MOTION-COMPENSATED FRAME INTERPOLATION FOR MULTIVIEW VIDEO USING INTER-VIEW AND INTRA-VIEW CORRELATIONS

Xiaohui Yang[1,2], *Zhiquan Feng*[1,2], *Tao Xu*[1,2], *Haokui Tang*[1,2], *Yan Jiang*[1,2]

[1]School of Information Science and Engineering, University of Jinan, Jinan 250022, China
[2]Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan 250022, China

ABSTRACT

A motion-compensated frame interpolation (MCFI) algorithm for multiview video based on inter-view and intra-view correlations is proposed in this paper. First, unidirectional motion estimation (ME) is implemented to obtain forward and backward motion vector fields (MVFs). Subsequently, occlusion blocks in previous and current frames are detected. Then, an inter-view-correlation based method is adopted for occlusion handling. After that, the motion vector (MV) outliers are detected and corrected by considering spatial, color and depth information of current viewpoint video. Finally, MVs are assigned to the interpolated frame for frame reconstruction. Experimental results demonstrate that the proposed algorithm provides better performance than existing 2D and 3D video MCFI methods.

Index Terms— Motion-compensated frame interpolation (MCFI), frame rate up-conversion (FRUC), multiview video, occlusion handling, motion vector processing

1. INTRODUCTION

Multiview video, which consists of color and depth video streams of multiple viewpoints, is one of the most popular 3D video formats [1, 2]. Compare to conventional 2D video, the data amount of multiview video is bigger, which limits the transmission frame rate. If the frame rate of the video cannot satisfy the the refresh rate of the liquid crystal displays (LCD), motion blur will occur. To deal with this problem, frame rate up-conversion (FRUC) is usually adopted at the receiver side to interpolate frames in the temporal domain.

FRUC has been widely applied in the fields of video format conversion, video compression and slow motion playback [3]. The simplest FRUC methods are frame repetition and frame averaging. Nevertheless, serious motion jerkiness and motion blur occur in the up-converted videos via these methods. By considering motion information during intermediate frames interpolation, motion jerkiness and motion blur can be alleviated effectively, and hence, motion-compensated frame interpolation (MCFI) is the mainstream of FRUC [4,5].

Numerous MCFI algorithms have been proposed for conventional 2D video. The motion vectors (MVs) used in MCFI can be retrieved from the received bitstreams directly. E.g., Rüfenacht *et al.* [6] proposed a temporal frame interpolation method upon their highly scalable video coding scheme. Instead, most of the other MCFI algorithms adopt motion reestimation (MRE) to get the true MVs. Block-based motion estimation (BME) is the most popular MRE method as it is simple and easy to implement [7]. For instance, Guo *et al.* [8] performed both forward and backward motion estimation (ME) to obtain the true MVs. Jeong *et al.* [9] developed a multihypothesis motion estimation (MHME) method to improve the ME performance for MCFI. Different from BME, in [10] and [11], the authors utilized pixel-based ME (i.e., optical flow [12]) for MCFI. Unfortunately, the computation complexity of pixel-based ME was high.

Recently, MCFI for 3D video has attracted increasing attention with the rapid development of 3D video systems. Lu *et al.* [13] proposed a MCFI method for depth-based 3D video. In [14], the authors employed depth constraint BME and depth-guided MV filtering in 3D video MCFI. Lee *et al.* [15] proposed a MCFI method for 3D video by motion and depth fusion. Yang *et al.* [16] introduced a 3D video MCFI method via adaptive hybrid motion estimation and compensation. However, these MCFI algorithms were basically designed for 3D video of singular viewpoint without considering the intre-view correlation.

In this paper, we propose a MCFI algorithm for multiview video using inter-view and intra-view correlations. The contributions of this paper are twofold. First, inter-view correlation is adopted for occlusion handling. To the best of our knowledge, the proposed algorithm is the first to deal with the occlusion problems in MCFI using inter-view correlation. Second, we propose an intra-view-correlation based motion vector field (MVF) refinement method to detect and correct MV outliers in depth-continuous area.

The rest of this paper is organized as follows. The details of the proposed algorithm are introduced in Section 2. Section 3 presents the experimental results. Finally, the paper is concluded in Section 4.

