RI-007

Reliable View Synthesis with Automatic Error Compensation for FTV

Lu Yang † Tomohiro Yendo † Mehrdad Panahpour Tehrani † Toshiaki Fujii ‡ Masayuki Tanimoto †

1. Introduction

Recently, there are increasing research interests in FTV [1] that offers arbitrary views of 3D scene. View synthesis is important for prediction-based Multiview Video Coding (MVC) and novel view display in FTV and other multiview imaging application, such as 3DTV [2]. Generally, two original images, which come from left and right sides respectively, are defined as the references to generate the intermediate view using their depth maps. Depth maps are obtained off-line by stereo matching with energy optimization technologies [3]. Unfortunately, current stereo-based depth estimation works are prone to generate errors especially at textureless and boundary areas. Boundary errors produce noticeable artifacts while smooth errors degrade the overall PSNR of the generated view.

In many view synthesis methods, the straightforward interpolation without any error correction was used to generate the virtual view. Depth maps were projected to the virtual view, followed by backward mapping the reference views. Mori et al. [4] assumed that depth estimation was faithful to provide high-quality depth maps. Median and bilateral filters were used to smooth the projected depth maps before the backward projection. Only random noises of depth maps were suppressed by naïve filters, while large depth errors still generated significant synthesis errors. Lee and Ho [5] proposed the Boundary Noise Removal (BNR) procedure to reduce the artifacts in the background. They detected occlusion holes by projecting depth maps and defined noise areas around occlusions. However, artifacts in the foreground were not eliminated and depth errors of smooth areas were ignored. Besides that, occlusion detection was not reliable since depth maps were erroneous. In [6], Yang et al. proposed systematic reliability reasoning for the boundary artifacts reduction. Although most visible artifacts of their generated virtual views were successfully eliminated, the binary reliability reasoning in [6] also ignored smooth areas, thus degraded PSNR values of virtual views.

We find that view synthesis errors in the smooth areas are symmetric for the two references. This property can be employed to compensate projected references by each other. The projected left reference can be automatically compensated by the projected right reference and vice versa. For the boundary areas, our cross-check based reliability reasoning assigns each projected pixel a reliability value, which adaptively blends the pixel intensity and avoids artifacts. Furthermore, our proposed continuous reliability can also suppress the noise on smooth areas. Finally, both smooth synthesis noise and boundary artifacts are reduced in our reliability based view synthesis framework.

The rest of this paper is organized as follows. In Sec. 2, we introduce the systematic error analysis for the view synthesis using depth maps. Our reliability based view synthesis algorithm is proposed in Sec. 3. We demonstrate experimental evaluations in Sec. 4. Sec. 5 concludes the paper.

2. Error analysis

2.1 The synthesis error of smooth areas

We start to present our idea by analyzing the synthesis error for smooth areas. Generally, there are two symmetric input references: I_L , I_R and their depth maps D_L and D_R , from left side and right side, respectively.

Using those two references and their depth maps, we can generate two projected views for the virtual intermediate viewpoint:

I_1 : Projected I_L using D_L

I_2 : Projected I_R using D_R

Those two projected views are considered as two observations and will be blended or averaged to generate the final virtual view. Unfortunately, the independence between D_L and D_R leads to independent synthesis errors in I_1 and I_2 . This means that, simply blending or averaging them can not suppress the natural synthesis errors except random noise. In order to efficiently compensate these errors, we introduce other two projected views as additional observations:

I_3 : Projected I_R using D_L

I_4 : Projected I_L using D_R

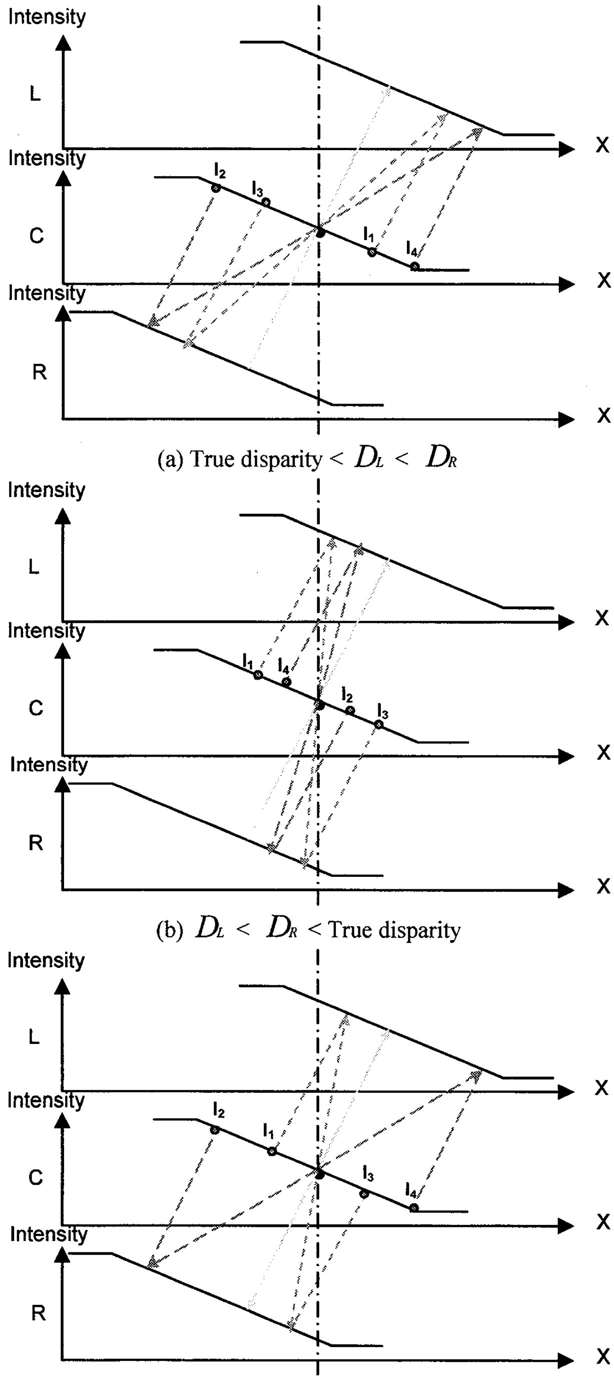
We can see I_1 , I_2 , I_3 and I_4 are the complete four combinations of two reference views with two depth maps.

The natural image can be considered as piece-wise continuous [7]. Since depth errors in smooth areas are usually small, the simplified piece-wise linear model is also suitable for our error analysis on local smooth parts of the virtual view. The four observations on the smooth area of synthesized view are simulated in Fig. 1. Note that we do not distinguish depth and disparity. Depth maps are projected to the virtual viewpoint. Here we assume D_L is smaller than D_R . Thus, we can obtain three different error patterns (Fig. 1(a)-(c)) dependent on the ground truth depth. The backward projection is utilized to synthesize each pixel on the center view. In order to clarify the presentation, the projected pixels of references are mapped to the coordinate of virtual viewpoint. In Fig. 1(a), both D_L and D_R are larger than the ground truth. In Fig. 1(b), both D_L and D_R are smaller than the ground truth. In Fig. 1(c), the ground truth is larger than D_L and smaller than D_R . If D_L is larger than D_R , we can obtain other three symmetric patterns. Therefore, we can include complete depth error cases for smooth areas.

It is interesting to see that for all depth error patterns in Fig. 1, observation errors have the symmetric property. I_1 and I_3 are always symmetrically distributed beside the ground truth pixel. Similarly, I_2 and I_4 are also symmetric to each other. On the other hand, I_1 and I_2 are not always symmetrically located, unless D_L and D_R are exactly same, which is not the usual case.

† Nagoya University

‡ Tokyo Institute of Technology



(a) True disparity < D_L < D_R

(b) D_L < D_R < True disparity

(c) D_L < True disparity < D_R

Green solid line: true disparity
Red short dashed line: D_L
Blue long dashed line: D_R

Fig. 1 Simulation of smooth errors

Besides that, from Fig. 1(c) we find that the blending of I_1 and I_2 may even generate worse synthesized view than the original observation. For example, since I_1 and I_2 in Fig. 1(c) are asymmetric located in the same side of ground truth pixel, the average of them will cause larger synthesis error than I_1. In contrast, the blending or averaging of I_1, I_2, I_3 and I_4 can compensate those errors automatically, by their symmetry property. As shown in Fig. 1, the average of these four observations should be the accurate value of ideal ground truth pixel (black dot in the middle line), as long as the linear surface assumption is true. We find that state-of-the-art depth estimation [8] usually generates piece-wise linear depth maps for local smooth areas. Thus, our error compensation with four observations is reasonable. The simulation of error cases when D_L is larger than D_R is similar to Fig. 1.

