

Depth guided image inpainting algorithm for free-viewpoint video

Citation for published version (APA):

Ma, L., Do, Q. L., & With, de, P. H. N. (2012). Depth guided image inpainting algorithm for free-viewpoint video. In *IEEE International Conference on Image Processing (ICIP), 30 September- 3 October 2012, Florida, USA* (pp. 1721-1724). Institute of Electrical and Electronics Engineers.

Document status and date:

Published: 01/01/2012

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

DEPTH-GUIDED INPAINTING ALGORITHM FOR FREE-VIEWPOINT VIDEO

Lingni Ma, Luat Do and Peter H. N. de With

Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, the Netherlands

ABSTRACT

Free-Viewpoint Video (FVV) is a novel technique which creates virtual images of multiple direction by view synthesis. In this paper, an exemplar-based depth-guided inpainting algorithm is proposed to fill disocclusions due to uncovered areas after projection. We develop an improved priority function which uses the depth information to impose a desirable inpainting order. We also propose an efficient background-foreground separation technique to enhance the accuracy of hole filling. Furthermore, a gradient-based searching approach is developed to reduce the computational cost and the location distance is incorporated into patch matching criteria to improve the accuracy. The experimental results have shown that the gradient-based search in our algorithm requires a much lower computational cost (factor of 6 compared to global search), while producing significantly improved visual results.

Index Terms— Free-viewpoint video, depth image, inpainting, exemplar-based algorithm, background-foreground separation

1. INTRODUCTION

Recently, substantial research has been devoted to the development of advanced 3DTV technologies. One interesting field is view synthesis, where multiple views are generated from existing cameras. A popular technique in view synthesis is Depth Image Based Rendering (DIBR) which uses depth information to generate virtual views by warping [1]. DIBR provides opportunities for new applications, such as Free-Viewpoint Video (FVV), which allows users to view a scene from multiple directions. However, DIBR also creates projection artifacts in virtual images. One major problem is holes or disocclusions resulting from uncovered areas after projection. Image inpainting [2] offers a promising solution to fill these holes.

In literature, there are mainly two types of inpainting algorithms. One type is based on Partial Differential Equations (PDE), which propagates structures into holes via diffusion [3]. The other type is based on exemplars, which copies known pixels for hole filling. Since the PDE-based approaches produce noticeable blurring for large holes, the exemplar-based algorithms are more attractive for our problem. A pioneering work for exemplar-based algorithms has been developed by Criminisi *et al.* [4], where a priority function is proposed to determine the inpainting order such that linear structures are propagated correctly to connect broken lines. To enhance the speed and accuracy, this algorithm has been further developed by other researchers [5–9]. However, these inpainting algorithms are insufficient for FVV inpainting, because the disocclusions in FVV usually reside in the background. To handle the depth ambiguities, the associated depth images can be used and this idea has been employed by others. Oh *et al.* [10] attempted to replace the foreground boundary of holes with the background one. Their intention to inpaint with background is not completely achieved, since foreground boundaries are not always well identified and properly

replaced. Daribo *et al.* [11] extended Criminisi’s algorithm by modifying the priority function and the patch matching criteria. Their results contain noticeable errors, since the assumption that patches of low depth variance are equal to the background, is not always valid. Gautier *et al.* [12] proposed prevention of foreground propagation by inpainting warped images whose projection direction is known. However, the foreground is still considered for hole filling, so that background disocclusions are not inpainted correctly. Köppel and Ndjiki-Nya *et al.* [13, 14] exploit the temporal information by building a background map. In our case, we are limited by frame-based processing to lower the complexity.

In this paper, we propose an improved exemplar-based algorithm for FVV inpainting using four advanced techniques. First, an enhanced priority function is constructed such that holes are inpainted from the background to the foreground. Second, a local foreground-background separation method is proposed which efficiently prevents the propagation of foreground. Third, a gradient-based searching is developed to limit the search-window size to reduce the computational cost. Last, the patch matching criterion is modified that optimizes the inpainting accuracy. The experimental results show that our algorithm significantly improves the FVV inpainting performance. The sequel of this paper is organized as follows. Section 2 gives background information for FVV inpainting and a brief overview of Criminisi’s algorithm. Section 3 explains our proposed algorithm in detail. Section 4 shows the experimental results and evaluations. Section 5 presents the final conclusions.

2. BACKGROUND

In this section, we provide background information related to FVV inpainting and Criminisi’s algorithm. The FVV inpainting problem is illustrated in Fig. 1 with the synthesized image and its associated depth image. We use synthesized images warped from two existing cameras. FVV inpainting differs from image inpainting in two aspects. First, disocclusions in FVV are usually holes residing in the background textures. Second, each FVV image is associated with a depth image, which provides information about the relative distances of the objects with respect to the camera.

