MULTIVIEW DEPTH MAP ENHANCEMENT BY VARIATIONAL BAYES INFERENCE ESTIMATION OF DIRICHLET MIXTURE MODELS

Pravin Kumar Rana, Zhanyu Ma, Jalil Taghia, and Markus Flierl

School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
Email: {prara, zhanyu, taghia, mflierl}@kth.se

ABSTRACT

High quality view synthesis is a prerequisite for future free-viewpoint television. It will enable viewers to move freely in a dynamic real world scene. Depth image based rendering algorithms will play a pivotal role when synthesizing an arbitrary number of novel views by using a subset of captured views and corresponding depth maps only. Usually, each depth map is estimated individually by stereo-matching algorithms and, hence, shows lack of inter-view consistency. This inconsistency affects the quality of view synthesis negatively. This paper enhances the inter-view consistency of multiview depth imagery. First, our approach classifies the color information in the multiview color imagery by modeling color with a mixture of Dirichlet distributions where the model parameters are estimated in a Bayesian framework with variational inference. Second, using the resulting color clusters, we classify the corresponding depth values in the multiview depth imagery. Each clustered depth image is subject to further sub-clustering. Finally, the resulting mean of each sub-cluster is used to enhance the depth imagery at multiple viewpoints. Experiments show that our approach improves the average quality of virtual views by up to 0.8 dB when compared to views synthesized by using conventionally estimated depth maps.

Index Terms— Multiview video; depth map enhancement; variational Bayesian inference; Dirichlet mixture model.

1. INTRODUCTION

Consistent and precise geometry information on natural scenes is highly desirable for high-quality free-viewpoint television (FTV) [1]. Scene geometry information such as depth maps significantly reduce the transmission requirements for the emerging FTV [2]. FTV will enable viewers to experience a dynamic natural 3D-depth impression while freely choosing their viewpoint of real world scenes. This has been made possible by recent advances in autostereoscopic multiview display technology which permits viewing of scenes from a range of perspectives for multiple viewers [3]. However, these multiview display require a high number of views at the receiver side to have a seamless transition among interactively selected stereo pairs and to maintain a high depth perception [2]. This requires to capture, store, and transmit an enormous amount of multiview video (MVV) [4]. MVV is a set of videos recorded by many video cameras that capture a dynamic natural scene from many viewpoints simultaneously. In recent years, many compression techniques have been proposed for MVV imagery [4], [5], [6]. These compression schemes exploit efficiently the inherent inter-view and temporal similarities in the MVV imagery. But the resulting transmission cost is approximately proportional to the number of coded views. Therefore, a large number of views cannot be efficiently transmitted using existing techniques. With only a limited subset of captured color information, high quality FTV is not feasible [2].

The transmission efficiency can be improved significantly by utilizing depth maps. A depth map is a single channel gray scale image. Each pixel in the depth map represents the shortest distance between the corresponding object point in the natural scene and the given camera plane. Usually, depth maps are compressed by existing video codecs as they contain large smooth areas of constant grey levels. Given a small subset of MVV imagery and its corresponding set of multiview depth images (MVD), an arbitrary number of views can be synthesized by using depth image based rendering [7]. However, the quality of depth maps affects significantly the quality of view synthesis as well as coding.

Usually, depth maps are obtained by establishing stereo correspondences between two or more camera images at different viewpoints by a matching criterion [8]. The accuracy of the stereo matching affects the resulting depth estimates. A number of optimization techniques are used to refine depth estimates, for example, graph-cut [9], belief propagation [10], and modified plane sweeping with segmentation [11]. Despite these optimizations, the resulting depth maps at different viewpoints usually lack inter-view consistency due to independent estimation as depicted in Fig. 1. Furthermore, depth estimation does not exploit temporal coherence in views, and this results in temporal inconsistency. These inconsistencies affect the quality of view synthesis negatively.

