Depth-Aided Exemplar-Based Disocclusion Filling for DIBR View Synthesis

Ryan Burke
Department of Electrical Engineering
Stanford University
Stanford, CA, USA
rburke2@stanford.edu

***Abstract*—Depth-image-based-rendering (DIBR) is a popular method of synthesizing new views of scenes using traditional images accompanied by depth maps. However these new view suffer from disocclusion holes – background areas revealed by the viewpoint change which have no image or depth data. The filling of disocclusion holes has been an area of development since 2011. This paper seeks to further develop one of the fundamental methods in disocclusion filling by adding small details from more recent (and much more complex) methods.**

Keywords—DIBR, Disocclusion, view synthesis, inpainting

I. Introduction

Virtual reality is a new and rapidly expanding area seeking to immerse the user in completely new surroundings. Total immersion is currently hindered in pre-recorded environments by the user's inability to move inside the scene. Additional views can be synthesized through depth-image-based-rendering (DIBR), but DIBR causes disocclusion holes – visual artifacts due to spatial regions that were not visible in the reference view. These holes can be filled with traditional inpainting (using known portions of the image to fill the unknown portions), but the additional depth data can be leveraged to produce higher quality, more accurate results.

II. Background

In this section I will be discussing the relevant works in image inpainting starting with simple rgb image inpainting and moving on to image plus depth data sets and disocclusion filling.

A. Criminisi et al.

In 2004, Criminisi, Pérez, and Toyama published a paper outlining an effective inpainting technique requiring minimal user guidance that simultaneously propagated structures into the masked region and believably filled textures [1]. Their work was based on the realization that exemplar-based texture synthesis is capable of propagating image structures. Exemplar-based image inpainting had already been implemented in many ways, but Criminisi et al. created a novel method of deciding which patch of the mask would be filled in first. Their calculation of priority for a patch centered at point p, *P*(p), is given by the equation

$$P(p) = C(p)D(p).$$

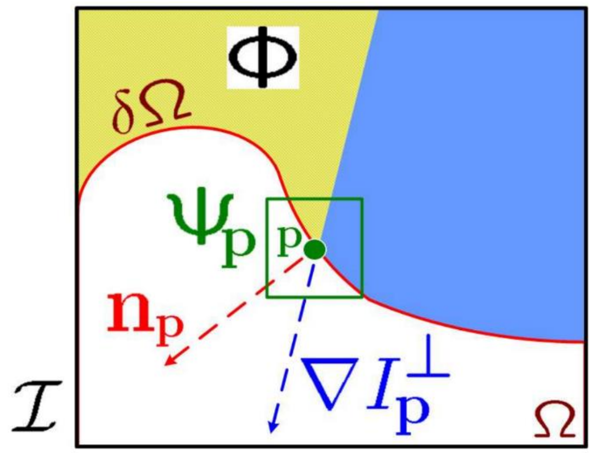

Figure 1: **Notation diagram.** "Given the patch Ψ_p, n_p is the normal to the contour $\delta\Omega$ of the target region Ω and $\nabla I_p^{\perp}$ is the isophote," or linear structure, "(direction and intensity at point p. The entire image is denoted with *I*" [1]

They refer to *C*(p) as the confidence term and *D*(p) as the data term. The confidence term is simply the ratio of known pixels surrounding a patch center p to the total number of patch pixels. This is shown mathematically as

$$C(p) = \frac{\sum_{q \in \Psi_p \cap \bar{\Omega}} C(q)}{|\Psi_p|}$$

Where $|\Psi_p|$ is the area of Ψ_p (all image segmentation and labelling is depicted in figure 1). At the start of the algorithm the confidence values of all pixels in Φ are set to one, and the confidence values of all pixels in Ω are set to 0.

The data term *D*(p) contributes a mathematical description of the structures surrounding the mask. The more pronounced the structure is, the larger the contribution of *D*(p) to *P*(p). *D*(p) is written as

$$D(p) = \frac{|\nabla I_p^{\perp} \cdot n_p|}{\alpha}$$

Where α is a normalization factor (255 for 8-bit images), $\nabla I_p^{\perp}$ is a vector orthogonal to the image gradient at a pixel p, and n_p is a unit vector orthogonal to the image contour $\delta\Omega$ at a pixel p.

