World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:11, No:4, 2017

Hybrid Temporal Correlation Based on Gaussian Mixture Model Framework for View Synthesis

Deng Zengming, Wang Mingjiang

***Abstract*—**As 3D video is explored as a hot research topic in the last few decades, free-viewpoint TV (FTV) is no doubt a promising field for its better visual experience and incomparable interactivity. View synthesis is obviously a crucial technology for FTV; it enables to render images in unlimited numbers of virtual viewpoints with the information from limited numbers of reference view. In this paper, a novel hybrid synthesis framework is proposed and blending priority is explored. In contrast to the commonly used View Synthesis Reference Software (VSRS), the presented synthesis process is driven in consideration of the temporal correlation of image sequences. The temporal correlations will be exploited to produce fine synthesis results even near the foreground boundaries. As for the blending priority, this scheme proposed that one of the two reference views is selected to be the main reference view based on the distance between the reference views and virtual view, another view is chosen as the auxiliary viewpoint, just assist to fill the hole pixel with the help of background information. Significant improvement of the proposed approach over the state-of –the-art pixel-based virtual view synthesis method is presented, the results of the experiments show that subjective gains can be observed, and objective PSNR average gains range from 0.5 to 1.3 dB, while SSIM average gains range from 0.01 to 0.05.

***Keywords*—**View synthesis, Gaussian mixture model, hybrid framework, fusion method.

I. Introduction

IN recent years, three-dimensional (3D) video has become very popular for its superior visual and sound effects to audiences. In contrast to the traditional two-dimensional (2D) video, 3D video includes texture image and its corresponding depth map, as the so-called Texture-plus-Depth format. It is obvious that the success of 3D video demonstrates the customers' requirements for a better visual experience [1]. To meet the demands for the potential market, in the past few decades, 3D video technology has already been a hot topic of research in the video processing field.

While viewers are not satisfied to watch the video from just one view location, FTV is about to be spotlighted for its incomparable interactivity [2]. This application will enable viewers to change the position or angle to watch the scenes, this multi-view video (MVV) composed of several (a set of *N*) video sequences representing the same scene, these *N* video sequences require *N* cameras positioned under different spatial configurations to capture the images simultaneously [3]. With *N* viewpoints for the viewers, more data storage and large transmission bandwidth is essentially required. Even though the associated depth sequences can be captured directly by depth camera or estimated from texture data, and video compression is efficient with multi-view video plus depth (MVD) format or HEVC standard, FTV still could not be practically applied widely. To solve this problem, view synthesis technology is proposed and has been developed in the past decade, this scheme uses reference views, which have already captured their texture images and corresponding depth maps, to interpolate or extrapolate virtual views. View synthesis is a popular research topic in the computer vision and video processing field because this application enables to synthesize more virtual view images using the limited numbers of reference viewpoints. Large numbers of methods for view synthesis have been developed. Among all these methods, depth-image-based rendering (DIBR) is no doubt the most common way [4]. DIBR utilizes the texture-plus-depth data format; it warps every pixel in the reference image to the corresponding position in the virtual view according to the intrinsic parameter of the camera and extrinsic parameter about the explicit geometry information [5].

A critical problem arises when generating the virtual view, the regions covered by the foreground objects in the reference views may be disoccluded in the virtual views. These areas will appear as holes in the virtual view, also referred to as disocclusions. As illustrated in Fig. 1, region A is the foreground region, and regions B, C, D, and E could be determined as the background regions. For the reference view, the real camera can only capture three regions: A, B and E, while regions A, D and E can be spotted from the target synthesized viewpoint. It is obvious that information in region D can hardly be rendered from the reference view, it may appear as holes in the virtual view image. Filling these holes after 3D warping is an important research in the synthesis technology.

Besides the disocclusion problem, there are still some other effects. Because of the mismatch between foreground object in the texture image and its corresponding depth level in the depth map, the foreground region may be reprojected into the background region, and the ghost effect will appear in the synthesized image. Small cracks may also be found because of the rounding errors in the warping process [6].

Deng Zengming is a Ph.D. candidate with Harbin Institute of Technology, Shenzhen Graduate School, Guangdong Province, 518055, China (e-mail: twolove06@163.com).

