IMPROVED VIEW SYNTHESIS BY MOTION WARPING AND TEMPORAL HOLE FILLING

Andrei I. Purica[⋆†]*, Elie G. Mora*[⋆]*, Beatrice Pesquet-Popescu*[⋆]*, Marco Cagnazzo*[⋆]*, Bogdan Ionescu*[†]

[⋆]Institut Mines-Telecom; Telecom ParisTech; CNRS LTCI; 75014, Paris
[†]University Politehnica of Bucharest,061071, Romania
Email:{purica, mora, pesquet, cagnazzo}@telecom-paristech.fr, bionescu@imag.pub.ro

ABSTRACT

View synthesis received increasing attention over the last years, as it offers a wide range of practical applications like Free Viewpoint Television, 3D video, video gaming, etc. The main issues in view synthesis are the filling of disoccluded areas and the warping of real views. In this paper we propose a new hole filling method, it uses temporal correlations in the real views to extract information on disoccluded areas from different time instants in the synthetic view. We also propose a sub-pixel warping technique that takes into account depth and can be used for both the warping of the real view as well as for motion compensation. Our method is proved to bring gains of up to 0.31dB in average over several multiview test sequences.

Index Terms— view synthesis, multiview video, hole filling, depth-image-based-rendering

1. INTRODUCTION

Recent advances in video coding, transmission and display technologies have raised the quality of 3D video representations to an acceptable level for usage in everyday applications. Typical application scenarios include 3D video, immersive teleconference systems and Free View Point Television (FTV) [1] [2]. A common 3D video format for these applications is MultiView Video (MVV), which is composed of several video sequences representing the same scene and acquired from different points of view. Another common representation is the Multiview-Video-plus-Depth format (MVD) [3], where each view is associated with depth information. This representation allows the synthesis of multiple view points at the receiver side. A typical class of synthesis methods based on this format are Depth-Image-Based-Rendering (DIBR) techniques [4].

The Moving Pictures Experts Group (MPEG) showed a high level of interest for MVD formats and their ability to support multiview video applications. An experimental framework for this format was developed in the standardization process of a 3D extension of the High Efficiency Video Coding standard [5]. The framework also defines a View Synthesis Reference Software (VSRS), which allows additional views to be rendered from MVD sequences using DIBR methods [6].

A common problem in view synthesis is caused by areas of the scene, which are occluded from one point of view but are visible in another. When rendering a new video from a new point of view, these areas will appear as "holes" in the synthetized image, also known as disocclusions. Traditionally this problem is solved with inpainting algorithms such as those described in [7] or [8]. Two popular inpainting algorithms were developed by Bertalmio and Sapiro [9] and Criminisi *et al.* [10]. However, the temporal correlations in a video sequence allow for a different approach, which can retrieve real scene information. Because disoccluded content is typically part of the background, a first class of methods propose to fill in disocclusions with background information extracted from multiple time instances of a real view [11] [12]. Other methods use temporal correlations in the synthesized view in order to retrieve information from different time instants. Shimizu and Kimata [13] use motion estimation directly in the rendered view to improve the synthesis. Chen *et al.* [14] use block-based motion estimation in adjacent views and retrieve information about disoccluded areas by warping the start and end point of the motion vectors.

In this paper we present a new temporal hole filling scheme. We use dense motion vector fields computed with optical flow [15] and warp them at the level of the synthetized view by imposing an epipolar constraint [16]. Furthermore we introduce a simple warping technique that can be used for both motion compensation and DIBR warping. This technique is used to warp the real adjacent views, and then to motion compensate a past or future frame using derived motion vectors from the left or right base views. This allows us to partly fill disocclusions with real background information from other temporal instants. The remaining holes can be filled with any inpainting algorithm.

The remainder of the paper is organized as following. In Sec. 2 we present our temporal hole filling scheme and Sec. 3 describes our warping technique. The experimental results are reported in Sec. 4 and Sec. 5 concludes the paper.

