OCCLUSION HANDLING FRAME RATE UP-CONVERSION

Yong Guo, Zhiyong Gao, Li Chen and Xiaoyun Zhang

Institute of Image Communication and Information Processing
Shanghai Jiao Tong University, Shanghai 200240, China
Shanghai Key Labs of Digital Media Processing and Communication, Shanghai 200240, China
Email: {guoyong1987, zhiyong.gao, hilichen, xiaoyun.zhang}@sjtu.edu.cn

ABSTRACT

Motion-compensated frame interpolation (MCFI) is a technique used extensively to enhance the temporal resolution of video sequences. In order to obtain a high quality interpolation, the motion vector field (MVF) between frames must be well-estimated. However, many current techniques for determining the MVF are prone to errors in occlusion regions. In this work, we propose an improved algorithm for improving the quality of MCFI by restoring the unreliable MVF and pixels in occlusion regions. We first utilize a dual motion estimation (DME) scheme which performs better in occlusion regions. Occlusion regions are determined by the ratio of two directional matching errors. Then, MVs in occlusion regions are refined using an orientation-based refinement (O-BR) method, which promotes occluded MVs with its orthogonal neighboring MVF. Finally, regional blending (RB) is proposed to restore the unreliable pixels in occlusion regions for further error concealment. Experimental results demonstrate that the proposed algorithm provides a better quality than previous benchmark frame rate up-conversion (FRUC) methods both objectively and subjectively.

***Index Terms*—** frame rate up-conversion (FRUC), motion-compensated frame interpolation (MCFI), dual motion estimation (DME), orientation-based refinement (OBR), regional blending (RB)

1. INTRODUCTION

Frame rate up-conversion (FRUC) is a technique that increases the frame rate of the video by inserting newly generated frames into the original sequence. FRUC is commonly applied in format conversion, and also used to reduce motion artifacts of videos in hold-type displays, such as liquid crystal displays (LCD). Motion-compensated FRUC (MC-FRUC) is composed of two major steps: motion estimation (ME) and motion-compensated frame interpolation (MCFI). The performance of a FRUC algorithm strongly depends on the accuracy of the estimated motion information as well as the design of the interpolation filter.

To improve the accuracy of motion vector (MV) between successive frames, many pioneering works have been done. In [1], Haan et al. proposed a 3-D recursive search (3-DRS) method using spatio-temporal related motion candidates to get the true motion. Choi et al. [2] proposed a bilateral ME method to avoid overlapped areas and holes. Huang et al. [3] employed the reliability information of MV and developed a hierarchical MV processing method to obtain a more reliable MVF for FRUC. Although true motion estimation techniques have developed in literature for FRUC, many current techniques are prone to errors in occlusion regions.

When block-based MCFI is used for FRUC, blocking artifacts are usually perceived due to the block-wise motion compensation. To produce better MCFI results, Wang et al. [4] regarded artifacts as MCFI noise and proposed a three-iterative trilateral filter to remove noise and achieved satisfactory visual effects.

Motivated by the above analysis, we propose in this paper an improved FRUC algorithm which makes use of dual ME (DME), orientation-based refinement (OBR) and regional blending (RB). The main reason why we use DME is that it not only offers more accurate motion trajectories, but also avoids overlapped areas and holes. The DME uses the matching ratio of forward and backward predictions in order to perform the validity check of MVF and refine them with OBR method. We use two MVFs to interpolate two additional frames, and estimate the pixel reliability from the difference between these two predictions. Subsequently, reliability-based regional blending is proposed for further error concealment.

The rest of this paper is organized as follows. Section II presents the proposed algorithm. Section III shows the experimental results and evaluates the performance of the proposed method. Finally, Section IV concludes the paper.

2. PROPOSED ALGORITHM

The proposed FRUC algorithm comprises four components: dual motion estimation (DME), MVF refinement, adaptive overlapped block motion compensation (ABOMC) [2] and re-

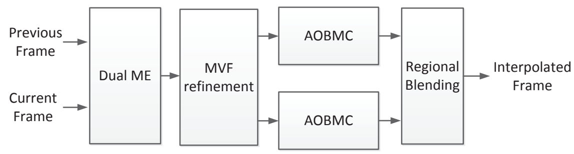


Fig. 1. Block diagram of proposed FRUC system

gional blending (RB). First, DME algorithm is used to predict the motion vectors of the intermediate frame both in the forward and backward directions. Second, the MVF is refined mainly focused on occlusion areas. Third, two predictions of the intermediate frame are produced using AOBMC with the post-processed MVF. Finally, regional blending is applied to unreliable pixels of two interpolated frames for further error concealment. Compared with existing methods [2], [4], the proposed method does not require extra iterative calculations to perform moving object segmentation and pixel-wise trilateral filtering. Fig.1 shows the overall framework of the proposed FRUC algorithm.