Wang Mingjiang is Ph.D. supervisor and professor at the Department of Electronics and Communication Engineering, Harbin Institute of Technology, Shenzhen, 518055, China (e-mail: mjwang@hit.edu.cn).

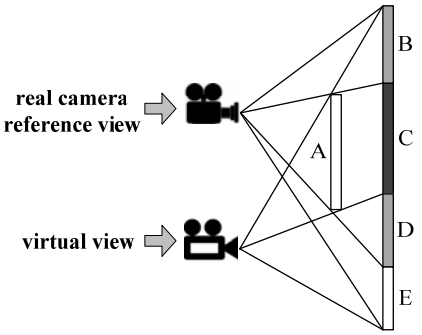

Fig. 1 Disocclusion in view synthesis

In this paper, a view synthesis approach which utilizes the traditional MVD sequences with spatial-temporal information is proposed. The rest of the paper is organized as follows. In Section II, the traditional synthesized algorithms are introduced. In Section III, proposed framework with hybrid blending method is discussed in detail. Experiment results are displayed in Section IV and the conclusions are given in Section V.

II. Related Works

Generally, the early approaches to solve the disocclusion problem could be separated into three categories. In the first category, preprocessing the depth maps before DIBR is proposed, aiming to reduce the disparity along the boundary between the foreground and background. After all these experiments, as given in [7]-[9], this kind of approach easily causes geometrically distorted foreground objects and other unwanted effects when the baseline requirement is large.

The second approach is filling the disocclusion with the texture information in the neighboring regions; these regions are selected in the same frame and same time instant. Criminisi et al. in [10] proposed an exemplar based method that iteratively fills the disocclusions using the neighboring information; this method is nowadays a classical inpainting algorithm for its theoretical basis. Experimental results show that inpainting obtains a good performance when the holes appear as narrow gaps, but it is observed that information is easily lost when the hole is larger. Applying inpainting when filling the large holes always results in blurring artifacts [11].

These two approaches mentioned above do not meet the research satisfactory for the synthesis image quality, and they are both processing the synthesis frame by frame, ignoring the temporal correlation of the disoccluded areas, this will lead to typical flickering artifacts in the virtual view. In the last category, image inpainting algorithm and texture synthesis technique are both inspired. Actually, inpainting always focused on filling the narrow gaps in a texture image, while texture synthesis is capable of filling the large scale holes [12], [13]. In [14], Scheming and Jiang firstly proposed to determine the background information using a background subtraction method, but this approach relies on a good performance of the foreground segmentation method, so it is not appropriate to be adopted in complex circumstances. Chen explored the motion vector of H.264/AVC bit stream to render the disocclusions in the virtual view [15]. In [16] and [17], a background sprite is generated by the original texture and synthesized images from the temporally previous frames for disocclusion filling, but the temporal consistency of the synthesized images need further investigation as described in [18].

In [19], Chao Yao proposed a disocclusion filling approach based on the temporal correlation information for the Single-View-plus-Depth (SVD) format. In the proposed approach, the background information is obtained from both the texture and depth sequences, by using the Gaussian Mixture Model (GMM) and Foreground Depth Correlation (FDC). On one hand, by using GMM, a temporally stable background sprite can be acquired. On the other hand, the FDC method is used to identify the covered background regions in different frames by detecting the movement of foreground regions. Finally, the obtained background image is used for disocclusion filling in the DIBR system. Experiments show that this approach yields very good subjective and objective performance, but it still has some shortage, some regions that can never be seen in the single reference view may easily be spotted in another distant virtual view position, this disocclusion will be filled by the existing inpainting algorithm, and the results turn out to be not very satisfied. In order to get the information about the unseen area, we still propose to utilize two reference views, one view which is nearer to the virtual view will be selected to be the main reference view, and another one is determined to be the auxiliary view, the details will be given in Section III.

III. Proposed Frameworks

A. Hybrid Framework

Among all these view synthesis technologies, the most popular one is VSRS. VSRS inputs the texture image and its corresponding depth map from two views, after the synthesis process, as illustrated in Fig. 2, a virtual view is generated. There are two modes to synthesizing frames: a general mode and a 1D mode. In general mode, the cameras on the view points are not aligned in a straight line (they are always set in an arc) and in 1D mode, the cameras are aligned strictly in a straight line, this line must be perpendicular to their optical axes.