2. TEMPORAL HOLE FILLING

In general, disocclusions in the synthesized view can be classified in two categories depending whether the area in the reference view is a border or non-border occlusion with respect to the image [17]. Border occlusions occur due to the reference image missing portions of the field of view that should be visible in the synthesized view. This types of occlusions are resolved by performing a synthesis from a left and a right reference view. The non-border occlusions are caused by objects in the foreground that obscure parts of the background that should be visible in the synthesis. Due to the motion of the foreground objects and camera, this types of occlusions vary over time and produce different holes at different time instants in the synthesized view. Thus, part of the missing information may be available in frames at different time instants. By exploiting the temporal correlation in the video sequence we can retrieve additional information and reduce the size and the number of holes in the synthesis.

In Fig. 1, a foreground object is represented in two views at two different time instants, black arrows represent the motion vector field and disparity fields for a past and current (c) time instant ($\mathbf{v}_l$, $\mathbf{d}_p$, $\mathbf{d}_c$)

[†]Part of this work was supported under ESF InnoRESEARCH POSDRU/159/1.5/S/132395 (2014-2015).

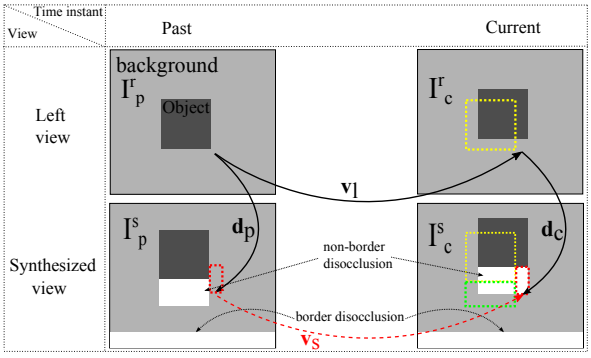


Fig. 1. Temporal retrieval of disoccluded area. Yellow dotted squares mark the position of the object in the previous frame, the green dotted square shows the disocclusions in the previous frame and red dotted squares mark the disoccluded area that was visible in a previous frame.

and the red dashed arrow represents the motion vectors in the synthesized view, which can be used to retrieve information about the disoccluded area. Yellow and green dotted lines show the position of the object and disoccluded area respectively, in the past frame. It can be observed that a part of the disoccluded area in the current frame was visible in a past frame due to the motion of the object (this is shown in the figure with a dotted red line). In Fig. 2 we show the relation between motion vectors field (MVF) and disparity maps for three views of a MVD sequence. Let us consider two base views, left and right, and an intermediate view, which is synthesized at the decoder side using classic DIBR algorithms. I^l, I^r and I^s denote frames from the left, right and synthesized views respectively at a past, current or future time instant (p, c, f). $\mathbf{v}$ and $\mathbf{d}$ are the MVF and disparity maps respectively.

The projection of a real world point in the synthesized view can be modeled using disparity maps. Let us consider a point $\mathbf{k} = (x, y)$ in frame I_p^l. Its projection in the synthesized frame is obtained by warping the point to a new position using the vector $\mathbf{d}_p^l(\mathbf{k})$. By compensating the motion in I_p^s with the vector $\mathbf{v}_{pl}^s(\mathbf{k} + \mathbf{d}_p^l(\mathbf{k}))$ we obtain the projection of the point at the current time instant in frame I_c^s. The same projection should be obtained by first compensating the motion in the left view and then warping the point with the associated disparity vector $\mathbf{d}_c^l(\mathbf{k} + \mathbf{v}_p^l(\mathbf{k}))$.

This condition defines a so-called *epipolar constraint* [16] for point $\mathbf{k}$, written as:

$$\mathbf{v}_p^l(\mathbf{k}) + \mathbf{d}_c^l(\mathbf{k} + \mathbf{v}_p^l(\mathbf{k})) = \mathbf{d}_p^l(\mathbf{k}) + \mathbf{v}_{pl}^s(\mathbf{k} + \mathbf{d}_p^l(\mathbf{k})) \quad (1)$$

Based on Eq. 1, $\mathbf{v}_{pl}^s$ can be derived from $\mathbf{v}_p^l$, $\mathbf{d}_p^l$ and $\mathbf{d}_c^l$ as:

$$\mathbf{v}_{pl}^s(\mathbf{k} + \mathbf{d}_p^l(\mathbf{k})) = \mathbf{v}_p^l(\mathbf{k}) + \mathbf{d}_c^l(\mathbf{k} + \mathbf{v}_p^l(\mathbf{k})) - \mathbf{d}_p^l(\mathbf{k}) \quad (2)$$