Many methods have been proposed to enhance the temporal inconsistency in MVD imagery, for example [12], [13], and [14]. Whereas [15], [16], and [17] address the inter-view depth inconsistency problem. To enhance the inter-view depth consistency, these methods warp multiple depth maps for spatial alignment from various viewpoints to a common viewpoint before applying enhancement algorithms. However, this warping causes errors due to the discrete values in the depth maps and affects enhancement algorithms negatively [18].

In [19], we propose a general model-based framework for depth map enhancement. First, the framework performs a color classification of the view imagery by making a generative model based on a mixture of Gaussian distributions. The model parameters are estimated by variational Bayesian inference. Next, for each resulting color cluster, we classify the corresponding depth values from multiple viewpoints. Finally, multiple depth levels are assigned to individual sub-clusters for depth enhancement at multiple viewpoints. The resulting improved depth maps are utilized to enrich the FTV user experience by synthesizing high-quality virtual views. In contrast to [19], this paper uses variational Bayesian inference for Dirichlet mixture models (VBDMM) to perform color classification in the view imagery [20]. As the vector of image pixels has nonnegative elements and is bounded, it can be efficiently modeled by utilizing non-Gaussian distributions such as the Dirichlet distribution [21]. Moreover, VBDMM reduces the model complexity significantly.

The paper is organized as follows: Section 2 describes our MVD image enhancement framework. Section 3 presents our experimental assessment. The conclusions are summarized in Section 4.

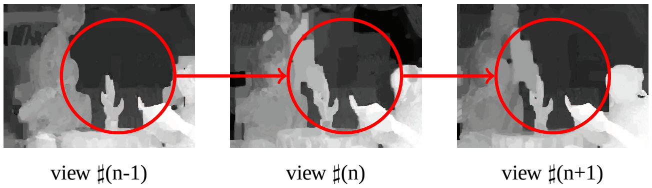


Fig. 1. Example of inter-view inconsistency among estimated depth maps at three viewpoints for Newspaper MVV [22]. The red circles mark prominent inconsistent areas in the depth maps.

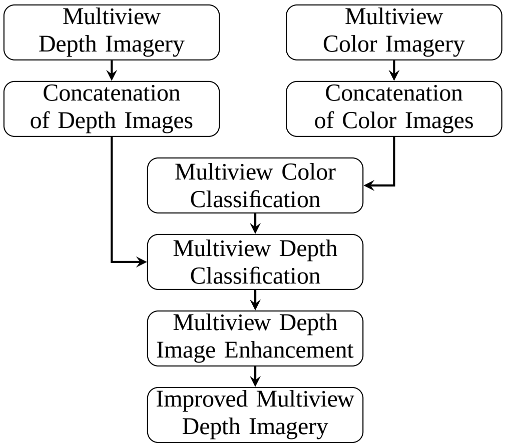

Multiview Depth Imagery
Multiview Color Imagery
Concatenation of Depth Images
Concatenation of Color Images
Multiview Color Classification
Multiview Depth Classification
Multiview Depth Image Enhancement
Improved Multiview Depth Imagery

Fig. 2. Block diagram of the proposed approach.

2. MULTIVIEW DEPTH ENHANCEMENT FRAMEWORK

The proposed algorithm is summarized in Fig. 2. We assume that the MVV imagery of resolution $H \times W$ is independently captured for a given natural dynamic scene using projective cameras at N viewpoints. Usually, each captured view of the scene is an image in YUV color space [23]. To make the procedure insensitive to the absolut luminance, we use the chromatic color representation [24], also known as the pure color space. We transform these views from YUV space to the chromatic color space. In this space, the virtual primary colors are denoted by X, Y, and Z, respectively. The chromaticity of a pixel in view $\widetilde{\mathbf{v}}_n \in \mathbf{R}^{H\times W\times 3}, n \in \{1,\dots,N\}$, is described by a vector of three coefficients, i.e., $\widetilde{\mathbf{v}}_n(p,q) = [x,y,z]^T$, whose entries sum to one. The chromaticity coefficients are defined as [25],

