EXEMPLAR BASED INPAINTING USING DEPTH MAP INFORMATION

Vahan Gevorgyan Vladimir, Aram Gevorgyan Vladimir, Gevorg Karapetyan Arakel

Abstract: *Image inpainting is the process of filling in missing or damaged parts. Exemplar based image inpainting algorithm fills in the missing parts using information from the known part of the image. Scanning the whole image for finding the most similar patch is a very time consuming task. Also the algorithm doesn't consider the depth information which harms the quality of the result. In this paper, we use depth information from the image depth map and this approach improves efficiency and quality of inpainted image. The paper includes experiment results and comparison with exemplar based inpainting algorithm.*

Keywords: *stereo image, inpainting, exemplar based, depth map, image completion*

ACM Classification Keywords: *I.4.4 Image Processing and Computer Vision - Restoration*

1. Introduction

The term inpainting comes from medieval art restoration, where pictures were restored by filling-in any gaps or scratches to bring them "up to date". In our application we mainly use inpainting for removing unwanted objects from an image. But there are also other applications for image inpainting such as repairing old photographs, text removal or improving quality of medical pictures. The goal of inpainting is to replace the missing region in such a way that it will be not observable for a viewer that there is something wrong with the image.

There are many known inpainting algorithms for 2D (ordinary) images, which trying to solve that problem. Those algorithms are usually classified into two main groups: diffusion based and texture based.

Diffusion based approach fills holes in images by propagating linear structures from known region into the missing region via process of diffusion. This problem is mainly modelled by Partial Differential Equation (PDE), so it is also called a PDE based approach. When the region to be inpaint is small, this approach provides good results, but it is not convenient to use it when the missing region is large, because some blur is noticeable due to diffusion process.

First texture based techniques [Igehy, 1997] produced good results if an image consisted of "pure" texture and were able to fill-in large regions. However, they were not able to propagate linear structures into the hole. In 2003 Criminisi *et. al* introduced an exemplar based algorithm [Criminisi, 2004], which is also referred to the texture based approach. In contrast with prior texture based approaches, exemplar based technique reconstructs both texture and structure features of the image and also it deals with large regions. There are also many other known algorithms from that class, for example PatchMatch [Barnes, 2009].

3D technologies have become very popular and are rapidly developing. With the development of 3D develops stereo vision. Stereo vision is a technique that gets 3D information of the world using two or more cameras. Stereo image is an image pair taken form left and right cameras, which show image from position of the left and right eyes respectively. Such approach enables users to feel depth. If we apply even best 2D inpainting approaches for stereo images, we will get unnatural results, so the appropriate algorithms for stereo images start to develop [Wang, 2008].

In this paper, we present an approach, which generates depth map for the image and improves exemplar based algorithm [Criminisi, 2004] using values from that map. In our approach, after removing some object, we use only background of it as a supporting region, because foreground objects may ruin the result.

In section 2.1 we will speak about depth maps. Section 2.2 is a brief review of exemplar based inpainting algorithm and our extension of that algorithm, using depth map values, is presented in section 3. Finally, section 4 shows the experimental results and comparisons with exemplar based algorithm [Criminisi, 2004], and conclusion is in section 5.

2.1 Depth map

A depth map contains information relating to the distance and position of the surfaces of scene objects from a viewpoint. Depth map can be constructed with special depth map cameras such as Kinect and active methods that use for example: radar, ultrasound, laser pulse or laser line scan. Nevertheless, these methods need special equipment and are expensive. The other more practical way of getting depth map is passive methods that use such techniques as stereovision. Multi-view stereovision computes depth map from several images of the same object or scene, taken from different angles and positions. Binocular stereovision produces depth map from a pair of images taken from two cameras, this pair is called stereo image.

Getting depth map usually consists of the following steps: calibration, rectification, stereo correspondence and triangulation. Calibration is a process of getting cameras external and internal parameters. More about calibration is discussed at [Bradski, 2008] Chapter 11, [Zhang, 2000], [Mann, 2004].