In other words $\mathbf{v}_{pl}^s$ is obtained by warping $\mathbf{v}_p^l$ with the disparity map $\mathbf{d}_p^l$ and then adjusting the motion intensity with the difference in disparity for point $\mathbf{k}$ at current and past time instants. This method can be applied for past and future time instants using either the left or right view. A dense motion vector field in the base views can be obtained from the current frame and either a future or past one with an optical flow algorithm [18]. Assuming we are dealing with a 1D parallel camera setup, the disparity maps only have an x component, which is easily computed from the corresponding depth maps of each base view [2] as:

$$\mathbf{d}(\mathbf{k}) = f \cdot B \left[\frac{Z(\mathbf{k})}{255} \left(\frac{1}{Z_{min}} - \frac{1}{Z_{max}} \right) + \frac{1}{Z_{max}} \right] \quad (3)$$

where $Z(\mathbf{k})$ is the inversed depth value of point $\mathbf{k}$, Z_{min} and Z_{max}

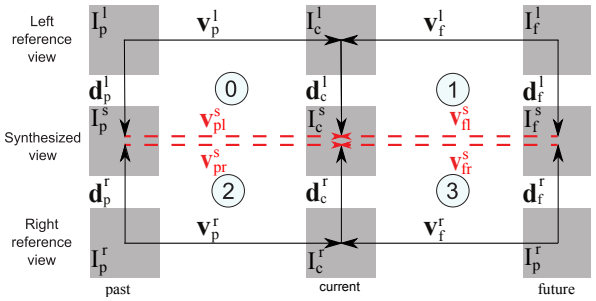


Fig. 2. Temporal hole filling scheme, for two base views and an intermediary synthesis, using past and future synthesized frames to retrieve information.

are the minimum and maximum depth values respectively, f is the focal length of the camera and B is the baseline between the synthesized and base views.

If we decompose Eq. 2 for the x and y components, we have:

$$\begin{aligned} &\mathbf{v}_{pl,x}^s(x + \mathbf{d}_{p,x}^l(x,y), y) = \\ &\mathbf{v}_{p,x}^l(x,y) + \mathbf{d}_{c,x}^l(x + \mathbf{v}_{p,x}^l(x,y), y + \mathbf{v}_{p,y}^l(x,y)) - \mathbf{d}_{p,x}^l(x,y) \\ &\mathbf{v}_{p,y}^l(x + \mathbf{d}_{p,x}^l(x,y), y) = \mathbf{v}_{p,y}^l(x,y) \end{aligned} \quad (4)$$

The MVF ($\mathbf{v}_{pl}^s$) may contain holes after warping it at the level of the synthesized view. Since I_c^s is obtained through warping I_c^l with $\mathbf{d}_c^l$, $\mathbf{v}_{pl}^s$ will not be defined in positions coinciding with disoccluded areas in I_c^s, thus making it impossible to fill this area. Using a different time instant will result in the same holes. The backward motion compensation ensures that the holes correspond with the disocclusions when warping I_p^l or I_f^l to I_p^s and I_f^s respectively, when using the left base view. Disocclusions at different time instants do not necessary coincide and so additional information can be obtained for the current frame using motion compensation. Finally, we are able to obtain 4 MVFs as shown in Fig. 2 and multiple temporal predictions of the same disoccluded area are averaged. Since, integer rounding of motion or disparity predictors is inefficient, we propose a sub-pixel precision warping to be used in parallel with this method.

3. SUB-PIXEL PRECISION WARPING

In addition to warping the side views as traditionally done in view synthesis, our method requires a warping of the motion vector fields in the adjacent views and a backward motion compensation of past or future frames. In practice our hole filling method requires a number of different warping operations, namely: typical DIBR image warping, a warping of the dense MVFs, a motion compensation of disparity and a backward motion compensation of the image. In order to take full advantage of dense motion vector fields and depth computed disparity we propose a simple technique for sub-pixel precision warping and backward motion compensation.

