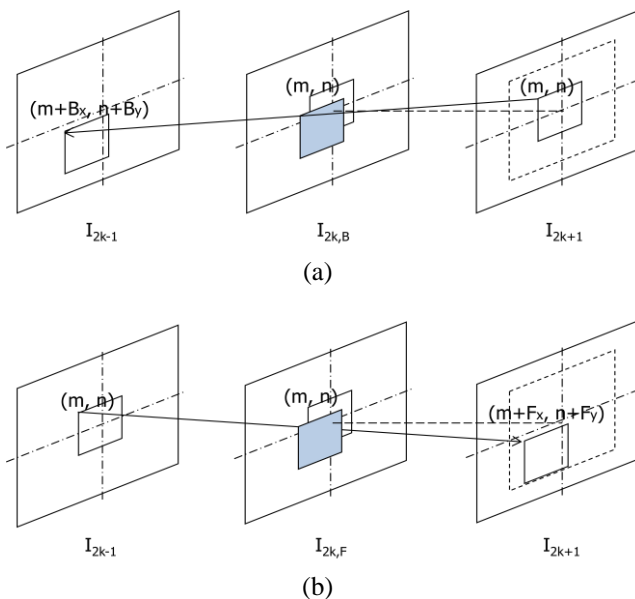


Fig. 2. Overview of motion-compensated frame interpolation between two adjacent frames. (a) backward MCFI. (b) forward MCFI.

composed of even frames is generated by using bi-directional motion estimation and compensation as shown in Fig. 2. Then, the MSE between the generated block and the original block is calculated and is classified by high MSE region (greater than 25) and low MSE region (less than 25). In these cases that the location with minimum SAD is not same with that of NCCC, it is shown that two regions can be easily partitioned. In the cases that the SAD value obtained by NCCC search is much greater than the value of SAD search, it is apparent that high MSE region can be easily partitioned from low MSE region. However, most samples overlap on same locations or concentrates on similar locations and so it is difficult to classify in these regions. Fig. 4 shows the conceptual diagram for these regions.

For these regions, the motion vector size for each sample is compared and shown in Fig. 5. Most samples in low MSE region concentrate on the region having less than 5, while most samples in high MSE region concentrate on the region having greater than 5. So, from this result, it shown that the size of motion vector is useful for classifying the samples whose MSE are same or similar.

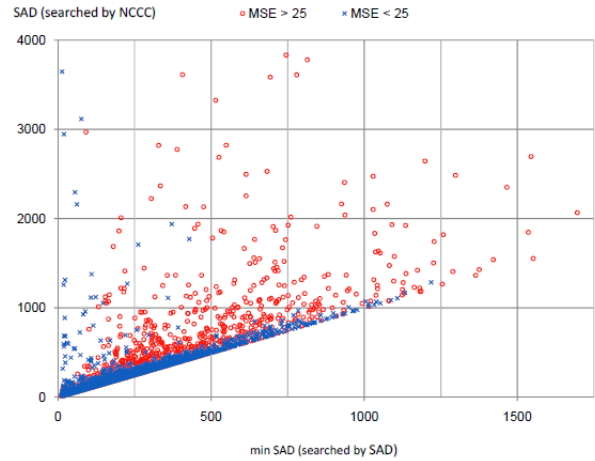


Fig. 3. Experimental result for high MSE and low MSE with Foreman sequence (QCIF, 30Hz).

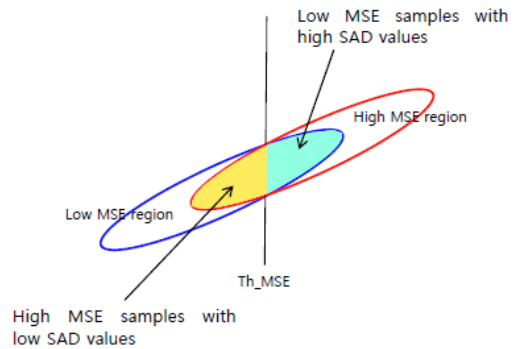


Fig. 4. Conceptual diagram for high MSE region and low MSE region.