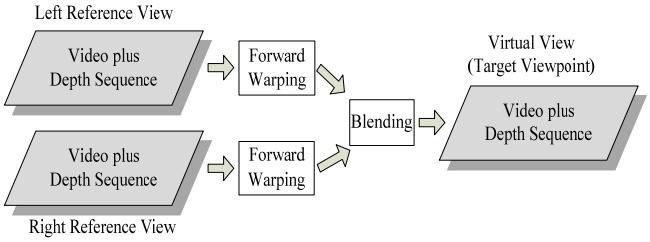

Fig. 2 Simple description for VSRS

Traditional rendering VSRS techniques synthesize an intermediate frame only from the left and right reference views at the same time instant. The texture image and depth map from each viewpoint get the same blending weight. In this method, two viewpoints obtain different roles according to the geometry

distance between the reference view and the virtual view. The closer one is determined as the main viewpoint and another viewpoint is named the auxiliary viewpoint. Temporal correlation will be explored to fill the disocclusion, while Gaussian Mixture Model (GMM) is utilized to process the texture sequences to obtain a stable background sprite. After the sprite is acquired, the depth map sequence will be used to check the mistaken pixels.

In SVD format, some background information can never be obtained in any time instant, the disocclusions are finally filled by the inpainting algorithm; the experiment results show an unsatisfied quality and the PSNR results are also not satisfactory enough. As mentioned in the beginning of this section, this approach makes different determination for the two reference views. The view which is closer to the target view is chosen to be the main view, and another view point is named auxiliary view point. Fig. 3 shows the main flow diagram of the proposed approach.

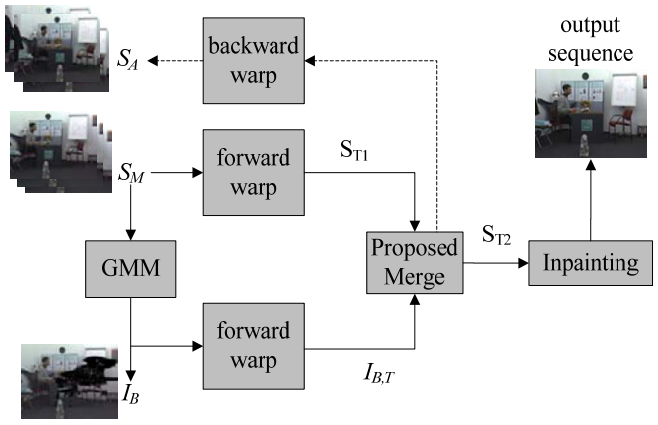

Fig. 3 Proposed framework for view synthesis

Firstly, the Gaussian Mixture Model is utilized to process S_M, the reference texture image sequence of the main view, to generate the background image I_B. S_M and I_B both go through the forward warp using their corresponding depth map independently, then S_{T1} and $I_{B,T}$ is obtained, respectively. At the next step, the proposed merge method is adopted to blend the S_{T1} and $I_{B,T}$, the detailed approach will be given in the following subsection. S_{T2} is the result after merging, and then the remaining holes will be filled by using the traditional inpainting algorithm.

B. Background Generation

The Gaussian Mixture Model is a commonly used function to detect the moving object in the video processing field [20], and it can also be applied to generate a stable background sprite in computer vision field. Another advantage for the GMM is that this model also works at the pixel level, just like the VSRS technology. Each pixel is modeled independently by a mixture of K Gaussian mixture distribution. K is usually set to be 3, and the distribution with K can be written as:

$$p(x_j)=\sum_{i=1}^{K}\omega_{j,t}^{i}\cdot\eta\left(x_j,\mu_{j,t}^{i},\sigma_{j,t}^{i}\right) \tag{1}$$

where $p(x_j)$ means the probability density of value x_j on pixel j, $\omega^i_{j,t}$ is the pixel j's ith Gaussian distribution's weight at time t, with $\sum^{K}_{i=1}\omega=1$, η is the GMM density function with three dependent variables: x_j denotes the pixel value at time t, $\mu^i_{j,t}$ denotes the mean value of pixel x_j and $\sigma^i_{j,t}$ is the variance value of the pixel. The function of η is given in (2).

