To Study the Effect of Varying Camera Parameters in the Process of Matching and Reconstruction

Anuj Kumar, Naveen Kumar, and Rakesh Kumar Saini

Abstract In computer vision, there are many methods of reconstructing the image point. 3D reconstruction from digital image sequence of scene or object is a difficult and important task in computer vision. However, such a reconstruction requires a large computational effort for finding the point of correspondence between different views. Furthermore, the accuracy should not be reduced in case of noisy data. There is one important technique to digitization of physical object which is binocular stereo vision. From two subsequent digital images of the physical object taken from different viewpoints, we can make a 3D virtual model for the physical object by using this approach. Basically, the common processes for binocular stereo vision comprise digital image acquisition, camera's calibration, feature point extractions, feature points matching and 3D reconstructions. In this paper, we have discussed the problem of varying camera parameter in the process of matching and reconstruction. we have the studied the problem and how it affects when the camera parameter varies in the process of matching; two types of parameter are as follows: intrinsic parameter and extrinsic parameter; image setup and some equation help to set the parameter, and a robust algorithm is proposed for reconstructing free-form space.

Keywords Binocular stereo · Epipolar constraint · Intrinsic parameter · Extrinsic parameter · Disparity map · Calibration

A. Kumar
Department of Computer Application, SRCM, Muzaffarnagar, India
e-mail: dixit.anuj12008@gmail.com

N. Kumar (✉) · R. K. Saini
School of Computing, DIT University, Dehradun, Uttarakhand, India
e-mail: naveen.it23@gmail.com

R. K. Saini
e-mail: rakeshcool2008@gmail.com

1 Introduction

Most recently working areas are camera calibration and stereo matching systems based on the binocular stereo vision camera. In camera calibration, we have to estimate the parameter of camera that is very useful to acquire good 3D positioning reconstructions accuracy. This paper investigates camera parameters that inference the system reconstruction accuracy. The camera parameters associated with a stereo imaging setup can be classified into two main categories: (i) intrinsic camera parameters and (ii) extrinsic camera parameters. The effect of extrinsic camera parameters in a stereo imaging system can be tackled by the epipolar geometry constraint. However, intrinsic camera parameters, like focal length, lens distortion, image resolution, etc., interfere directly in the 3D reconstruction. In this context, it will be interesting to study the effects of these parameters in the 3D reconstruction.

In set of 2D digital image that considered as an input (arbitrary perspective views) and parameters of the viewing geometry, the output represents 3D object that defines by structural parameter of the model. This type of reconstruction is performed by input in the form of individual point (pixels), and groups of pixels represent some geometric feature in the image like as line, curve, etc. However, it is very difficult to process because the relative depth is missing during the projection of 3D scene on to the projection plane. In traditional method, 3D real-time depth information was extracted from image using sensor, whereas passive stereo vision model represents a significant error in processing un-texture region which is frequent in indoor world.

The 3D points lie on the line which starts from the center of projection, and it moves in the direction of its corresponding projected points in an image plane. We have only one image or view; the corresponding unique 3D points cannot be reconstructed as different 3D points lie on the same projected line. So if we want unique reconstruction, then we use at least two images. It provides the intersection of two projection lines through several projection planes that yield the unique 3D point.

There are following basics steps for stereo view 3D reconstruction process:

- Check the corresponding 2D features of the projected 3D scene or object onto the view planes.
- Find out the parameters of these projected 2D geometrical features in both view planes.
- Establish the correspondence between these projected 2D features.
- Use inverse perspective equations for computing the 3D structural parameters.
- Show the 3D object using line drawings or shading model.

The 2D projections or input image generate hurdles in the problem of reconstruction of a 3D object. These projections mainly depend on the relative position of the object with effect to the cameras, line of sight and other parameters of the cameras. The 3D reconstruction problem is divided into two subproblems: correspondence and triangulation. In the problem of correspondence includes the corresponding features in different images. It is formulated as: Given different images or views of a 3D

object, find different points or features in one view which may be referred as the same points or features in other views. While the triangulation problem yields the location of 3D points from its two or more projections. The main difficulties in the field of multiview reconstruction involve: establishing the correspondence between pair of images and obtaining inverse mathematical functions based on camera models (perspective, weak perspective, affine, etc.)

