MOTION COMPENSATED FRAME RATE-UP CONVERSION BASED ON MULTIPLE FRAME INTERPOLATION ALGORITHM

Hyeongchul Oh, Sang-Jun Park, Hanjin Park, and Jechang Jeong

Department of Electronics and Computer Engineering, Hanyang University
17 Hangdang-dong, Sungdong-gu, 133-791, Seoul, Korea
email: ohc8203@gmail.com, sjp21c@gmail.com, airjin999@naver.com, jjeong@ece.hanyang.ac.kr

ABSTRACT

This paper presents a frame rate-up conversion algorithm utilizing two frame interpolation methods and adaptive threshold to generate the four initial frames. The first frame interpolation method interpolates the block using the block co-located in the reference frame. The second frame interpolation method interpolates the block using the half value of the motion vector in interpolated frame. In this process, the occlusion regions are occurred in four initial frames since the optimal blocks are only selected by threshold. Hence, the four initial frames are merged in order to interpolate the occlusion regions. Finally, we perform the re-search using the adjacent pixels and the available data in the occlusion regions to fill the occlusion regions. The experimental results show that the proposed algorithm provides better PSNR results and the visual quality than the conventional algorithms.

1. INTRODUCTION

The frame rate up-conversion (FRUC) plays the key role in image coding technology in order to transmit high resolution video of huge amounts of data and the video format of various sizes without the missing data in the limited bandwidth. The high resolution video needs higher bit-rate to keep the visual quality. However, the missing data occur when the image compression is performed. In other words, if huge amounts of data are transmitted using the conventional compression method, the missing data increase. Thus, the decoded video formats have annoying jerky or blurry artifacts phenomenon. Various FRUC algorithms have been studied to enhance the frame rate and minimize the missing data. FRUC is able to improve the subjective quality of video and satisfy the transmittable maximum value through the conversion of bit-rate [1]-[4]. The conventional FRUC algorithms only use the information that exist in the reference frames such as the co-location of the backward or the forward direction [5]-[6].These algorithms have a low complexity and are easy to realize because it takes the same blocks from the reference frames. Also, they provide good performance when there is no motion of object. However, the ghost artifacts can be observed because they cannot predict the objects having large motion.

To solve these problems, a number of works have focused on motion compensation frame interpolation (MCFI) algorithm in order to accurately interpolate the blocks because most of videos include the objects having motion [7], [8]. MCFI using the motion information can be categorized into two approaches: The first approach obtains the optimal motion vector from the co-located block of reference frame [9]-[13]. It is based on a hypothesis that the object move linear. If the co-located block in the previous frame moves in a single direction, then the block in the following frame have the same direction as well. Here, the overlapped regions do not occur because the processing is at regular intervals. However, when the obtained motion vector is irregular, the error increases seriously. Girod *et al*. proposed Overlapped Block Motion Compensation (OBMC) that extends the block size to reduce the blocking artifacts [12]. The optimal blocks are searched by using information of the motion vector including the overlapped regions in the estimated block. It is interpolated respectively by the divided block which has the overlapped regions so that OBMC reduces the blocking artifacts. But, OBMC has a tendency to over-smooth the reconstructed image due to the excessive overlap effect on the edges of the block. The second approach interpolates the estimated block on half of the obtained motion vector [14]-[17]. The interpolated frame has the overlapped areas as well as the occlusion regions without the blocking artifacts. However, the overlapped regions occur irregularly in the interpolated frame, because the blocks have different motion vectors. Also, the generated frame includes the occlusion areas which are the regions of nonexistent data. Thaipanich *et al*. proposed to apply the obtained motion information from several reference frames using the bilateral motion estimation in order to interpolate the image efficiently [15]. The improvement of visual quality is slightly achieved by using the multiple reference frames. However, it has the high complexity and the excessive overlapped regions. Oh *et al*. proposed adaptive occlusion area interpolation method using the bilateral motion estimation and the threshold [16]. It provides the improved results using interpolated frames which are generated by the threshold and the process of the re-search. However, the unnecessary searching process is added and the available blocks are counted as the error block by using higher threshold values.

