DEPTH-BASED IMAGE COMPLETION FOR VIEW SYNTHESIS

Josselin Gautier[(1)]*, Olivier Le Meur*[(1)]*, Christine Guillemot*[(2)]

[(1)] IRISA / Université de Rennes 1 - [(2)]INRIA, Campus de Beaulieu - 35042 Rennes, France

ABSTRACT

This paper describes a depth-based inpainting algorithm which efficiently handles disocclusion occurring on virtual viewpoint rendering. A single reference view and a set of depth maps are used in the proposed approach. The method not only deals with small disocclusion filling related to small camera baseline, but also manages to fill in larger disocclusions in distant synthesized views. This relies on a coherent tensor-based color and geometry structure propagation. The depth is used to drive the filling order, while enforcing the structure diffusion from similar candidate-patches. By acting on patch prioritization, selection and combination, the completion of distant synthesized views allows a consistent and realistic rendering of virtual viewpoints.

Index Terms— DIBR, FTV, 3DTV, view synthesis, image completion, exemplar-based inpainting

1. INTRODUCTION

3DTV and FTV are promising technologies for the next generation of home and entertainment services. Depth Image Based Rendering (DIBR) are key-solutions for virtual view synthesis on multistereoscopic display from any subset of stereo or multiview plus depth (MVD) videos. Classical methods use depth image based representations (MVD, LDV [1]) to synthesize intermediate views by mutual projection of two views. Then, disoccluded areas due to the projection of the first view to the new one could be filled in with the remaining one.

However, in freeviewpoint video (FVV) applications, larger baseline (distance or angle between cameras) involves larger disoccluded areas. Traditional inpainting methods are not sufficient to complete these gaps. To face this issue the depth information can help to guide the completion process.

The use of depth to aid the inpainting process has already been considered in the literature. Oh et al. [2] based their method on depth thresholds and boundary region inversion. The foreground boundaries are replaced by the background one located on the opposite side of the hole. Despite the use of two image projections, their algorithm relies on an assumption of connexity between disoccluded and foreground regions, which may not be verified for high camera baseline configurations. Indeed, upon a certain angle and depth, the foreground object does not border the disoccluded part anymore. Daribo et al. [3] proposed an extension to the Criminisi's [4] algorithm by including the depth in a regularization term for priority and patch distance calculation. A prior inpainting of the depth map was performed.

Our approach relies on the same idea. However, our contributions are threefold. The relevance of patch prioritization is improved by first using the depth as a coherence cue through a 3D tensor, and then by using a directional term preventing the propagation from the foreground. A combination of the K-nearest neighbor candidates is finally performed to fill in the target patch.

We present in Section 2 contributions to this priority calculation, based on tensor and then on depth, before describing depth-based patch matching. Section 3 describes the implementation of the method in a MVD context. Results are given in Section 4, as well as a comparison with existing approaches. Conclusions are drawn in Section 5.

2. ALGORITHM

The motivation to use a Criminisi-based algorithm resides in its capacity to organize the filling process in a deterministic way. As seen in fig.1, this technique propagates similar texture elements $\Psi_{\hat{q}}$ to complete patches Ψ_p along the structure directions, namely the isophotes. Their algorithm basically works in two steps. The first step defines the higher order patch priorities along the borders $\delta\Omega$. The idea is to start from where the structure is the strongest (in term of local intensity, with $D(p)$) and from patches containing the highest number of known pixels, $C(p)$. The priority is then expressed as $P(p) = D(p) \times C(p)$. The second step consists in searching for the best candidate in the remaining known image in decreasing priority order.

In the context of view synthesis, some constraints can be added to perform the inpainting and improve the natural aspect of the final rendering. The projection in one view will be along the horizontal direction. For a toward-right camera movement the disoccluded parts will appear on the right of their previously occluding foreground (Fig.1a), and oppositely for a toward-left camera movement. Whatever camera's movement, these disoccluded areas should always be filled in with pixels from the background rather than the foreground.

978-1-61284-162-5/11/$26.00

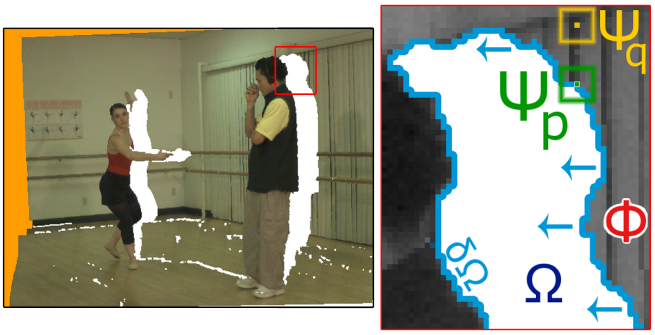

Fig. 1. Illustration of principle. On (a) a warped view, (b) a zoom on the disoccluded area behind the person on the right, with the different elements overlaid.