Each object in real world is 3D, and this is one of the main issues in computer vision to acquisition of 3D information of real-world scene from 2D image. The acquisition is divided into two methods: active method and passive method. The passive method calculates the inverse problem of process of projecting a 3D scene onto a 2D image plane, and here is an advantage to find the three-dimensional information with no effect to the scene. And the active method emits the radio or light energy from source. The passive technique is referred to as motion, stereo, shading, texture, contours and so on. Basically, there are two categories of inverse problem: photometric (optical) and geometrical. And all these types of inverse problems are used for image formation process.

1.1 Binocular Stereo

Binocular stereo or two-view stereo is based on the person stereo vision, and it is a non-contact passive sensing method. A pair of images of a 3D object (scene) is obtained from two different viewpoints under perspective projection as illustrated in Fig. 1. To obtain a 3D object from these perspective images, a distance measurement (depth) from a known reference coordinate system is computed based on triangulation. The main problem in binocular stereo is to find the correspondences in stereo pair called the stereo correspondence problem. In this section, some concepts

**Fig. 1** Binocular stereo image formations
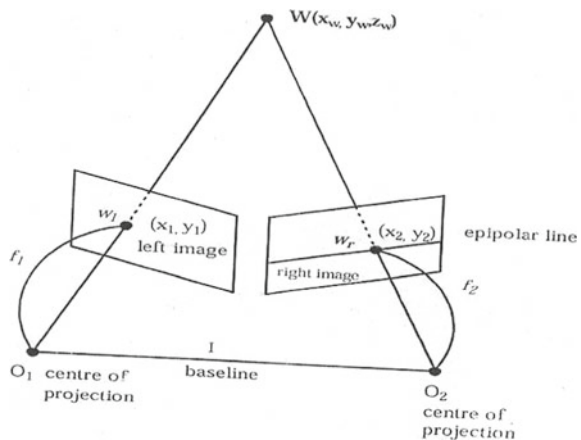
(camera model, epipolar geometry, rectification and disparity map) related to stereo correspondence problem have been introduced.

2 Literature Review

Several studies [1, 13, 16, 20] were carried out for a surface reconstruction from the scanned data. However, the study split into two methods: Voronoi based and mesh free. Voronoi methods are used with Delaunay triangle [5] to reconstruct the surfaces, and according to mesh-free algorithm, the surfaces are reconstructed by using B-splines [22], radial basis functions [18], and PDE and MLS [9] techniques. Along with these techniques, data point acquisition is proving to be a very useful step. In most of the hardware system [7, 21], a light wave technique was utilized in [7] for 3D surface point reproduction. A MS Xbox kinect sensor is used to get the volumetric depiction, and a similar kind of MS kinect sensor camera was used in [21] 3D inside scene recreation. Computational geometry devices were created in [1] for the surface demonstration from the valuable information. To again construct surfaces and to get correspondences all the while, some notable finite set of instruction [6, 11, 12] which depends on the volumetric reproduction is displayed from the 2D pictures existed in writing. A strategy for spaces carving and voxel's coloring dependent on the perceive ability and the stability of the voxels in the picture was created [12]. In the direct discrete minimization (DDM), details of graph cut methodologies were likewise proposed in [8, 11, 17] for building up the corresponding within the images. The entirety of the proposed strategies acquires disparity maps with exact shapes [2, 19] but limited depth precision. Be that as it may, this limitation was removed in [4]. Other calculation for the reconstruction is of 3D scenes from just two perspectives and depends on minimum line correspondences introduced in [15]. Another method for a precise 3D shape estimation of different separate items was exhibited in [10] utilizing stereo vision methodology. Actually, the excellence of the system is capacity of the system of playing out full-field three-dimensional shape estimation with high precision even within the sight of discontinuities and different separate districts. Numerous applications like acknowledgment, robot vision and activity require a substitute portrayal of three-dimensional questions in a minimized structure. Curve skeleton is an alternate representation from average surface representation [3]. Curve skeleton incorporates a few models: virtual route, enrollment, liveliness, transforming, logical investigation, shape acknowledgment and shape recovery. On account of 2D, skeleton is named as medial axis, but in 3D, these are called mean surface. Curve skeleton is an uncommon 1D representation of a 3D object whose recreation is additionally a troublesome subject because of a not-well-presented issue. Aside from the above representation techniques, endeavors were additionally made for reproducing the 3D surface utilizing skeletons of the object [7, 14].
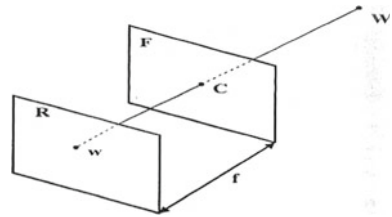
The Pinhole Camera Model.