$$\eta\left(x_j,\mu_{j,t}^{i},\sigma_{j,t}^{i}\right)=Q\cdot\exp\left[-\frac{1}{2}\left(\sigma_{j,t}^{i}\right)^{-1}\left(x_j-\mu_{j,t}^{i}\right)^{2}\right] \tag{2}$$

where,

$$Q=\frac{1}{(2\pi)d/2\cdot\left|\sigma_{j,t}^{i}\right|1/2} \tag{3}$$

In this paper, the detailed process to model the stable reference background image is described as follows [20]:

1) **Initialization.** Model is initialized at the beginning of the generation (time t_0). The mean value of the first Gaussian model $\mu^1_{j,t0}$ is set equal to the pixel value of the current frame, that of other models is set to 0. The variance value $\sigma^i_{j,t}$ of each Gaussian model is set to a large number, here it is set to 900. The weight value of the first Gaussian model $\omega^1_{j,t0}$ is set to 1, other weight value is set to 0.
2) **Update.** In the next frame at time instant t_1, the pixel in this frame is used to match with all the K Gaussian models, the condition $\left|x_{j,t_1}1-\mu^{i}{}_{j,t_1}\right|\le 2.5\sqrt{\sigma^{i}_{j,t_1}}$ will be examined.

– If the condition is satisfied, the matching process will stop, this matched Gaussian model will be updated and other models remain unchanged. The update equation is as:

$$\omega_{j,t_1}^{i}=(1-\alpha)\omega_{j,t_0}^{i}+\alpha \tag{4}$$

$$\mu_{j,t_1}^{i}=(1-\rho)\mu_{j,t_0}^{i}+\rho\cdot x_j \tag{5}$$

$$\sigma_{j,t_1}^{i}=((1-\rho)\sqrt{\sigma_{j,t_0}^{i}}+\rho\cdot(x_j-\mu_{j,t_0}^{i})^2)^2 \tag{6}$$

where, α is the learning rate of the model which is set to 0.005, and ρ is the learning rate of the Gaussian distribution parameters with $\rho=\alpha/\omega^i_{j,t_0}$. These two parameters reflect the rate of model convergence. If pixel x_{j,t_1} is not satisfied with the condition, it means that all K-Gaussian distribution is not matched, then the last N Gaussian distribution will be replaced.

– If all Gaussian models fail to satisfy the condition, a new Gaussian model is proposed with $\mu=x_{j,t_1}$, ω = 0.001, σ = 900. The model with smallest $\omega/\sqrt{\sigma}$ value will be discarded. The mean value and variance value of other Gaussian models remain unchanged and the weight value of K Gaussian models are normalized to $\sum_i^K\omega_{i,t_1}=1$.

3) **Convergence.** The remaining frames will be processed by repeating step 2). The value of background pixel will be

derived in μ, the most stable pixels on time domain will be modeled as the background image.

C. Proposed Fusion Method

In VSRS 3.5 version, merge method has already been proposed in mainly three ways. The first method is to apply Z-Buffer only, the second is to blend the two reference views into one weighted average target view, and the last one is the most popular choice, it adaptively chooses Z-Buffer or weighted average according to the pixel value distribution. This proposed merge method is described in Fig. 4.

Every pixel in the merged image will be processed to get a value in turn, and *P[MERG]* is the flag which indicates the value of the processing pixel in the blended image. Firstly, judgment is taken whether this pixel is warped from the main reference view or not. If the pixel from the S_{T1} is not filled, the pixel in the generated background image is selected if the pixel information is valid. If the pixel in the main reference view S_{T1} and generated image $I_{B,T}$ are both holes, reverse warp is conducted to search for the valid value in S_A. After reverse warping, remaining holes will be filled by using inpainting algorithm.