2.1. Dual motion estimation

Conventional bilateral ME (BME) estimates a motion vector by exploiting the temporal symmetry between the corresponding blocks in the previous and current frames. As shown in (1), the candidate motion vector with the minimum sum of absolute differences (SAD) value is selected as the final motion vector. That is

$$SAD(d\vec{v}) = \sum_{\vec{x}\in B} |f_{n-1}(\vec{x} - d\vec{v}/2) - f_n(\vec{x} + d\vec{v}/2)|$$
$$\vec{x} = arg \underset{d\vec{v}\in S}{min} SAD(d\vec{v}) \tag{1}$$

where $d\vec{v}$ denotes the candidate motion vector, and f_{n-1} and f_n are the previous and current frames, respectively. B denotes the range of a block and S stands for the search window.

The BME works well in general, and it sometimes performs better than unilateral ME (UME) because there is no holes or overlapped areas arising. However, it is difficult to accurately estimate the bilateral motion because of the lack of information on occlusions. UME helps to improve the accuracy of ME of occlusion regions [5]. The forward estimation can provide the uncovering regions reliably and vice versa for the backward estimation. Hence, we utilize the dual motion estimation (DME) which takes advantages of both BME and UME.

In DME, we first conduct conventional UME in the viewpoint of the previous and current frames respectively. The blocks MVF passes through are not contiguous in the interpolated frame as illustrated in Fig.2, which results in overlaps and holes. To avoid overlaps and holes, BME is then conducted using two MVFs of UME as the motion candidate respectively. As in Fig.2, in interpolated frame, for blocks with several UME blocks overlapped, the candidate motion vector with the minimum SAD value in (1) is selected as the final motion vector. For blocks without UME blocks overlapping, the final motion vector is determined using the median filter of the neighborhood MVs.

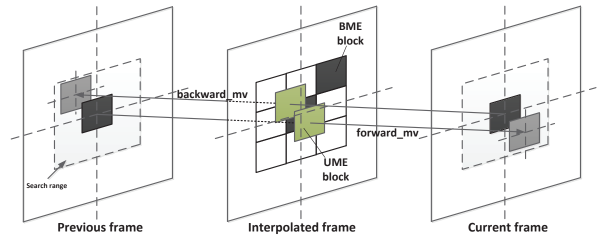


Fig. 2. The proposed dual motion estimation

2.2. Orientation-based refinement

In order to improve the accuracy of the MVF in occlusion regions, OBR is computed subsequently to the DME. The method is based upon the idea that the true motion in the occlusion region is likely to be similar to motions in neighbors of the orthogonal direction, rather than the same direction with the boundary.

Validity check of MVF is firstly performed to determine the occlusion region. Two directional match errors are used: sum of forward absolute differences (SFAD) and sum of backward absolute differences(SBAD),which are computed as (1) in correspondence to forward ME and backward ME, respectively. In the regions without occlusions, MV can be correctly estimated, and SFAD and SBAD of the same block should be similar (or identical). However, the SAD match distortion is not reliable in occlusion regions. SFAD tends to lead a wrong prediction in covering regions which are not visible at current frame, and SBAD tends to lead a wrong prediction in uncovering regions which are not visible at previous frame. Motion vector validity (V) is evaluated using the SFAD and SBAD as follows:

$$V = \begin{cases} 1, & 1-\alpha < \frac{SFAD}{SBAD} < 1+\beta \\ 0, & otherwise \end{cases} \tag{2}$$

where α is the ratio for the margin of variation between SFAD and SBAD. Assuming $(1-\alpha)*(1+\beta) == 1$ and we set α to 0.3 and β to 0.5. If the ratio of SFAD to SBAD is in the range between $1-\alpha$ and $1+\beta$, the selected motion vector is decided to be valid, and the validity value is set 1. Otherwise, the two predictions are considered very different and we set the value to 0.

