SEGMENTATION OF STEREO IMAGES WITH THE USE OF THE 3D HOUGH TRANSFORM

Ye.V. Goshin, G.E. Loshkareva

Image Processing Systems Institute - Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Samara, Russia
Samara National Research University, Samara, Russia

Abstract. A technology for the segmentation of stereo images using the Hough transform is proposed in this paper. The first stage of the process is the formation of a 3D model of a scene in the form of a point cloud - where information about camera parameters is lacking. In the second stage, with the help of the Hough space, the most appropriate set of planes is found. The segmentation of the generated scene is then carried out, based on these. Experimental research results relating to simulated scenes are given.

Keywords: image processing, 3D reconstruction, segmentation, Hough transform

1   Introduction

The problem of image analysis and, in particular, the problem of image segmentation is common to many different research areas. However, information useful for a reliable segmentation of images is often lacking. This may occur, for example, when the texture of the scene objects consists of large areas of different colors. In this case, if there is more than one image, it makes sense to analyze not the intensity distribution of individual images but the three-dimensional structure of the scene instead.
There are a large number of algorithms which construct a three-dimensional model of a scene from stereo images [1, 2]. Typically, in the case where the images have been obtained from arbitrary viewpoints, the camera parameters are unknown, and thus it is necessary to determine these parameters. The problem of stereo image matching under these conditions is considered in several articles [3, 4].
One of the approaches to the segmentation of three-dimensional scenes consists, first, in the detection of some specific objects in the scene [5, 6]. For example, [7] considers the scene segmentation based on the detection of planes in the point cloud, ob-

tained by scanning the surface with a laser ranger (LIDAR). The 3D Hough transform is used for the detection of the planes [8].

The aim of this work is to study a 3D scene segmentation method using a two-step procedure. The first step is the formation of a 3D model of a scene in the form of a point cloud - where the camera parameters are unknown. In the second step, with the use of the Hough space, we find the most appropriate set of planes on which to base a segmentation of the generated scene.

The main difference between the proposed technology and the one described in [7] is the use of stereo images. It should be emphasized that these images are not only used as initial data for the 3D model construction (in the form of a point cloud), but also act as a meaningful result in themselves, since the result of a 3D point cloud segmentation can be used for the segmentation of the initial images.

Generally, the proposed method can be used for the segmentation of the scene and the images in relation to multiple objects which belong to different planes in three-dimensional space as is also the case with [7]. However, the present paper considers a restricted problem of segmentation; this is the division of the image pixels and the constructed 3D model points into two classes – background points and object points.

The results of these experimental studies demonstrate the segmentation result in relation to a simulated scene.

2   The segmentation method

The main stages of the stereo image segmentation method, using the 3D Hough transform, are shown in Fig. 1.

In accordance with the scheme, the 3D scene model is constructed from two images. Then, a Hough transform is carried out on all points of the resulting 3D scene. Among all the planes we search for the maximum and thus select this as a background plane. By calculating the distance from the points to the selected plane, the model is divided into background points and object points. The initial image can be segmented with the use of the segmented scene obtained.

The key stages of the process will be considered in detail further on in this paper.

Camera parameters estimation

In this paper, we use the algorithm for camera parameters determination described in [9]. It is assumed that the matrices of the first and the second cameras are the same and known. Also the global system of coordinates has been made to coincide with the coordinate system of the first camera: i.e., the camera view direction coincides with the direction of the axis OZ. In addition, the set of corresponding points $\left\{ \mathbf{m}_i \mid i = \overline{1,N} \right\}$ and $\left\{ \mathbf{m}_i' \mid i = \overline{1,N} \right\}$ is given, where each pair of points $\mathbf{m}_i = \begin{pmatrix} x & y \end{pmatrix}^T$ and $\mathbf{m}_i' = \begin{pmatrix} x' & y' \end{pmatrix}^T$ are projections of the same point in three-dimensional space in the first and second images, respectively.
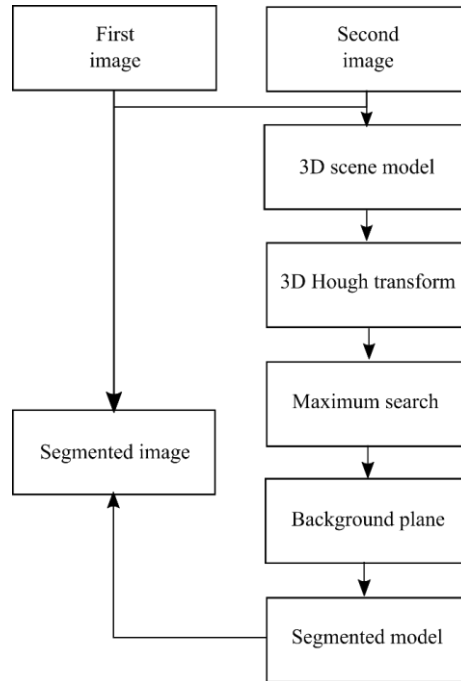
Fig. 1. General scheme of the technology stages