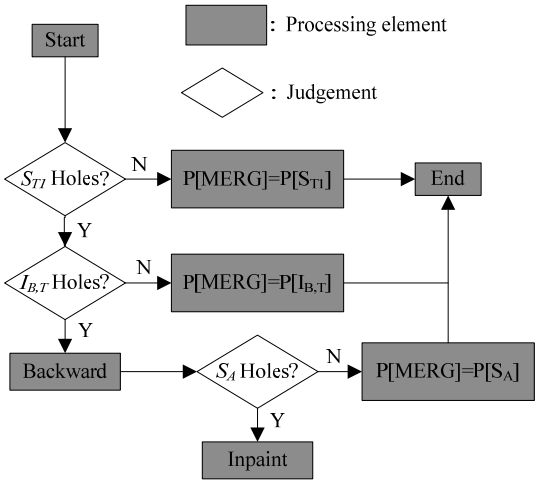

Fig. 4 Proposed fusion method

In contrast to the VSRS weighted average blending method, this scheme is proposed based on the theory that the information of texture images from the nearer view has much more similarities and influences on the rendered view, such as illumination impacts. While this blending method is adopted, computation complexity is reduced significantly. Traditional VSRS needs to warp every pixel in the two reference views with the corresponding depth maps to the target view position, while in this framework, only pixels in main reference view are all directly warped to the target virtual view, background image and its corresponding depth map also needs forward warping. After this, only the several remaining hole pixels are finally processed with backward warping to the search within the auxiliary reference image. In addition, this framework improves the reliability of the background texture by giving higher priority to warp and use of the background image. The inpainting algorithm or the interpolation method is efficient enough to fill the disocclusion when the disocclusion regions in the virtual view appear as cracks, whereas in case of large baseline, disocclusion may appear as large hole areas, and inpainting will fail to fill these regions perfectly. Under some circumstances, the background image may also fail to contain the required information because it may be never observed from the main reference view position, backward warping is utilized to search the corresponding pixel in the auxiliary reference view. This framework obviously makes the disocclusion filling much more efficient and reliable.

IV. Experiment Results

In this section, the presented algorithm is validated by the experiments in C++ using a PC, with the configuration of Intel(R) Core(TM) i5-3470 CPU (3.20GHz) with 4.00GB RAM running Windows 7. The tested video sequences include: *Bookarrival* (1024 × 768, 100 frames) and *Newspaper* (1024 × 768, 100 frames). For the *Bookarrival* video sequences, the chosen two reference view include view 6 and view 10, the target synthesis view is view 7. Based on the distance between each reference view to the virtual view, view 6 is selected as the main reference view in the proposed synthesis algorithm, with its baseline set as 6.5 cm, and the baseline between the auxiliary view and target view is 19.5 cm. For the *Newspaper* video sequences, the multi-view includes main view 2 and auxiliary view 6, the gene rated location is view 3. The baseline for these two reference views to the target view is 5 cm and 10 cm, respectively. The depth maps of the tested sequences are generated with the MPEG depth estimation reference software (DERS), the camera parameters are provided with the sequences.

TABLE I
AVERAGE PSNR (DB)

Sequence	Criminisi	VSRS 3.5	Proposed
Bookarrival(P1)	29.9767	31.2793	31.3569
Bookarrival(P4)	31.2775	32.2893	33.5369
Newspaper(P1)	26.5795	27.6301	28.1568
Newspaper(P4)	27.2438	27.5265	28.0191

TABLE II
AVERAGE SSIM

Sequence	Criminisi	VSRS 3.5	Proposed
Bookarrival(P1)	0.81675	0.85193	0.86539
Bookarrival(P4)	0.85311	0.87351	0.89478
Newspaper(P1)	0.85704	0.87762	0.91298
Newspaper(P4)	0.87029	0.88014	0.92291

The presented approach is compared with the classical Criminisi inpainting algorithm and the commonly used VSRS 3.5 algorithm. In VSRS 3.5, the adaptive blending method is adopted for its better objective performance. An integer pixel may be mapped to a non-integer pixel position when processing the forward warping, in order to reduce the effect from rounding errors, the algorithm is pre-set to process on the integer pixel, half-pixel and quarter-pixel level. In this paper, the algorithm is validated on integer pixel precision and quarter-pixel precision. Up sampling the reference view prior to

the warping process may be used to improve the forward mapping quality [21].

