Visual-Weighted Motion Compensation Frame Interpolation With Motion Vector Refinement

Wei Bai, Jiaying Liu*, Jie Ren, and Zongming Guo
Institute of Computer Science and Technology, Peking University, Beijing, P. R. China 100871

***Abstract*— In this paper, we propose a novel frame rate up-conversion algorithm based on joint motion vector refinement and visual-weighted motion compensation interpolation (MCI). It utilizes a hierarchical motion vector refinement to correct inaccurate motion vectors (MVs), which is composed of the global level and the local level. In the global level, distinct inaccurate MVs are detected by global controlling and then corrected by neighborhood information. Afterwards, the local level performs the local controlling to pick out local outliers and re-estimate them with the maximum likelihood method. Finally, plausible weights for each block in the interpolated frame, computed by the similarity index(SSIM), are applied for visual compensation. The experimental results demonstrate that compared with the conventional algorithm EBME, the proposed algorithm achieved the average PSNR by up to 2.7dB while the visual quality improvement is also remarkable.**

I. Introduction

Frame rate up-conversion (FRUC) refers to the technique that generates a higher frame rate from the video with a lower frame rate by producing new frames and inserting them into the original one. It is widely used to convert two display formats with different frame rates, or to remove the temporal redundancy in video coding. The easiest way is frame repeat or temporal averaging. Nevertheless, they fail to handle sequences with high motion. Motion-compensated interpolation (MCI) is soon adopted to employ FRUC using the unidirectional or bidirectional motion estimation(ME) and compensation. It provides better temporal visual experience than frame repeat or temporal averaging. However, the MVs from the bitstream are estimated to maximize the coding efficiency instead of finding true motion. As a result, directly using the received motion vector field (MVF) for MCI may generate blockiness and ghost artifacts in the interpolated results.

In order to find the true motion for the missing frames, Choi *et al.* [1] performed bidirectional ME for each interpolated blocks with additional spatial smoothness constraint. To reduce complexity, Zhai *et al.* [2] presented an adaptive overlapped block bidirectional motion estimation. Both approaches aim to find the reliable motion. Kang *et al.* [3] extended the bidirectional motion estimation by recursively smoothing the motion vector field with a median filter. The drawback of this method is that the smoothing process revises excessive MVs, resulting in even more errors. In [4], Lee *et al.* explicitly classified the types of MVs and then handle them respectively. However, compared with the previous frame, the current frame always has occlusion problems, *i.e.*, some areas that did not appear in the previous frame appeared in the current frame. Consequently, it becomes difficult for the ME methods that we mentioned above to find the right corresponding blocks in the previous and current frames. Therefore artifacts or block effects usually occur in the predicted frame, leading to distinct visual quality deterioration. Moreover, occlusion problems also bring about ambiguity when predicting the middle blocks. It is hard to decide which of the two blocks, the ones in the previous frame and the current frame, should the predicted block be more similar.

Considering the occlusion problem, although no more reference can be obtained from its adjacent frames, neighboring blocks provide important information. For occlusion regions, regardless of which part they belong to, the background or the moving foreground, they are consistent with their neighbors. Hence, we predict the trajectory of occlusion regions without explicit classification. The proposed MV refinement and similarity index-based MCI algorithm, defined as MRS-MCI, initializes with the bidirectional estimated MVs, and then the motion vector refinement process is performed to enhance the accuracy of the predicted motion vectors. Inaccurate MVs are detected with SAD threshold controlling and corrected by coherent MVs. Instead of SAD, the structural similarity(SSIM) [5] between the to-be-interpolated block and the previous/current block, interpreted as visual weight, is computed to work as a weight function in later use. With the refined MVs and weights, MRS-MCI is performed to obtain a visual pleasing whole frame.

The rest of this paper is organized as follows: Section II describes the framework of the proposed algorithm, whereas Section III is devoted for the detection and correction of inaccurate MVs. The similarity index-based weighted MCI process is given in Section IV. Experimental results are shown in Section V. Finally, concluding remarks are given in Section VI.

II. Framework of the Proposed Algorithm

The existence of occlusion problem results in an inaccurate estimation of the MVF, producing artifacts in interpolated frames. In order to achieve a reasonable and favorable intermediate frame, the proposed scheme excludes incorrect MVs hierarchically and applies weight to the motion compensation

*Correspongding author
This work was supported by National Natural Science Foundation of China under contract No.61101078, Doctoral Fund of Ministry of Education of China under contract No. 20110001120117 and the National Basic Research 973 Program of China under Contract 2009CB320907.