We designate the elements of the translation matrices and the rotation vectors of the first and second cameras as: $\mathbf{R}$, $\mathbf{R'}$, $\mathbf{t}$ and $\mathbf{t'}$

We obtain an expression correlating the corresponding points in the two images through the parameters of the rotation matrix and translation vector up to the accuracy of an unknown parameter, Z :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \left(r'_{11}x + r'_{12}y + r'_{11} + t'_x/Z \right)\left(r'_{31}x + r'_{32}y + r'_{31} + t'_z/Z \right)^{-1} \\ \left(r'_{21}x + r'_{22}y + r'_{21} + t'_y/Z \right)\left(r'_{31}x + r'_{32}y + r'_{31} + t'_z/Z \right)^{-1} \end{pmatrix}$$

The resulting expression can be represented as a sequence of rotations and translations:

$$\mathbf{m}\left(\tilde{x}, \tilde{y}, \tilde{z} \right) = \mathbf{R}'\mathbf{m}\left(x, y \right),$$

$$\begin{cases} x'\left(\tilde{z} + t'_z/Z \right) = \tilde{x} + t'_x/Z, \\ y'\left(\tilde{z} + t'_z/Z \right) = \tilde{x} + t'_y/Z, \end{cases}$$

where Z is the point coordinate in three-dimensional space.

The proposed decomposition enables an iterative procedure to determine the parameters of the camera. This consists of two stages: the procedure for determining the rota-

tion, in the course of which the rotation matrix $\mathbf{R}'$ is constructed, and the procedure for determining the translation, in the course of which the translation vector $\mathbf{t}' = \begin{pmatrix} t_x' & t_y' & t_z' \end{pmatrix}$ is constructed.

Using the Lucas–Kanade method [10], we form an optical flow which matches points between the first and second images. The point cloud based on the matches obtained is formed by triangulation [11].

Segmentation of a 3D scene model

The aim of this stage is to separate the objects from the background.

To detect planes we use the 3D Hough transform. The Hough transform is a method for detecting parametrized objects which is generally used to detect circles and lines. For example, paper [12] describes the detection of a variety of two-dimensional objects with the reference contour using the generalized Hough transform.

In this paper, a set of points in space $\mathfrak{R}^3$ is used as the input and output values of the parameterized plane. The plane is represented by the distance from the origin of coordinates to the plane and a normal vector to this plane. Let $\mathbf{p}$ be a point in the plane, $\mathbf{n}$ is a normal vector which is perpendicular to the plane and ρ is the distance

$$\rho = \mathbf{p} \cdot \mathbf{n} = p_x n_x + p_y n_y + p_z n_z$$

After substituting the angles between the normal vector and the selected coordinate system the equation of the plane can be written as follows:

$$p_x \cdot \cos\theta \cdot \sin\phi + p_y \cdot \sin\phi \cdot \sin\theta + p_z \cdot \cos\phi = \rho \tag{1}$$

where θ is the angle of inclination of the normal vector to the plane xy and ϕ is the angle between the plane xy and the normal vector in the direction of axis z. The coordinates, ϕ, θ and ρ define, as such, a 3D Hough space, each point of which has a corresponding plane in $\mathfrak{R}^3$. In turn, each point (x_0, y_0, z_0) of space $\mathfrak{R}^3$ has some surface in the Hough space corresponding to it. At the same time, each point of this surface (ϕ, θ, ρ) characterizes some plane, passing through the desired point, (x_0, y_0, z_0).

In this paper, we consider the problem of finding the background plane containing the highest number of points from the point cloud. After determining the parameters $(\hat{\phi}, \hat{\theta}, \hat{\rho})$ of the background plane, for all points of the initial cloud we determine whether this point belongs to this background plane. To do this, we substitute the point's coordinates into the plane equation and then compare the resulting value with a certain threshold:

$$p_x \cdot \cos\hat{\theta} \cdot \sin\hat{\phi} + p_y \cdot \sin\hat{\phi} \cdot \sin\hat{\theta} + p_z \cdot \cos\hat{\phi} - \hat{\rho} < \Delta \, .$$

All the points that satisfy this inequality belong to the background plane, while the rest are considered to belong to objects of the scene.

Since there remains a one-to-one correspondence between the points of the 3D model and the image pixels, the results of the model segmentation can be used for the segmentation of the initial images.

3   The implementation of the 3D Hough transform algorithm

In order to implement the 3D Hough transform the following algorithm is used. As the Hough accumulator space, we used a three-dimensional array of integers wherein each cell of the array corresponds to some plane - with the parameters set by the co-ordinates of this cell.

Each point of the point cloud formed in the previous stage increments the value of those cells of the accumulator space, which correspond to planes passing through the given point or near it - an exact match is not possible due to the discrete nature of the array indices.