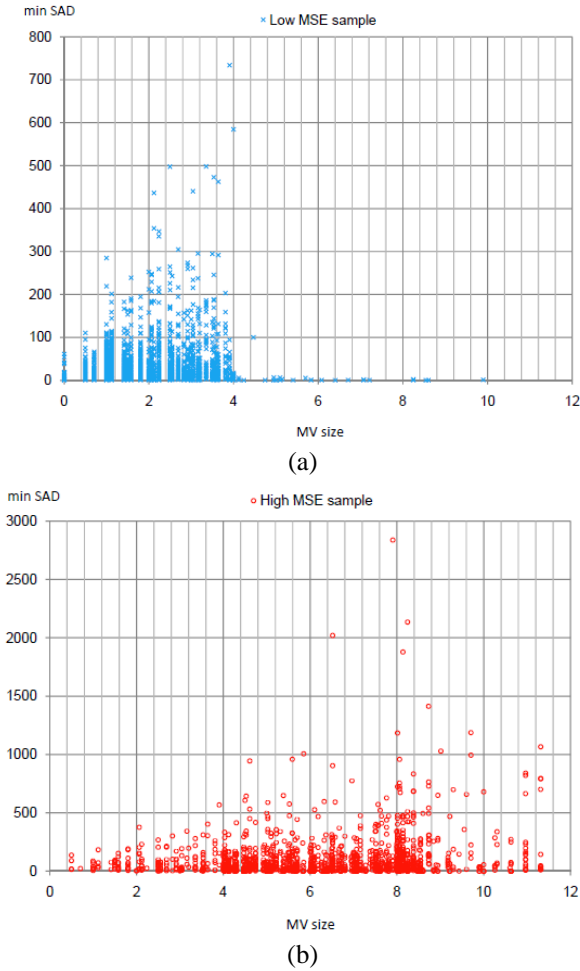


Fig. 5. Experimental result for size of motion vector vs. minimum SAD value. (a) low MSE region. (b) high MSE region.

#### IV. PROPOSED ALGORITHM FOR DETECTING OCCLUDED REGION

##### A. Proposed Algorithm

As mentioned above, even if the current block has a low SAD value, it does not mean that the interpolated block matches well the original block. In this section, we propose an efficient detecting occluded region based on SAD, NCCC, and the size of motion vector. Fig. 6 denotes the flow chart of the proposed algorithm. At first, by using (3) and (4), respectively, the motion vectors are found. Then their positions and SAD values are compared. If the difference between the SAD value found by NCCC search and the minimum SAD value is less than threshold value of SAD difference (DSAD), this block is determined by the motion vector size. If the difference is greater than DSAD and the SAD value of the current block is compared with the threshold of SAD ( $Th\_SAD$ ), this block is classified into occluded region. If the difference is not greater than DSAD, this block is

classified by the size of motion vector. In this case, the size of motion vector is not greater than the threshold of motion vector size and this block is determined to be interpolated linearly. This procedure is applied for the forward predictive frame and the backward predictive frame, independently. Accordingly, two predictive frames are obtained.

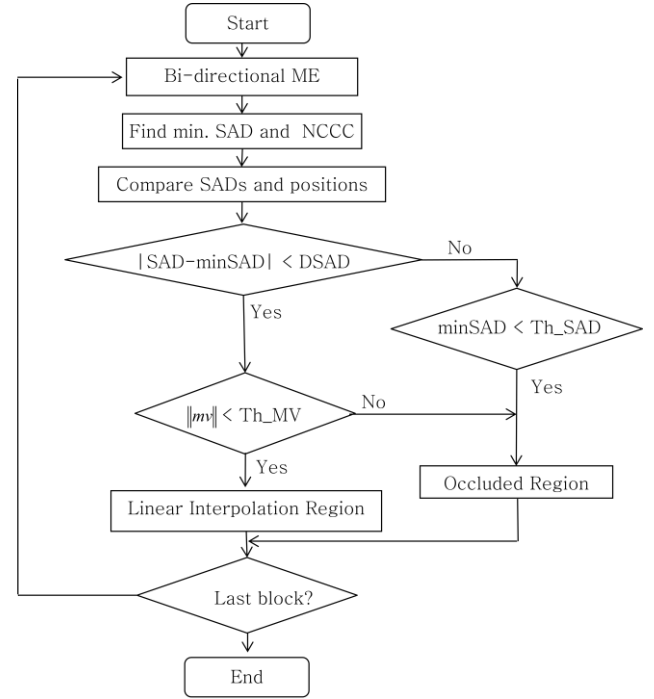


Fig. 6. Flow chart of detecting occluded region and linear interpolation region.