To alleviate the aforementioned drawbacks of the conventional methods, the proposed algorithm utilizes multiple MCFIs to generate four reference frames. Then, merging process is performed to yield one reference frame. In addition, we perform spatial interpolation and re-search processes to fill the occlusion regions. The rest of this paper is organized as follows. In Section 2, we describe the proposed algorithm in details. The experimental results and analysis are provided in Section 3. Finally, we conclude in section 4.

2. PROPOSED ALGORITHM

In this section, we propose the FRUC algorithm that generates the optimal image by using the bilateral motion information in order to enhance the performance. The overall block diagram of the proposed FRUC algorithm is illustrated in Fig. 1. When the minimum sum of absolute difference (SAD) of a block during the block matching phase is less than the pre-defined threshold value, we use a multiple MCFI process. Here, the multiple MCFI technique generates four initial frames. These frames are merged to an initial frame by the image merging technique. Then the occlusion region is interpolated by the spatial interpolation technique which takes average of the neighboring pixels around the occlusion block that

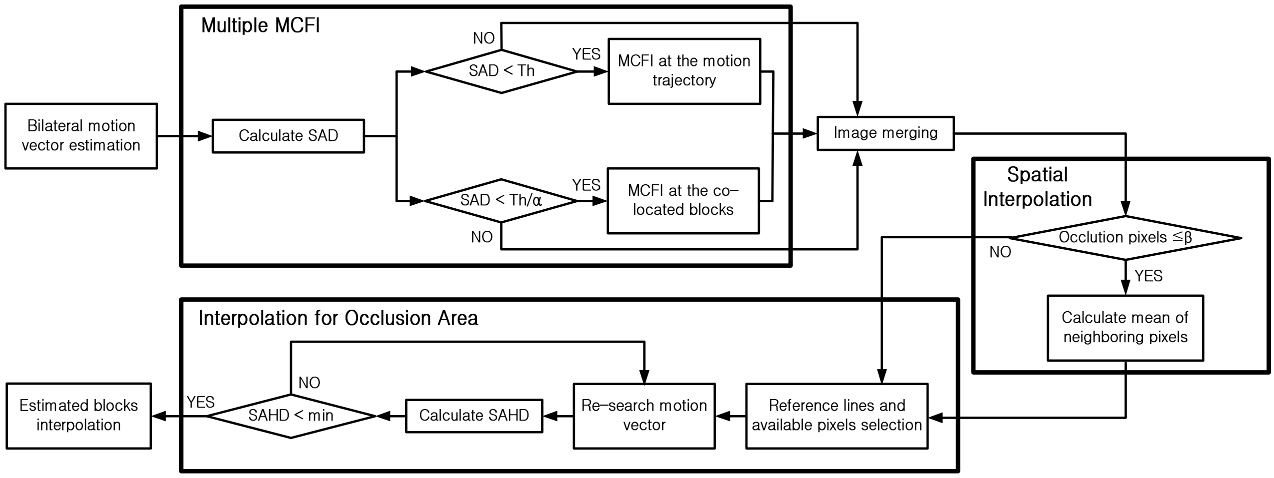


Figure 1 – The block diagram of the proposed FRUC algorithm.

has fewer pixels than a fixed value. The rest of the occlusion regions are finally interpolated by the re-search technique.