The pinhole camera model describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane.

As shown in Fig. 2, basically it consists of two screens: retinal plane and focal plane. The retinal plane (Re) is where the 2D image is formed; the focal plane (F) placed in center is called optical center C. Here f is the focal length of the camera based on both planes which are parallel from certain distance. A straight line connects the point W. It is world point. This point W mapped 2D image by using perspective projection. Since retinal plane and focal plane are parallel to each other, the points lie on focal plane and it has no image on retinal plane.

Epipolar Constraints:

The epipolar constraint makes sure that the epipolar lines are corresponded to the horizontal scan lines and here shifted horizontally all the corresponding point in both the images and reduced the two-dimensional search space into one-dimensional image.

If there are two cameras, every points of these cameras are $W = [x_w\ y_w\ z_w,\ 1]^T$ of the real-world reference frame can be projected to the suitable image frame (w_l and w_r) using the transformation matrix $P = TiT_l$ as known form above equation.

$$W_r = P_r W,$$
$$W_l = P W$$

In order to correct the image according to the epipolar constraints, these left projection matrix P_l & right projection matrix P_r have to be used in the following particular conditions:

1. Focal length must be equal for both cameras which is necessary.
2. Focal plane must be same for both cameras which is also necessary
3. And the optical centers of camera need to be constant.
4. Each point in left image and right image needs to be same for correspondence of the vertical coordinate.

P_l and P_r can be further calculated using these, and obtained image can be transformed according to the epipolar constraint.

Rectification Process:

The rectification process is used to determine distance of an object in triangular-based stereo vision. However, binocular disparity is the process of corresponding depth of an object to change its position, especially in case of different camera view, when relative position is known for the camera.

It is a transform process that is used onto same plane image for two or more images. Generally, ratification is used for correspondence analysis. Correspondence processes the used rectification in very simple way. It defines two new perspective matrices that store the optical centers with the baseline contained in the focal planes. This ensures that the epipoles are at infinity, and hence, these epipolar lines are parallel.

3   Disparity for Three-dimensional Reconstructions

The disparity is the distance between two points in the rectified object that is image. It is actually applied in three-dimensional reconstruction, because it is proportional to the 3D world coordinate and distance between the cameras.

3.1   Disparity Map

Once both the stereo images are rectified, the next step is to find the correspondence between them. For this, between left image and the right image, it shows the similarity or different measures. **Sum of Square Differences** (SSD) used measure for are based, the **Sum of Absolute Differences** (SAD), the Normalized Cross Correlation (NCC) and the census. Here coordinates are defined as
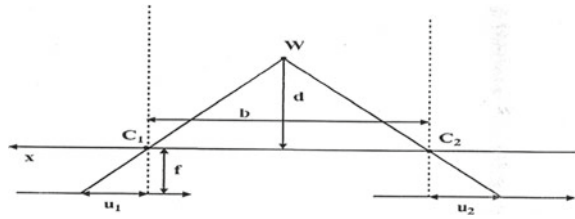
$$\text{SSD (d)} = \sum_i \sum_j [I_l(x + d + i, y + j) - I_r(x + i, y + j)]$$

$$\text{SSD (d)} = \sum_i \sum_j [I_l(x + d + i, y + j) - I_r(x + i, y + j)]$$

$$\text{NCC } (d) = \frac{C(I_l I_r) - \sum_i \sum_j \mu_l \mu_r}{\sum_i \sum_j \sigma_l \sigma_r}$$

Here d represents the disparity between left and right images, and I_1 and I_r represent the stereo pair in left and right images at a point (x, y) in the right image. μ_1 and μ_r represent the mean intensities of left image as well as right image in the corresponding windows, and σ_1 and σ_r are standard deviations in the windows, respectively. $C(I_l, I_r)$ is a cross-correlation between the corresponding windows:

Fig. 3 Disparity

$$C(I_l, I_r) = \sum_i \sum_j I_l(x + d + i, y + j) I_r(x + i, y + j)$$

In Fig. 3, the disparity d can be found by minimize difference measures or by maximize similarity measures, but due to ill-posed nature of the problem, one cannot find the unique correspondence.