$$x = \frac{X}{X+Y+Z},\ y = \frac{Y}{X+Y+Z},\ z = \frac{Z}{X+Y+Z}. \qquad (1)$$

2.1. Concatenation of View Imagery

To have a unique model for the captured natural scene, we first exploit this inherent inter-view similarity of the MVV imagery by concatenating views from N viewpoints to a single view $\mathbf{v} \in \mathbf{R}^{H\times NW\times 3}$,

$$\mathbf{v} = [\mathbf{v}_1,\dots,\mathbf{v}_N]. \qquad (2)$$

For simplicity, we transform

$$\mathbf{v} \in \mathbf{R}^{H\times NW\times 3} \longmapsto \overline{\mathbf{v}} \in \mathbf{R}^{3\times M}, \qquad (3)$$

where $\overline{\mathbf{v}} = [\overline{\mathbf{v}}_1,\dots,\overline{\mathbf{v}}_M]$, with $M = HWN$. Each $\overline{\mathbf{v}}_m, m \in \{1,\dots,M\}$, is a point in chromaticity space with the chromaticity coefficients, x, y, and z.

2.2. Multiview Color Classification

In this section, we classify the color pixels of the captured MVV imagery in the chromaticity space. As mentioned in [26], the goal of classification is to partition the image into regions each of which has a reasonably homogeneous visual appearance or which corresponds to objects or parts of objects. By assuming the chromaticity space is partitioned into K clusters, the pixels of the captured MVV imagery can be classified into K clusters. The best classification should provide a high intra-cluster similarity and a low inter-cluster similarity.

By considering the spatial proximity, statistical models can be applied to classify the MVV imagery efficiently. As the pixel vector in the chromaticity space has nonnegative elements which are bounded by the interval $[0,1]$ and sum to one, it is obvious that the pixel vectors are not Gaussian distributed. For such, we can utilize non-Gaussian distributions to efficiently model the data [21]. Based on the pixel vector's properties, a natural and reasonable choice is to assume that the pixel vectors of each cluster are Dirichlet distributed [27]. Hence, we use a Dirichlet mixture model (DMM) to capture the underlying distribution of all clusters. Thus, for one pixel $\overline{\mathbf{v}}_m$, its probability density function (PDF) can be expressed by

$$f(\overline{\mathbf{v}}_m) = \sum_{k=1}^{K} \pi_k \mathbf{Dir}(\overline{\mathbf{v}}_m; \mathbf{u}_k), \qquad (4)$$

where K is the number of mixture components (clusters), π_k represents the weighting factor of the k^{th} mixture component, and $\mathbf{u}_k$ denotes the parameter vector in the k^{th} mixture component [28]. For a single L-dimensional Dirichlet distribution, the PDF is

$$\mathbf{Dir}(\overline{\mathbf{v}}_m; \mathbf{u}_k) = \frac{\Gamma\left(\sum_{l=1}^{L+1} u_{lk}\right)}{\prod_{l=1}^{L+1}\Gamma(u_{lk})} \prod_{l=1}^{L+1} \overline{v}_{lm}^{u_{lk}-1},\ u_{lk} > 0, \qquad (5)$$

