

Modified Depth-Map Inpainting Method Using the Neural Network

Nikolay Gapon^{1,*}, *Roman Sizyakin*¹, *Marina Zhdanova*^{1,2}, *Oksana Balabaeva*¹ and *Yigang Cen*³

¹Don State Technical University, RU-344000, Rostov-on-Don, Russia

²Moscow State Technological University "STANKIN", RU-127055, Moscow, Russia

³School of Computer & Information Technology, Beijing Jiaotong University, Beijing, China

Abstract. This paper proposes a method for reconstructing a depth map obtained using a stereo pair image. The proposed approach is based on a geometric model for the synthesis of patches. The entire image is preliminarily divided into blocks of different size, where large blocks are used to restore homogeneous areas, and small blocks are used to restore details of the image structure. Lost pixels are recovered by copying the pixel values from the source based on the similarity criterion. We used a trained neural network to select the "best like" patch. Experimental results show that the proposed method gives better results than other modern methods, both in subjective and objective measurements for reconstructing a depth map.

1 Introduction

Assessing the true values of image pixels is necessary in most digital image processing tasks. This problem is especially relevant for the automatic processing of images obtained in photosensitive matrices in digital photo and video cameras and machine vision systems. Machine vision is part of information technology that is directly related to robotics. The field of robotics is becoming increasingly important in our society. Today, robots can be found in almost all aspects of our daily lives. When using robots in industry, the relevant task of determining the distance from the end effector to the surfaces that it processes. To perform movements without crashing into obstacles, the robot must see the environment. In robotics, to determine the distance from the robot to the object, methods are required that allow for a minimum time to determine the distance with acceptable accuracy. A similar method is to build a depth map by distance. The depth map can be obtained using a special depth camera, or a stereo pair of images. If high accuracy is not required or a compromise between price and quality is required, a depth map is constructed using a pair of stereo images. This repeats the principle of binocular vision in humans and eliminates the use of numerous and expensive sensors, radars or lidars. A depth map by distance can be built using two cameras, which, taking a photo, create a stereo pair. The system that allows you to build a depth map by distance based on stereo images includes a pair of cameras set on one line, which is parallel to the image plane, with known distances between cameras and focal length. Accuracy, quality, and map creation time. In most cases, when constructing a depth map, lost areas appear on it that interfere with distinguishing one object from another.

This problem can be solved using image reconstruction methods, restoring lost areas on the depth map, making it more accurate.

Simplified methods for reconstructing pixel values of images can be divided into the following groups:

- 1) Methods based on the solution of partial differential equations.
- 2) Methods based on orthogonal transformations.
- 3) Methods based on texture synthesis.
- 4) Methods based on the neural network

An analysis of the existing processing methods shows that the area of their use, in the context of a limited amount of information about the components of the processed process, is extremely limited. The use of methods for reconstructing pixel values of images based on the solution of partial differential equations leads to blurring of sharp changes in brightness and contours and requires a priori information to select the parameters of the methods and minimize the functional. The inability to restore the texture of images and curved contours limits the use of these methods, which are mainly applicable when removing scratches and small defects on the image structure. To use methods based on orthogonal transformations, a priori information is required to select a threshold value, an orthogonal basis, and a block size of a spectral representation. It should also be noted that these methods lead to blurring of the texture and structure when recovering large areas with lost pixels, and a large number of iterations leads to significant computational costs. The use of methods based on texture synthesis requires a priori information about the size and shape of the restoration region and the geometric properties of the image to select the parameters of the methods.

This paper presents a method for reconstructing a depth map based on the operation of a neural network. The main stages of the algorithm are presented in Fig. 1.

* e-mail: gapon.nv@gmail.com

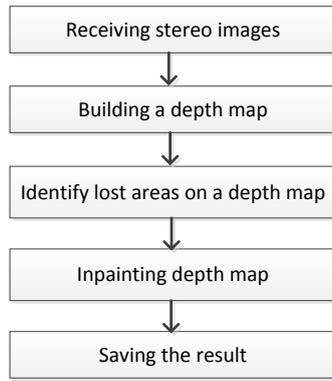


Fig. 1. Block diagram of the proposed algorithm

2 Proposed method

2.1. Depth Map

The idea underlying the construction of a depth map using a stereopair is that for each point on one image a pair of points is searched for it on another. For a pair of such points, you can perform triangulation and determine the coordinates of the pre-image already in three-dimensional space. Knowing these coordinates, the depth is calculated as the distance to the camera plane.