In addition to the objective experimental data, Figs. 6 and 7 show some subjective results. In Fig. 6, partial enlarged detail clipped from Frame 5 and Frame 37, which belong to the sequence, *Bookarrival*, is shown to reveal the effects from different rendering approaches, whereas, Fig. 8 reveals Frame 56 and Frame 63, which belong to the sequence, *Newspaper*.

It is obvious that the presented method has better subjective performance than the other two methods. In fact, as shown in Figs. 6 (a) and 7 (a), some severe blurring effects are spotted along the transition regions between the foreground objects and the background area, whereas the VSRS adaptive blending yields ghost effects because of its blending theory, important background information is missing and pixel values are covered by the pixels nearby, the coarse texture detail could be spotted near the boundary in Figs. 6 (b) and 7 (b). The proposed framework warrants mentioning for its obvious effects in reducing ghost effects meanwhile preserving the background information near the foreground regions (Figs. 6 (c) and 7 (c)).

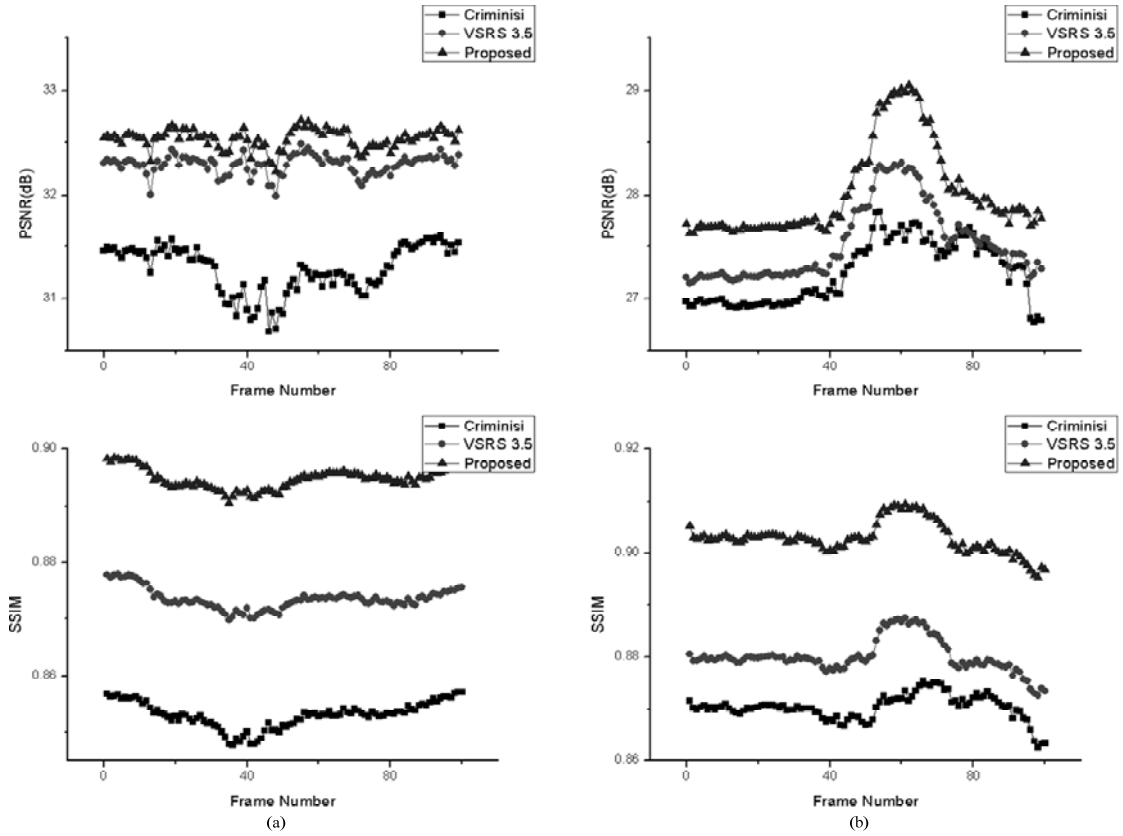

Fig. 5 Objective results frame by frame for (a) Bookarrival and (b) Newspaper

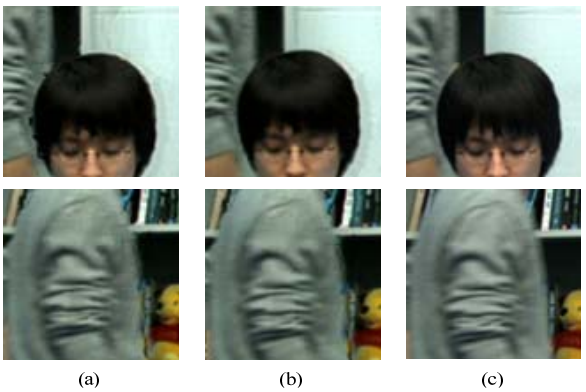

Fig. 6 Subjective results (top: frame 5. Bottom: frame 37): (a) Criminisi (b) VSRS 3.5 (c) Proposed

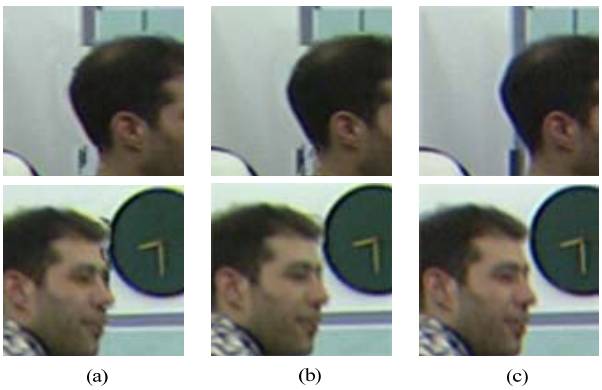

Fig. 7 Subjective results (top: frame 63. Bottom: frame 56): (a) Criminisi (b) VSRS 3.5 (c) Proposed

V. Conclusions and Future Works

In this paper, a synthesis framework utilizing the temporal correlations of texture and depth frames is presented, different roles are determined based on the baseline distance for the MVD format, the synthesized view image is blended mainly from information of the main reference view, the disocclusions are filled with three approaches: The background image, the information backward searched from the auxiliary view