4 Camera's Parameters

As to change a point of the 3D world coordinate into a 2D purpose of the image planet, the information of quality camera is needed. Basically, two types of camera parameter, that is intrinsic and extrinsic, are also called internal and external, respectively. Intrinsic parameter defines the interior geometric and optical attributes of the camera, and extrinsic defines the position and direction of the camera in the world framework. The consequences for the framework reconstruction exactness are analyzed using results of simulation. At the end, we analyzed three parameter errors and system reconstruction errors, and the affected accuracy of the system is given. However, intrinsic camera parameters, like focal length, lens distortion, image resolution, etc., interfere directly in the 3D reconstruction accuracy which is somewhat bigger, and relationship of them is direct. The mistake in outer parameter straightforwardly impacts remaking exactness by influencing pattern separation and the point between cameras.

As observed in Fig. 4, the system for displaying at least two cameras comprises three separate coordinate scheme, (x_w, y_w, z_w) is the world reference frame and (x_c, y_c, z_c) is the camera frame with the focuses on origin and the imaging frame (X, Y). A 3D point, given in same world directions, can be changed over into the camera frame, by a revolution r_{ij} and an interpretation t_j which is communicated by the outward parameter as

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = T_l \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} \text{Where } T_l = \begin{pmatrix} r_{11}\ r_{12}\ r_{13}\ t_1 \\ r_{21}\ r_{22}\ r_{23}\ t_2 \\ r_{31}\ r_{32}\ r_{33}\ t_3 \end{pmatrix}$$
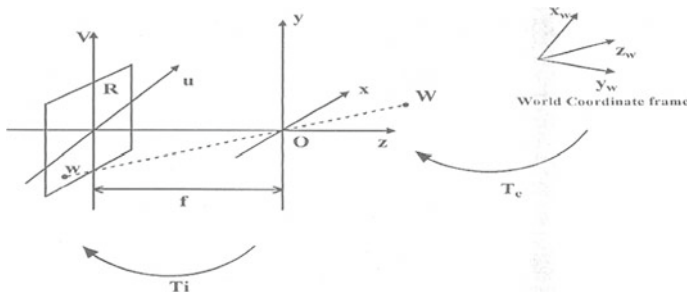
Fig. 4 Intrinsic and extrinsic parameters of camera

(I) Intrinsic Parameters:

It is the most important parameter of the camera that for each camera, it characterizes the change from picture plane coordinate to pixel coordinate. If the intrinsic camera parameters are known, it is possible to obtain a metric reconstruction. If you obtain matrix reconstruction, it is necessary know intrinsic parameter. This metric has five parameters; this calibration can be obtained through off-line calibration with a calibration object.

At that point, it is changed over to the two-dimensional picture planes utilizing the intrinsic coordinates. These are specifically the central length f, the rule point (u_0, v_0), which is the focal point of the picture plane, and (k_0, k_1) the size of pixels in millimeter (mm) or $\alpha = f/k_0$ and $\beta = f/k_1$. The change utilizing inherent parameter is as per the following:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = T_i \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \text{ Where } T_i = \begin{pmatrix} \alpha 0 u_0 \\ 0 \beta v_0 \\ 0 0 1 \end{pmatrix}$$

Since (X, Y, Z) is uniform, and Z variables are divided X, Y, Z variable in order to find out coordinate of pixels X' & Y' and on the focal plane the points are $Z = 0$ and $z_c = 0$. These pixels cannot be changed into image plane coordinate because it is divided by zero that is not define and this point connected by in straight; by this reason does not intersect by optical center to image plane, whereas image plane is parallel to it. There are following equations that given a point in world coordinate to convert into 2D image plane.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = T_l T_i \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

The information given by internal and external parameter of camera permits for rectifying of images and ensures the epipolar constraint.