2.2 The synthesis error of boundary areas

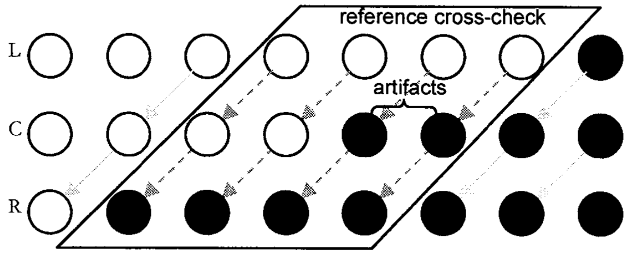
Now we consider boundary synthesis errors, which are visible artifacts and would significantly degrade the virtual view quality. The challenge is that the real synthesis error is between the projected reference and the unknown virtual view, which should be properly inferred. Here we follow [6] and use the reference cross-check to approximate the boundary synthesis error. The cross-check is carried out as following steps:

Step 1. The left (right) reference I_L (I_R) is projected to the right (left) viewpoint, using D_L (D_R).

Step 2. The intensity difference between the projected I_L (I_R) and original I_R (I_L) are computed.

Step 3. The cross-check differences e_L and e_R are projected back to the virtual viewpoint.

The cross-check example for left reference is shown in Fig. 2. We can see the boundary foreground pixels in the left reference have incorrect disparities (background disparity). Two of the foreground pixels are wrongly projected to the background area of the center view, thus artifact happens. The cross-check can successfully detect four erroneous projections, including the two real artifacts pixels. Although this cross-check enlarges the erroneous candidates, it can be considered as a minimum coverage for view interpolation. This means that we avoid to project artifacts pixels and their neighbor pixels. We find that the artifacts from references also compensate each other [6]. For example, the artifacts areas detected in left reference will not happen in right reference, and vice versa. Thus, we can compensate boundary artifacts by adaptively blending left and right references using reliability. The cross-check errors are used to formulate our reliability for view synthesis, which will be presented in next section.



Dashed line: cross-check detection. White circle: foreground pix.
Solid line: correct pixel projection. Black circle: background pix.

Fig. 2 Simulation of boundary errors

3. Reliable view synthesis

We have analyzed potential synthesis errors for both smooth areas and boundaries. However, those two kinds of errors should be properly combined to automatically suppress errors in our view synthesis framework. The main problem is how to adaptively synthesize each pixel. Our basic idea is to assign each pixel a reliability value, which is the weight for blending each pixel from each observation. In other words, reliability is the factor that determines how much contribution each observation pixel has. One important advantage is that the reliability can automatically integrate smooth error compensation and boundary artifacts reduction. For smooth areas, all pixels from observations will have comparable reliability values. Around boundaries, reliable observation pixels have large reliability values and dominate the blending. The occlusion holes will be considered as the most unreliable areas and should not be blended at all. Thus, we can quantify reliability by the cross check errors e_L and e_R , which are the approximate versions of real synthesis errors.

We define that the reliability for each non-occluded pixel of references should be inverse proportional to the corresponding cross-check error. This definition heuristically matches the truth that large synthesis errors should have small weight or reliability in the interpolation. For occluded pixels, we define their reliability as zero. Thus, for each pixel p in the virtual view, we have reliability values r_{1p} , r_{2p} , r_{3p} and r_{4p} , for its corresponding pixels in observations I_1 , I_2 , I_3 and I_4 , respectively.

$$r_{1p} = \begin{cases} \dfrac{1}{e_{Lp}^2 + t} & non-occl(D_L) \\ 0 & occl(D_L) \end{cases} \quad (1)$$

$$r_{2p} = \begin{cases} \dfrac{1}{e_{Rp}^2 + t} & non-occl(D_R) \\ 0 & occl(D_R) \end{cases} \quad (2)$$

$$r_{3p} = \begin{cases} \dfrac{1}{e_{Lp}^2 + t} & non-occl(D_L \,\&\, D_R) \\ 0 & occl(D_L \mid D_R) \end{cases} \quad (3)$$

$$r_{4p} = \begin{cases} \dfrac{1}{e_{Rp}^2 + t} & non-occl(D_L \,\&\, D_R) \\ 0 & occl(D_L \mid D_R) \end{cases} \quad (4)$$

Where e_{Lp} and e_{Rp} are the cross-check errors for the particular synthesized pixel p . The tuning parameter t is used to control the adaptivity of the reliability. For example, when t is large, the cross-check error can be ignored and all reliability values tend to be constant. In contrast, if t is zero, the reliability will be highly spatial varying.