Let us now briefly describe Criminisi’s method, which forms the basis for our algorithm. Criminisi’s method is an iterative process, which consists of three main steps. In the first step, a target patch with the highest priority is located to ensure a desirable inpainting order. The second step searches the entire image for a proper candidate that matches the target patch. This patch matching is determined by finding the smallest *texture distance*, which is the sum of squared errors between the target patch and a candidate patch. The third step updates the image by copying the candidate patch to fill the target patch. Criminisi’s algorithm is limited to normal images, since it does not distinguish background from foreground textures. In the next section, we show that exemplar-based algorithms can be improved using the available depth information.

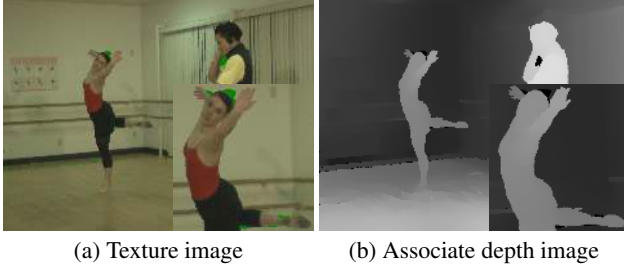


Fig. 1: FVV inpainting problem: synthesized texture from two cameras and the associated depth image, with disocclusions denoted by green and black (near the heads of the persons).

3. DEPTH-GUIDED INPAINTING ALGORITHM

In this section, we describe our proposed algorithm in detail. For each stage of Criminisi’s algorithm, we have provided new techniques for improvement. In particular, we have used the available depth information to enhance the priority function. In addition, a new step is inserted after the priority computation that separates the background from the foreground. The complete diagram for our algorithm is depicted in Fig. 2. In the following subsections, we present the new techniques involved in the diagram, which are *depth-based priority computation*, *background-foreground segmentation*, *gradient-based searching* and *distance-related patch matching*.

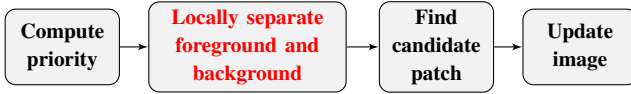


Fig. 2: Principal steps of our proposed algorithm for one iteration.

3.1. Depth-Based Priority Computation

Since Criminisi’s priority function does not distinguish the foreground and background, it leads to the propagation of foreground into background. To achieve a better inpainting order, the patch of larger depth which represents the relative background should be filled first. With this assumption, a new priority function is proposed and specified by

$$P(p) = C(p) \times D(p) \times Z(p),$$

where the confidence term $C(p)$ and data term $D(p)$ are defined according to Criminisi’s algorithm [4]. The new *depth term* Z_p is specified by

$$Z(p) = 1 - \frac{\bar{z}_p}{z_{max}},$$

where \bar{z}_p is the average depth of patch Ψ_p based on all known pixels, while z_{max} denotes the maximum depth of the entire image. Since the foreground usually contains a much higher depth value than the background, this new definition of the priority function ensures a correct inpainting from the background to the foreground as shown in Fig. 3(b). In contrast to the technique in [12], we arrange the proper inpainting order using the depth term without prior knowledge about camera positions.

3.2. Background-Foreground Segmentation

A desirable inpainting order is not always sufficient to prevent foreground propagation, as we have found that it is better to prevent

the use of these foreground pixels for hole filling. This explains why existing algorithms fail to produce satisfying results. To solve this problem, we propose a method to exclude foreground objects in the patch updating processing by background-foreground segmentation. From our experiments, we have observed that although both foreground and background cover a wide range of depth values, the high contrast in depth is still preserved when limited to a local region. This observation promotes a local segmentation technique by thresholding using the Ostu’s method [15]. This is a nonparametric and unsupervised algorithm to select an optimal threshold by maximizing the discriminant measure of separability, i.e. the between class variance. Ostu’s method works especially well when the entire distribution is a mixture of two distinguishable clusters, which is usually the case when a local region is considered. Fig. 4 demonstrates the efficiency of local segmentation for the image content, as shown in Fig. 1.

It should be noted that we only exclude foreground pixels from filling disocclusions. Patches containing foreground pixels are still considered when searching for the optimal candidate, because these patches are also very likely to contain desirable background information. The difference between existing algorithms and ours is that we only copy background pixels to fill holes, thereby effectively preventing the propagation of foreground pixels.