This work was supported by the National Natural Science Foundation of China (Nos. 61603151, 61472163), by the National Key Research & Development Plan of China (No. 2016YFB1001403), by the Science and Technology Project of Shandong Province (No. 2015GGX101025), and by the Doctor Fund of University of Jinan (Nos. XBS1621, XBS1653).

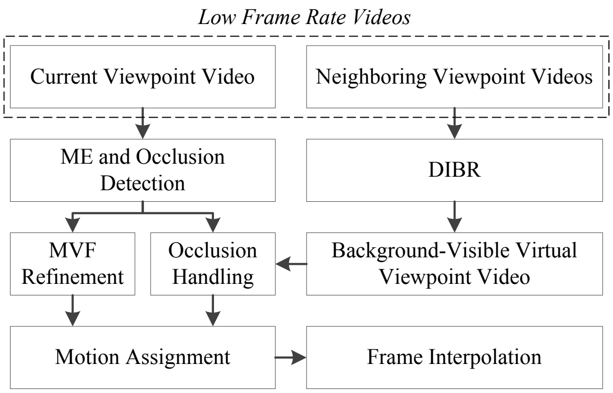

Fig. 1. Flowchart of the proposed multiview video MCFI algorithm.

2. THE PROPOSED ALGORITHM

The flowchart of the proposed multiview video MCFI method is shown in Fig. 1. In our description, the current viewpoint video is the target video for MCFI, and the neighboring viewpoint videos locate at the left and right sides of the current viewpoint. The details of our algorithm are described in the following subsections.

2.1. MV estimation and occlusion detection

In the proposed algorithm, we first utilize unidirectional ME to re-estimate forward and backward MVFs for the previous frame f_{t-1} and current frame f_{t+1} of the current viewpoint video. Let d_{t-1} and d_{t+1} be the depth frames corresponding to f_{t-1} and f_{t+1}. Then, the occlusion blocks in f_{t-1} (or f_{t+1}) can be detected based on their depth variances. Furthermore, these occlusion blocks in f_{t-1} (or f_{t+1}) are classified into covering and uncovering blocks using depth and motion information [16]. The MVs of the occlusion blocks and non-occlusion blocks are named as occlusion MVs and normal MVs in this paper, respectively.

2.2. Inter-view-correlation based occlusion handling

Generally, the occlusion MVs do not reflect the true motion. Let B^c_{t-1} (or B^u_{t-1}) denote a covering block (or an uncovering block) in f_{t-1}, and B^c_{t+1} (or B^u_{t+1}) is a covering block (or an uncovering block) in f_{t+1}. Then, every pixel of B^u_{t-1} (or B^c_{t+1}) has a corresponding pixel in f_{t+1} (or f_{t-1}). On the contrary, some pixels in B^c_{t-1} (or B^u_{t+1}) are occluded by the foreground object in f_{t+1} (or f_{t-1}). In our algorithm, quadtree-based partition is adopted to deal with the occlusion problem. We split the occlusion blocks into four quadrants. Recursively, the quadrants are split into four subquadrants until (a) the block size is 4×4, or (b) the depth variance of the split block is smaller than a threshold.

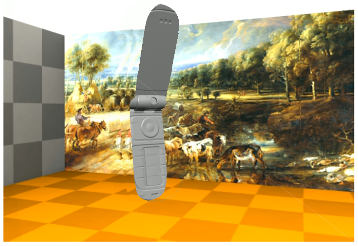

Fig. 2. Background-visible virtual viewpoint video of *Mobile*.

Meanwhile, the neighboring viewpoint videos are used to generate the background-visible virtual viewpoint video via depth image based rendering (DIBR) technology. Different from conventional view synthesis [2], if more than one points in 3D space are projected to the same location in the virtual view frame plan, the point which is furthest to the virtual camera is chosen as the best candidate. In this way, the background regions which are occluded by the foreground object boundaries are exposed in the virtual view frame, as shown in Fig. 2. f^v_{t-1} and f^v_{t+1} are frames in the background-visible virtual viewpoint video corresponding to f_{t-1} and f_{t+1} respectively.