2.1 Selective motion estimation

The SAD measure is widely used to estimate the reliable block. However, there is a possibility of choosing wrong blocks, even though the block has the minimum SAD. Hence, we set up the threshold to reduce the error and select only the reliable blocks. The process of the block estimation is defined as:

$$SAD_{t\pm1} = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\left|f_{t\pm1}(k,l) - f_{t\mp1}(k+mv_x, l+mv_y)\right|$$
$$mv_{x,y} = (B_x, B_y) \quad at \;\; f_{t+1}(k,l) \;\; of \;\; MCFI^{TR,CO} \qquad (1)$$
$$mv_{x,y} = (F_x, F_y) \quad at \;\; f_{t-1}(k,l) \;\; of \;\; MCFI^{TR,CO}$$

where $MCFI^{TR}$ represents the method which interpolates the block using the half value of the obtained motion vector. $MCFI^{CO}$ denotes the method that obtains the motion vector of the co-located block in the reference frame. $mv_{x,y}$ represents either the estimated motion vector $B_{x,y}$ when the backward frame f_{t-1} is considered as the reference frame or $F_{x,y}$ when the forward frame f_{t+1} is considered as the reference frame. $B_{x,y}$ and $F_{x,y}$ denote the motion vectors pointing the reference frames f_{t-1} and f_{t+1} in the horizontal and vertical directions, respectively. We can get the motion vectors that are created differently by each reference frame using the following equations:

$$MCFI^{TR} = \begin{cases} \hat{f}^{tr}_{B,F} & if \quad SAD_{t\pm1} \le T \\ 0 & else \quad SAD_{t\pm1} > T \end{cases} \qquad (2a)$$

$$MCFI^{CO} = \begin{cases} \hat{f}^{co}_{B,F} & if \quad SAD_{t\pm1} \le T/\alpha \\ 0 & else \quad SAD_{t\pm1} > T/\alpha \end{cases} \qquad (2b)$$

$$\alpha = 2 \qquad (2c)$$

where $\hat{f}^{tr}_{B,F}$ denotes the initial frame which is created by $MCFI^{TR}$ when the minimum SAD is less than the threshold (T) in (2a). $\hat{f}^{co}_{B,F}$ denotes another initial frame which is created by $MCFI^{CO}$ when the minimum SAD is less than the threshold (T/2) in (2b). α represents a scaling factor for accurate estimation in (2c).

Two of the initial frames utilize $MCFI^{TR}$ technique shown in Fig. 2. In Fig. 2, the estimated block from the backward frame f_{t-1} is generated on half of the motion vector by using in the following equations:

$$\hat{f}^{tr}_B(k+\frac{1}{2}B_x, l+\frac{1}{2}B_y) = \frac{1}{2}\left\{f_{t-1}(k+B_x, l+B_y) + f_{t+1}(k,l)\right\} \qquad (3a)$$

$$\hat{f}^{tr}_F(k+\frac{1}{2}F_x, l+\frac{1}{2}F_y) = \frac{1}{2}\left\{f_{t+1}(k+F_x, l+F_y) + f_{t-1}(k,l)\right\} \qquad (3b)$$

where $\hat{f}^{tr}_B$ represents first frame which is created by the backward motion vector $B_{x,y}$ in (3a) and $\hat{f}^{tr}_F$ represents second initial frame created by the forward motion vector $F_{x,y}$ in (3b). Fig. 2(b) shows the generation of the second initial frame $\hat{f}^{tr}_F$.

The block position in two initial frames $\hat{f}^{tr}_B$ and $\hat{f}^{tr}_F$ is irregular because it is influenced by different motion vectors. In other word, either the overlapped regions occur between adjacent blocks or the occlusion areas exist in the initial frame. To avoid the loss of data, the overlapped region is calculated by the average of pixels within this region. Fig. 3 shows two initial frames $\hat{f}^{tr}_B$ and $\hat{f}^{tr}_F$ using $MCFI^{TR}$. These frames have different occlusion regions. The created image is similar to that of the original frame.