where $\Gamma(\cdot)$ is the gamma function as defined by

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. \qquad (6)$$

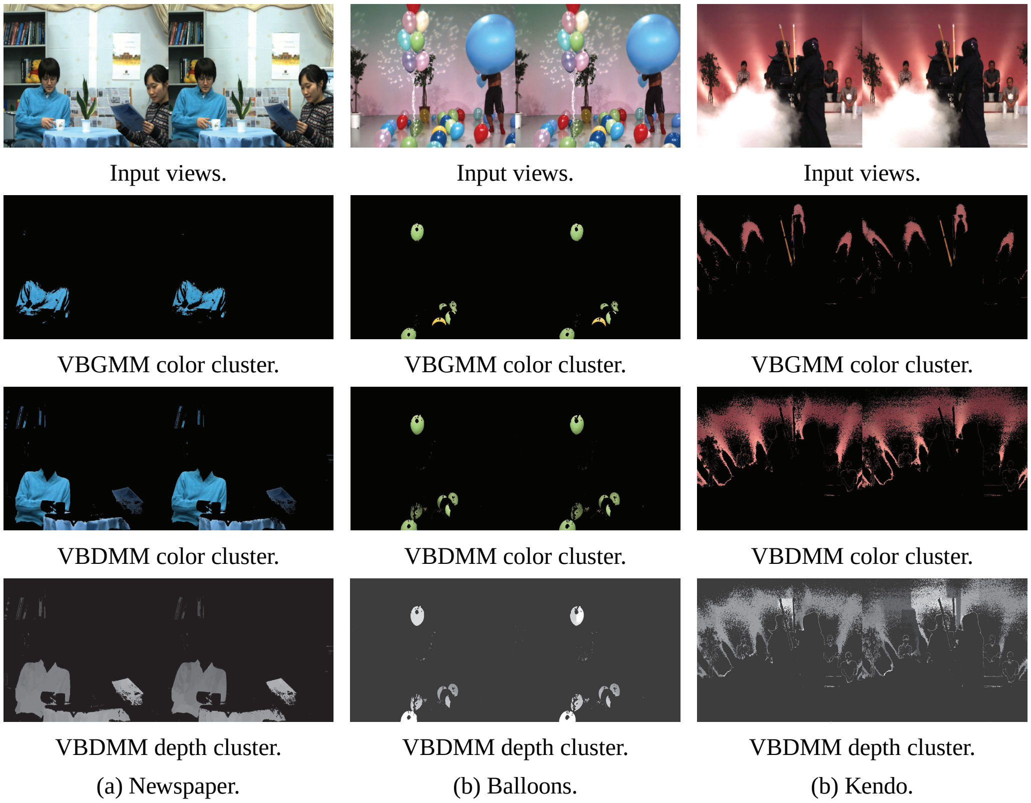
If the number of clusters K is known in advance, the Expectation Maximization (EM) algorithm [28] can be used to fit a DMM with K mixture components to the pixel vectors. However, the number of components (clusters) is in general unknown and should be chosen empirically by the used algorithm.

An alternative way of learning the number of clusters is to employ the Bayesian framework to estimate the DMM, as the Bayesian method can determine the number of mixture components automatically from the data. A fully Bayesian DMM approach is proposed in [20], in which the variational Bayesian (VB) method [28, 29] was applied to deal with the intractable integration expression appearing in the Bayesian approach. With the extended factorized approximation approach [30], an analytically tractable solution was derived. This general optimization method has been used in a number of recent works. With the Bayesian approach, the mixture model is initialized by a relatively large number of mixture components K. After convergence, the mixture components with extremely small weights will be discarded from the model and only I mixture components (clusters) are kept afterwards as

$$\underline{I} = \{i : \pi_i \geq \delta\}, \qquad (7)$$

where $\underline{I} = \{1,\dots,i,\dots,I\}$ and $I \leq K$. In this paper, we choose empirically the threshold as $\delta = 0.01$. The remaining I clusters are of significant weight and explain the underlying data sufficiently. Hence, the most efficient number of mixture components (clusters) is learned by the model itself after the algorithm has converged.

Let $\mathbf{R} = [\mathbf{r}_1,\dots,\mathbf{r}_M]$ denote the responsibility matrix in the Bayesian estimation [20], where $\mathbf{r}_m = [r_{m1},\dots,r_{mI}]^T$. Each element r_{mi} represents the probability that $\overline{\mathbf{v}}_m$ is generated from the i^{th} cluster. Thus, we assign each pixel to the component (cluster) which



Input views. Input views. Input views.