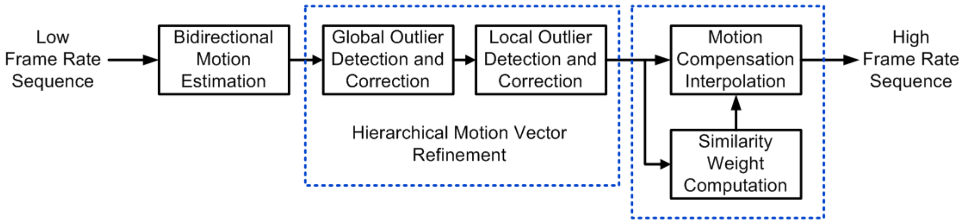

Fig. 1. Flow diagram of the proposed MRS-MCI algorithm.

process to enhance visual quality. Fig.1 shows the entire processing flow of the proposed method including three parts.

- **Bidirectional ME:** The to-be-interpolated frame is divided into non-overlapped blocks of same size and then the block-based motion estimation is performed. Various ME methods can be adopted. In this paper, we utilize bidirectional ME [1] to acquire the initial motion vector field. The bidirectional ME process is depicted in (1). For each block in the to-be-interpolated frame f_n, we find its aligned blocks in the previous frame f_{n-1} and the current frame f_{n+1} by integer-pel ME.

$$\hat{\vec{v}} = \arg\min_{\vec{v}} \sum_{\vec{p} \in B} (|f_{n+1}(\vec{p}+\vec{v}) - f_{n-1}(\vec{p}-\vec{v})|), \quad (1)$$

  where $\vec{v}$ denotes the MV of block B and $\vec{p}$ is the center location of B.
- **Hierarchical motion vector refinement:** The bidirectional estimated motion field is not so accurate that we have to do some post processing to refine the incorrect motion vectors. Without loss of generality, we consider motion refinement in a global-local two-level first. The solution can be easily generalized to a multi-level one by considering some specific refinement, such as the border block handling.
- **Similarity index-based weighted MCI:** The refined MVs are used to perform a similarity weight computation for each block to decide it is more "like" the corresponding block in the previous frame or the current frame. Hence, motion compensation can be performed with the similarity weight applied to the refined MVs.

To well address the occlusion issue, we will employ the last two parts of the MRS-MCI as elaborated in the next sections.

III. Hierarchical Motion Vector Refinement

The MVs acquired with bidirectional motion estimations may not represent the true motion vectors, making it necessary to further refine the estimated motion vector. The proposed MV refinement method performs hierarchically in both global and local levels, or rather from coarse-to-fine scales. As to the coarse-granularity scale, we use global information to correct obvious inaccurate MVs for the whole frame to reduce major errors, such as missing MVs for occlusion areas. And the fine-granularity scale refers to minor adjustments in each local area to improve the quality of the inserted frame.

A. *Global Outlier Detection and Correction*

Generally, a motion vector can be regarded as the inlier, only if it has good matching properties together with spatial coherence with the vectors assigned to the neighboring blocks. Oppositely, an outlier is quite different from its neighboring blocks, either in content or in MVF.

SAD is widely used as a measure to determine the difference between two blocks. As computed in Eq.(2), where f_{n+1}, f_{n-1} are two adjacent original frames, $SAD(B_i)$ denotes the SAD of the i-th block B_i.

$$SAD(B_i) = \sum_{\vec{p} \in B} (|f_{n+1}(\vec{p}+\vec{v}) - f_{n-1}(\vec{p}-\vec{v})|). \quad (2)$$

Then those MVs, which propagate over large SAD, can be considered as outliers, like block B_i in Eq.(3):

$$SAD(B_i) > T, \; T = \alpha \cdot \frac{\sum_{j=1}^{n} SAD(B_j)}{n} (\alpha > 1). \quad (3)$$

where $\frac{\sum_{j=1}^{n} SAD(B_j)}{n}$ indicates the average SAD of all the blocks within a frame. T, the threshold, is then computed as the average number multiplied by α (usually we set α=2). Most probably the more motion the sequence contains, the more outliers to correct, which means a smaller α. After the outliers are detected, we take three steps to dispose them.