To create the rest of four initial frames, we use $MCFI^{CO}$. It interpolates the blocks with the co-located block in the reference frame. $\hat{f}^{co}_B$ and $\hat{f}^{co}_F$ are interpolated at (k,l) from the estimated motion vectors in the reference frame f_{t-1} and f_{t+1} as shown in Figs. 4(a) and (b), respectively. $\hat{f}^{co}_B$ and $\hat{f}^{co}_F$ are generated as following equations:

$$\hat{f}^{co}_B(k,l) = \frac{1}{2}\left\{f_{t-1}(k+B_x, l+B_y) + f_{t+1}(k-B_x, l-B_y)\right\} \qquad (4a)$$

$$\hat{f}_F^{co}(k,l) = \frac{1}{2}\{f_{t+1}(k+F_x, l+F_y) + f_{t-1}(k-F_x, l-F_y)\} \tag{4b}$$

where $\hat{f}_B^{co}$ and $\hat{f}_F^{co}$ represent third and forth initial frames which are created by the backward and forward motion vectors $B_{x,y}$ and $F_{x,y}$ respectively.

Fig. 5 shows the other initial frames $\hat{f}_B^{co}$ and $\hat{f}_F^{co}$ by using $MCFI^{CO}$ which has lower threshold than $MCFI^{TR}$. So, the third and forth frames include a lot of the occlusion blocks. However, the interpolated images are trustworthy.

Initial frames $\hat{f}_B^{tr}$, $\hat{f}_F^{tr}$, $\hat{f}_B^{co}$, and $\hat{f}_F^{co}$ are merged as an merged frame $\hat{f}_M$ for the next process. Because four initial frames are generated by using different motion vectors and MCFI techniques, these frames have the occlusion areas. If the data in one or more initial frames are unavailable, we calculate the average of the rest of initial frames having the data. When all of initial frames cannot consider the data, the corresponding region is remained as the occlusion area. The merged frame $\hat{f}_M$ has fewer occlusion regions as shown in Fig. 6.

2.2 Adaptive threshold

The two reference frames have high correlation because these frames are temporally close. So we utilize the covariance considering the correlation between two frames in order to determine the threshold. If a fixed threshold is used by the block estimation, the initial frame includes incongruent blocks with high SAD or excludes congruent blocks with low SAD. To prevent having incorrect information, we use an adaptive threshold (T) determined as.

$$T = \frac{1}{\gamma}\left\{\frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}(f_{t+1}(k,l)-\sigma_{t+1})(f_{t-1}(k,l)-\sigma_{t-1})\right\} \tag{5a}$$

$$\sigma_{t\pm1} = \sqrt{\frac{1}{N(N-1)}\left\{N\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}f_{t\pm1}(k,l)^2 - (\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}f_{t\pm1}(k,l))^2\right\}} \tag{5b}$$

$$\gamma = 3, \qquad N = 8 \tag{5c}$$

$\sigma_{t\pm1}$ represents the standard deviation between two reference frames f_{t+1} and f_{t-1} in (5). The covariance subtracts the average value of two factors, but we applied the standard deviation to estimate distribution of pixels in a block. Let γ denotes a scaling factor to adjust the values of T to satisfy the range of SAD. Therefore, it is possible to extract the adaptive threshold.

2.3 Concealment of occlusion regions

The occlusion regions are distributed irregularly widespread in the merged frame $\hat{f}_M$. Particularly, parts of the occlusion regions are very small, as shown in Fig. 6(a). Therefore, the spatial interpolation technique is applied to these regions to efficiently reduce the com-

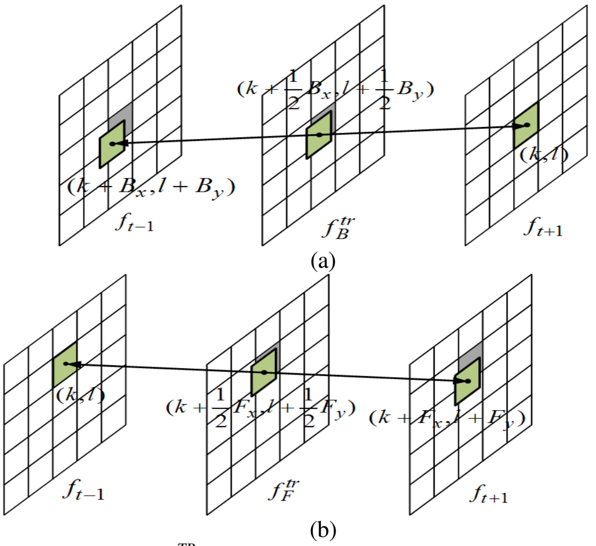


Figure 2 – $MCFI^{TR}$ scheme of (a) the backward and (b) forward estimation for generation of initial frames.