Occlusion areas are then classified into five categories in terms of direction of unreliable blocks, as in Fig.3. Specifically, we separate motion vectors into four sub-parts (top left, top right, bottom left, bottom right) as in Fig.4. In the first

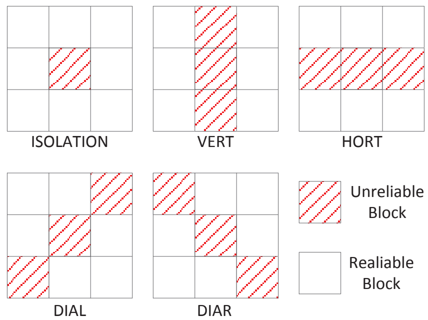
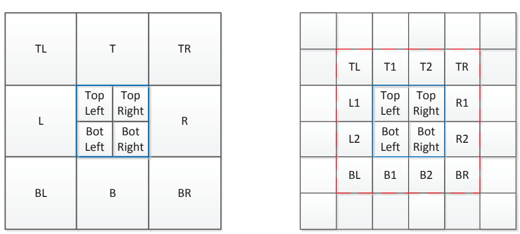


Fig. 3. Categories of occlusion patterns determined by directional templates

stage, for the reliable block whose validity value is 1, we use the its nearest reliable neighbors (up to 4) to refine the four sub-blocks. In the second stage, occlusion blocks are refined using motion vectors of 12-sided sub-blocks. Each part is refined with its nearest orthogonal neighbors, as illustrated in Table.1. The refinement is conducted in sub-blocks, expressed as

$$d\vec{x} = \underset{d\vec{v} \in C_{RN}}{arg\, min} SAD(d\vec{v}) \tag{3}$$

where $d\vec{v}$ denotes the candidate motion vector, C_{RN} stands for MV set of reliable neighbors.



(a) The first stage (b) The second stage

Fig. 4. Block refinement using 9-sided parent-block and 12-sided sub-block neighbors

2.3. Regional blending

AOBMC [2] can effectively reduce blocking artifacts and provide good visual quality. In this work, we adopt AOBMC as the interpolation filter to generate two interpolated frames. Because the forward and backward motion vectors are not always identical, especially in occlusion regions, their interpolated frames may be different. Regional blending is used to further occlusion error concealment. First, pixel validity is defined by the variance of corresponding values in two interpolated frames, given by

$$var = (f_{n-0.5,F} - f_{avg})^2 + (f_{n-0.5,B} - f_{avg})^2 \tag{4}$$

Table 1. Adjacent motion candidates selection for unreliable sub-block refinement

Categories	Top-Left	Top-Right	Bot-Left	Bot-Right
ISOLA -TION	TL T1,L1	TR T2,R1	BL L2,B1	BR B2,R2
VERT	L1,R1	L1,R1	L2,R2	L2,R2
HORT	T1,B1	T2,B2	T1,B1	T2,B2
DIAL	TL,BR	T1,R2	L1,B2	TL,BR
DIAR	T2,L2	TR,BL	TR,BL	B1,R1

where $f_{n-0.5,F}$, $f_{n-0.5,B}$ denote a pixel of the forward and backward MCFI frames, and f_{avg} is their average. If var of pixel (x, y) equals 0, (x, y) is considered most reliable. On the contrary, unreliable pixels would have a lager $var(x, y)$ value.

Second, RB is applied to the unreliable pixels. The weight of RB is based on the reliability of pixels in the two interpolated frames, given by

$$\begin{aligned} w_F &= (f_{n-0.5,B} - f_{n-0.5,E})^2 / var; \\ w_B &= (f_{n-0.5,F} - f_{n-0.5,E})^2 / var; \end{aligned} \tag{5}$$

The final interpolated frames are reconstructed by the weighted sum of two frames.

3. EXPERIMENTAL RESULT

We demonstrate the performance of the proposed algorithm using several test sequences, which are in the standard CIF (352×288) format. To get the ground truth, we down-sample the test sequences of 30 fps to 15 fps and then up-convert them back to 30 fps. The proposed FRUC algorithm is compared with two benchmark algorithms: Choi et al. algorithm [2] and Wang et al. algorithm [4]. As peak signal-to-noise ratio (PSNR) is not always consistent with the quality perceived by human visual system, in our experiments, the quality of the interpolated frame is evaluated by both subjective and objective comparisons.