3.3. Gradient-Based Searching

Most exemplar-based algorithms use a global search to find the optimal candidate. However, the global searching is not only computationally expensive, but it is also unnecessary in many situations. Therefore, we have developed a method that adapts the search window size proportional to the gradient magnitude of the patch. We observe that disocclusions surrounded by intricate details usually need a large search window to match their variation, while holes in smooth texture can be well inpainted with a neighboring patch. Since the magnitude of image gradient serves as a good indication for the texture variation, we set the radius r of the symmetric search window as a function of the gradient magnitude g by specifying

$$r = R^{\alpha g + \beta},$$

where R is the maximum search radius and g is measured as the highest gradient magnitude of neighboring known pixels. The variables α and β are coefficients constrained by two conditions such that $R^{\alpha G_L + \beta} = R_L$ and $R^{\alpha G_U + \beta} = R_U$, where R_L and R_U are the lower and upper bound for the search radius, respectively. Likewise, variables G_L , G_U are the lower and upper bound for gradient magnitude, respectively. This mapping correlates the search-window size with the texture variation, i.e. the higher the variation, the larger

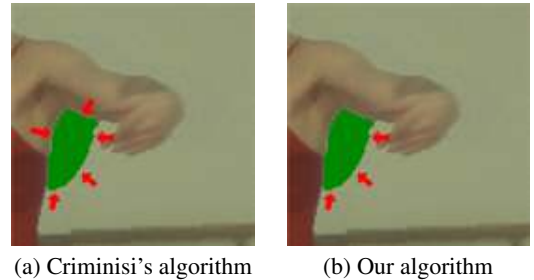


Fig. 3: Inpainting direction determined by Criminisi’s priority function and our proposed priority function.

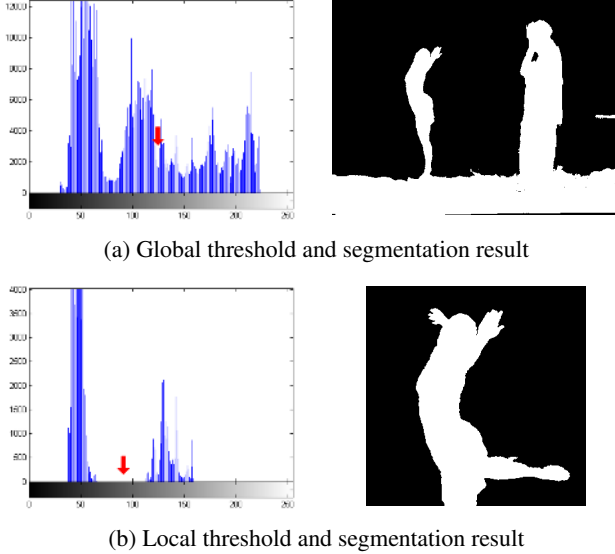


Fig. 4: Segmentation of foreground (white region) and background. The global segmentation fails to extract the complete foreground object, while the local segmentation overcomes this limitation.

the search window. From our experiments we have observed that most holes in FVV are surrounded by low-frequency textures and thus our modification highly reduces the computational cost without sacrificing the performance.

3.4. Distance-Related Patch Matching

We have observed that a patch is more likely to resemble its neighboring patches than its far away counterparts. This observation stimulates the use of location distance as a penalty for the patch matching. We propose the following procedure to select the optimal patch. First, we rank the patches according to their texture distance as defined in Criminisi’s algorithm. Second, we select the first n patches with the shortest texture distance. Third, the location distance is added as the penalty to the texture distance and these patches are re-ranked. The patch with the shortest distance is selected as a candidate. We define the location distance penalty by

$$D(\Psi_p, \Psi_q) = \gamma \cdot (|x_p - x_q| + |y_p - y_q|),$$

where γ is the weight and (x_p, y_p) , (x_q, y_q) are the centers of patch Ψ_p and Ψ_q , respectively. To prevent propagation of inpainting errors, we do not reuse partially inpainted patches. The modified patch matching produces better results in two situations. First, when several candidates have the same texture distance, the location distance helps to select the correct patch. Moreover, when a candidate with minimum texture distance is further away from the target, the higher penalty reduces the risk to include undesired details.

3.5. Preprocessing Step for Implementation

In our implementation, we have observed that using the depth information to guide the filling process is not always necessary. For example, small holes (containing pixels less than half of the patch size) and boundary holes can be sufficiently inpainted with their neighboring pixels. In addition, small holes do not demand a critical

inpainting order. Therefore, we classify the disocclusions into *regular holes*, *small holes* and *boundary holes*. Since the background-foreground separation requires more computation, we only apply it to regular holes to achieve a better efficiency. The classification of holes also plays a crucial role in producing accurate results.

4. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the performance of our proposed algorithm, two series of experiments have been carried out in MatLab. In the first series of experiments, the visual quality is assessed by using the well-known 3D video sequence “Ballet”. We take a patch size of 15×15 pixels and limit the maximum and minimum search window to 151×151 pixels and 15×15 pixels, respectively. For comparison, we have also implemented the Criminisi’s and Daribo’s algorithms. The former algorithm offers the reference for patch-based inpainting without using depth, while the latter does. The results in Fig. 5 show that in a worst-case scenario, our algorithm clearly outperforms the other algorithms, because the propagation of foreground objects into background texture is prevented. The average PSNR of 100 frames in the disoccluded areas for Criminisi’s, Daribo’s and our algorithm are 28.5, 28.7 and 28.8 dB, respectively. Generally, our system shows quality improvements in two situations. First, the contours of foreground objects are well preserved, while the shape of foreground objects in both Criminisi’s and Daribo’s algorithm are deformed. Second, our algorithm produces improved perceptual results for holes that are completely surrounded by foreground objects. While the Criminisi’s and Daribo’s algorithms introduce artifacts, our algorithm fills the disocclusions with desirable information.

In the second series of experiments, we compare the execution time of inpainting with global searching and our gradient-based searching for images of various size. The test set consists of images with sizes varying from 320×240 , 640×320 , 768×576 , 800×600 to 1024×768 pixels. For each size, 10 images are tested and the average iteration cycle time is given in Table 1. It can be observed that the gradient-based searching speeds up the inpainting process drastically, especially for images with high resolution. In particular, for images with 1024×768 pixels, the gradient-based searching is almost 6 times faster than the global searching.

| Image Size (pixels) | Patch Size (pixels) | GS per iteration (ms) | GBS per iteration (ms) |
|---------------------|---------------------|-----------------------|------------------------|
| 320×240 | 09×09 | 34.2 | 11.6 |
| 640×480 | 11×11 | 86.6 | 24.4 |
| 768×576 | 11×11 | 123.3 | 29.9 |
| 800×600 | 13×13 | 129.0 | 30.5 |
| 1024×768 | 15×15 | 200.7 | 34.8 |

Table 1: Comparison of computational cost between Global Searching (GS) and Gradient-Based Searching (GBS).

5. CONCLUSIONS

In this paper, we have proposed a depth-guided inpainting algorithm for Free-Viewpoint Video. Four new techniques have been developed to improve existing exemplar-based inpainting algorithms. First, the depth information is added to the priority function, which helps to impose the desirable inpainting order. Second, an efficient local segmentation approach is proposed to prevent the propagation of foreground objects into background texture. Third, a gradient-based searching is developed to lower the computational cost by