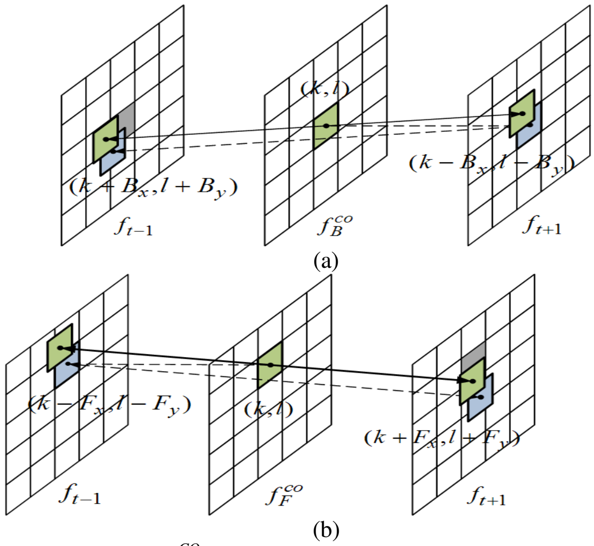


Figure 4 – $MCFI^{CO}$ scheme of (a) the backward and (b) forward estimation for generation of initial frames.

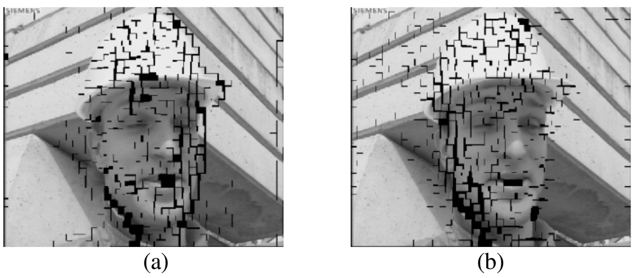


Figure 3 – Generation of two initial frames, (a) $\hat{f}_B^{tr}$ and (b) $\hat{f}_F^{tr}$, using $MCFI^{TR}$ scheme.

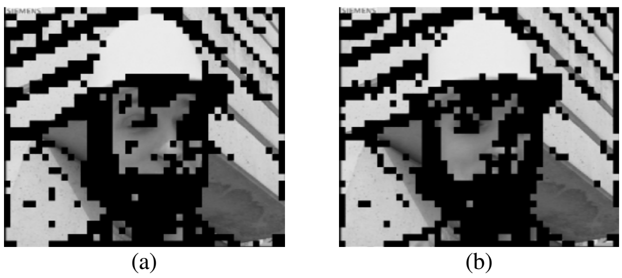
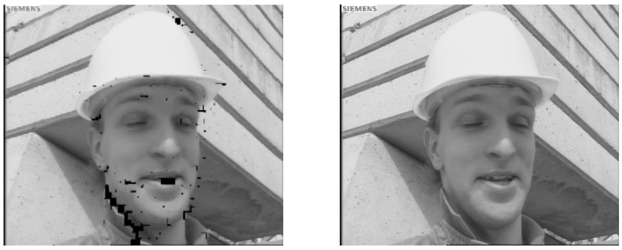
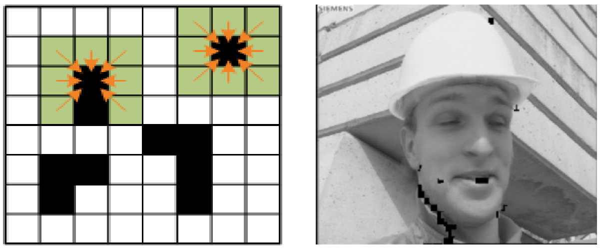


Figure 5 – Generation of two initial frames, (a) $\hat{f}_B^{co}$ and (b) $\hat{f}_F^{co}$, using $MCFI^{TR}$ scheme.