If the two cameras are coplanar, the two images have the same image plane, so the corresponding points have the same row coordinates and stereo correspondence problem is reduced from 2D to 1D. However, in real world it is impossible to place two cameras ideally coplanar so rectification is needed. Rectification is a process of reprojecting image planes onto the common plane parallel to the line between optical centers.

Stereo correspondence is a problem of finding matching pixels in left and right images corresponding to the same points on the 3D surface. The output of stereo correspondence algorithms is disparity map. Disparity shows the difference of point location in corresponding left and right images. There are many stereo correspondence algorithms. State-of-the-art of stereo matching algorithms and their taxonomy is described at [Sharstein, 2002]. One of the most commonly used stereo correspondence algorithms is Stereo-Global Matching (SBM) [Hirschmuller, 2005] that is fast enough and gives good results.

Knowing disparity map and geometric arrangement of the cameras and their internal parameters, such as focal length, we can calculate depth map by triangulation [Marshall, 1997], [Bagga, 2013].

2.2 Exempla based algorithm

This algorithm was proposed by Criminisi *et al*. [Criminisi, 2004]. For input, it has an image I. A user selects a target region Ω, which should be removed and filled in and its contour denotes as $\delta\Omega$. This contour also called as "fill front" and it evolves inward during the algorithm. The supporting region from where the algorithm will take information, which is called source region and denotes as Φ, should be specified. By default it is defined as whole image minus the target region $\Phi = \mathrm{I} / \Omega$ This algorithm is fragment based, which means that it fills the missing area by patches not by pixels. The default size of a patch that authors suggest is 9×9 pixels, but this parameter may be specified by the user, depending on the image size and color features. After setting all input parameters, the algorithm starts to inpaint the removed region. The steps of the algorithm can be briefly described as:

- Detection of fill front (boundary points) and calculation of priority values for that points,
- Selecting a target patch (patch which center is a point with highest priority value),
- Finding a source patch (the most similar patch to the target patch),
- Filling values for unknown points from the target patch
- Update terms values for filled pixels which participate in priority function.

This steps are executing iteratively until all pixels are filled in. For every boundary point priority function is computed, which has the following representation:

$$P(p) = C(p)D(p), \forall p \in \delta\Omega,$$

where $C(p)$ is called confidence term and $D(p)$ – data term and they have following form:

$$C(p) = \frac{\sum_{q \in \Psi_p \cap (\mathrm{I}/\Omega)} C(q)}{|\Psi_p|}, D(p) = \frac{|\nabla \mathrm{I}_p^{\perp} \cdot n_p|}{\alpha},$$

where $|\Psi_p|$ is the area of a patch Ψ_p centered at point p, α is a normalization factor (its value is 255 for a typical grayscale image), n_p is a unit vector orthogonal to the front $\delta\Omega$ in the point p and $\Delta \mathrm{I}_p^{\perp}$ - the isophote (a line of equal intensity value) at p.

The confidence term shows how much is reliable the information surrounding the pixel p. The idea is to give preference to those patches, which have more of their pixels already filled and the earlier pixel filled (or it was never part of the target region) the more reliable it is. The function $C(p)$ initially is set to the following values:

$C(p) = 0, \forall p \in \Omega$ and $C(p) = 1, \forall p \in \mathrm{I} / \Omega$.

The data term $D(p)$ shows the strength of an isophote hitting the front $\delta\Omega$ at point p. This term gives preference to those patches, which prolongs isophote direction. Due to it, linear structures are encouraged to be inpainted first and therefore propagate in a plausible way into the target region.