Note that observations I_3 and I_4 contain both occlusions and disocclusions. Thus their valid areas should not be occluded in either left or right depth maps.

After the reliability computation, we can obtain the reliable view synthesis result for the intermediate virtual pixel f_p :

$$f_p = \frac{r_{1p}I_{1p} + r_{2p}I_{2p} + r_{3p}I_{3p} + r_{4p}I_{4p}}{r_{1p} + r_{2p} + r_{3p} + r_{4p}} \quad (5)$$

Where I_{1p} , I_{2p} , I_{3p} and I_{4p} are the corresponding pixels of p in observations I_1 , I_2 , I_3 and I_4 , respectively.

From (5), we can see our view synthesis algorithm is the pixel-wise weighted view blending. The reliability will automatically benefit correct projected pixels while preventing wrongly projected ones and compensating errors.

Finally, for the residual holes which can not be synthesized from either left or right references, we utilize inpainting [9] to fill them.

4. Experimental evaluations

The proposed method will be evaluated on two standard MPEG video sequences [10]. The first one is the "Champagne Tower" (Fig. 3) with 200 temporal frames. The view size of each single frame is 1280 x 960. "Champagne Tower" is a challenging sequence for view synthesis. Due to the sharp contrast between different object colors, boundary depth errors would generate visible artifacts. The textureless background and foreground also cause depth errors in smooth areas and add noise in the observations. Another test sequence is the "Book arrival" (Fig. 3) with 100 temporal frames. Each single frame has the size of 1024 x 768. For both two sequences, we utilize MPEG depth estimation reference software [11] to generate corresponding depth maps. The depth estimation is based on stereo matching using graph cuts [8]. Thus, depth errors on the textureless smooth areas and boundaries are inevitable.

All frames are well calibrated and rectified. We use left and right references and their depth maps to generate the center view. The tuning parameter t is heuristically set to be 160. The depth maps used in our experiments are shown in Fig. 3. Since human eyes are much more sensitive to the changes of luminance than chrominance, we compute the reliability in luminance channel for each pixel of observations. The improvements of the proposed method are measured by the comparisons against the

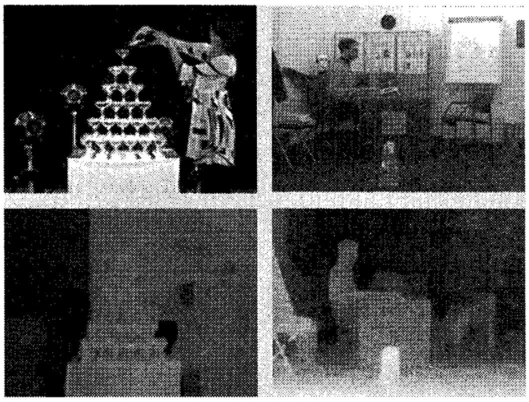


Fig. 3 Left references (top) and the depth maps (bottom). From left to right: Champagne Tower and Book Arrival

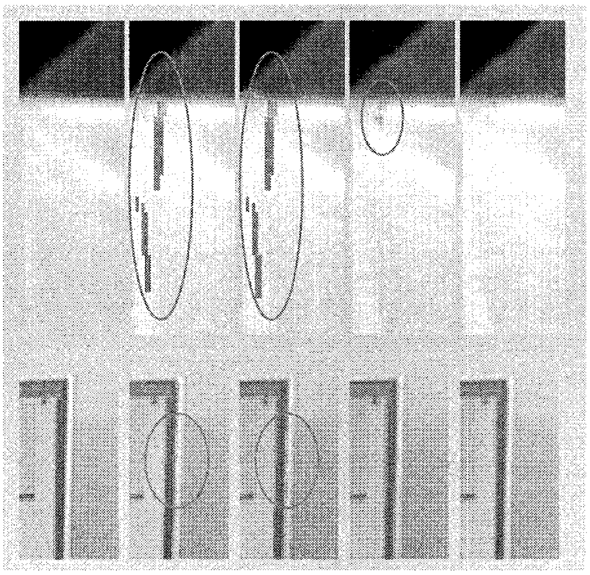


Fig. 4 Local synthesis results for "Champagne" (top) and "Book" (bottom). From left to right: ground truth, synthesis without reliability, BNR, binary reliability method and proposed method