image and inpainting algorithms. The stable background image is obtained mainly with the classical Gaussian Mixture Model, and calibrated with its corresponding depth map. The reported results show that the proposed framework with novel blending approach yields good subjective and objective results. During the research, a better background sprite generation that can be adopted in the moving scene or captured with moving cameras will be discussed. In addition, the coarse depth map video generated with DERS leads to inevitable synthesis artifacts, mainly in the areas along the boundary between the background and foreground objects, a complex algorithm to reduce the impact before 3D warping will be explored.

Acknowledgment

The authors would like to thank the Fraunhofer Heinrich Hertz Institute (HHI) and Gwangju Institute of Science and Technology (GIST) to provide the test video sequence *Bookarrival* and *Newspaper,* respectively.

This work is supported by the Basic Research Distribution Project of Shenzhen (JCYJ20150827165024088) and Supporting Platform Project of Guangdong Province (2014B0909 -B001).

References

[1] M.C. Frederic Dufaux, Beatrice Pesquet-Popescu, *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*, JOHN WILEY & SONS INC, 2013.

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

[3] A.I. Purica, E.G. Mora, B. Pesquet-Popescu, M. Cagnazzo, and B. Ionescu, "Multiview plus depth video coding with temporal prediction view synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.26, no.2, pp.360–374, Feb. 2016.

[4] C.M. Cheng, S.J. Lin, S.H. Lai, and J.C. Yang, "Improved novel view synthesis from depth image with large baseline," *Proc. 19th Int. Conf. Pattern Recognition ICPR 2008*, pp.1–4, Dec. 2008.

[5] Z. w. Liu, P. An, S. x. Liu, and Z. y. Zhang, "Arbitrary view generation based on dibr," *Proc. Int. Symp. Intelligent Signal Processing and Communi -cation Systems ISPACS 2007*, pp.168–171, Nov. 2007.