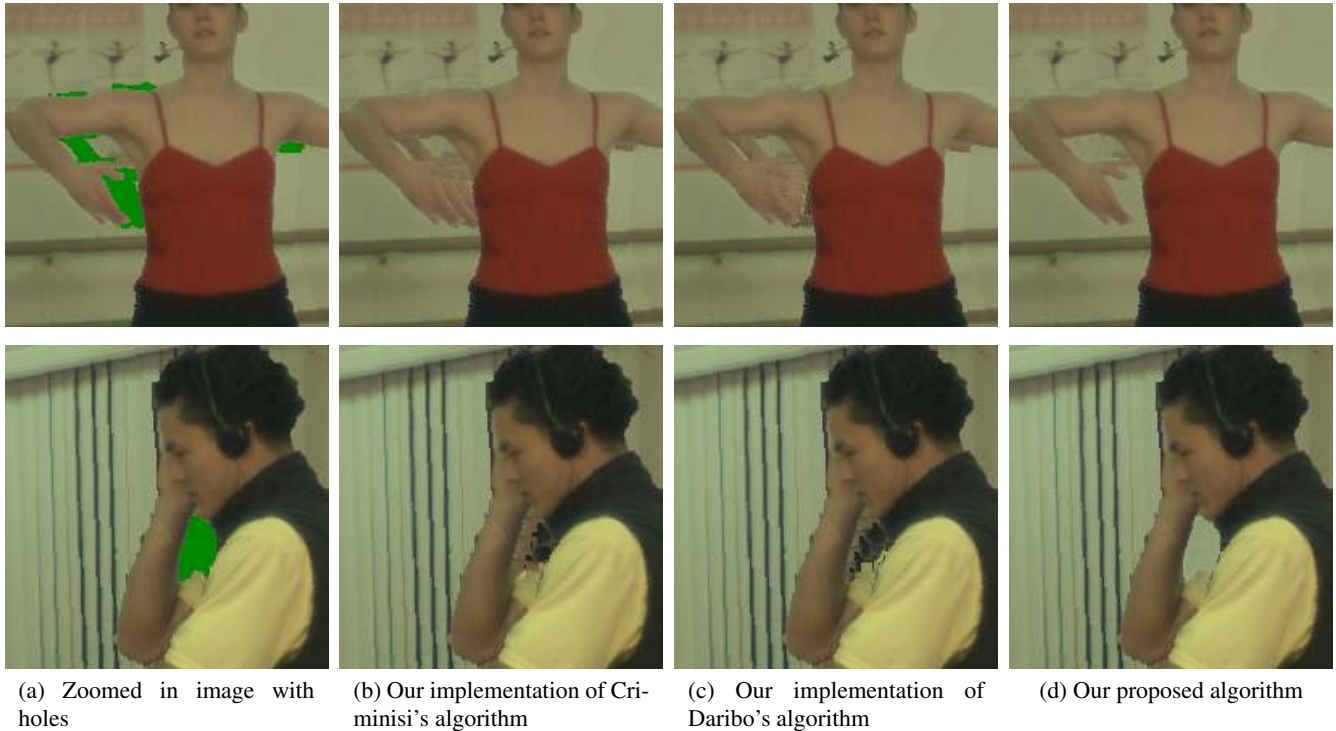


Fig. 5: Experimental results of our proposed algorithm, compared to our implementation of Criminisi's and Daribo's algorithms.

adapting the search window size. Fourth, the accuracy of patch matching is improved by using the location distance as a penalty. Experiments have shown that better perceptual results are produced with a good preservation of object contours and accurate filling of disocclusions in the foreground. The objective results measured in PSNR appear to be comparable to the existing algorithms, but our algorithm is more efficient, since it speeds up the inpainting process substantially. The speed enhancement is especially noticeable for larger images.

Despite the good performance of our algorithm in general, inpainting artifacts occur when segmentation fails to properly separate the background from the foreground. In future studies, better segmentation algorithms can be developed to reduce classification errors in order to further improve the quality of FVV inpainting.

6. REFERENCES

- [1] L. McMillan, Jr., *An image-based approach to three-dimensional computer graphics*. PhD thesis, Chapel Hill, NC, USA, 1997. UMI Order No. GAX97-30561.
- [2] L. Do, S. Zinger, and P. de With, "Quality improving techniques for free-viewpoint DIBR," in *Stereoscopic displays and appl. XXI*, 2010.
- [3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Transactions on Image Processing*, vol. 12, pp. 882–889, Aug. 2003.
- [4] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, pp. 1200–1212, Sept. 2004.
- [5] W.-H. Cheng, C.-W. Hsieh, S.-K. Lin, C.-W. Wang, and J.-L. Wu, "Robust algorithm for exemplar-based image inpainting," in *International Conference on Computer Graphics, Imaging and Vision*, pp. 64–69, July 2005.
- [6] Q. Chen, Y. Zhang, and Y. Liu, "Image inpainting with improved exemplar-based approach," in *Proc. International Conference on Multimedia Content Analysis and Mining*, pp. 242–251, 2007.
- [7] Anupam, P. Goyal, and S. Diwakar, "Fast and enhanced algorithm for exemplar based image inpainting," in *Proc. Fourth Pacific-Rim Symp. Image and Video Technology (PSIVT)*, pp. 325–330, 2010.
- [8] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing*, vol. 19, pp. 1153–1165, May 2010.
- [9] T.-H. Kwok, H. Sheung, and C. C. L. Wang, "Fast query for exemplar-based image completion," *IEEE Transactions on Image Processing*, vol. 19, pp. 3106–3115, Dec. 2010.
- [10] 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 *Proc. Picture Coding Symp. PCS 2009*, pp. 1–4, 2009.
- [11] I. Daribo and H. Saito, "A novel inpainting-based layered depth video for 3d tv," *IEEE Transactions on Broadcasting*, vol. 57, pp. 533–541, June 2011.
- [12] J. Gautier, O. Le Meur, and C. Guillemot, "Depth-based image completion for view synthesis," in *Proc. 3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, May 2011.
- [13] 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," in *ICIP'10*, pp. 1809–1812, 2010.
- [14] P. Ndjiki-Nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-d video," *Multimedia, IEEE Transactions on*, vol. 13, pp. 453–465, June 2011.
- [15] N. Ostu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, Jan. 1979.