(a) (b)

Figure 6 – Comparison of (a) the merged frame $\hat{f}_M$ and (b) the original frame.



(a) (b)

Figure 7 – The occlusion pixel interpolation. (a) Spatial interpolation method and (b) the interpolated frame $\hat{f}_M$.

plexity and easily interpolate the pixels. It can be done as

$$\hat{f}_M = \begin{cases} S, & if \quad Occlusion\ pixels \le \beta \\ 0, & else\ Occlusion\ pixels > \beta \end{cases} \tag{6a}$$

$$S = \left\{ \sum_{m=-1}^{1} \sum_{n=-1}^{1} \hat{f}_M(k+m, l+n) \right\} / (k \times l - 1 - N) \tag{6b}$$

Let β denotes the threshold to perform spatial interpolation, and S denotes the average value of neighboring pixels. We set β to 16. In Fig. 7(a), if the numbers of unavailable pixels are smaller than β, we carry out spatial interpolation. Fig. 7(b) shows the merged frame $\hat{f}_M$ by using the spatial interpolation method.

We additionally perform the temporal interpolation for the remaining occlusion regions in the merged frame $\hat{f}_M$. The previously generated data in $\hat{f}_M$ are applied to this process because these data are trustable. When the adjacent pixels within the neighbor blocks are available, we use two boundary reference lines. Otherwise, we adaptively alter the number of considering boundary reference lines in order to adjust amounts of the reference pixels, because little information is used if the adjacent blocks include many occlusion areas. It can prevent wrong estimation from the less information of reference pixels to process the optimum block search.

Finally, the re-search is performed with both the existing data in the occlusion block and the selected reference lines. Even if different frames include the same object, it can be recognized as different object due to the specific circumstances when we use SAD measure for block matching. Therefore, we use the sum of absolute histogram difference (SAHD) measure that calculates the cumulative value to estimate the optimal block instead of the SAD. The optimal block is selected by SAHD that bin is set to limited range. The block having minimum SAHD is used for the frame interpolation. SAHD is calculated as

$$SAHD = \sum_{i=0}^{b} \left| h^{s}_{\hat{f}_M}[i] - h^{s+mv}_{f_{t\mp1}}[i] \right|, \quad s \in b_{k,l} \tag{7}$$

where b presents a number of bin for histogram. Let $h^{s}_{\hat{f}_M}$ and $h^{s+mv}_{\hat{f}_{t\mp1}}$ denote the cumulative values of bins for $\hat{f}_M$ and the reference frames f_{t-1} and f_{t+1}, respectively. We finally create the interpolated frame $\hat{f}$ as

$$\begin{aligned} \hat{f}(k,l) = \frac{1}{2}\{ & f_{t-1}(k + B_x + B'_x, l + B_y + B'_y) \\ & + f_{t+1}(k + F_x + F'_x, l + F_y + F'_y)\} \end{aligned} \tag{8}$$

where $B'_{x,y}$ and $F'_{x,y}$ denote the re-searched motion vectors using the reference lines in the previous and following frames, respectively. Then, the block, which is calculated by the averaging with the selected blocks in f_{t-1} and f_{t+1}, is employed so that we can create the interpolated frame.

3. EXPERIMENTAL RESULTS

In this section, we compared the PSNR results and the subjective perception with those of the conventional methods, including the MCFI, the OBMC [12], the Thaipanich *et al.* [15], and the Oh *et al.* algorithm [16], to demonstrate the performance of the proposed algorithm. The ten raw sequences, listed in Table 1, in CIF (352×288) format were encoded and estimated with a block size of 8×8, with the initial search range ($\pm$ 16 pixels) for extracting motion vectors for both the horizontal and vertical directions. Since the median predictor is used to predict the motion vector among the neighbor blocks during the re-search process, the re-search range was set to be $\pm$ 2.