The target patch (patch which center is a point with highest priority) are fixed and the algorithm searches for the source patch calculating similarity between patches by the following formula:

$$\Psi_q = \arg\min_{\Psi_l \in \Phi} d(\Psi_l, \Psi_p),$$

where Ψ_q is the source patch, Ψ_p – the target patch, d - a function of distance (similarity) between two patches. The distance function that algorithm uses is sum of squared difference (SSD) of known pixel color values from the target patch, with their corresponding pixels from the source patch.

Then for each unfilled pixel in Ψ_p the algorithm copies pixel values from the corresponding position inside Ψ_q. Fig. 1 illustrates the inpainting process described above.

Figure 1 : **Inpainting process**

Finally, the confidence values for the already filled pixels should be updated. The confidence term $c(p)$ is updated as follows: $C(t) = C(p), \forall t \in \Psi_p \cap \Omega$.

3. Modification of exemplar based algorithm using depth information

We take stereo pair (left and right image) and generate disparity map using Semi-Global Matching stereo correspondence algorithm [Hirschmuller, 2005] that was mentioned above. Then from disparity map by triangulation, we calculate depth value for each pixel and due to it a depth map is generated, which shows objects positions related to each other.

Now we have an image with its depth map. The main idea of our approach is the following, that after removing some object from a stereo image, the information for inpainting that region should be taken from its background.

After user selects the target region that should be removed and inpainted, we generate a mask for inpainting. In contrast with standard 2D inpainting, this mask is ternary not binary. First value stands for pixels from a region that should be inpainted (target region), second value is for pixels from supporting region from which the information will be taken (source region) and third value is for other points that will not participate in the inpainting process. While a user selects an object to remove, he can also touch neighborhood points. Because of it, we calculate the average depth of the target region, which is almost equal to the object depth if the object is homogeneous. As the source region, we consider pixels that have disparity great or equal than this average value minus some threshold. This threshold is for the case when removing object is the furthest object from the camera.

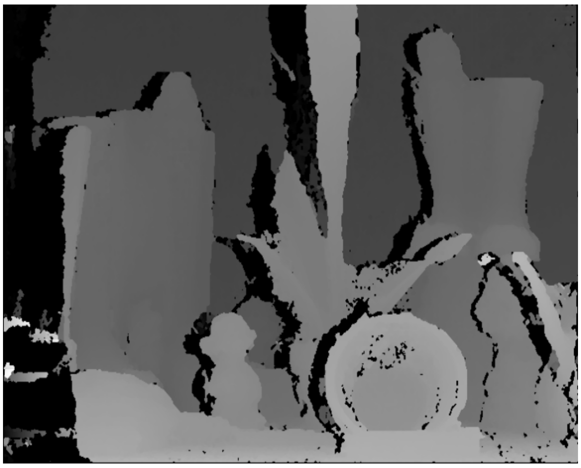
Our approach improves quality of the result because information from impeding foreground objects is not taken and after removing some object, only its background will remain. Execution time of the algorithm is also noticeably increased, because area of source region is limited only by its background, while searching for the most similar patch for every point.

4. Experimental results

We implement Criminisi's exemplar based algorithm and our approach on C++ using OpenCV library [Opencv]. In our experiments, we use Middlebury Stereo Datasets [Middlebury, 2014]. For comparing the original algorithm result with our modified result, we use **PSNR** (peak signal-to-noise ratio) [PSNR]**.** PSNR is an approximation to human perception of reconstruction quality. Generally, a higher PSNR value means that reconstruction result is of higher quality. We test algorithms on more than 100 examples and in average, our algorithm is 4-5 times faster than the original one and it gives about 3% increase of PSNR value. Fig. 2,3 show some of examples. We take a default patch size of 9x9 pixels. In inpainting mask from below examples, gray colored region means the region that should be inpainted, black – is the background of it and white – foreground.