The search for paired points occurs on the epipolar line. Accordingly, to simplify the search, the images are aligned so that all the epipolar lines are parallel to the sides of both images.

For each pixel of the left picture with coordinates (x_0, y_0) , a pixel is searched in the right picture. In this case, it is assumed that the pixel in the right picture should have coordinates $(x_0 - d, y_0)$, where d is a quantity called disparity. The search for the corresponding pixel is performed by calculating the maximum of the response function, which can be, for example, the correlation of neighborhoods of pixels. The result is a disparity map.

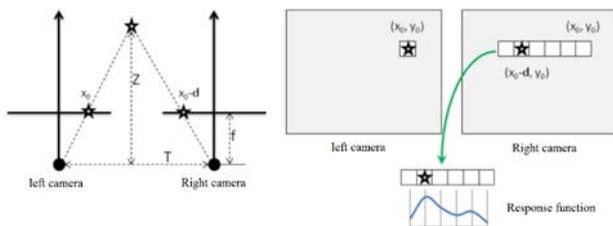


Fig. 2. Calculate the depth map.

Actually, the depth values are inversely proportional to the pixel offset. If you use symbols from the left half of the Fig. 2, the relationship between disparity and depth can be expressed in the following way:

$$\frac{T-d}{Z-f} = \frac{T}{Z} \rightarrow Z = \frac{fT}{d} \quad (1)$$

Due to the inverse relationship of depth and displacement, the resolution of stereo vision systems that

work on the basis of this method is better at close distances, and worse at far distances. The depth map can be seen in the Fig. 3.

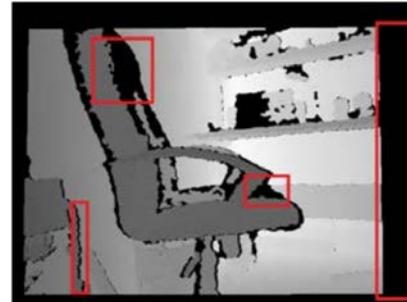


Fig. 3. Depth map with defects.

Next, the resulting depth map is processed by the reconstruction method using a neural network.

2.2 Reconstruction method

At the first step, for each pixel of the boundary $\delta S_{i,j}$ using the inversion method the shape of the domain for similarity search is adaptively determined by combining two adjacent homogeneous sub regions in the direction of the maximum gradient [2,5,7].

The second step calculates the priority value $P(\delta S)$ for each pixel value of the boundary, which consists of two factors [3,9]:

$$P(\delta S) = C(\delta S) \cdot D(\delta S) \quad (2)$$

$$C(\delta S) = \frac{\sum_{l \in \Psi_{\delta S}} C(l)}{|\Psi_{\delta S}|}, D(\delta S) = \frac{|\nabla I_{\delta S}^{\perp} \cdot n_{\delta S}|}{\alpha} \quad (3)$$

Using the method continued in the work [1] the entire image is preliminarily divided into blocks of $5 \times 5, 7 \times 7, \dots, 21 \times 21$, where large blocks are used to reconstruct homogeneous areas, and small blocks are used to reconstruct structure image details (Fig. 4).



Fig. 4. Example of splitting an image into patches

Next, in the third step, we determine blocks $\Psi_q^{(h)}$, $h = \overline{1, R}$ in the area of available pixels S for which the Euclidean metric is minimal [4,5,7]:

$$\sqrt{\sum (\Psi_p - \Psi_q)^2} \rightarrow \min. \quad (4)$$

The pixel values in the region η adjacent to the pixel with the highest priority p are reconstructed by averaging the corresponding pixels from chosen areas $\Psi_{q^{(h)}}$ in the area of available pixels S using a neural network, in particular, a multi-layer perceptron[10].

The confidence coefficient C for restored pixels is assigned to the current value $C(p)$.

After that, the procedure of priority correction and search of similar areas with subsequent replacement is repeated.

2.3 Neural network

In this work, a neural network of direct signal propagation, which was trained using the algorithm for back propagation of the error was developed.