To better describe our method let us consider the motion vector field $\mathbf{v}_s(\mathbf{k})$, the disparity field $\mathbf{d}_c(\mathbf{k})$, the images I_p^s,I_c^r and a warped image defined on possibly fractional positions I^{fg}. $\mathbf{u} = (x, y)$ represents a set of coordinates in I^{fg}. Each position $\mathbf{k} = (c, r)$ in the image, MVF or disparity field will correspond to a position $\mathbf{u}$ in I^{fg} through the function τ as shown in Eq. 5:

$$\tau(\mathbf{k}) = \mathbf{u}, \qquad \tau(\mathbf{k}) = (c/\alpha, r/\alpha) \tag{5}$$

where α is defined as $1/t$ and $t \in \mathbb{N}$ is used to indicate the precision of the warping. Considering that $\mathbf{v}_s$ and $\mathbf{d}_s$ contain fractional values, the goal is to perform a sub-pixel warping of I_c^r with the disparity field $\mathbf{d}_c$ and to backward motion compensate I_p^s image using the derived MVF. A first step is to quantize the values in $\mathbf{d}$ in function of the precision parameter α as shown in Eq. 6:

$$\Phi_\alpha(x, y) = (\lfloor \frac{x}{\alpha} + \alpha \rfloor \alpha, \lfloor \frac{y}{\alpha} + \alpha \rfloor \alpha) \tag{6}$$

where Φ_α is a rounding operation and "$\lfloor \rfloor$" indicates a floor operation. The quantized values of disparity and motion vectors are obtained by applying Φ over the two vector fields. The actual synthesis is performed in three steps, a warping of the inter-view reference image I_c^r in I^{fg}, a filtering step and a temporal hole filling.

The I_c^r image is warped in I^{fg} as shown in Eq. 7:

$$I^{fg}(\tau(\mathbf{k} + \Phi_\alpha(\mathbf{d}_c(\mathbf{k})))) = I_c^r(\mathbf{k}) \tag{7}$$

Overlapping values in I^{fg} will be dealt with by using the disparity information, which relates to depth as shown in Eq. 3. High disparity indicates an object in the foreground and should be considered over a point with low disparity value. Nevertheless, overlaps should be marked and both values should be considered in the filtering step described in what follows.

In Fig. 3 we show an example of sub-pixel precision warping using our proposed method. We have a luminance matrix (top-left) with the corresponding disparity or MV field for X-axis (top-right) and Y-axis (bottom-left). On the right side of the image a fractional grid is displayed after displacing the pixels from the luminance image using Eq. 7. With dotted lines we represent 2 examples of filtering windows. Green indicates a hole and red an overlapping between foreground and background. The final luminance image (bottom-right) is obtained by centering a filtering window in each position $u = \tau(\mathbf{k})$. The output of the filter is obtained in two steps. First we identify the foreground luminance values by creating a list of pixels found in the filtering window and ordering them with respect to their associated depth in the reference image I_c^r. All $\{s, .., n\}$ positions in our list are then interpolated to obtain the final value, s is obtained as the smallest value that satisfies $\Delta(s) > \beta$ and Δ is defined as:

$$\begin{aligned} \mathcal{L} = \{d_1, .., d_i, .., d_n\}, \quad \mathcal{L}_{dif} = \{\delta_1, \delta_2, .., \delta_{n-1}\} \\ \Delta(i) = \frac{\delta_i - \delta_{i-1}}{\delta_{i-1}} \end{aligned} \tag{8}$$

where d_i are depth values, $\mathcal{L}$ is the list, $\delta_i = d_{i+1} - d_i$ and β is an empirically determined threshold. Finally, we apply the temporal hole filling algorithm for unknown areas. We use derived motion vectors from the adjacent views as shown in Sec. 2 to backward motion compensate a past or future synthesized frame and extract additional information about the disoccluded area in the current frame. Note that past and future motion reference frames do not have an associated depth map, in this case when we derive a vector from the left or right MVFs to $\mathbf{v}_s$ we retain the corresponding depth from left and right, future and past frames, see Fig. 2. Additional unfilled disocclusions are marked for inpainting.