In example 1, execution time of Criminisi's algorithm takes 1.01667, our modification – 0.266667 minutes. Average disparity of target region is 362.306 mm. PSNR-s are respectively 35.7465 and 38.4639

In example 2, execution time of Criminisi's algorithm takes 1.08333, our modification – 0.166667 minutes. Average disparity of target region is 34.247027 mm. PSNR-s are respectively 35.7465 and 38.4639.

a) b)


c) d)


e) f)

Figure 2: Example 1. Comparison of exemplar based inpainting and our approach. Image size is 695x555, a) the original image , b) depth map, c) inpainted image d) inpainting mask, e) result of Criminisi's algorithm, f) our result

Figure 3: Example2. Comparison of exemplar based inpainting and our approach. Image size is 640x480, a) the original image , b) depth map, c) inpainted image d) inpainting mask, e) result of Criminisi's algorithm, f) our result

5. Conclusion

In this paper, we introduced a new approach for digital image inpainting, which extended exemplar based inpainting algorithm [Criminisi, 2004], using 3d features of images, such as depth map. We try to improve the case of removing unwanted object from an image. We reduce the source region area of the algorithm, taking in account only background points of removing object. This approach significantly improves execution time of the algorithm and gives better results in most of the cases.

Bibliography

[Igehy, 1997] H. Igehy, L. Pereira. Image replacement through texture synthesis. In Proc. Int. Conf. Image Processing, pp. III:186–190, 1997

[Criminisi, 2004] A. Criminisi, P. Perez, 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

[Barnes, 2009] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graphics,* vol. 28, no. 3, 2009

[Wang, 2008] L.Wang, H.Jin, R. Yang, M. Gong "Stereoscopic inpainting: Joint color and depth completion from stereo images ", Proc. of CVPR, pp.1–8, 2008

[Bradski, 2008] Gary Bradski, Adrian Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library", 2008.

[Zhang, 2000] Z. Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000 [4] Z. Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000

[Mann, 2004] http://cs.nyu.edu/courses/fall14/CSCI-GA.2271-001/06StereoCameraCalibration.pdf

[Sharstein, 2002] Daniel Scharstein, Richard Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", IJCV 7, 1/3, 2002.

[Marshall, 1997] http://www.cs.cf.ac.uk/Dave/Vision_lecture/node11.html#SECTION00132300000000000000

[Bagga, 2013] Prashant Jagdishchandra Bagga, "Real Time Depth Computation Using Stereo Imaging. Journal Electrical and Electronic Engineering" Vol. 1, No. 2, 2013, pp. 51-54. doi: 10.11648/j.jeee.20130102.13

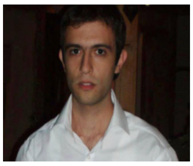
[Hirschmuller, 2005] H. Hirschmuller, “Accurate and Efficient Stereo Processing by Semi Global Matching and Mutual Information.” CVPR, 2005.

[Opencv] http://opencv.org/

[Middlebury, 2014] Datasets of stereo images: http://vision.middlebury.edu/stereo/data/

[PSNR] http://www.ni.com/white-paper/13306/en/

Authors' Information



Vahan Gevorgyan – *Institute for informatics and automation problems of NAS RA, 1, P. Sevak street, Yerevan 0014, Armenia; Russian-Armenian University, Hovsep Emin steet, 123, Yerevan, Armenia; e-mail: vahangev8@gmail.com*

Major Fields of Scientific Research: Digital image inpainting, Image reconstruction, Computer vision

Aram Gevorgyan – *Institute for informatics and automation problems of NAS RA, 1, P. Sevak street, Yerevan 0014, Armenia; Russian-Armenian University, Hovsep Emin steet, 123, Yerevan, Armenia; e-mail: aramgv@gmail.com*

Major Fields of Scientific Research: Computer Vision, Stereo Vision, 3D reconstruction

Gevorg Karapetyan – *Institute for informatics and automation problems of NAS RA, 1, P. Sevak street, Yerevan 0014, Armenia, e-mail: gevorgka@gmail.com*

Major Fields of Scientific Research: Computer Vision, Medical Image Processing, Digital Image Inpainting, Image Classification