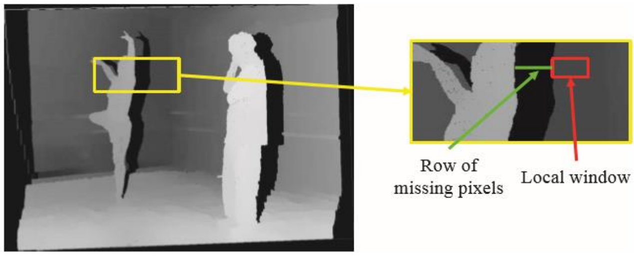

Figure 2: "Depth map with remaining disocclusions"… "and zoomed in part indicating in green the row of missing pixels whose depth values are determined on the depth values in the red rectangle."[3]

After the priority of each contour pixel is found the patch corresponding to the pixel with the highest priority is inpainted. The algorithm compares the available pixels of the target patch with every known patch in the image and replaces the missing pixels with those of the most similar patch. The sum of squared differences (SSD) is used to determine similarity through the equation

$$\Psi_{\hat{q}} = \arg\min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q)$$

When a patch is filled, the confidence terms of all patch pixels contained within the mask are set to the confidence of the center pixel, *C*(p). This process is iterated until the originally masked area has been entirely filled.

The algorithm created by Criminisi et al. was a simple and effective method for image inpainting, but it did not consider potential depth information that could accompany the image.

B. *Daribo et al.*

In 2011, Daribo and Saito took the work that Criminisi et al. had done and adapted it to 3D video data [2]. They made this transition in a very simple manner. They simply added another term to the two main equations used in the Criminisi et al. algorithm. Priority *P*(p) went from the product of the confidence term *C*(p) and the data term *D*(p) to

$$P(p) = C(p)D(p)L(p).$$

L(p) is the level regularity term, or the inverse variance of the depth patch Z_p, described by

$$L(p) = \frac{|Z_p|}{|Z_p| + \sum_{q \in \Psi_p \cap \Phi} (Z_p - \overline{Z_p})^2}$$

Where $|Z_p|$ is the area of Z_p and $\overline{Z_p}$ is the mean of Z_p. The level regularity term gives higher priority to patch overlaying at the same depth level, which naturally favors background pixels in the case of holes caused by disocclusions.

The best exemplar calculation went from minimizing the SSD of the known pixels of the target patch and patches of the known portion of the image to

$$\Psi_{\hat{q}} = \arg\min_{\Psi_q \in \Phi} \{d(\Psi_{\hat{p}}, \Psi_q) + \beta \cdot d(Z_{\hat{p}}, Z_q)\}.$$

Daribo et al. added the SSD of the target depth patch and all possible exemplar depth patches. With this new equation, we can control the importance of the SSD of the depth by changing the value of β.

Daribo et al. added depth considerations to the algorithm by Criminisi et al. but they only considered depth pixels with values in their calculations, and their final product leaves the depth map full of holes.

C. *Ružić et al.*

The disocclusion inpainting method employed by Ružić et al. uses highly advanced mathematics to achieve its goal, however there is one portion of the algorithm which can be extracted for our use. Their algorithm fills the depth map prior to inpainting the image. This is done for two reasons: first, depth maps are usually constant or slowly changing, making them easier to fill, and second, the greater quantity of depth data assists in the image inpainting process. Ružić et al. found that filling in entire rows of the depth map holes reasonably interpolates the depth data. Their algorithm fills the rows by extracting a suitable value from a local window located on the background side of the disocclusion hole (see figure 2). The extracted value is found through an algorithm described in [4], developed by Jain et al.

III. Algorithm & Method

The proposed algorithm is a combination of the algorithms described in the Related Works section. The backbone of the function is the method proposed by Daribo et al. The following changes were made. The depth map is filled using a modified Ružić et al.'s method before any inpainting is performed. Ružić et al.'s method is changed by generalizing it so that no prior knowledge is needed concerning which direction the camera moved to cause the disocclusion holes. During inpainting, all depth pixels are considered, and the top L best exemplars are averaged when filling in the target patch. Averaging the top exemplars adds a slight blur to the details added to the image, but it also helps prevent the algorithm from creating and propagating false structures in the image. The final change made is to limit the search area while the algorithm is checking for the best exemplar.