(II) Extrinsic Parameters (R, T)

The extrinsic parameter describes the relative position and orientation of the two cameras. It defines the 3D world coordinate to 3D camera coordinates. It also defines the position of the camera's heading in world coordinates and camera center. R represents the rotation matrix, and T represents the position of the origin of the world coordinate.
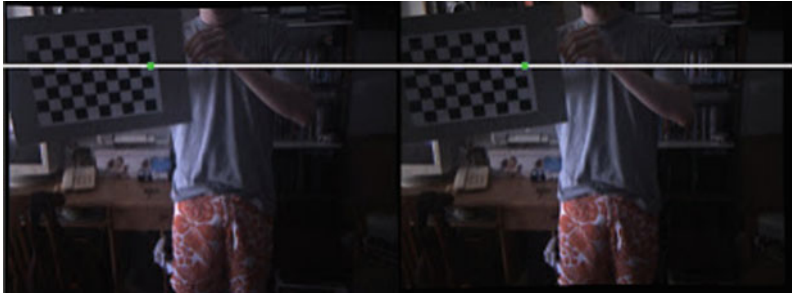
5 Calibration

Stereo vision is used to find out the depth from camera image when we compare one or more views of same scene. When we compute the result, each 3D point corresponds to a pixel in one of the images. Binocular stereo process applies on only two images, which taken with cameras that are differentiated by a horizontal distance called "baseline." Calibrating the stereo camera defines the actual unit of 3D world points such as millimeters, relative to the cameras. So many applications of this process include measurement, aerial photogrammetric and robot navigation. The calibrated stereo camera combination is of following steps.

1. To the calibrated stereo camera system
2. To correct a pair of stereo images
3. To compute disparity
4. And reconstructing the three-dimensional point cloud.

5.1 Calibration of a Stereo Camera

The calibration of stereo camera provides the intrinsic estimation of every camera. It involves the translation and rotation of the second camera corresponding to the first camera. All the parameters of camera are used to rectify a stereo pair of image and to make the two images parallel. The rectified image calculates the disparity map required to reconstruct 3D view. When you calibrate the stereo method, you drew multiple pairs of a calibration design from different views. The whole design of calibration can be shown in every image. You must save the image in a format that not has use lossy compression such as PNG format. JPEG image format is not in lossy format. The lossy format reduces the calibration accuracy. When the camera is set in parallel position, then epipolar line turns into parallel. This means that, the left image point is exactly same in the right image on the line (Fig. 5).

When you want to calibrate your stereo camera, then the camera lenses make some distortion. This distortion in the image makes straight lines in the real world appear curved (Fig. 6).

Fig. 5 Calibrated image
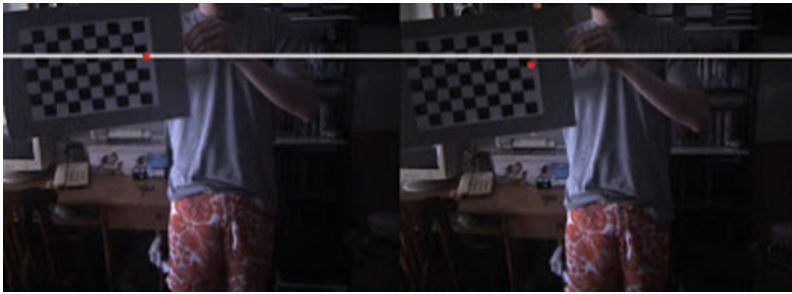
Fig. 6 Uncalibrated stereo images

The above two images are taken from uncalibrated stereo. In this picture, the cameras are not perfectly parallel that mean more or less parallel.

6 Experimental Results

A stereo imaging system was simulated in order to observe the effect of stereo rig parameters in the reconstruction error. We have given our emphasis on the two parameters mainly:

1. Baseline distance
2. Focal length.

In order to see the effect of these two parameters, we have considered the reconstruction of a quadratic curve, in which these parameters have been perturbed to see the effect on the 3D shape of the curve in stereo reconstruction.

