Depth Mapping Method Based on Stereo Pairs

Vasiliy E. Gai*, Igor V. Polyakov, Olga V. Andreeva

*Nizhny Novgorod State Technical University n.a. R.E. Alekseev,
Minin St., 24, Nizhny Novgorod, Russia
iamuser@inbox.ru

Abstract. The paper proposes a new method for solving the problem of constructing a depth map based on a stereo pair of images. The result of the depth information recovery can be used to capture the reference points of objects in the film industry when creating special effects, as well as in computer vision systems used on vehicles to warn the driver about a possible collision. Proposed method consists in using the theory of active perception at the stage of segmentation and image matching.

To implement the proposed method, a software product in the C# language was developed. The developed algorithm was tested on various sets of input data. The results obtained during the experiment indicate the correct operation of the proposed method in solving the problem of constructing a depth map.

The accuracy of depth mapping using the described method turned out to be comparable with the accuracy of the methods considered in the review. This suggests that this method is competitive and usable in practice.

Keywords: Theory of active perception, Depth map, Stereo pair of images.

1 Introduction

One of the important tasks of the computer vision is the transformation of a stereo pair of images into a three-dimensional scene. As a result of this process, the depth information of each image point is restored. Obtaining an accurate depth map is the ultimate goal of three-dimensional image recovery.

The depth information received as the result of this process can be used in many other areas. For example, depth maps are used to capture the reference points of objects in film production when creating special effects, as well as in computer vision systems used on vehicles to warn the driver about a possible collision.

Based on this, we can conclude that the development of new models and methods for solving the problem of constructing a depth map based on a stereo pair is relevant.

2 The general principle of the methods for constructing depth maps using a stereo pair

The general algorithm of depth mapping using stereo images includes the following steps [1]: camera calibration, image rectification, image segmentation, search for matches between points of a pair of images, conversion of a discrepancy map into a depth map. In this paper, the first two stages are not considered, since these stages are simple geometric transformations and they are solved at the hardware level in most computer vision systems.

When analyzing the algorithms that implement the steps described above, the following problems were identified. The problem of segmentation. This problem lies in the fact that part of the segmentation algorithms does not have sufficient accuracy, and therefore, at the stage of the search for correspondences, there are multiple errors associated with incorrect segmentation. The other part of the algorithms provides sufficient accuracy, but has a high computational complexity. There is also a problem with the correlation of segments of two images [2]. The problem of finding matches. This problem lies in the imperfection of matching algorithms, as a result of which the accuracy of depth map construction is reduced [3]. The problem of handling errors after matching. This problem based on the fact that usually after the stage of the search for correspondences the discrepancy map contains a number of erroneously determined points and their additional processing is necessary.

3 Methods of depth mapping

The proposed method for solving the problem of depth mapping applies the theory of active perception (TAP) at the stage of segmentation and the search for correspondence of points [4].

To solve the problem of depth mapping in this paper, the following algorithm is proposed:

1. Image input – receiving images from cameras or from files;
2. Pre-processing – converting images to a brightness function.
3. Segmentation – the selection of objects in the first image in order to reduce the search area in the future;
4. Search for matching segments – search for segments of the left image on the right image;
5. Discrepancy mapping – the formation of a matrix containing information on how much each point of the first image differs in its position in space from the same point in the second image;
6. Depth mapping – the final stage of the restoration of depth information with the subsequent visualization of the results.

4 Segmentation

The next step is to divide the image into segments. This is necessary to reduce the search area at the stage of matching.

Due to the fact that there is a search for matching points on the same objects, the best solution would be to divide the image into a set of objects, thereby narrowing the search area to the inner area of objects. Also, the two source images are epipolar, which allows the image to be divided into horizontal segments without the accuracy loss.

Based on this, it was decided to produce segmentation in two stages:

1. Division into horizontal segments;
2. Selection in horizontal segments, segments based on the boundaries of objects.

Since the images included in the system are epipolar, all the horizontal lines of one image coincide with the horizontal lines of the other. Therefore, you can divide the image into horizontal segments without compromising the accuracy of the search.

After the image is divided into horizontal segments, it is necessary to select the boundaries of objects inside each horizontal segment. For this process, we will use filters that allow us to find the change in brightness in different directions (see Fig. 1).

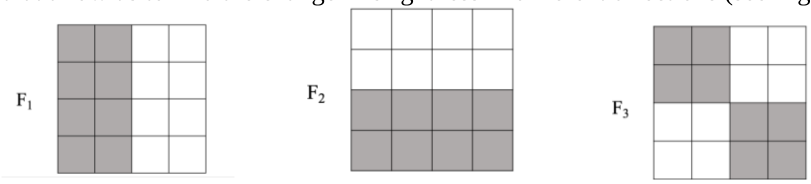

Fig. 1. Filters used to select borders.

Filter $\mathbf{F}_1$ is used to select vertical borders, $\mathbf{F}_2$ – to select horizontal borders, $\mathbf{F}_3$ – to select diagonal borders. These filters are applied to each point in the horizontal segment. From the obtained values, the greatest in absolute value is selected, i.e. denoting the greatest difference in brightness (see Fig. 2).