The Criminisi method and the proposed modified Daribo method are implemented in MATLAB both functions accept the original image, the mask, the patch size and the exemplar search area as inputs. The modified Daribo function also requires the depth map of the image. Both functions output the inpainted image.

IV. Results

Both the Criminisi and modified Daribo functions were tested for a set of images. The first image is a simple geometric shape that demonstrates the utility of depth information in image inpainting. Figure 3 shows how providing depth information can drastically improve algorithm functionality. Because the Criminisi and proposed algorithms propagate structures, the line at the top of the white box is a likely point to begin inpainting. But, by instructing the proposed method that the white box is part of the foreground, the background is automatically targeted first for inpainting, resulting in perfect reconstruction of the original image.

The second set of results were gathered using the Middlebury college dataset [5]. Both of the implemented algorithms were tested for a small portion of the Adirondack,

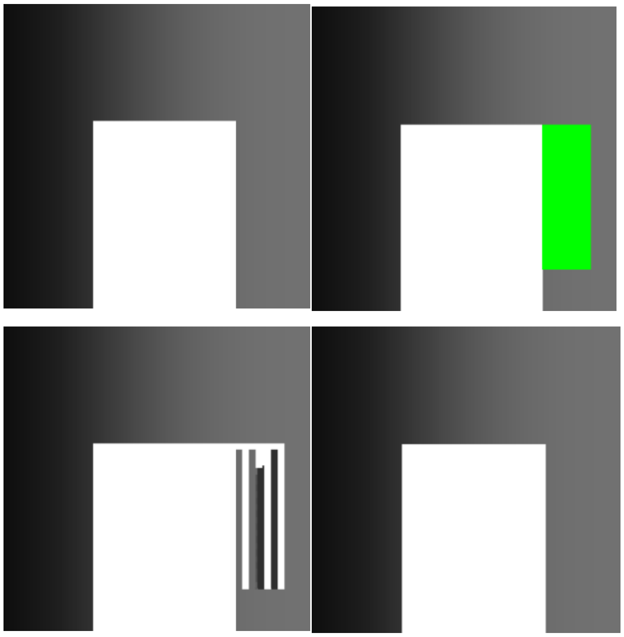

Figure 3: An example of the benefits of depth information. Top left: The original image, top right: the original image with mask applies (in green), bottom left: the inpainted image with the Criminisi algorithm, bottom right: the inpainted image with the proposed method (provided depth map indicating the white was foreground and the grayscale was background).

Jadeplant, Motorcycle, Piano, and Pipes images. The peak signal-to-noise ratios (PSNRs) are found for both methods for all images, where a premade mask was applied as the target area. Table 1 shows the results. The proposed method outperforms the depth-less Criminisi algorithm on in 4/5 cases, and underperforms by a very small margin in the fifth case. The images used for these calculations and the proposed method outputs are shown in figure 4.

Image	Criminisi PSNR (dB)	Proposed method PSNR (dB)
Adirondack	31.45	30.64
Jadeplant	39.88	39.84
Motorcycle	28.96	29.12
Piano	36.82	35.24
Pipes	29.70	29.30

Table 1: Comparison of the PSNRs of the Criminisi method and the proposed method for various images

V. Conclusion

In this paper a simple method of depth image inpainting was presented. The proposed method utilizes aspects of several disocclusion inpainting methods to achieve desirable and visually believable results. The proposed function consistently outperforms a similar function which does not consider depth information.

References

[1] Criminisi, Antonio; Perez, P.; Toyama, K., "Region filling and object removal by exemplar-based image inpainting," in *Image Processing, IEEE Transactions on* , vol.13, no.9, pp.1200-1212, Sept. 2004J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Daribo, I.; Saito, H., "A Novel Inpainting-Based Layered Depth Video for 3DTV," in *Broadcasting, IEEE Transactions on* , vol.57, no.2, pp.533-541, June 2011K. Elissa, “Title of paper if known,” unpublished.

[3] Ruzic, T.; Jovanov, L.; Hiep Quang Luong; Pizurica, A.; Philips, W., "Depth-guided patch-based disocclusion filling for view synthesis via Markov random field modelling," in *Signal Processing and Communication Systems (ICSPCS), 2014 8th International Conference on* , vol., no., pp.1-9, 15-17 Dec. 2014.