Step 1: If the MV $\vec{v_o}$ is an outlier, we search its 8-connectivity MVs as Fig.2(a) to find a MV with the minimum SAD. Let $\vec{v_m}$ denote the found MV, then $\vec{v_m}$ should be considered reliable.

$$\vec{v_m} = \arg\min_{\vec{v_i}} SAD(B_i), \quad (4)$$

where $\vec{v_i}$ is the i-th MV of the outlier's 8-connectivity area.

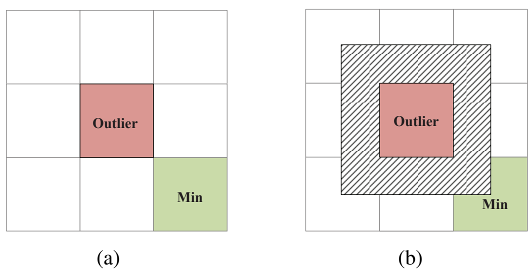

Fig. 2. Global outlier detection and correction. (a) Find the MV with a minimum SAD around the outlier. (b) Enlarge the outlier's block size to research in adjacent frames.

Step 2: The outlier block is so close to the minimum block that $\vec{v_m}$ is probably the true MV for the outlier, hence, we can set the outlier's MV as below: $\vec{v_o}$=$\vec{v_m}$.

Step 3: With the initializing MV $\vec{v_o}$, the outlier block is enlarged of half block size to re-search in the previous and current frames for a new MV. Enlarged block contains more information, which reduces mismatch errors, especially in smooth regions. In this way, the motion vector field is updated with less erroneous MVs.

B. Local Outlier Detection and Correction

In the local level, we use a window W to slide through the whole image with the similar detection method in Sec.3.1 to find the local outliers. After the outliers are detected, they are corrected by selecting the most appropriate vector from a candidate set in terms of the maximum likelihood described as follows.

Centered at the outlier block as in Fig.3, we divide its surrounding 8-connectivity blocks into 16 blocks by half-block step. Because 8 blocks partition are not adequate to reflect the trend of the local area. Then re-search the newly built blocks in the previous and current frames initialized with its neighboring MVs. Choose an optimal MV by trying different initials. Thus, we get enough coherent references for the outlier.

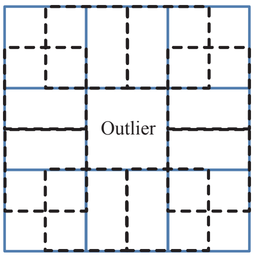

Fig. 3. Local outlier detection and correction. The outlier's neighborhood is divided into overlapped blocks and their MVs contribute to the refinement of the outlier.

As stated before, no matter what kind of motion the outlier belongs to, we can utilize its neighbor to predict its MV. Let $\{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ denote the MVs of the neighboring blocks, $\hat{\vec{v}}$ denotes the outlier's MV. According to the maximum likelihood estimation, under the assumption of Gaussian stationary local MVF, $\hat{\vec{v}}$ is predicted by the mean value of $\{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ as depicted in Eq.(5).

$$\hat{\vec{v}} = \frac{1}{n}\sum_{i=1}^{n} \vec{v_i}. \tag{5}$$

IV. Similarity Index-based Weighted MCI

Although we interpolate the frame in the middle of the previous and current frames, it does not mean the moving object locates in the middle of the motion trajectory. In general cases, the middle frame should be more similar to the block either in the previous or the current frame so as to be visually smooth.For example, Fig.4 shows successive frames from one sequence. Note the corresponding blocks, marked by white squares. In the previous frame,the man's face is obstructed by an arm while in the current frame it is revealed. Fig.4(e) presents the problem, from the adjacent frames, we do not know whether the man's face is still hid behind the arm. If we still interpolate by averaging the corresponding blocks, there would be artifacts or fuzzy boundaries. The proposed MRS-MCI implements a weighted process to avoid such ambiguity. Since we need to compute the "structural" similarity between blocks, SAD is no longer proper. Thus, we utilize the SSIM to measure the similarity.