##### B. Experimental Results

First, Foreman sequences with QCIF, 30 frames (block size =  $8 \times 8$ ) is tested. Fig. 7 shows one experimental result, which simply classifies into the occluded region and the linear interpolated region, based on the proposed algorithm denoted in Fig. 6. In this experiment, the threshold values  $Th\_DSAD=50$ ,  $Th\_SAD=150$ , and  $Th\_MV=5$  were used. This result shows the similar result for Fig. 3 and it makes sure that the proposed scheme can be used for dividing two regions based on the SAD, NCCC, and the size of motion vector.

In order to evaluate how well is classified, two parameters,  $Er\_HL$  and  $Er\_LH$ , are introduced.  $Er\_HL$  denotes the portion of low distorted block in high MSE region, while  $Er\_LH$  means the portion of high distorted block in low MSE region. That is, the  $Er\_HL$  is the ratio of the low MSE blocks having greater than  $Th\_MSE$  in the high MSE region. Table I shows the experimental result for Foreman sequence. The proposed algorithm shows better performance than the  $SAD\_only\_method$ , which uses only SAD in classifying the distorted

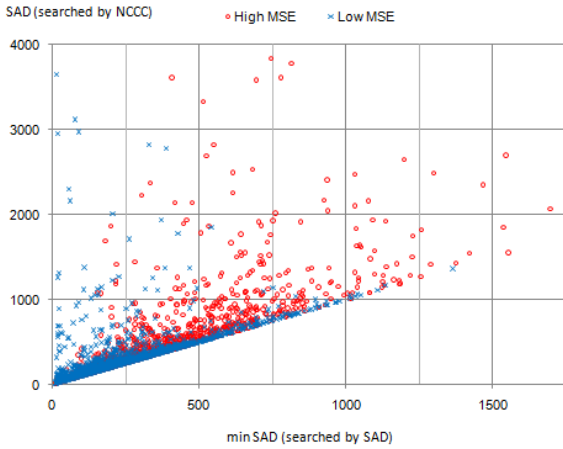


Fig. 7. Classification result ( $Th\_SAD=150$ ,  $Th\_DSAD=50$ ,  $Th\_MV=5$ ).

TABLE I  
CLASSIFICATION RESULTS FOR FOREMAN  
SEQUENCE

	SAD_only_method		Proposed	
	$Er\_HL$	$Er\_LH$	$Er\_HL$	$Er\_LH$
$Th\_SAD=120$	34.6%	26.3%	20.1%	16.3%
$Th\_SAD=180$	30.5%	32.3%	18.7%	19.2%
$Th\_SAD=240$	27.7%	37.8%	16.5%	20.7%

TABLE II  
CLASSIFICATION RESULTS FOR HALL MONITOR  
SEQUENCE

	SAD_only_method		Proposed	
	$Er\_HL$	$Er\_LH$	$Er\_HL$	$Er\_LH$
$Th\_SAD=100$	29.5%	26.7%	15.8%	9.4%
$Th\_SAD=140$	27.2%	27.5%	12.2%	10.5%
$Th\_SAD=180$	20.3%	29.1%	9.6%	12.3%

block. Particularly, by using large threshold values, the performance becomes obvious. Table II shows the experimental result for Hall Monitor sequence. In this sequence, the overall SAD values are smaller than the SAD\_only\_method.

## V. ADAPTIVE OCCLUDED REGION INTERPOLATION

### A. Proposed algorithm

As explained in (2) and Fig. 6, bi-directional motion-compensated interpolated frames (two initial frames) are obtained, respectively, by denoting  $I_{2k,F}$  and  $I_{2k,B}$ .

These frames are merged in order to interpolate the occluded areas as follows.

$$I_{2k}(m,n) = \begin{cases} I_{2k,B}(m,n), & \text{if } I_{2k,F}(m,n)=0 \text{ and } I_{2k,B}(m,n) \neq 0 \\ I_{2k,F}(m,n), & \text{if } I_{2k,B}(m,n)=0 \text{ and } I_{2k,F}(m,n) \neq 0 \\ \frac{1}{2}\{I_{2k,B}(m,n)+I_{2k,F}(m,n)\}, & \text{if } I_{2k,B}(m,n) \neq 0 \text{ and } I_{2k,F}(m,n) \neq 0 \\ \text{Occluded region,} & \text{otherwise} \end{cases} \quad (5)$$

This approach will lead to three phenomena. There are three cases for a given pixel:

- uniquely defined by a single motion vector;
- defined by more than one motion vector (an overlapping occurred);
- not defined by any motion vector (it is left blank).