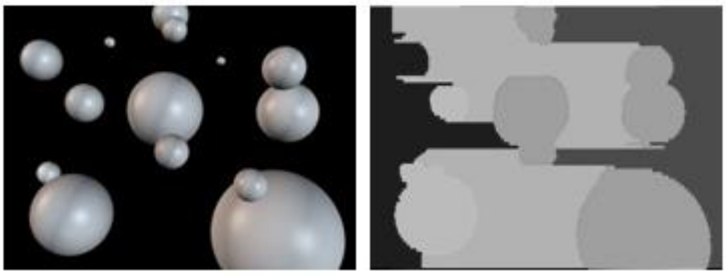

Fig. 2. Segments

For the subsequent use it is necessary to form a segment model. It consists of the following elements:

1. The starting point of the segment and its description with the help of TAP.
2. The end point of the segment and its description using TAP.

TAP filters are used to describe points. The description of the points is formed by applying to them all 16 TAP filters.

5 Segment matching

At the moment, one image is divided into segments. The next step is to search for the segments of the first image on the second one. To do this, the second image searches for the most similar points for the beginning and end of the segment using the following algorithm:

1. The response for the reference point is calculated for all 16 filters.
2. A 4x4 window is passed through the pixels of the horizontal segment of the second image. The current pixel is the coordinate of the upper left corner of a 4x4 window. As the window passes through the image, the response is calculated for all 16 filters.
3. The difference modulus ("delta") of each response is found with the reference response that was found at the beginning.
4. All sixteen differences are summed up and saved together with the coordinates of the current position of the window.
5. From all the obtained differences the minimum difference is found, which determines the minimum difference of the found point from the original one.
6. This point is set in accordance with the original.

This algorithm is performed for the starting and ending points of the segment. Thus, pairs of segments of the first and second images are formed.

6 Discrepancy mapping

The first main stage of the algorithm is discrepancy mapping – a matrix containing information about how much each point of the first image differs in position in space from the same point in the second image. For each point of the segment, the corresponding point is searched for in the second image. The scope of the search in this case is limited by the size of the segment.

When the desired point is found, its discrepancy is calculated by the formula:

$$D = |X_1 - X_2| \, , \tag{1}$$

where X_1 – coordinates of the point on the first image, X_2 – coordinates of a point on the second image.

7 Depth mapping

Depth mapping is the final stage of solving the problem. At this stage, the discrepancy map is converted to a depth map. It is also necessary to solve the problem of possible errors made at the stage of the search for matches. Therefore, it was decided to apply the following formula to all points of the discrepancy map:

$$D_{x,y} = \begin{cases} D_{x,y}, D \le Max, \\ \dfrac{\sum_{i=x-\frac{n}{2}}^{x+\frac{n}{2}} D_{i,y}}{n}, D > Max, \end{cases} \tag{2}$$

where $D_{x,y}$ – depth map value at point, Max – maximum possible depth map value, n – the size of the area on which the average value is calculated.

This formula is a filter. In other words, if the value of a point is greater than expected, replace its value with an average value from neighboring points. This completes the depth recovery. The following formula is used to visualize the results:

$$G_{x,y} = 255 \cdot \left(D_{x,y} / D_{\max} \right), \tag{3}$$

where $D_{\max}$ – maximum depth map value, $G_{x,y}$ – point value in grayscale.

8 Computational experiment

To conduct a computational experiment, a database of stereo images was formed. The database consists of 2000 different pairs of images. For each of the pairs of images in the database there is also a reference depth map (see Fig. 3).

Fig. 3. An example of images used in a computational experiment (left, right and depth map)

During the experiment, each point of the reference depth map is compared with the corresponding points of the depth map obtained by the algorithm proposed in this paper.

The proposed method for solving the problem of depth mapping has a set of input parameters. Therefore, in the course of the experiment, different sets of values of the input parameters of the algorithm were investigated in order to identify the set that allows depth mapping with the greatest accuracy. As a result of a combination of all the specified values of the input parameters of the algorithm, nine launch configurations were obtained. For each configuration of the launch of the algorithm, the following values were obtained: the accuracy of depth mapping, the average processing time of a single image. The test results of the algorithm are given in Table 1.

Table 1. Algorithm testing results

Maximum amount of segments	Minimum size of segments	Accuracy, %	Average processing time, *s*
1	10	82,4	10
	50	83,7	9
	70	82,8	9
4	10	90,2	7
	50	90,4	6
	70	90,6	6
8	10	90,3	5
	50	90,7	5
	70	90,7	5

Table 2 presents the results of the known methods for depth mapping [1].

Table 2. The results of work of the known methods of mapping depth

Method	Accuracy, %
SAD without segmentation	87,6
MeanShift и SAD	90,7
Trust Distribution Algorithm and SSD	91,8

Comparing the data from Table 2 and the obtained results of testing the algorithm (see Table 1), we can conclude that the developed method has a depth map construction accuracy, which is quite comparable with the accuracy of the known methods considered. As a result of testing the algorithm in normal conditions, the accuracy of constructing a depth map equal to 90.7% was obtained.

References

1. Kamencay P., Breznan M., Jarina R., Lukac P., Zachariasova M. Improved Depth Map Estimation from Stereo Images Based on Hybrid Method. Radioengineering Journal, 2012. P. 70-78.
2. Comaniciu D., Meer P. Mean Shift: A Robust approach towards feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. N. 5, May 2002.
3. M. B. Hisham. Template Matching using Sum of Squared Difference and Normalized Cross Correlation. Research and Development (SCOReD). 2015 IEEE Student Conference
4. Utrobin V.A. Physical Interpretations of the Elements of Image Algebra // Uspekhi Fizicheskikh Nauk (UFN). 2004. V. 174, No. 10. pp. 1089-1104.