Suppose Q is a final sub-block after quadtree-based partition, and the maximum depth value of the pixels in Q is $\bar{d}'$. $\bar{d}$ is the average depth value of the block before quadtree-based partition. If $\bar{d}' < \bar{d}$, Q is a background sub-block. Otherwise, Q is a foreground sub-block. The MV of Q is estimated as:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\in S} \{\mathrm{SAD}\left(\mathbf{v}\right)\} \tag{1}$$

where $\hat{\mathbf{v}}$ denotes the estimated MV for Q. $\mathbf{v}$ represents a MV candidate within the search range S. If Q is split from B^u_{t-1}, or Q is a foreground sub-block split from B^c_{t-1}, the SAD value is calculated as (2). Then, if Q is a background sub-block split from B^c_{t-1}, (3) is selected to calculate the SAD value. On condition that Q is split from B^c_{t+1}, or Q is a foreground sub-block split from B^u_{t+1}, (4) is used for SAD value calculation. Otherwise, if Q is a background sub-block split from B^u_{t+1}, we calculate the SAD value as (5).

$$\mathrm{SAD}\left(\mathbf{v}\right) = \sum_{\mathbf{p}\in Q} \left| f_{t-1}\left(\mathbf{p}\right) - f_{t+1}\left(\mathbf{p}+\mathbf{v}\right) \right| \tag{2}$$

$$\mathrm{SAD}\left(\mathbf{v}\right) = \sum_{\mathbf{p}\in Q} \left| f_{t-1}\left(\mathbf{p}\right) - f^v_{t+1}\left(\mathbf{p}+\mathbf{v}\right) \right| \tag{3}$$

$$\mathrm{SAD}\left(\mathbf{v}\right) = \sum_{\mathbf{p}\in Q} \left| f_{t-1}\left(\mathbf{p}-\mathbf{v}\right) - f_{t+1}\left(\mathbf{p}\right) \right| \tag{4}$$

$$\mathrm{SAD}\left(\mathbf{v}\right) = \sum_{\mathbf{p}\in Q} \left| f^v_{t-1}\left(\mathbf{p}-\mathbf{v}\right) - f_{t+1}\left(\mathbf{p}\right) \right| \tag{5}$$

where vector $\mathbf{p}$ represents pixel location.

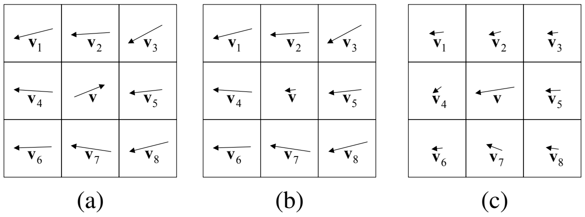

Fig. 3. Different types of MV outliers.

2.3. Intra-view-correlation based MVF refinement

2.3.1. MV outlier detection

It is definitely that some normal MVs in the MVF are unreliable, which causes visual quality degradation in the interpolated frame. In order to improve the performance of frame interpolation, an intra-view-correlation based MVF refinement method is proposed.

The reliability of a MV can be measured based on the spatial correlation in MVF, which can be reflected in two aspects: direction and magnitude. Generally, the direction or magnitude of the MV outlier is different from its spatial neighboring MVs', as shown in Fig. 3. In Fig. 3, $\mathbf{v}$ is the MV being processed, and $\mathbf{v}_i$, $i = 1, \cdots, 8$, are the neighboring MVs. The outlier can be classified into two types: the first type is shown as Fig. 3 (a), the direction of $\mathbf{v}$ is different from the directions of all or most of the neighboring MVs; the second type is shown as Fig. 3 (b) and (c), the magnitude of $\mathbf{v}$ is different from the magnitudes of all or most of the neighboring MVs. Based on this observation, an efficient spatial correlation based MV outlier detection method is proposed.

Direction correlation based outlier detection: The similarity of $\mathbf{v}$ and its neighboring MV $\mathbf{v}_i$, $i = 1, \cdots, 8$ are calculated as:

$$\gamma_i = \frac{\mathbf{v}^T \mathbf{v}_i}{\|\mathbf{v}^T\| \, \|\mathbf{v}_i\|}, \quad i = 1, \cdots, 8 \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean distance of a vector. Obviously, γ_i reflects the direction correlation between $\mathbf{v}$ and its neighboring MV $\mathbf{v}_i$, $i = 1, \cdots, 8$. Then, the numbers of positive and non-positive value in $\gamma_i, i = 1, \cdots, 8$ are denoted as n_p and n_n. If $n_p < n_n$, $\mathbf{v}$ is marked as an outlier. Otherwise, $\mathbf{v}$ will be further detected according to magnitude correlation.