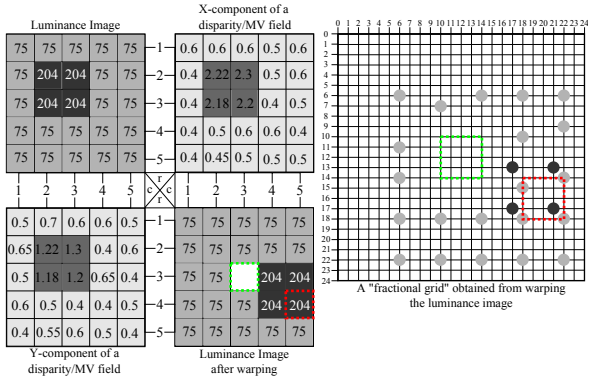


Fig. 3. A simple sub-pixel precision warping example with our proposed technique. Dotted lines represent filtering windows and the corresponding result in the warped image; green indicates a hole and red a case of foreground and background overlapping.

4. EXPERIMENTAL RESULTS

We test our method on four multiview sequences defined in the Common Test Conditions (CTCs) for conducting experiments with the reference software of 3D-HEVC [19]: Balloons, Kendo, Newspaper and PoznanHall2. For each sequence we consider two non-adjacent reference views and we synthesize a middle view with our method and the reference VSRS1D-Fast rendering used in 3D-HEVC [6] experiments. In order to have a fair comparison the remaining disocclusions in our synthesis use the same filling as the reference. Each of the tested sequences is encoded using the configuration described in the CTCs. Four different QPs (25 30 35 40) are used for the texture encoding, the depth maps are encoded using corresponding QPs (34 39 42 45) as indicated by the CTCs. For more details on the sequences check [20].

We evaluate the PSNR of the synthesis against original views for each sequence at each of the tested QPs. The encoding is performed with 3D-HEVC, the left view is set as base view, and the right as dependent view. The GOP size is set to 8, and the first frame of each GOP is used as a reference frame for temporal hole filling of the other frames of the GOP, inside the synthesized view. These reference frames are synthesized with VSRS1D-Fast. In our experiments we set β parameter to $1/10$ and α to $1/4$, and the size of the filtering window to 5, we found these values to provide best gains. The dense MVFs are computed using the optical flow algorithm in [18] between frames of the reference views. The optical flow parameters used in our experiments along with more details can be found in [21]. Tab. 1 shows the PSNR results for our method and the reference, for each tested sequence and QP. We can see that the proposed method outperforms the reference on all tested sequences, obtaining and overall average gain of 0.31dB. Using a different metric like SSIM will yield similar results for the proposed (0.9283) and reference (0.9276) methods, on average over tested sequences.

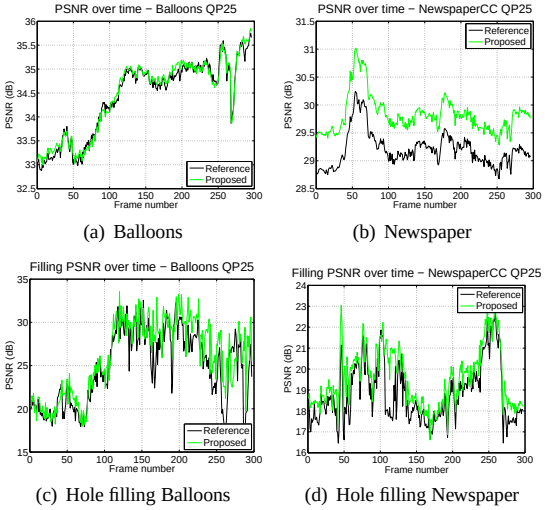
Tab. 2 shows the PSNR results on disoccluded areas. The same filling was used for both methods for remaining holes. This results reflect the improvement achieved only through temporal hole filling. We can see that even though only a part of the disoccluded areas is completed with temporal predicted pixels (as described in Sec. 2, see Fig. 1) we are able to achieve a good PSNR improvement. Note that these gains only reflect disoccluded areas, which represent a small

Table 1. Average PSNR and gain for each sequence and each QP.