VBGMM color cluster. VBGMM color cluster. VBGMM color cluster.

VBDMM color cluster. VBDMM color cluster. VBDMM color cluster.

VBDMM depth cluster. VBDMM depth cluster. VBDMM depth cluster.

(a) Newspaper. (b) Balloons. (b) Kendo.

Fig. 3. Example of color and corresponding depth classification results.

gives the largest probability. Members of the i^{th} cluster are denoted by $\underline{\mathbf{Y}}^{(i)}$ which can be extracted from the observation set $\overline{\mathbf{v}}$ as

$$\underline{\mathbf{Y}}^{(i)} = \{\mathbf{y}_1^{(i)}, \ldots, \mathbf{y}_M^{(i)}\}, \tag{8}$$

$$\mathbf{y}_m^{(i)} = \mathcal{M}_m^{(i)}\, \overline{\mathbf{v}}_m, \tag{9}$$

with the definition

$$\mathcal{M}_m^{(i)} = \begin{cases} 1, & \text{if } r_{mi} > r_{mj}, \forall i \neq j \;\; (i, j \in \{1, \ldots, I\}); \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

2.3. Multiview Depth Classification

For each view $\mathbf{v}_n$, we assume that the associated per-pixel depth map $\mathbf{d}_n \in \mathbf{R}^{H \times W}$ exists. Each pixel in the depth map $\mathbf{d}_n$ has a discrete value, where the value zero represents the farthest point and 255 the closest. In order to enhance inter-view consistency, we concatenate depth maps from N viewpoints to a single depth $\mathbf{d} \in \mathbf{R}^{H \times NW}$,

$$\mathbf{d} = [\mathbf{d}_1, \ldots, \mathbf{d}_N]. \tag{11}$$

Again, for simplicity, we consider the following mapping

$$\mathbf{d} \in \mathbf{R}^{H \times NW} \longmapsto \overline{\mathbf{d}} \in \mathbf{R}^{1 \times M}, \tag{12}$$

where $\overline{\mathbf{d}} = [\overline{\mathbf{d}}_1, \ldots, \overline{\mathbf{d}}_M]$ is such that for each color pixel $\overline{\mathbf{v}}_m, m \in \{1, \ldots, M\}$, we have an associated depth value $\overline{\mathbf{d}}_m \in \{0, \ldots, 255\}$. We therefore utilize this per-pixel depth value association with the color values by using $\mathcal{M}_m^{(i)}$ in order to obtain members of the i^{th} depth cluster $\underline{\mathbf{X}}^{(i)}$,

$$\underline{\mathbf{X}}^{(i)} = \{\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_M^{(i)}\}, \tag{13}$$