The activation function which is used in this network is the sigmoidal nonlinear function (sigmoidal nonlinearity), namely the hyperbolic tangent function (5) [6,10].

$$f(x) = \tanh\left(\frac{ax}{2}\right) = \frac{1 - e^{-ax}}{1 + e^{-ax}} \quad (5)$$

a - is the slope parameter of the sigmoidal activation function.

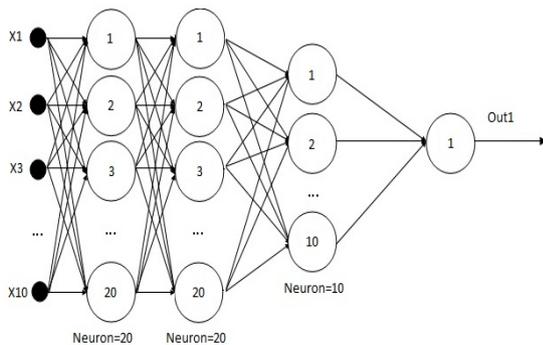


Fig. 5. Architecture of the neural network.

At the training stage of the neural network, pre-prepared data was fed to the input: a blocks with random coordinates $3 \times 3, 5 \times 5, 7 \times 7, \dots, 21 \times 21$, in size was allocated to the image, then the central pixel was removed, and five most similar blocks were found on the whole image comparing them by MSE.

Then the procedure was repeated on 35 images, and 100,000 blocks were obtained, they were used as a training sequence for this network.

This network contains three layers [10]: the first layer contains 20 neurons, the second layer also contains 20 neurons, the third layer contains ten neurons (Fig. 5). The network created ten inputs, 5 of them were fed only the central pixels in the blocks found, and the other five inputs fed the MSE of these blocks. As an output, there were the central pixels of the original blocks.

3 The effectiveness of the proposed method

Figure 6 shows an example of the depth map reconstruction by the proposed method (a - images obtained from the left and right cameras, b - depth map, c - reconstructed depth map). The proposed method allows to correctly restore borders without blur.

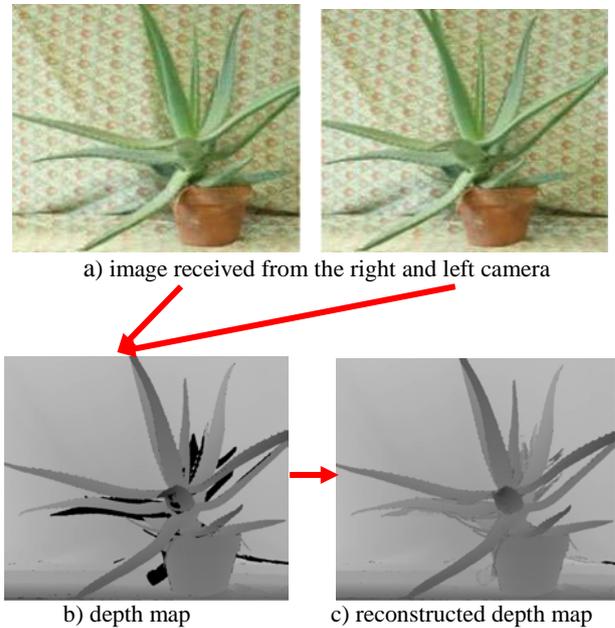


Fig. 6. Processing example.

Figure 7-8 shows an example of reconstruction of a depth map by the proposed method (a - depth map, b - reconstructed depth map)

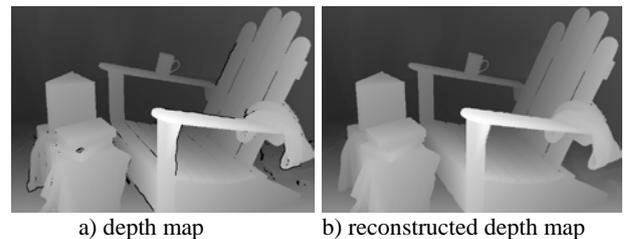


Fig. 7. Processing example.