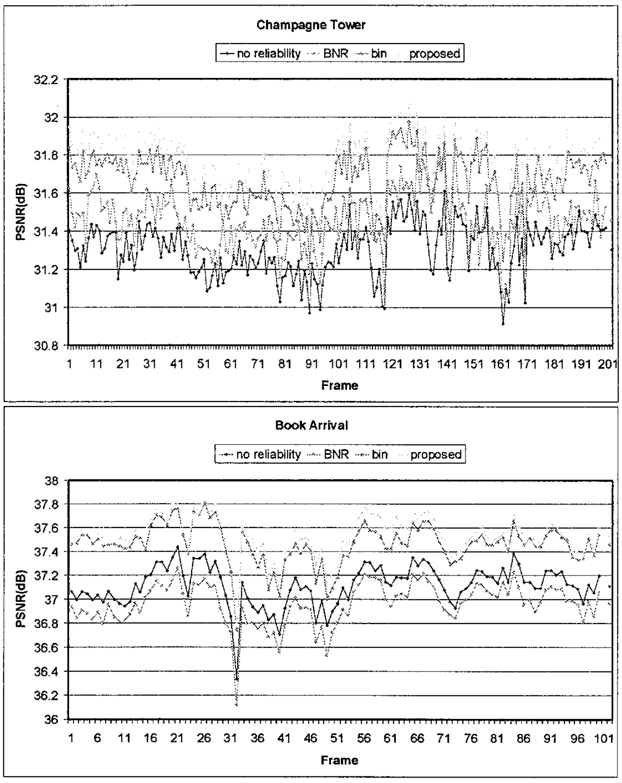


Fig. 5 PSNR comparisons

view synthesis without reliability [12], view synthesis with BNR [5], view synthesis with binary reliability [6] and the ground truth. We magnify the local areas for visual purpose. Fig. 4 shows the comparisons of different view synthesis results. We can see the proposed method generates the fewest artifacts in comparison to other methods.

We find that both binary reliability based view synthesis and our proposed method can significantly reduce the artifacts around boundaries (red circles). However, our method still generates fewer artifacts in smooth areas. The objective PSNR comparisons are also illustrated in Fig. 5. It is clear that the proposed method consistently outperforms all other methods on all test frames.

5. Conclusions

In this paper, we introduce a new reliability based view synthesis method which automatically suppresses the synthesis errors for FTV. The proposed reliability is quantitively estimated based on the reference cross-check, which gives each pixel an individual weight for the view synthesis. We demonstrate our experimental results on MPEG sequences and show the outperformance of our method both at subjective artifacts suppression and objective PSNR improvement.

ACKNOWLEDGMENTS

This work is partially supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) 093106002 of the Ministry of Internal Affairs and Communications of Japan and China Scholarship Council (CSC). We would like to thank Heinrich-Hertz-Institut (HHI) for providing "Book Arrival".

REFERENCES

[1] M. Tanimoto, "Overview of free viewpoint television", Signal Process.: Image Commun., Vol. 21, No. 6 (2006).

[2] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, C. Zhang, "Multi-view imaging and 3DTV", IEEE Signal Process. Mag., Vol. 24, No. 6 (2007).

[3] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors", IEEE Trans. Pattern Anal. & Mach. Intell. Vol. 30, No. 6 (2008).

[4] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, M. Tanimoto, "View generation with 3D warping using depth in-formation for FTV", Signal Process.: Image Commun., Vol. 24, No. 1-2 (2009).

[5] C. Lee, Y. S. Ho, "Boundary filtering on synthesized views of 3D video", FGCN 2008 (2008).

[6] L. Yang, T. Yendo, M. Panahpour Tehrani, T. Fujii, M. Tanimoto, "Artifact reduction using reliability reasoning for image generation of FTV", J. Vis. Commun., doi:10.1016/j.jvcir.2009.09.009, In Press (2009).

[7] Stan Z. Li, "Markov Random Field Modeling in Image Analysis", Tokyo: Springer-Verlag (2001).

[8] V. Kolmogorov, R. Zabih, "Computing visual correspondence with occlusions using graph cuts", ICCV 2001 (2001).

[9] M. Bertalmio, A. L. Bertozzi, G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting", CVPR 2001 (2001).

[10] ISO/IEC JTC/SC29/WG11, "Coding of moving pictures and audio", N9783 (2008).

[11] ISO/IEC JTC/SC29/WG11, "Coding of moving pictures and audio", M15377 (2008).

[12] ISO/IEC JTC/SC29/WG11, "Coding of moving pictures and audio", M16090 (2009).