$$\mathbf{x}_m^{(i)} = \mathcal{M}_m^{(i)}\, \overline{\mathbf{d}}_m. \tag{14}$$

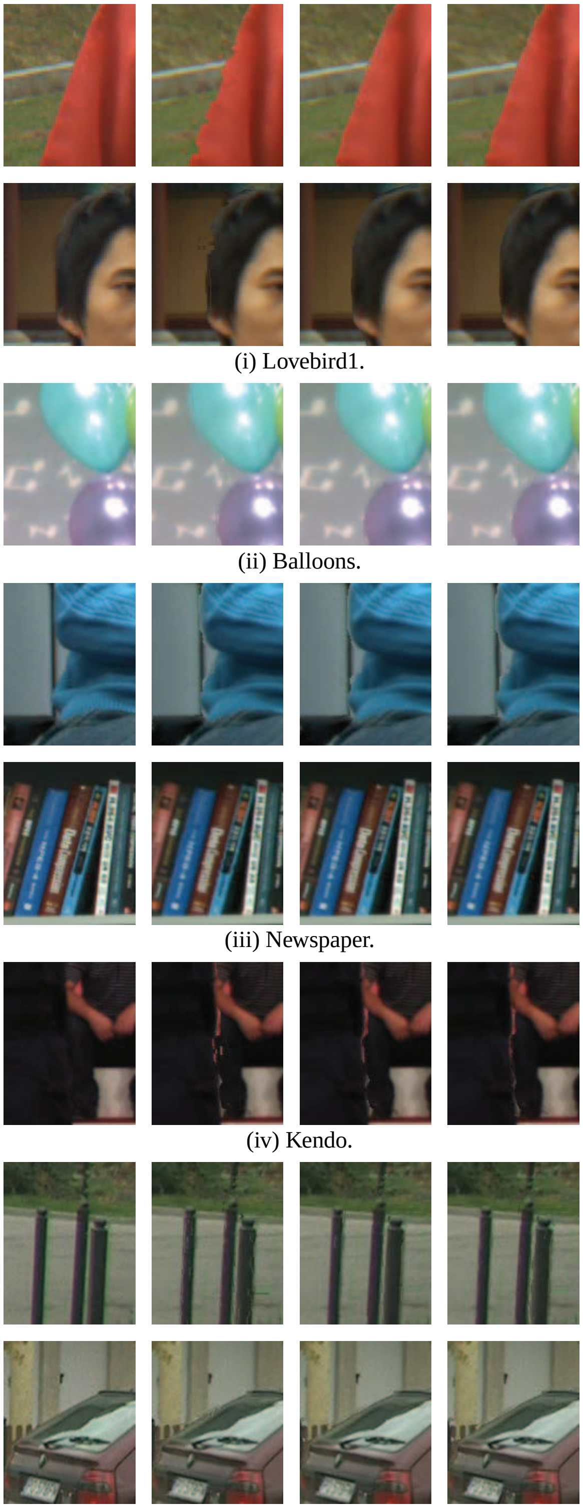
Fig. 3 shows such color clusters and associated depth clusters for concatenated color images and depth maps, respectively.

2.4. Multiview Depth Image Enhancement

The members of the cluster $\underline{\mathbf{Y}}^{(i)}$ have similar colors, whereas members of the cluster $\underline{\mathbf{X}}^{(i)}$ may have different depth values. This is because a foreground and a background object point can have a similar color, but foreground object points have different depth values compared to background object points. As we assume 1D parallel camera arrangements, an object point with a given color which is visible from N viewpoints should have the same depth value in all N depth maps. However, such points usually have different depth values in the cluster $\underline{\mathbf{X}}^{(i)}$ due to the inconsistency across multiple viewpoints. This motivates us to consider further sub-clustering of each $\underline{\mathbf{X}}^{(i)}$, where the variance of each sub-cluster reflects the inconsistency of depth values at various viewpoints. Here, we apply the mean-shift algorithm for the purpose of sub-clustering [31] instead of K-means as used in [19]. The K-means clustering algorithm is computationally fast but it suffers from two main drawbacks: 1) it does not consider the spatial proximity of different pixels and 2) it requires a good guess for the number of actually present clusters. Therefore, an incorrect guess of the number of actual clusters may lead to erroneous K-means clustering results. However, mean-shift clustering does not require prior knowledge of the number of clusters [32], and hence, is a good choice for this sub-clustering problem. We may use again the Bayesian mixture model of non-Gaussian in order to perform this sub-clustering stage. This would result in a more accurate clustering, but it would also entail a higher computational complexity. Finally, we assigns the mean of each sub-cluster to all depth pixels which fall into the specified depth sub-cluster, irrespective of the originating viewpoint.

3. EXPERIMENTAL RESULTS

The Moving Picture Experts Group (MPEG) uses the view synthesis reference software (VSRS) for view synthesis [33], [34]. It uses a DIBR approach to synthesize a virtual view at an arbitrary intermediate viewpoint by using two reference views, left and right, the two



(i) Lovebird1.

(ii) Balloons.

(iii) Newspaper.

(iv) Kendo.

(v) Poznan Street.

(a) Original. (b) MPEG. (c) [19]. (d) Proposed.