Sequence	**Reference PSNR (dB)**				**Proposed PSNR (dB)**				**Gain (dB)**				**Avg. Gain (dB)**
QPs	25	30	35	40	25	30	35	40	25	30	35	40	
Balloons	34.41	34.12	33.47	32.45	34.45	34.19	33.57	32.55	0.04	0.08	0.1	0.1	0.08
Kendo	35	34.53	33.79	32.77	35.4	34.92	34.17	33.1	0.4	0.39	0.38	0.32	0.37
Newspaper	29.2	29.05	28.78	28.31	29.83	29.71	29.4	28.84	0.63	0.66	0.62	0.53	0.61
PoznanHall2	36.25	35.87	35.36	34.55	36.35	36.03	35.62	34.78	0.11	0.15	0.26	0.23	0.18

Table 2. Average PSNR and gain for disoccluded areas for each sequence and each QP.

Sequence	**Reference PSNR (dB)**				**Proposed PSNR (dB)**				**Gain (dB)**				**Holes (%)**	**Avg. Gain (dB)**
QPs	25	30	35	40	25	30	35	40	25	30	35	40		
Balloons	24.73	24.89	24.77	24.27	26.08	26.01	25.86	25.3	1.34	1.12	1.09	1.03	0.11	1.14
Kendo	25.51	25.73	25.99	26.03	26.51	26.72	26.52	26.41	1	0.99	0.53	0.38	0.08	0.72
Newspaper	18.83	19.13	18.98	19.5	19.57	19.74	20.01	20.13	0.74	0.61	1.04	0.63	0.3	0.755
PoznanHall2	28.68	27.85	28.52	28.67	30.94	29.77	28.67	29.71	2.26	1.92	0.15	1.04	0.04	1.34



(a) Balloons (b) Newspaper

(c) Hole filling Balloons (d) Hole filling Newspaper

Fig. 4. PSNR variation of the middle synthesized view over time for the reference and proposed method at QP 25 in Balloons and Newspaper sequences. 4(a), 4(b):full frame; 4(c), 4(d):disoccluded areas.

percentage of the image, as shown in the table. The gain obtained for the entire frame comes from both temporal hole filling and proposed warping.

In Fig. 4 we show the PSNR comparison between our proposed method and the reference one for Balloons and Newspaper sequences. Out of the four tested sequences our method has the lowest gain on Balloons sequence and the highest gain on Newspaper sequence. For brevity reasons we only show the result for QP25, the behavior is similar across all QPs. In Figs. 4(a) and 4(b) the PSNR is computed over the entire frame and in Figs. 4(c) and 4(d) the PSNR is computed over the disoccluded areas. We can see that our method outperforms the reference throughout the sequences on both full frame and disocludded areas. In Fig. 5 we show an example of the difference between the absolute errors of VSRS1D-Fast and our proposed synthesis on frame 15 of Newspaper sequence. Green to red colors indicate our method has a lower error. We can see high error pixels around the edges of object from both our method (red) and the reference (blue), however, it is easily noticeable that for most areas of the image our method offers a better prediction with a lower error.

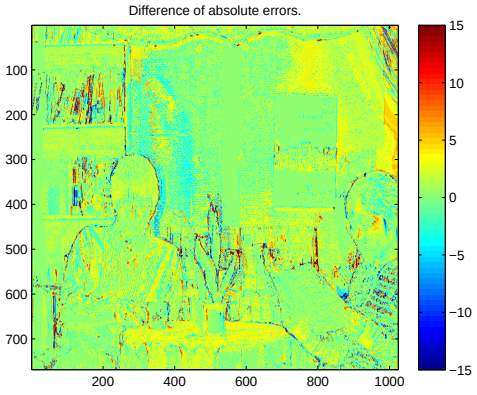


Fig. 5. Difference between absolute errors for frame 15 from Newspaper sequence.

5. CONCLUSION

In this paper, we presented a temporal hole filling method based on motion derivation and a sub-pixel precision warping technique that can be applied for both DIBR warping and motion compensation. Real information on disoccluded areas is retrieved from previously synthesized past or future frames in order to reduce holes in the synthesis. This method is very robust and can be used with any motion estimation technique and a variety of schemes for the reference past and future frames. Our method brings gains of up to 0.31dB PSNR in average over the VSRS1D-Fast rendering software in 3D-HEVC for several test sequences.