Magnitude correlation based outlier detection: If $\mathbf{v}$ cannot be determined as an outlier by direction correlation based outlier detection, then, the magnitudes of $\mathbf{v}$ and its neighboring MVs that satisfy $\gamma_i > 0$ are calculated. Let $\mathcal{V} = \{\|\mathbf{v}\|, \|\mathbf{v}_i\| \,|\gamma_i > 0\}$, if $\|\mathbf{v}\|$ is the minimum or maximum in $\mathcal{V}$, $\mathbf{v}$ is marked as a possible outlier. Otherwise, $\mathbf{v}$ is marked as a reliable MV.

2.3.2. MV outlier correction

Next, the outliers and possibly outliers are refined by jointly considering the spatial, color and depth information as follows:

$$E = E_{spatial} + \lambda_c E_{color} + \lambda_d E_{depth} \tag{7}$$

$$\mathbf{v} = \arg\min_{\mathbf{v}_j \in \Omega} \{E(\mathbf{v}_j)\} \tag{8}$$

where $\mathbf{v}$ is the refined MV. $\mathbf{v}_j$ is a MV candidate in set Ω. If the target MV is an outlier, Ω is composed by reliable MVs within a neighborhood. Otherwise, if the target MV is an possible outlier, Ω contains the possible outlier itself in addition. λ_c and λ_d are the regularization parameters.

The spatial term $E_{spatial}$, the color term $E_{temporal}$ and the depth term E_{depth} in (7) can be calculated as follows:

$$E_{spatial}(\mathbf{v}_j) = \frac{1}{K} \sum_{\mathbf{v}_k \in R} \|\mathbf{v}_j - \mathbf{v}_k\|_2$$

$$E_{color}(\mathbf{v}_j) = \frac{1}{N^2} \sum \sum \left| B_{t-1}\left(\mathbf{p} + \frac{\mathbf{v}_j}{2}\right) - B_{t+1}\left(\mathbf{p} - \frac{\mathbf{v}_j}{2}\right) \right|$$

$$E_{depth}(\mathbf{v}_j) = \frac{1}{N^2} \sum \sum \left| D_{t-1}\left(\mathbf{p} + \frac{\mathbf{v}_j}{2}\right) - D_{t+1}\left(\mathbf{p} - \frac{\mathbf{v}_j}{2}\right) \right|$$

where $\mathbf{v}_k$ is a one of $\mathbf{v}_j$'s surrounding normal MVs within a neighborhood R, and the block size of B_{t-1} is $N \times N$.

2.4. MV assignment and frame interpolation

In MCFI, unidirectional interpolation usually results in holes and overlaps in the interpolated frame. Thus, bidirectional interpolation is applied to generate the interpolated frame f_t in this paper. In our implementation, first, f_t is divided into blocks with the same size as the blocks used for initial unidirectional ME. Then, the forward and backward MVs are treated as candidates for MV assignment.

Suppose B_t is a block in f_t, and we first find all the MVs pass through it. If only normal MVs pass through it, then the MV with minimum SAD value is chosen as the final MV of B_t. If there is none MV pass through it, we use median filter on normal MVs of neighboring blocks to calculate the final MV for B_t. Denote the final MV as $\mathbf{v}$, and we utilize $\mathbf{v}$ to find the block pair $\{B_{t-1}, B_{t+1}\}$ in f_{t-1} and f_{t+1}. Then, B_t is interpolated as $B_t = \frac{1}{2}B_{t-1} + \frac{1}{2}B_{t+1}$.

Otherwise, if one or more occlusion MVs pass through it, B_t will be divided into 4×4 sub-blocks. Denote Q_t a sub-block, then the occlusion MV with minimum SAD value is chosen as the final MV of Q_t. If there is none occlusion MV pass through it, we use median filter on occlusion MVs of neighboring sub-blocks to calculate the final MV for Q_t. Assume the final MV is $\hat{\mathbf{v}}$, if the SAD of $\hat{\mathbf{v}}$ is calculated using (2) or (4), then Q_t is interpolated as $Q_t = \frac{1}{2}Q_{t-1} + \frac{1}{2}Q_{t+1}$. $\{Q_{t-1}, Q_{t+1}\}$ is the sub-block pair in f_{t-1} and f_{t+1}, respectively. If the SAD of $\hat{\mathbf{v}}$ is calculated using (3), then Q_t is interpolated as $Q_t = Q_{t-1}$. Otherwise, if the SAD of $\hat{\mathbf{v}}$ is calculated using (5), then Q_t is interpolated as $Q_t = Q_{t+1}$.