The original 3D plot of the synthetic shapes has been shown in Fig. 7. Generally, the projected images in both views correspond to the different samplings of 3D object. Thus, we have evaluated the reconstruction algorithm with those synthetic objects whose data points correspond to different samplings of the 3D free-form objects. We
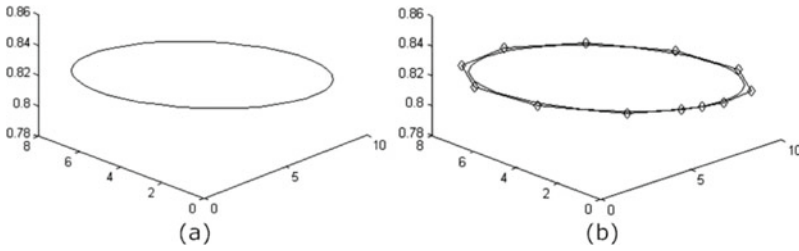
Fig. 7 Original and reconstructed elliptical curve in space with perturbed stereo parameters

Table 1 Effect of focal length

Change in focal length (in percentage)	Absolute error in reconstruction of an ellipse (voxels)	Absolute error in reconstruction of a helix (voxels)
1%	0.25	1.40
2%	0.98	3.58
5%	3.5	7.88
10%	17.25	32.40

have added the white Gaussian noise with varying standard deviations to the original data and perturbed the focal length and baseline distance a little bit in order to access the effect on reconstructed result.

In this experiment, the original and reconstructed objects are found to have a similar shape. The proposed algorithm is equally effective for the reconstruction of curve in case of a small variation in the stereo rig parameters. Moreover, from a qualitative point of view, we consider the following two tables to evaluate the effect of stereo parameters in the reconstruction of an ellipse and a helix using stereo vision technique.

Table 1 represents absolute error in reconstruction of an ellipse and helix when changing focal length parameter of camera (Fig. 8).

In Table 2, we have shown the mean error in reconstruction of an ellipse and helix when changing the baseline distance parameter of camera (Fig. 9).
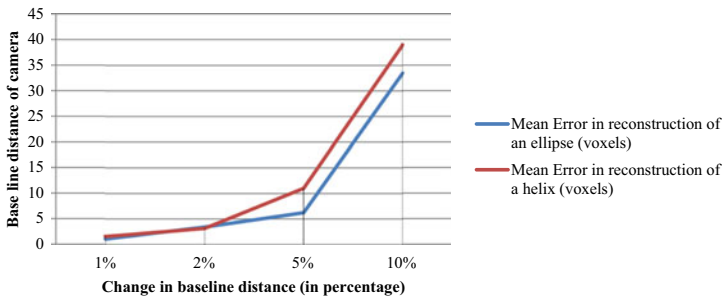
7 Conclusion

In this paper, we studied the effect of varying camera parameters in the process of matching and reconstruction and gave some results. This paper shows if the recent methods are adopted using inverse geometric reconstruction, then the problem can be modeled as a stereo vision where the effect of stereo rig parameters is negligible if the deviation in the parameters is quite small. However, the error increases significantly
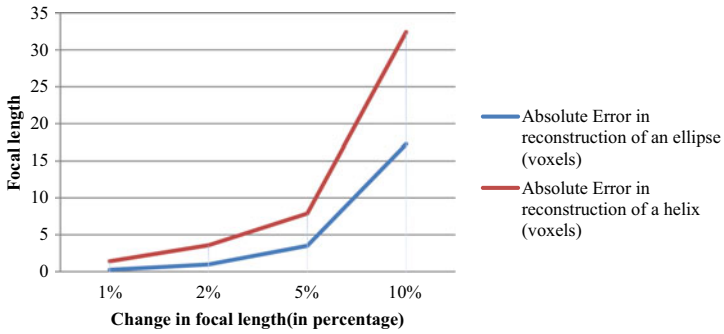
Fig. 8 Focal length versus change in focal length percentage

Table 2 Effect of baseline distance

Change in baseline (in percentage)	Mean error in reconstruction of an ellipse (voxels)	Mean error in reconstruction of a helix (voxels)
1%	1.08	1.58
2%	3.42	3.14
5%	6.22	11.00
10%	33.44	38.96

Fig. 9 Baseline distance of camera versus baseline distance of camera in percentage

if there is a large perturbation in the stereo rig parameters. In future, this problem can be studied in real-time stereo. Such an implementation will be useful in many applications such as ball trajectory tracking in sports event, missile and robot path planning.