6. REFERENCES

[1] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Tomohiro Yendo, "Free-Viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, pp. 67–76, 2011.

[2] Frederic Dufaux, Beatrice Pesquet-Popescu, and Marco Cagnazzo, Eds., *Emerging technologies for 3D video: content creation, coding, transmission and rendering*, Wiley, May 2013.

[3] Philipp Merkle, Aljoscha Smolic, Karsten Muller, and Thomas Wiegand, "Multi-view video plus depth representation and coding," *IEEE International Conference on Image Processing*, vol. 1, pp. 201–204, 2007.

[4] Christoph Fehn, "A 3D-TV approach using depth-image-based rendering," in *3rd IASTED Conference on Visualization, Imaging, and Image Processing*, Benalmadena, Spain, 8-10 September 2003, pp. 482–487.

[5] "High Efficiency Video Coding," ITU-T Recommendation H.265 and ISO/IEC 23008-2 HEVC, April 2013.

[6] Li Zhang, Gerhard Tech, Krzysztof Wegner, and Sehoon Yea, "3D-HEVC test model 5," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JCT3V-E1005, July 2013.

[7] Christine Guillemot and Oliver Le Meur, "Image inpainting: Overview and recent advances," *IEEE Signal Processing Magazine*, vol. 31, pp. 127–144, 2014.

[8] Ismael Daribo and Beatrice Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE MMSP*, Saint Malo, France, 4-6, October 2010.

[9] Marcel Bertalmio and Guillermo Sapiro, "Image inpainting," in *SIGGRAPH*, New Orleans, USA, Jully 2000, pp. 417–424.

[10] Antonio Criminisi, Patrick Perez, and Kentaro Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[11] Wenxiu Sun, Oscar C. Au, Lingfeng Xu, Yujun Li, and Wei Hu, "Novel temporal domain hole filling based on background modeling for view synthesis," in *IEEE International on Image Processing (ICIP)*, Orlando, FL, 30 Sept. - 3 Oct. 2012, pp. 2721 – 2724.

[12] Katta Phani Kumar, Sumana Gupta, and K. S. Venkatesh, "Spatio-temporal multi-view synthesis for free viewpoint television," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, Aberdeen, 7-8 October 2013, pp. 1 – 4.

[13] S. Shimizu and H. Kimata, "Improved view synthesis prediction using decoder-side motion derivation for multiview video coding," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2010, pp. 1–4.

[14] Kuan-Yu Chen, Pei-Kuei Tsung, Pin-Chih Lin, Hsing-Jung Yang, and Liang-Gee Chen, "Hybrid motion/depth-oriented inpainting for virtual view synthesis in multiview applications," *3DTV-CON*, pp. 1–4, 7-9 June 2010.

[15] Frederic Dufaux, Marco Cagnazzo, and Beatrice Pesquet-Popescu, *Motion Estimation - a Video Coding Viewpoint*, vol. 5: Image and Video Compression and Multimedia of *Academic Press Library in Signal Processing*, Academic Press, 2014.

[16] Ismael Daribo, Wided Milded, and Beatrice Pesquet-Popescu, "Joint Depth-Motion Dense Estimation for Multiview Video Coding," *Journal of Visual Communication and Image Representation*, vol. 21, pp. 487–497, 2010.

[17] Shafik Huq, Andreas Koschan, and Mongi Abidi, "Occlusion filling in stereo: theory and experiments," *Computer Vision and Image Understanding*, vol. 117, pp. 688–704, June 2013.

[18] C. Liu, "Optical flow Matlab/C++ code," .

[19] Dmytro Rusanovsky, Karsten Muller, and Anthony Vetro, "Common Test Conditions of 3DV Core Experiments," ITU-T SG16 WP3 & ISO/IEC JTC1/SC29/WG11 JCT3V-D1100, April 2013.

[20] "Call for Proposals on 3D video coding technology," ISO/IEC JTC1/SC29/WG11 N12036, March 2011.

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