Based on this a priori knowledge, we propose a depth-based image completion method for view synthesis based on robust structure propagation. In the following, $D(p)$ is described.

2.1. Tensor-based priority

First, the data term $D(p)$ of the inpainting method proposed by [4] involving the color structure gradient is replaced with a more robust structure tensor. This term is inspired by partial differential equation (PDE) regularization methods on multi-valued images and provides a more coherent local vector orientation [5]. The Di Zenzo matrix [6] is given by:

$$J = \sum_{l=R,G,B} \nabla I_l \nabla I_l^T = \sum_{l=R,G,B} \begin{pmatrix} \frac{\partial I_l}{\partial x}^2 & \frac{\partial I_l}{\partial x}\frac{\partial I_l}{\partial y} \\ \frac{\partial I_l}{\partial x}\frac{\partial I_l}{\partial y} & \frac{\partial I_l}{\partial y}^2 \end{pmatrix}$$

with ∇I_l the local spatial gradient over a 3x3 window. This tensor can also be smoothed with a gaussian kernel G_σ to give robustness to outliers, without suffering from cancellation effects. We call it $J_\sigma = J * G_\sigma$. Finally, the local vector orientation is computed from the structure tensor J_σ. Its eigenvalues $\lambda_{1,2}$ reflect the amount of structure variation, while its eigenvectors $v_{1,2}$ define an oriented orthogonal basis. Of particular interest is v_2 the preferred local orientation and its "force" λ_2. Based on the coherence norm proposed in [7], the data term $D(p)$ is then defined as:

$$D(p) = \alpha + (1-\alpha) exp\left(\frac{-C}{(\lambda 1 - \lambda 2)^2}\right)$$

with C a constant positive value and $\alpha \in [0,1]$. Flat regions (when $\lambda_1 \approx \lambda_2$) do not favor any direction, it is isotropic, while with strong edges ($\lambda_1 >> \lambda_2$) the propagation begins along the isophote.

2.2. Depth-aided and direction-aided priority

The priority computation has been further improved by exploiting the depth information, first by defining a 3D tensor product, secondly by constraining the side from where to start inpainting.

2.2.1. *3D tensor*

The 3D tensor allows the diffusion of structure not only along color but also along depth information. It is critical to jointly favor color structure as well as geometric structure. The depth-aided structure tensor is extended with the depth map taken as an additional image component Z:

$$J = \sum_{l=R,G,B,Z} \nabla I_l \nabla I_l^T$$

2.2.2. *One side only priority*

The second improvement calculates the traditional priority term along the contour in only one direction. Intuitively, for a camera moving to the right, the disocclusion holes will appear to the right of foreground objects, while out-of-field area will be on the left of the former left border (in orange in Fig.1a). We then want to prevent structure propagation from foreground by supporting the directional background propagation, as illustrated in Fig.1b with the blue arrows.

The patch priority is calculated along this border, the rest of the top, bottom and left patches being set to zero. Then for disocclused areas, the left border possibly connex to foreground will be filled at the very end of the process. For out-of-field areas, even if left borders are unknown, we will ensure to begin from the right border rather than possible top and bottom ones.

These two proposals have been included in the prioritization step.

2.3. Patch matching

Once we precisely know from where to start in a given projected image, it is important to favor the best matching candidates in the background only. Nevertheless, starting from a non-foreground patch does not prevent it from choosing a candidate among the foreground, whatever the distance metric used. Thus, it is crucial to restrict the search to the same depth level in a local window: the background. We simply favor candidates in the same depth range by integrating the depth information in the commonly used similarity metric, the SSD (Square Sum of Differences):

$$\Psi_{\hat{q}} = \underset{\Psi_q \in \Phi}{\arg\min}\, d(\Psi_{\hat{p}}, \Psi_q) \text{ with } d = \sum_{p,q \in \Psi_{p,q} \cap \Phi} \alpha_l \left\| \Psi_{\hat{p}} - \Psi_q \right\|^2$$

The depth channel is chosen to be as important as the color one ($l \in R,G,B,Z$ with $\alpha_{R,G,B} = 1$ and $\alpha_Z = 3$). Then it will not prevent the search in foreground patches, but will seriously penalize and unrank the ones having a depth difference above, i.e in front of the background target patch.

As proposed by [8], a combination of the best candidates to fill in the target patch shows more robustness than just duplicating one. We use a weighted combination of the K-best patches depending on their exponential SSD distances to the original patch. ($K = 5$ in our experiments).