References

1. Amenta, N., Choi, S., Kolluri, R.: The power crust, unions of balls and the medial axis transform. Computational Geometry: Theory and Applications **19**, 127–153 (2001)
2. P. K. Atrey, A. De, and N. Rajpal. Polynomial representation of 2-D image boundary contours. In IEEE Region 10 Conference on Computers, Communications,Control and Power Engineering (TENCON '02), volume 1, pages257–260, Beijing, China, 2002.
3. Bhatt, D., Mohammad, S.: Validation on medial axis transforms objects. Computer-Aided Design & Applications **9**(4), 517–529 (2012)
4. Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In International Conference on Computer Vision, 2003.
5. F. Cazals and J. Giesen. Delaunay triangulation based surface reconstruction: ideas and algorithms. Springer, 2006.
6. O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In European Conference on Computer Vision (ECCV), pages 379–393,
7. E. Frigerio, M. Marcon, and S. Tubaro. Surface reconstruction using 3D morphological operators for objects acquired with a multi-kinect system. In MIRAGE, volume 9, 2013.
8. H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines instereo.In European Conference on Computer Vision (ECCV), 1998.[63] I. Isuppli. The teardown: The kinect for xbox 360. IET Engineering Technology, 6(3):94–95,
9. M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In Proceedings of the Fourth Euro graphics Symposium on Geometry Processing, 2006.
10. Kieu, H., Pan, T., Wang, Z., Le, M., Nguyen, H., Vo, M.: Accurate 3D shape measurement of multiple separate objects with stereo vision. Measurement Science Technology **25**, 1–7 (2014)
11. V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In European Conference on Computer Vision (ECCV), 2002.176
12. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. International Conference on Computer Vision, Kerkyra **1**, 307–314 (1999)
13. M. S. Lee, G. Medioni, and P. Mordohai. Inference of segmented overl aping surfaces from binocular stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2002.
14. J. Liu, X. Zhang, and F. Blaise. Distance contained skeleton for volume reconstruction. In Proceedings of Asian Conference on Computer Vision (ACCV2004), pages 581–586, 2004.
15. Mosaddegh, S., Fofi, D., Vasseur, P.: Two view line-based motion and structure estimation for planar scenes. Electronic Letters on Computer Visionand Image Analysis **11**(1), 28–40 (2012)
16. P. J. Narayanan, P. W. Rander, and T. Kanade.Constructing virtual worlds using dense stereo. In European Conference on Computer Vision (ECCV), Bombay, 1988.
17. S. Roy. Stereo without epipolar lines: A maximum-flow formulation.International Journal of Computer Vision, 34:147–161, 1999.
18. M. Samozino, M. Alexa, P. Alliez, and M. Yvinec. Reconstruction with voronoi centered radial basis functions. In Proceedings of the Fourth Euro graphics Symposium on Geometry Processing, pages 51–60, 2006.
19. Singh, V.K., Atrey, P.K., Kankanhalli, M.S.: Coopetitive multicamera surveillance using model predictive control. Springer Journal of Machine Vision and Applications **19**(5–6), 375–393 (2008)
20. C. K. Tang and G. Medioni.Curvature-augmented tensor voting for shape inference from noisy 3D data. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2002.
21. Tiangang, S., Zhou, L., Xinyang, D., Yi, W.: 3D surface reconstruction based on kinect sensor. International Journal of Computer Theory and Engineering **5**(3), 567–573 (2013)

22. H. Wendland.Scattered data approximation, volume 17.Cambridge University Press, 2005.
23. G. Zeng, P. Sylvain, L. Quan L., and L. Maxime. Surface reconstruction by propagating 3D stereo data in multiple 2D images. In European Conference on Computer Vision (ECCV), Berlin, pages 163–174. Springer-Verlag, 2004