The proposed algorithm provided better image quality, as shown in Fig. 8. In Figs. 8(a)-(d), a lot of blocking artifacts could be perceived easily in the frames produced by the conventional algorithms; however, our method reduced the blocking artifacts and improved visual quality as shown in Fig. 8(e). In addition, we found that the background and the face were clearly separated because the proposed algorithm removed the remaining error block. Fig. 9 shows that *Stefan* sequence was applied by the FRUC algorithms. The conventional algorithms, as shown in Figs. 9(a)-(d), could not interpolate the player's legs properly; however, the proposed method interpolated the most similar image to the original frame, as shown in Fig. 9(e).

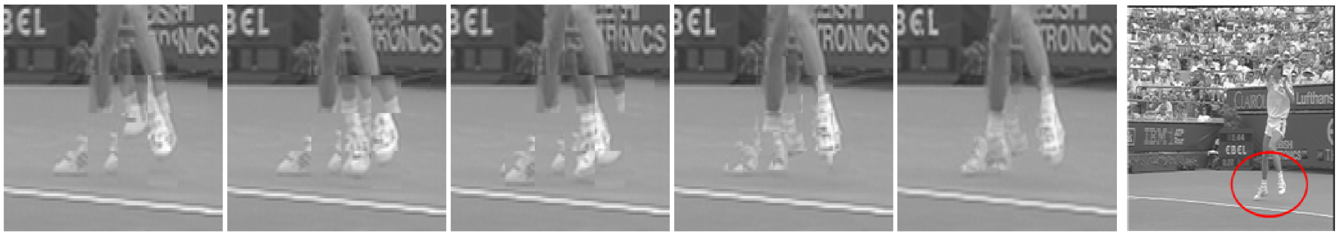
Table 1 presents the PSNR results showing that the proposed algorithm outperformed the conventional methods. The best PSNR results of the proposed algorithm were observed in *Foreman* and *Monitor* sequences (0.7 dB and 0.6 dB higher than that of the [16]). In *Foreman*, *Flower* and *Mobile* sequences, 1 dB higher PSNR results were achieved when compared to [15] method.

4. CONCLUSIONS

In this paper, we proposed a multiple FRUC algorithm to create interpolated frames by using the adaptive threshold and the bilateral predictions of the backward and forward frames. The proposed algorithm does not depend on the motion information of the bilateral direction; rather it interpolates the occlusion region using a re-search technique based

(a) (b) (c) (d) (e) (f)

Figure 8 – Comparison of the original frame and the interpolated frames of the 14th frame within *Forman* (CIF). (a) MCFI, (b) OBMC, (c) Thaipanich *et al*., (d) Oh *et al*., (e) proposed algorithm, and (f) original frame.



(a) (b) (c) (d) (e) (f)

Figure 9 – Comparison of the original frame and the interpolated frames of the 112th frame within *Stefan* (CIF). (a) MCFI, (b) OBMC, (c) Thaipanich *et al*., (d) Oh *et al*., (e) proposed algorithm, and (f) original frame.

Table 1. PSNR [dB] performance comparison

Sequences	MCFI	OBMC	[15]	[16]	Proposed
Foreman	32.73	33.14	33.49	34.48	**35.19**
Bus	21.67	21.91	22.66	22.87	**23.21**
Flower	31.21	31.38	31.80	32.62	**32.98**
Mobile	29.29	29.94	30.81	31.80	**32.25**
Tempete	29.56	29.68	29.84	29.87	**30.02**
Football	22.67	22.87	23.41	23.98	**24.22**
Ice	26.17	26.32	26.86	26.90	**27.29**
Monitor	36.82	36.96	37.34	37.68	**38.31**
Stefan	25.96	26.11	26.72	26.75	**27.11**
Table	31.18	31.33	32.05	32.22	**32.24**
Average	28.73	28.97	29.49	29.92	**32.58**

on the reliabilities of the motion vectors and the adjacent pixels. The experimental results showed that the proposed algorithm provides better performance than the conventional algorithms in terms of both the objective and subjective image qualities.