Fig. 4. Selected regions of synthesized virtual views of the test sequences as generated by VSRS 3.5 using (b) MPEG depth maps, (c) improved depth maps from [19], and (d) enhanced depth maps from the proposed VBDMM+Mean-shift-based algorithm for a detailed comparison. Full resolution synthesized virtual views are available at http://www.ee.kth.se/ prara/research/icassp.zip

Table 1. Objective quality of the synthesized virtual views

Test Sequence	Input Views	Virtual View	VSRS 3.5 [dB]		
			MPEG Depth (a)	VBGMM +K-Means Depth (b)	VBDMM +Mean-Shift Depth (c)
Lovebird1	6, 8	7	28.50	28.68	29.04
Balloons	3, 5	4	35.69	35.93	36.02
Newspaper	4, 6	5	32.00	32.10	32.11
Kendo	3, 5	4	36.54	36.72	39.35
Poznan Street	3, 5	4	35.56	35.58	35.72

corresponding reference depth maps, and camera parameters. The proposed algorithm is evaluated in two steps. First, the depth imagery at two viewpoints is improved by choosing a large number of mixture components K, for example $K = 100$. For this, we concatenate only two views and the two corresponding depth maps as input to our algorithm. Second, a virtual view for a given viewpoint is synthesized by VSRS 3.5 using the improved depth maps. We synthesize these virtual views by using the 1D parallel synthesis mode with half-pel precision. Further, we measure the objective quality of the synthesized views in terms of the peak signal-to-noise ratio (PSNR) with respect to the captured view of a real camera at the same viewpoint.

Table 1 shows a comparison of the luminance signal Y-PSNR (in dB) of the virtual views as synthesized by VSRS 3.5 with the help of (a) MPEG depth maps, (b) enhanced depth maps from VBGMM [19], and (c) enhanced depth maps from the proposed VBDMM approach. The presented enhancement algorithm offers average improvements of up to 0.8 dB. The improvement in quality depends on the input reference depth maps at various viewpoints. For the Balloons test data, the mean quality and standard deviation of ten experiments with different initialization is 35.86 ± 0.09 dB. This compares to 35.69 dB when using MPEG depth maps. In Table 1 the best results are presented. Fig. 4 shows that our proposed depth enhancement algorithm noticeably improves the visual quality of virtual views when compared to using MPEG depth maps. Specially, artifacts around the edges in synthesized virtual views have been significantly reduced. This demonstrates the efficiency of our multiview depth imagery enhancement algorithm. Hence, it is a promising algorithm for improving the visual quality of FTV.

Besides improving the quality of FTV, VBDMM introduces less model complexity than the VBGMM approach. When modeling a D-dimensional vector by a VBDMM with I mixture components, the number of free parameters is $s_D = I(D+2) - 1$. The number of free parameters for the VBGMM with diagonal covariance matrix is $s_G = I(2D+1) - 1$. Thus, by measuring the model complexity in terms of the number of free parameters, the VBDMM requires a smaller model complexity than the VBGMM with the same initial number of mixture components.

4. CONCLUSIONS

We have proposed a MVD image enhancement algorithm that improves inter-view depth consistency. With that, we are able to enhance the visual quality of FTV. The presented algorithm is based on multiview color classification by variational Bayesian inference for Dirichlet mixture models. It uses the resulting color clusters to classify depth values from various viewpoints. Here, a per-pixel association between depth and color has been exploited for the classification. Both objective and subjective results demonstrate the advantage of the presented algorithm. Furthermore, our approach has potential to improve temporal depth consistency by concatenating temporally successive frames from multiple viewpoints.

5. REFERENCES

[1] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 67 –76, Jan. 2011.

[2] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. IEEE*, vol. 99, no. 4, pp. 643 –656, Apr. 2011.

[3] H. Urey, K.V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540 – 555, April 2011.

[4] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66 –76, Nov. 2007.