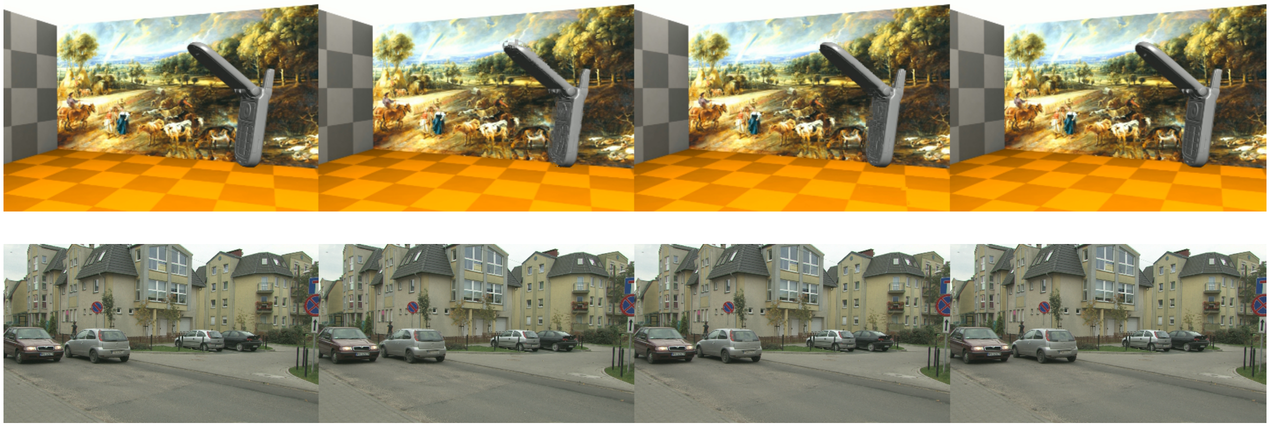

Fig. 4. Interpolated frames of *Mobile* (top) and *PoznanStreet* (bottom), from left to right: Original, Wang's [5], Lee's [15] and Ours.

3. EXPERIMENTAL RESULTS

We evaluate the proposed MCFI algorithm on four multiview video sequences: *BookArrival* (1024×768, 99 frames), *Mobile* (720×540, 199 frames), *Newspaper* (1024×768, 201 frames), and *PoznanStreet* (960×540, 53 frames). Current viewpoints and neighboring viewpoints of these four multiview video sequences are shown in Table 1. The even frames of these sequences are dropped for frame interpolation. A conventional 2D video MCFI algorithm [5] and a 3D video MCFI algorithm [15] are simulated for performance comparison. In our experiment, a fast subpixel ME method [17] is used for unidirectional MVF estimation. The block size for ME is 8×8, and the search range is 32×32 for all of the test sequences.

The averaged PSNR and SSIM values are illustrated in Table 2. From Table 2, we can see that the averaged PSNR values are higher than those of existing methods on the sequences of *BookArrival*, *Mobile*, and *PoznanStreet*. For *Mobile* and *PoznanStreet* sequences, the averaged SSIM values of our algorithm are best. Obviously, the proposed MCFI algorithm provides better performance in objective evaluation. Please note that the depth map qualities of *BookArrival* and *Newspaper* are low, which suggests that the depth quality of the multiview video affects the performance of the proposed algorithm to some extent. And this is one limitation of the proposed method.

Moreover, Fig. 4 shows the interpolated frames of *Mobile* and *PoznanStreet*. It can be observed that, for the benchmark algorithms, obvious motion blur and ghost artifacts occur around the foreground objects, i.e., the occlusion areas on the interpolated frame. However, these motion blur and ghost artifacts are suppressed significantly in the frames interpolated using the proposed method.

Table 1. Multiview video sequences used in our experiments.

	BookArrival	Mobile	Newspaper	PoznanStreet
Current View	Cam8	Cam6	Cam4	Cam4
Left View	Cam10	Cam4	Cam2	Cam5
Right View	Cam6	Cam8	Cam6	Cam3

Table 2. Objective evaluations in averaged PSNR values and SSIM values.