In order to perform filling process for every pixel, the first case is trivial. For the second case, when more than one option for a pixel exist, a simple averaging method is applied to solve the problem. The last case is more designed in this paper. If no motion vector points to a pixel, then it is not easy to guess its value. One might use the co-located pixel in the previous frame and the next frame. However, it may not be very efficient since it might be that the motion vector of that block is not zero. So, in this paper, we propose an adaptive occluded region interpolation method, which uses adjacent pixel information. That is, by combining the frame generated by the forward estimation with the one generated by backward estimation will result in a frame with less blank areas. After the motion estimation and compensation, and after averaging the overlapping areas, the interpolated frame already looks like a frame, with some blank areas. At this point, there is enough information available about the current frame to perform motion estimation using the current frame.  $2N \times 2N$  block is selected as shown in Fig. 8, and then, to improve the method, bi-directional motion estimation is performed. To fill the blank areas, a reference block is searched in both the previous and next frames.

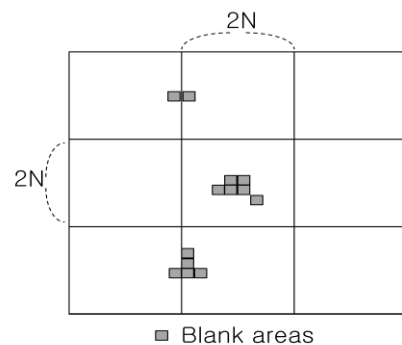


Fig. 8.  $2N \times 2N$  blocks division for filling blank areas.

### B. Experimental Results

To evaluate the performance of the proposed occlusion detection algorithm, three QCIF sequences (Foreman, Stefan, Table Tennis) are tested. As shown in Table III, the proposed algorithm consistently outperforms the conventional methods, the MCI method and Choi et al [3], for these sequences, which are highly active. The performance improvements are from 0.23 to 0.42 dB.

TABLE III  
PSNR [dB] performance comparison

Seq.	MCFI[1]	Choi et al[3]	Proposed
<i>Foreman</i>	32.65	33.28	33.51
<i>Stefan</i>	24.98	26.37	26.72
<i>Table Tennis</i>	30.91	32.11	32.53

## VI. CONCLUSIONS

This paper proposes an effective detection algorithm for occluded regions and an adaptive interpolation algorithm, which can be applied for FRUC technique. The proposed algorithm introduces NCCC as well as SAD and the size of motion vector and exploits to determine the occluded regions. Then, this paper proposes an adaptive interpolation method for occluded regions. Experimental results show that the proposed method effectively can classify and interpolate the occluded regions, and then achieve better performance than conventional methods.

It is expected that the proposed scheme can be efficiently used for FRUC technique and DVC schemes. For practical implementations, also we are analyzing the overall system complexity and we are searching for reducing computational burdens.

## 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. 2010-0016880 and No.2010-0022529)

## REFERENCES

- [1] R. Castagno, P. Haavisto, and G. Ramponi, "A method for motion adaptive frame rate upconversion," *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] B.D.Choi, J.W.Han, C.S.Kim, and S.J.Ko, "Motion-Compensated Interpolation Using Bilateral Motion Estimation and Adaptive Overlapped Block Motion Compensation," *IEEE Trans. CSVT*, Vol. 17, No.4, pp.407-416, Apr. 2007.
- [4] T.Thaipanich, P.-H.Wu, and C.-C. J. Kuo, "Low Complexity Algorithm for Robust Video Frame Rate Up-Conversion(FRUC) Technique," *IEEE Trans. Consumer Electronics*, Vol. 55, No. 1, pp. 220-228, Feb. 2009
- [5] 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.
- [6] J. Ascenso, C. Brites, and F. Pererira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," in *Proc. EURASIP Conf. Speech and Image Processing*, July 2005. Pp. 311-316
- [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.



**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 assistant 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.