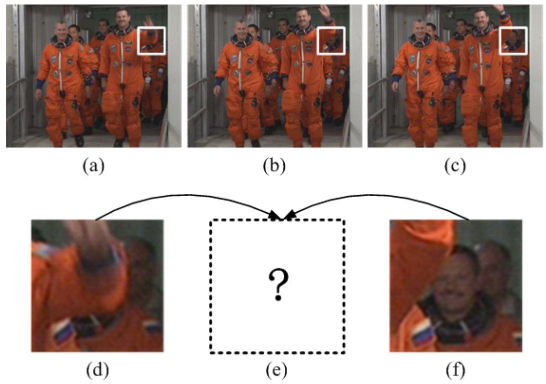

Fig. 4. The ambiguity during the interpolation of the middle block. (a) and (c) are successive frames of *Crew* sequence. (b) is the reference for the intermediate frame we want to generate. (d) and (f) zoom into the areas highlighted in (a) and (c). (e) to-be-interpolated area.

S_{n-1} denotes the similarity between the previous block and the predicted block. S_{n+1} denotes the similarity between the current block and the predicted block. Thus, the weights of the two temporally adjacent blocks can be computed like this:

$$\omega_{n-1} = \frac{S_{n-1}}{S_{n-1}+S_{n+1}}, \omega_{n+1} = \frac{S_{n+1}}{S_{n-1}+S_{n+1}}, \tag{6}$$

where ω_{n-1} and ω_{n+1} denote the weight of the previous block and the current block respectively.

As we have obtained the refined MVs and similarity weights, the interpolation process can be performed to get a whole frame. A weighted MCI method for motion compensation is used as bellow:

$$\hat{B_n}(\vec{p}) = \omega_{n-1}(\vec{p}) f_{n-1}(\vec{p}-\vec{v}) + \omega_{n+1}(\vec{p}) f_{n+1}(\vec{p}+\vec{v}), \vec{p} \in \hat{B_n}, \tag{7}$$

where $\hat{B_n}$ is the predicted block.

V. Experimental Results

The performance of the proposed MRS-MCI algorithm has been evaluated through the objective and subjective evaluations. In the objective evaluation, we compared the PSNR values of 50 interpolated frames. The image qualities of interpolated frames constructed using the proposed and existing algorithms are assessed in the subjective evaluation. For experiments, we set the block size to 16×16 and the search range to ± 12. *Flower, Foreman, and Football* are used in the CIF (352×288) format as test sequences.

Table I summarizes the average PSNR for the test video sequences, obtained by the proposed algorithm, MWCI [6] and EBME [3]. The table indicates that the proposed algorithm provides the PSNR improvement of up to 6.9dB compared to MWCI and 2.7dB compared to EBME. The PSNR improvement of the proposed algorithm comes from the use of motion vector refinement and the visual-weighted MCI. The table also

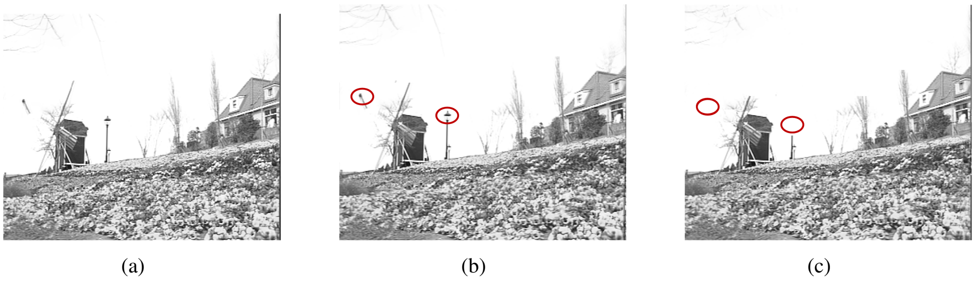

Fig. 5. Comparison of the performance of conventional FRUC and the proposed MRS-MCI. (a) An original frame of *Flower*. Interpolated frame obtained using (b) the conventional MCI (PSNR: 28.12dB), (c) the proposed motion vector refinement (PSNR: 29.75dB).

reflects that sequence *Football* sees no apparent improvement in PSNR by the proposed algorithm. It is because the sequence *Football* itself contains abundant high-motion caused artifacts, which is not suitable for block-based ME.

Further experiments were conducted to verify the effectiveness of each step in the proposed algorithm. The steps include the global MV refinement, the local MV refinement and the SSIM-based weighted MCI. For simulation, we implemented the algorithm step by step. Moreover, results are output after each step was taken. Fig.6 shows that the PSNR gain is improved as a step is added. Note that for most frames the local level refinement does correct some outliers missed by the global level, proving the necessity of this step.

TABLE I

AVERAGE PSNR (dB) COMPARISON OF TEST SEQUENCES.