REFERENCES

[1] 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.

[2] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blanchfield, "Towards optimal rate control: A study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate," in *Proc. VCIP*, pp. 198-209, July, 2003.

[3] Q. Huynh-thu and M. Ghanbari, “Temporal aspect of perceived quality in mobile video broadcasting,” *IEEE Trans. Broadcasting*, vol. 54, no 2, pp. 641-651, Sep. 2008.

[4] I. P. Gunawan and M. Ghanbari, “Efficient reduced-reference video quality meter,” *IEEE Trans. Broadcasting*, vol. 54, no. 5, pp. 669-679, Sep. 2008.

[5] K. A. Bugwadia, E. D. Petajan, and N. N. Puri, “Progressive-scan rate up-conversion of 24/30 source materials for HDTV,” *IEEE Trans. Cons. Elect.*, vol. 42, no. 3, pp. 312-321, Aug. 1996.

[6] G. Dane and T. Q. Nguyen, “Optimal temporal interpolation filter for motion-compensated frame rate upconversion,” *IEEE Trans. Image Process*., vol. 15, no. 4, pp. 978-991, Apr. 2006.

[7] K. Hilman, H. W. Park, and Y. Kim, “Using motion-compensated frame rate conversion for the correction of 3:2 pull down artifacts in video sequences,” *IEEE Trans. Circuit Syst. Video Technol*., vol. 10, no. 6, pp. 867-877, Sep. 2000.

[8] R. Castagno, P. Haavisto, and G. Ramponi, “A method for motion adaptive frame rate upconversion,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 6, no. 5, pp. 436-446, Oct. 1996.

[9] C.–L. Huang and T.-T. Chai, “Motion-compensated interpolation for scan rate up-conversion,” *SPIE Optical Engineering*, vol. 35, pp. 166-176. Jan. 1996.

[10] T.-Y. Kuo and C.-C. J. Kuo, “Motion-compensated interpolation for low-bit-rate video quality enhancement,” *in Proc. ADIP XXI*, San Diego, CAa, pp. 277-288, Jul. 1998.

[11] S. Liu, C.-C. J. Kuo, and J. W. Kim, “Hybrid global-local motion compensated frame interpolation for low bit rate video coding,” *Journal of Visual Communication and Image Representation*, vol. 14, pp. 61-79, Oct. 2003.

[12] B. Girod, “Efficiency analysis of multihypothesis motion-compensated prediction for video coding,” *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 173-183, Feb. 2000.

[13] B. D. Choi, J. W. Han, C. S. Kim, and S. J. Ko, “Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17, no. 4, pp. 407-416, Apr. 2007.

[14] J. O. Lee, S. K. Jamg, Q. S. Chen, Y.-H. Kim, and C.-W. Kim, "An efficient frame rate up-conversion method for mobile phone with projection functionality," *IEEE Trans. Cons. Elect.*, vol. 53, no. 4, pp. 1615-1621, Nov. 2007.

[15] T. Thaipanich, P.-H. Wu, and C.-C. J. Kuo, “Low complexity algorithm for robust video frame rate up-conversion (FRUC) technique,” *IEEE Trans. Cons. Elect.*, vol. 55, no. 1, pp. 220-228, Feb. 2009.

[16] H. C. Oh, J. H. Lee, C. K. Min, and J. C. Jeong, “Frame interpolation method based on adaptive threshold and adjacent pixels,” in *Proc. IIH-MSP*, pp. 264-267, Oct. 2010.

[17] S.-J. Kang, K.-R. Cho, and Y. H. Kim, “Motion compensated frame rate up-conversion using extended bilateral motion estimation,” *IEEE Trans. Cons. Elect.*, vol. 53, no. 4, pp. 1759-1767, Nov. 2007.