		BookArrival	Mobile	Newspaper	PoznanStreet
Wang's	PSNR	31.6190	35.1748	36.6314	34.1164
	SSIM	0.9460	0.9826	0.9774	0.9584
Lee's	PSNR	31.5683	35.5157	35.8603	34.8967
	SSIM	0.9268	0.9840	0.9650	0.9507
Ours	PSNR	31.7112	36.1144	36.2640	34.9608
	SSIM	0.9323	0.9848	0.9700	0.9621

4. CONCLUSION

In this paper we proposed a MCFI algorithm for multiview video using inter-view and intra-view correlations. First, unidirectional ME is implemented to obtain forward and backward MVFs for the previous and current frames, respectively. Thereafter, occlusion blocks are detected based on the motion and depth information. Then, the occlusion areas are post-processed based on the inter-view correlation. Subsequently, we adopt an intra-view correlation based method to detect and correct the MV outliers by considering spatial, color and depth information of current viewpoint video. Finally, MVs are assigned to the interpolated frame for frame reconstruction. Experimental results show that the proposed algorithm efficiently improves the quality of the interpolated frames.

5. REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Process. Mag.*, vol. 24, no. 7, pp. 10–21, Nov. 2007.

[2] X. Yang, J. Liu, J. Sun, X. Li, and W. Liu, "DIBR based view synthesis for free-viewpoint television," in *Proc. IEEE 3DTV Conf. (3DTV-CON'11)*, May 2011, pp. 1–4.

[3] 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, no. 4, pp. 407–416, Apr. 2007.

[4] Q. Lu, N. Xu, and X. Fang, "Motion-compensated frame interpolation with multiframe-based occlusion handling," *J. Display Technol.*, vol. 12, no. 1, pp. 45–54, Jan. 2016.

[5] C. Wang, L. Zhang, Y. He, and Y.-P. Tan, "Frame rate up-conversion using trilateral filtering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 886–893, Jun. 2010.

[6] D. Rüfenacht, R. Mathew, and D. Taubman, "Bidirectional, occlusion-aware temporal frame interpolation in a highly scalable video setting," in *Proc. Picture Coding Symp. (PCS'15)*, May/Jun. 2015, pp. 5–9.

[7] Y. Dar and A. M. Bruckstein, "Motion-compensated coding and frame rate up-conversion: Models and analysis," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2051–2066, Jul. 2015.

[8] Y. Guo, L. Chen, Z. Gao, and X. Zhang, "Frame rate up-conversion method for video processing applications," *IEEE Trans. Broadcast.*, vol. 60, no. 4, pp. 659–669, Dec. 2014.

[9] S.-G. Jeong, C. Lee, and C.-S. Kim, "Motion-compensated frame interpolation based on multihypothesis motion estimation and texture optimization," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4497–4509, Nov. 2013.

[10] K. Chen and D. A. Lorenz, "Image sequence interpolation using optimal control," *J. Math. Imag. Vision*, vol. 41, no. 3, pp. 222–238, Nov. 2011.

[11] M. Werlberger, T. Pock, M. Unger, and H. Bischof, "Optical flow guided TV-L1 video interpolation and restoration," in *Proc. Energy Minimization Methods Comput. Vis. Pattern Recognit. (EMMCVPR'2011)*, Jul. 2011, vol. 6819, pp. 273–286.

[12] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, Massachusetts Institute of Technology, May, 2009.

[13] Q. Lu, X. Fang, C. Xu, and Y. Wang, "Frame rate up-conversion for depth-based 3D video," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME'12)*, Jul. 2012, pp. 598–603.

[14] Y. Liu, X. Fan, X. Gao, Y. Liu, and D. Zhao, "Motion vector refinement for frame rate up conversion on 3D video," in *Proc. IEEE Vis. Commun. Image Process. (VCIP'13)*, Nov. 2013, pp. 1–6.

[15] Y. Lee, Z. Lee, and T. Nguyen, "Frame rate up conversion of 3D video by motion and depth fusion," in *Proc. IEEE Image, Video, Multidimensional Signal Process. Workshop (IVMSP'13)*, Jun. 2013, pp. 1–4.

[16] X. Yang and Z. Feng, "3D video frame interpolation via adaptive hybrid motion estimation and compensation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'16)*, Mar. 2016, pp. 1691–1695.

[17] S. Chan, D. Vo, and T. Nguyen, "Subpixel motion estimation without interpolation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'10)*, Mar. 2010, pp. 722–725.