[6] I. Ahn and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Transactions on Broadcasting*, vol.59, no.4, pp.614–626, Dec. 2013.

[7] L. Zhang, W.J. Tam, and D. Wang, "Stereoscopic image generation based on depth images," *Proc. Int. Conf. Image Processing ICIP '04*, pp.2993–2996 Vol. 5, Oct. 2004.

[8] L. Zhang and W.J. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Transactions on Broadcasting*, vol.51, no.2, pp.191–199, June 2005.

[9] P.J. Lee and Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Transactions on Multimedia*, vol.13, no.2, pp.246–254, April 2011.

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

[11] Y. Zhao, C. Zhu, Z. Chen, D. Tian, and L. Yu, "Boundary artifact reduction in view synthesis of 3D video: From perspective of texture depth alignment," *IEEE Transactions on Broadcasting*, vol.57, no.2, pp.510–522, June 2011.

[12] M. Bertalmio, A.L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition CVPR 2001*,pp.I–355–I–362 vol.1, 2001.

[13] M. Bertalmio, "Strong-continuation, contrast-invariant inpainting with a third-order optimal pde," *IEEE Transactions on Image Processing*, vol.15, no.7, pp.1934–1938, July 2006.

[14] M. Schmeing and X. Jiang, "Depth image based rendering: A faithful approach for the disocclusion problem," *Proc. Transmission and Display of 3D Video 2010 3DTV-Conf.: The True Vision - Capture*, pp.1–4, June 2010.

[15] K.Y. Chen, P.K. Tsung, P.C. Lin, H.J. Yang, and L.G. Chen, "Hybrid motion/depth-oriented inpainting for virtual view synthesis in multi-view applications," *Proc. Transmission and Display of 3D Video 2010 3DTV-Conf.: The True Vision - Capture*, pp.1–4, June 2010.

[16] M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering," *Proc. IEEE Int. Conf. Image Processing*, pp.1809–1812, Sept. 2010.

[17] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-D video," *IEEE Transactions on Multimedia*, vol.13, no.3, pp. 453–465, June 2011.

[18] E. Bosc, M. Köppel, R. Pépion, M. Pressigout, L. Morin, P. Ndjiki-Nya, and P.L. Callet, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," *Proc.18th IEEE Int. Conf. Image Processing*, pp.2597–2600, Sept. 2011.

[19] C. Yao, T. Tillo, Y. Zhao, J. Xiao, H. Bai, and C. Lin, "Depth map driven hole filling algorithm exploiting temporal correlation information," *IEEE Transactions on Broadcasting*, vol.60, no.2, pp.394–404, June 2014.

[20] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," *Proc. IEEE Computer Society Conf Computer Vision and Pattern Recognition*, p.252 Vol. 2, 1999.

[21] D. Tian, P.L. Lai, P. Lopez, and C. Gomila, "View synthesis techniques for 3D video," *Applications of Digital Image Processing XXXII*, ed. A.G. Tescher, SPIE-Intl Soc Optical Eng, aug 2009.

Deng Zengming received Bachelor degree in Microelectronics from Harbin Institute of Technology (HIT), Harbin, China, in 2010. After receiving Master degree in Microelectronics from HIT, Shenzhen Graduate School, he continues to pursue the Ph.D. degree in the same academy. His recent research interests include video processing, multi-view video coding and computer vision.

Wang Mingjiang received Bachelor and Master degree in Microelectronics from Harbin Institute of Technology (HIT), Harbin, China, in 1990 and 1993, respectively. From 1993 to 1995, he was a Teaching Assistant in National Integrated Circuit System Engineering Technology Research Center at Southeast University, Nanjing, China. In 1998, he received the Ph.D. degree in Microelectronics from Fudan University, Shanghai, China. After several years of senior engineer in Huawei Technologies Co Ltd, in 2003 he became an Associate Professor at HIT, Shenzhen Graduate School, where he became a Professor in 2009. He is currently the Doctoral Supervisor at HIT, and director of the HIT IOT terminal technology provincial key laboratory. He is currently leading and participating several provincial industry-university-research cooperation and national scientific research program. His recent research interests include video/audio processing, voice recognition algorithm and related hardware and related medical apparatus and instruments.