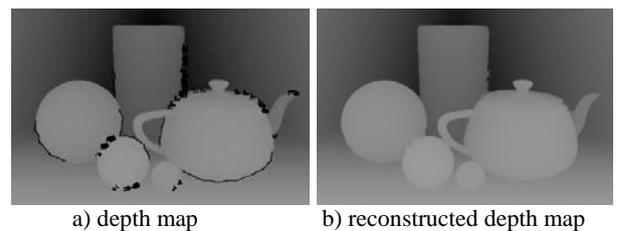


Fig. 8. Processing example.

Analysis of the processing results shows that the proposed method allows you to correctly restore the details and background on the image, and there is also the absence of artifacts when restoring lost pixels. It should be noted that the proposed method does not blur the texture and structure when restoring large areas with

lost pixels. The proposed method allows you to correctly restore the boundaries, does not lead to their blurring.

4 Conclusion

We proposed an inpainting algorithm for depth map reconstruction using a stereo pair image. The analysis of the method of reconstructing the depth map was carried out, which showed that the proposed method allows you to effectively restore occlusions on the depth map. The proposed method allows to correctly restore the details and background in the image

Acknowledgments. This work was supported by Russian Ministry of Education and Science in accordance to the Government Decree № 218 from April 9, 2010 (project number № 074-11-2018-013 from May 31, 2018 (03.G25.31.0284)). This work was carried out using equipment provided by the Center of Collective Use of MSUT "STANKIN".

References

1. N.N. Ponomarenko, V.V. Lukin, K.O. Egiazaryan, J.T. Astola, 5th Northern Symposium on Signal Processing on board Hurtigrunen, Tromsø-Trondheim, Norway, (2002).
2. A. Criminisi, P. Perez, and K. Toyama, Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 721–728 (2003).
3. A. Criminisi, IEEE transactions on image processing, (2004).
4. V.V. Voronin, V.I. Marchuk, N.V. Gapon, R.A. Sizyakin, A.I. Sherstobitov, K.O. Egiazarian, Proceedings of SPIE - The International Society for Optical Engineering Cep. "Proceedings of SPIE-IS and T Electronic Imaging - Image Processing: Algorithms and Systems XII", 901907 (2014).
5. N.V. Gapon, R.A. Sizyakin, M.M. Pismenskova In the collection: Modern trends of development of science and education materials of the international (correspondence) scientific and practical conference. Scientific and publishing center "The World of Science.", 113-117, (2016).
6. M.M. Pismenskova, N.V. Gapon, R.A. Sizyakin Modern structures of neural networks. In the collection: modern trends of development of science and education materials of the international (correspondence) scientific and practical conference. Scientific and publishing center "The World of Science", 161-167, (2016).
7. V.V. Voronin, V.I. Marchuk, N.V. Gapon, A.V. Zhuravlev, S. Maslennikov, S. Stradanchenko, Proceedings of SPIE - The International Society for Optical Engineering, **9497**, article No. 94970Y, (2015).
8. V.V. Voronin, V.I. Marchuk, V.A. Franz, N.V. Gapon Successes of modern radio electronics. **6**, 31-34 (2012).
9. V.V. Voronin, V.I. Marchuk, N.V. Gapon, R.A. Sizyakin Successes of modern radio electronics. **9**, 16-20 (2011).
10. N. Gapon, M. Pismenskova, O. Tokareva, M. Ponamorenko, MATEC Web of Conferences 13. Cep. "13th International Scientific-Technical Conference "Dynamic of Technical Systems", DTS 2017". P. 05015, (2017)
11. S.N. Grigoriev, V.D. Gurin, M.A. Volosova, and N. Y. Cherkasova, Development of residual cutting tool life prediction algorithm by processing on CNC machine tool, Materialwiss. Werkstofftech., 44(9), 790-796 (2013)
12. S.N. Grigoriev, G.M. Martinov, Proc. CIRP, **41**, 858-863 (2016)
13. S.N. Grigoriev, G.M. Martinov, Proc. CIRP, **14**, 517-522 (2014)
14. A.V. Oleynik, L.V. Kuznetsova, V.E. Petrov, 2018 Proc. Int. Conf. "Quality Management, Transport and Information Security, Information Technologies", IT and QM and IS, 704-708 (2018)
15. S.N. Grigoriev, M.P. Kozochkin, F.S. Sabirov, and A.A. Kutin, Proc. CIRP, **1**, 599-604 (2012)