3. IMPLEMENTATION

Experiments are performed on an unrectified Multiview Video-plus-Depth (MVD) sequence "Ballet" from Microsoft [9]. The depth maps are estimated through a color segmentation algorithm [9] and are supplied with their camera parameters. The choice of this sequence is motivated by the wide baseline unrectified camera configuration as well as its highly depth-and-color contrast resulting in distinct foreground-background.This makes the completion even more visible and the issue even more challenging.

First, the central view 5 is warped in different views. Standard cracks (unique vacant pixels) are filled in with an average filter. We then suppress certain ghosting effects present on the borders of disoccluded area in the background: the background ghosting. Indeed, as we start the filling process by searching from the border, it is of importance to delete ghostings containing inadequate foreground color values. A Canny edge detection on the original depth map, followed by a deletion of color pixels located behind that dilated border successfully removes this ghosting.

Finally, our inpainting method is applied on each warped image, using the depth of the final view. The depth inpainting issue is out of the scope of this paper, but encouraging methods are proposed in the literature [3]. In the context of MVD applications, it is realistic to consider a separate transmission of depth information through geometric representation (currently under investigation).

4. RESULTS

Fig.2 illustrates the results obtained with the proposed method, comparatively with methods from the literature [4], [3], when rendering views located at varying distances from the reference viewpoint. The three versions take in input the same color and depth information, except for the approach in [4] using color only. Our method not only preserves the contour of foreground persons, but also successfully reconstructs the structure of missing elements of the disoccluded area (i.e. edges of the curtains and bars behind the person on the right, background wall behind the left one).

Thanks to our combination term, we can even extend the synthesis to very distant views, without suffering of aliasing effects. As illustrated, the view 5 is projected to view 2 ($V_{5\rightarrow 2}$) and the out-of-field blank areas occupying one quarter width of the warped image are reconstructed. The counterpart of the patch combination is the smoothing effect appearing on the bottom part of this area. By taking different numbers of patches for combination, it is possible to limit this effect. We encourage people to refer to additional results available on our webpage[1] with videos illustrating the priority-based progressive inpainting principle. The results can indeed be essentially address visually, as argued by [10].

[1] http://www.irisa.fr/temics/staff/gautier/inpainting

5. CONCLUSION AND PERSPECTIVES

A robust depth based completion method for view synthesis has been presented. We address the disocclusion issue by going beyond the limitations of scene warping. To start inpainting, coherent depth and color structures are favored along contour through a robust tensor-based isophote calculation while directional inhibition prevents to start from foreground borders. For target patch propagation, a combination of closest geometric and photoconsistent candidates manages effective natural filling. Future works will focus on completion of synthesized views extremely far from the reference view. The natural aspect of this filling in situation, i.e for video, will also be investigated.

6. ACKNOWLEDGMENT

The authors would like to thank I. Daribo and B. Pesquet for providing their source code, and P. Pérez for inspiring discussions. This work was supported by ANR through the Persée project.

7. REFERENCES

[1] J. Shade, S. Gortler, L. He, and R. Szeliski, "Layered depth images," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM New York, NY, USA, 1998, pp. 231–242.

[2] K.J. Oh, S. Yea, and Y.S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-d video," in *PCS*, 2009, pp. 1–4.

[3] I. Daribo and B. Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE International Workshop on Multimedia Signal Processing*, 2010.

[4] A. Criminisi, P. Pérez, 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, 2004.

[5] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving pde's," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 65–82, 2006.

[6] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 116–125, 1986.

[7] J. Weickert, "Coherence-enhancing diffusion filtering," *International Journal of Computer Vision*, vol. 31, no. 2, pp. 111–127, 1999.

[8] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE transactions on PAMI*, pp. 463–476, 2007.

[9] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *ACM SIGGRAPH*. ACM New York, NY, USA, 2004, pp. 600–608.

[10] N. Kawai, T. Sato, and N. Yokoya, "Image inpainting considering brightness change and spatial locality of textures," in *Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, 2008, vol. 1, pp. 66–73.

(a) $V_{5\to4}$ after warping and background antighosting

(b) $V_{5\to2}$ after warping and background antighosting

(c) $V_{5\to4}$ inpainted with Criminisi's method

(d) $V_{5\to2}$ inpainted with Criminisi's method

(e) $V_{5\to4}$ inpainted with Daribo's method

(f) $V_{5\to2}$ inpainted with Daribo's method

(g) $V_{5\to4}$ inpainted with our method

(h) $V_{5\to2}$ inpainted with our method

Fig. 2. Illustration of different methods of inpainting. Our approach relying on 3D tensor and directional prioritization shows efficient filling.