[5] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV–A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1606 –1621, Nov. 2007.

[6] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626 –642, Apr. 2011.

[7] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," 2004, vol. 5291, pp. 93–104, SPIE.

[8] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Computer Vision*, vol. 47, pp. 7 –42, Apr. 2002.

[9] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124 –1137, Sept. 2004.

[10] P.F. Felzenszwalb and D.R. Huttenlocher, "Efficient belief propagation for early vision," in *Proc. IEEE CS Conf. CVPR*, Jun. 2004, vol. 1, pp. I–261 –I–268.

[11] C. Cigla, X. Zabulis, and A.A. Alatan, "Region-based dense depth extraction from multi-view video," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2007, vol. 5, pp. V–213 –216.

[12] C. Cigla and A.A. Alatan, "Temporally consistent dense depth map estimation via belief propagation," in *3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video*, May 2009, pp. 1 –4.

[13] S. Lee and Y. Ho, "Temporally consistent depth map estimation using motion estimation for 3DTV," in *Int. Workshop Adv. Image Technol.*, Jan. 2010, pp. 149(1–6).

[14] D. Fu, Y. Zhao, and L. Yu, "Temporal consistency enhancement on depth sequences," in *Picture Coding Symp.*, Dec. 2010, pp. 342 –345.

[15] P. K. Rana and M. Flierl, "Depth consistency testing for improved view interpolation," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, St. Malo, France, Oct. 2010, pp. 384 –389.

[16] P. K. Rana and M. Flierl, "Depth pixel clustering for consistency testing of multiview depth," in *Proceedings of European Signal Processing Conference*, Bucharest, Romania, Aug. 2012, pp. 1119–1123.

[17] E. Ekmekcioglu, V. Velisavljević, and S.T. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 352 –361, Apr. 2011.

[18] Luat Do, S. Zinger, and P.H.N. de With, "Objective quality analysis for free-viewpoint DIBR," in *Proceedings of IEEE International Conference on Image Processing*, Hong Kong, Sept. 2010, pp. 2629 –2632.

[19] P. K. Rana, J. Taghia, and M. Flierl, "A variational Bayesian inference framework for multiview depth image enhancement," in *Proc. IEEE Int. Symposium Multimedia*, Irvine, California, USA, Dec. 2012, pp. 183 –190.

[20] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of Dirichlet mixture model with variational inference," *Submitted*, 2013.

[21] Z. Ma, *Non-Gaussian statistical models and their applications*, Ph.D. thesis, KTH Royal Institute of Technology, Stockholm, 2011.

[22] MPEG, "Call for proposals on 3D video coding technology," Tech. Rep. N12036, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Mar. 2011.

[23] Charles A. Poynton, *A technical introduction to digital video*, John Wiley & Sons, Inc., New York, NY, USA, 1996.

[24] G. Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Wiley classics library. John Wiley & Sons, 2000.

[25] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ, second edition, Jan. 2002.

[26] D. A. Forsyth and J. Ponce, *Computer vision: A modern approach*, Prentice Hall, Englewood Cliffs, NJ, first edition, 2003.

[27] N. Bouguila, D. Ziou, and J. Vaillancourt, "Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1533–1543, Nov. 2004.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, first edition, 2006.

[29] Z. Ghahramani and M. J. Beal, "Variational inference for Bayesian mixtures of factor analysers," in *Adv. Neural Inf. Process. Syst. 12*. 2000, pp. 449 –455, MIT Press.

[30] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2160–2173, 2011.

[31] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603 –619, May 2002.

[32] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *Proceedings of IEEE International Conference on Computer Vision*, Oct. 2003, pp. 456 –463 vol.1.

[33] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," Tech. Rep. M15377, ISO/IEC JTC1/SC29/WG11, Archamps, France, Apr. 2008.

[34] MPEG, *View Synthesis Software Manual*, ISO/IEC JTC1/SC29/WG11, Sept. 2009, release 3.5.