3.1. Subjective evaluation

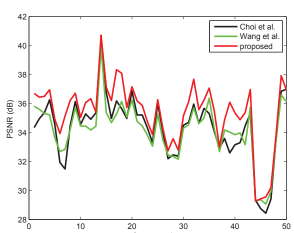
Fig. 5 show the interpolated images of the Foreman, Tennis and Mobile sequences, respectively. The regions located by red circles are occluded artefacts around motion boundaries. For all these sequences, the proposed algorithm reduces blocking artifacts effectively, especially in occlusion regions, and provides significantly better image quality. The quality improvement can be easily observed in the motion boundaries.

(a) Choi et al. (b) Wang et al. (c) Proposed

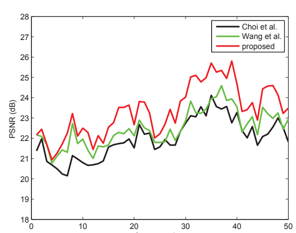
Fig. 5. Visual comparison of interpolated frames by different FRUC algorithms. (a) Choi et al. algorithm, (b) Wang et al. algorithm, (c) proposed algorithm.

3.2. Objective evaluation

Objective quality of the proposed algorithm compared with two benchmark algorithms is evaluated in terms of PSNR. Fig.6 and Table.2 shows the PSNRs of the first 50 interpolated frames the Foreman, Football, Tennis and Mobile sequences. It is seen that the proposed algorithm provides better PSNR performance than the benchmark algorithms. The performance achieves about 1.38dB improvement.

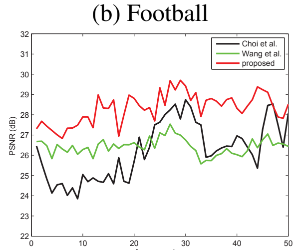


(a) Foreman



(b) Football

(c) Tennis



(a) Mobile

Fig. 6. Comparison of the PSNR performances on the (a) Foreman, (b) Football, (c) Tennis, (d) Mobile sequences.

Table 2. Average PSNR(dB) obtained by proposed algorithm and the benchmark algorithms

Sequence	Proposed	Choi et al.	Wang et al.	Gain
Foreman	35.34	34.28	34.18	1.06
Football	23.36	21.98	22.49	0.87
Tennis	30.87	28.57	29.20	1.67
Mobile	28.33	26.22	26.43	1.90
Average	29.48	27.76	28.10	1.38

4. CONCLUSION

In this paper, we proposed an improved FRUC method based on the dual ME and regional blending. To get the true motion in occlusion regions, our proposed algorithm adopt DME from previous and current frames respectively and performed an effective correction of motion vectors. Subsequently, two motion compensation interpolated frames are produced and the accuracy of its pixels is measured. Finally, we proposed RB which controls weighting coefficients adaptively based on the reliability of interpolated pixels. Compared with existing FRUC algorithms, the proposed algorithm can handle occlusion regions effectively and achieves 1.38 dB PSNR improvement in average.

5. ACKNOWLEDGMENT

This paper was supported by National Natuaral Science Founding of China (NSFC) key project (61133009),(61102098), and Chinese National Key Science and Technology Special Program (2009ZX01033-001-002-2).

6. REFERENCES

[1] G. de Haan, P. W. Biezen, H. Huijgen, and O. A. Ojo, “True-motion estimation with 3-d recursive search block matching,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 368–379, October 1993.

[2] 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. Circuits Syst. Video Technol.*, vol. 17, pp. 407–416, April 2007.

[3] A. H. Huang and T. Q. Nguyen, “A multistage motion vector processing method for motion-compensated frame interpolation,” *IEEE Trans. Image Process.*, vol. 17, pp. 694–708, May 2008.

[4] Ci Wang, Lei Zhang, Yuwen He, and Yap-Peng Tan, “Frame rate up-conversion using trilateral filtering,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 886–893, June 2010.

[5] B. Ozkalayci and A. A. Alatan, "Occluison handling frame rate up-conversion," *Int. Conf. on Image and Signal Process. and Anal. (ISPA)*, pp. 149–152, September 2011.