[4] Jain, A.K.; Tran, L.C.; Khoshabeh, R.; Nguyen, T.Q., "Efficient stereo-to-multiview synthesis," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* , vol., no., pp.889-892, 22-27 May 2011.

[5] Middlebury Dataset, Available online: vision.middlebury.edu/stereo/data/scenes2014/

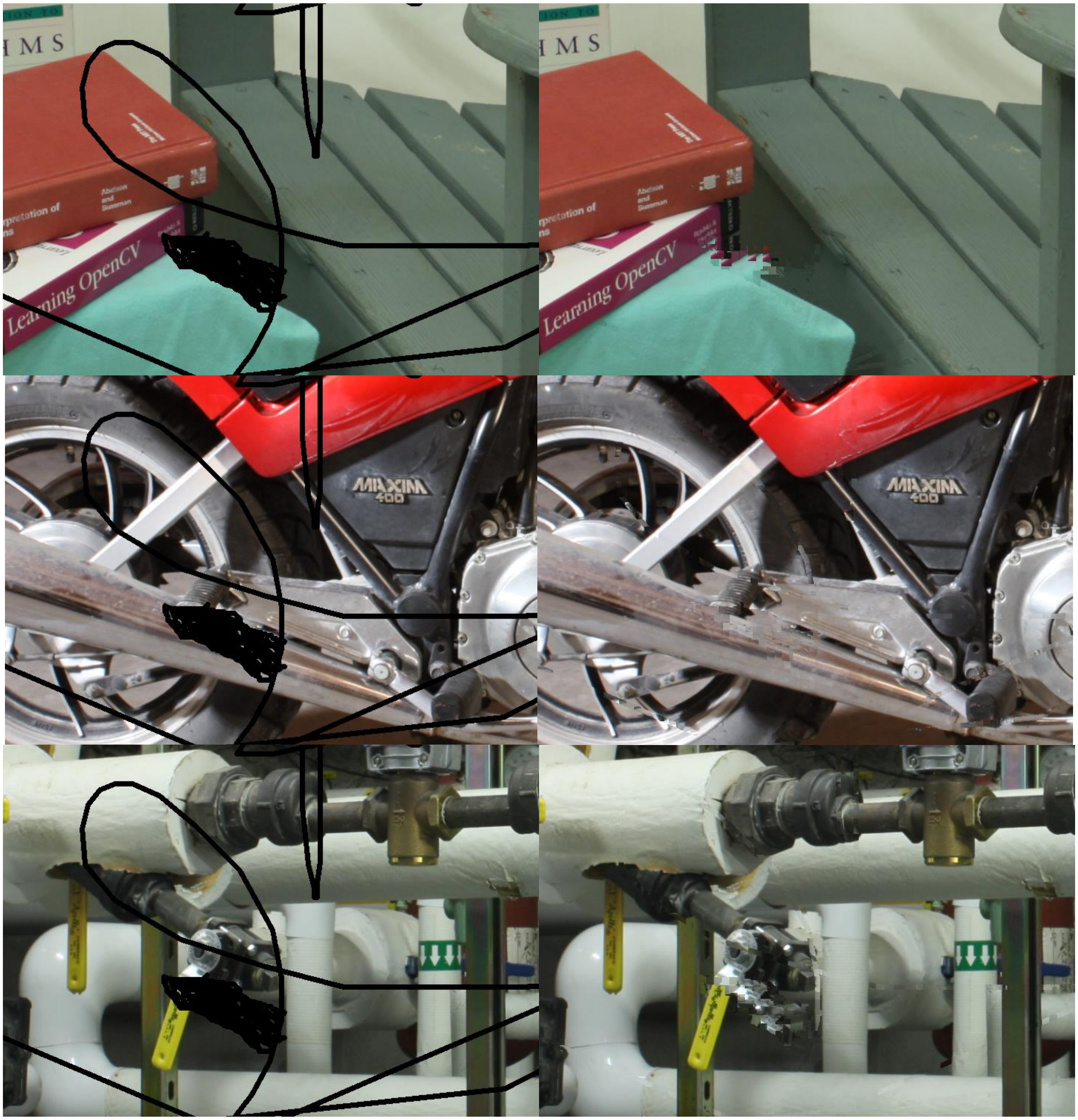

Figure 4: On left: The original image with overlaid mask. On right: the inpainted image with the proposed method.