This algorithm can be written as pseudocode in the following way:

For each point of the point cloud:

└For each value of the angle coordinate θ:

└ For each value of the angle coordinate ϕ:

├Calculate ρ using formula (1)

└Increment $A(\theta, \phi, \rho)$

As a result of this algorithm, each cell of the array is assigned a value, which is the number of points of the point cloud lying near the plane which is represented by this cell. The cell of the array that contains the maximum value represents the required background plane.

4   Experimental results

The findings of the experimental study of the proposed technology are given below. For these purposes, a simulated scene was set up and two images were obtained at various angles (Fig. 2, 3).
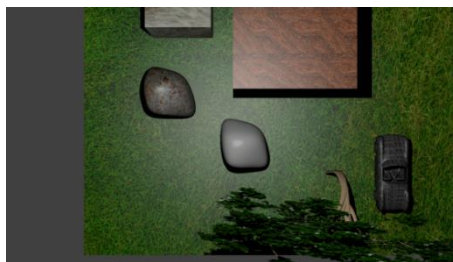

Fig. 2. First image of the scene

Fig. 3. Second image of the scene

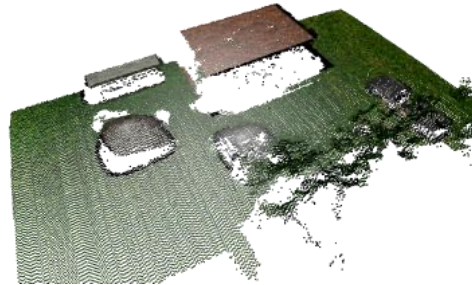Fig. 4 illustrates the constructed 3D model of the scene as a point cloud.

Fig. 4. Constructed scene model

To form the Hough space, the following parameters were selected:

$$\theta = k^\circ, \quad k = \overline{0,179},$$
$$\phi = k^\circ, \quad k = \overline{0,359},$$
$$\rho = 0.001k, \quad k = \overline{0,1000}.$$

With the use of these parameters and the generated point cloud, the background plane was determined, and the points of the cloud were divided into those belonging (Fig. 5) and those not belonging (Fig. 6) to this plane.

Fig. 5. Background plane of the scene

Fig. 6. Objects of the scene

The initial images were segmented in accordance with the segmentation which had been carried out on the 3D model. The first segmented images of the scene are shown in Figs. 7 and 8.
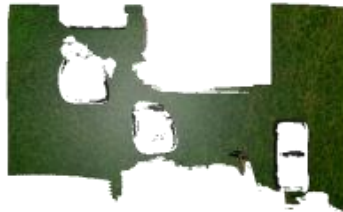

Fig. 7. Background of the first image

Fig. 8. Objects of the first image

It is evident that the algorithm separated (though not perfectly) the background plane from the objects not belonging to that plane.

5   Conclusion

It has been demonstrated that the proposed information technology, relating to stereo images segmentation and based on the application of the 3D Hough transform to a generated point cloud, provides a good quality of segmentation. This experimental work on a simulated scene allowed us to separate the background from the other planes, of an image.

Acknowledgements

References

1. Pollefeys M, Nistér D, Frahm JM, Akbarzadeh A. Detailed real-time urban 3d reconstruction from video. International Journal of Computer Vision, 2008; 78(2-3): 143-167.
2. Baillard C, Maıtre H. 3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy. Computer Vision and Image Understanding, 1999; 76(3): 244-258.
3. Pollefeys M, Koch R, Van Gool L. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. International Journal of Computer Vision, 1999; 32(1): 7-25.
4. Eisert P, Steinbach E, Girod B. Automatic reconstruction of stationary 3-D objects from multiple uncalibrated camera views. Circuits and Systems for Video Technology, IEEE Transactions on, 2000; 10(2): 261-277.
5. Reitberger J, Schnörr C, Krzystek P, Stilla U. 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 2009; 64(6): 561-574.
6. Tarsha-Kurdi F, Landes T, Grussenmeyer P. Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. Proceedings of the ISPRS Workshop on Laser Scanning, 2007; 36: 407-412.
7. Zhang J, Lin X, Ning X. SVM-based classification of segmented airborne LiDAR point clouds in urban areas. Remote Sensing, 2013; 5(8): 3749-3775.
8. Borrmann D, Elseberg J, Lingemann K, Nüchter A. The 3D Hough Transform for plane detection in point clouds: A review and a new accumulator design. 3D Research, 2011; 2(2): 1-13.
9. Goshin YeV, Fursov VA. 3D scene reconstruction from stereo images with unknown extrinsic parameters. Computer Optics, 2015; 39(5): 770-776. DOI: 10.18287/0134-2452-2015-39-5-770-775.
10. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. IJCAI, 1981; 81: 674-679.
11. Hartley RI, Sturm P. Triangulation. Computer vision and image understanding, 1997; 68(2): 146-157.
12. Fursov VA, Bibikov SA, Yakimov PYu. Localization of objects contours with different scales in images using hough transform. Computer Optics, 2013; 37(4): 496-502.