Method	Test Sequences		
	Flower	Foreman	Football
MWCI [6]	24.48	28.43	19.10
EBME [3]	29.32	31.82	21.31
MRS-MCI	**31.41**	**34.54**	**21.32**

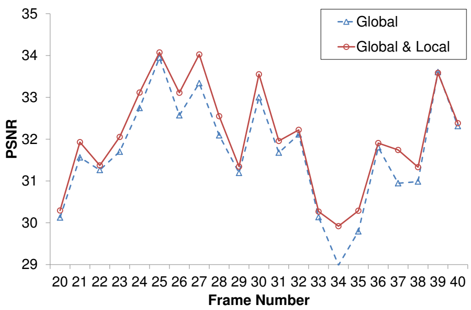

Fig. 6. PSNR value as a function of the frame number with the proposed two-level refinement separately for *Flower* sequence.

We also compared the performance of the proposed algorithm with the conventional MCI [1] method in subjective quality. Fig.5 illustrates the interpolated frames of the test images, *Flower*. Areas that are highlighted by circles should appear based on the original image, Fig.5(b) and Fig.5(c) indicate that the proposed method works more effectively than the conventional method.

Fig.7 demonstrates the effect of visual-weighted MCI, which is one frame of Foreman. Fig.7(c) and Fig.7(d) are two aligned blocks in its adjacent frames, and the proposed algorithm considers the predicted block more "structurally" similar to the previous block Fig.7(c). The interpolated result of the highlighted patch by the proposed algorithm shown in Fig.7(f) is closer to the original in visual quality.

VI. CONCLUSION

In this paper, we presented the MRS-MCI algorithm which differs from the existing algorithms in that it uses a hierarchical MV refinement method and a visual-weighted MCI. In the subjective evaluation, the image quality improvement by the proposed algorithm is clearly visible. The objective evaluation results indicate that the proposed FRUC algorithm outperforms the EBME method by up to 2.7dB in the average PSNR for the test sequences.

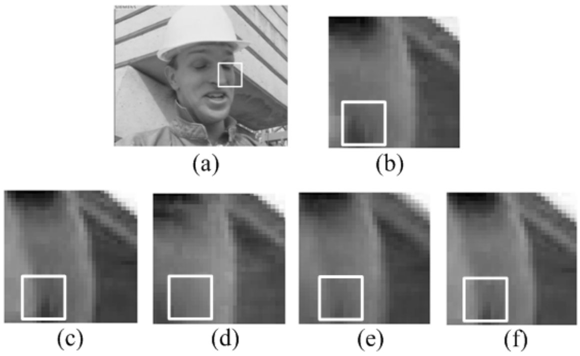

Fig. 7. Zoom of interpolation comparison between conventional method [1] and MRS-MCI. (a) An original frame of *Foreman*, with a patch selected for comparison. (b) Zoom-in of the patch. (c) and (d) are two aligned blocks in adjacent frames. (e) Interpolated by the conventional method. (f) Interpolated by the proposed MRS-MCI algorithm.

REFERENCES

[1] B.-D. Choi, S.-H. Lee, and S.-J. Ko, *New Frame Rate Up-conversion Using Bi-directional Motion Estimation*. IEEE Trans. on Consumer Electronic, vol.46, pp. 603-609, Aug. 2000.

[2] J. Zhai, J. Li K. Yu, and S. Li, *A Low Complexity Motion Compensated Frame Interpolation Method*. Proc. of International Sym. on Circuits and Systems (ISCAS-2005), Japan, Sep. 2005.

[3] S.-J. Kang, K.-R. Cho, and Y. H. Kim, *Motion Compensated Frame Rate Up-conversion Using Extended Bidirectional Motion Estimation*. IEEE Trans. on Consumer Electronics, vol.53, pp.1759-1767, Nov. 2007.

[4] S.-H. Lee, O. Kwon, and R.-H. Park, *Weighted-adaptive Motion Compensated Frame Rate Up-conversion*. IEEE Trans. Consumer Electronics, vol.49, pp.485-492, 2003.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. Simoncelli, *Image Quality Assessment: From Error Visibility to Structural Similarity*. IEEE Trans. on Image Processing, vol.13, pp.600-612, Apr. 2004.

[6] T. Ha, S. Lee, and J. Kim, *Motion Compensated Frame Interpolation by New Block-based Motion Estimation Algorithm*. IEEE Trans. Consumer Electronics, vol.50, pp.752-759, May 2004.