Using Image Stereo Pair to Determine the Displacement between Two Corresponding Cameras

Lukáš Šroba
Slovak University of Technology
Institute of Electrical Engineering
Bratislava, Slovakia
e-mail: lukas.sroba@stuba.sk

Rudolf Ravas
Slovak University of Technology
Institute of Electrical Engineering
Bratislava, Slovakia
e-mail: rudolf.ravas@stuba.sk

Ján Grman
Slovak University of Technology
Institute of Electrical Engineering
Bratislava, Slovakia
e-mail: jan.grman@stuba.sk

ABSTRACT

This paper deals with the procedure of determination of the mutual displacement between two corresponding cameras using image stereo pair and prior information coming from camera calibration. This whole process of distance estimation consists of multiple steps which are explained on the next pages. To evaluate the described approach, an experimental test was designed and performed. All the obtained results were statistically analyzed and appropriately presented in the form of the corresponding graphs and tables.

Keywords

Image stereo pair, epipolar geometry, camera displacement estimation

1. INTRODUCTION

Nowadays the camera is becoming a valuable sensor suitable for various kinds of measurements. The requirement of distance measuring or 3D position reconstructing is very often considered in computer or machine vision area and therefore widely used in many practical applications. In case of camera displacement estimation, the whole procedure is quite complex and theory behind camera stereo pair configuration has to be applied. This theory divided into several algorithm steps is explained in the next chapter.

2. CAMERA DISTANCE ESTIMATION

As it was already mentioned, there have to be multiple particular algorithms employed to obtain a real distance between two corresponding cameras in SI units.

2.1. Camera calibration

First step is generally known as camera calibration. By definition it is primarily looking for quantities that affect the imaging process [1]. For our purposes the knowledge of camera intrinsic matrix and distortion coefficients, as part of camera calibration outputs, is crucial.

Camera intrinsic matrix C is used to denote the projective mapping from 3D camera coordinates to 2D pixel coordinates, as you can see in the following formula:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f_x & s & u_c \\ 0 & f_y & u_c \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (1)$$

The elements f_x and f_y represent focal length expressed in pixel units and s is skew coefficient between both x and y axis. Symbols u_c and v_c mark the coordinates of camera principal point, u and v are 2D image coordinates, X, Y and Z are coordinates of point in camera 3D space.

There are calibration patterns in the shape of chessboard image fragments used very often [2]. The algorithm considers various images of chessboard pattern positions to be able to determine all output characteristics. Fig. 1 illustrates the usual camera calibration technique.

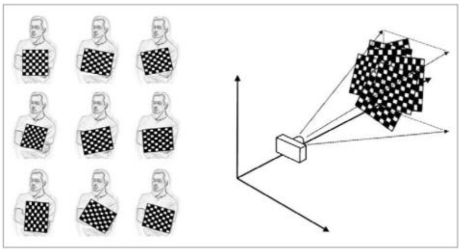

Figure 1. Illustration of camera calibration technique

2.2. Feature detection and matching

The next step is an essential component of many computer vision applications. Basically it is about the searching for interested parts of an image, also known as features. In case of image stereo configuration, the images from both cameras contain the same part of the scene. Therefore these features can be corresponding matches using the so called descriptors. There is a multiple algorithm to use the descriptors to match the features, one of the most famous is the SIFT algorithm [3]. There is the example of found SIFT corresponding features shown in Fig. 2.

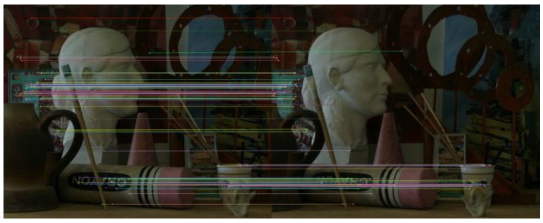

Figure 2. The corresponding features in stereo pair images

Another approach is to use a corner detector algorithm and find the corresponding corner point using, for example, RANSAC algorithm [4] together with the homography matrix [5].

As it was published in [6], the accuracy of corner point's localization has a significant impact on the quality of results in 3D reconstruction. That's the reason why the subpixel corner detector [7] was chosen to detect the corner point's features.

2.3. Image undistortion

In photography the distortion [8] is generally an abbreviation for deformation and bending physically straight lines and making them appear curly in images. Most often they are divided into radial and tangential distortions. The radial deformation is the result of optic lenses shape and usually occurs in cheap cameras using the lower quality optics. The examples of this distortion types are in Fig. 3. Tangential deformation is presented due to manufacturer errors when the lenses are not exactly parallel to imaging plane.

To describe the camera distortion there are distortion coefficients usually used and are one of the outputs of camera calibration. Because of these distortion coefficients, the reversed mathematical operation called image undistortion can be applied. Previously found corresponding points are undistorted to get more proper set of coordinates and then are processed in this form for the rest of the procedure.

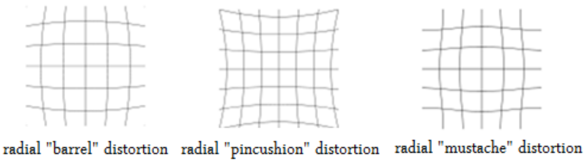

Figure 3. The types of radial distortion

2.4. Epipolar geometry

The epipolar geometry is an inner projective geometry between two cameras views and is based only on intrinsic camera parameters and the relative positions of cameras [9]. The situation around epipolar geometry is illustrated in Fig. 4 [10]. As it is possible to see, the 3D point *P* which is projected into left camera 2D point p_1 has to lie somewhere on the line l_2 as point p_2 in the image of the second camera. The lines mark as *l* are epipolar lines and *e* stands for the camera epipoles.

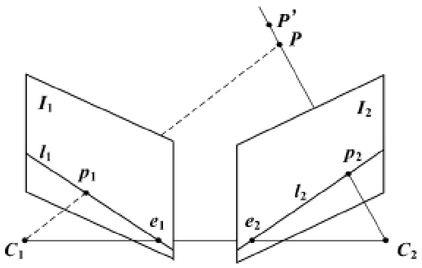

Figure 4. The epipolar geometry in camera stereo pair

This basic characteristic of epipolar geometry is described by the so called fundamental matrix *F*. Depending on the algorithm, but usually being able to compute the fundamental matrix there are at least 8 corresponding pairs of points needed. The elements *f* of matrix *F* can be calculated by using formula 2 for example.

$$\begin{bmatrix} x'x & x'y & x' & y'x & y'y & y' & x & y & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \quad (2)$$

The points *x* and *y* belong to the first camera and the points x' and y' are related to the second camera respectively.

The more suitable matrix to investigate the mutual position between the cameras in stereo pair is the essential matrix *E*. This matrix describes the relation between 3D points in stereo pair and can be determined by the following formula:

$$E = C^T \cdot F \cdot C \quad (3)$$

where *C* represents the already mentioned intrinsic camera matrix coming from the camera calibration.

2.5. Decomposition of essential matrix

The reason why the stating of essential matrix is convenient is that it can be decomposed into the rotation matrix *R* and the translation vector *t*, that can fully describe the relative movement from one camera to the another one in 3D Cartesian coordinates system.

To achieve this, usually SVD technique is used [11]. But the problem is that using this approach 4 possible combinations of rotation matrix and translation vector are found. The way how to find the right combination is to use linear triangulation and triangulate the arbitrary 2D point into the 3D space using combination of *R* matrix and *t* vector [12]. Only in one case the *Z* depth coordinate of triangulated point is positive in both camera views and that's the proper solution. The situation is shown in Fig. 5.

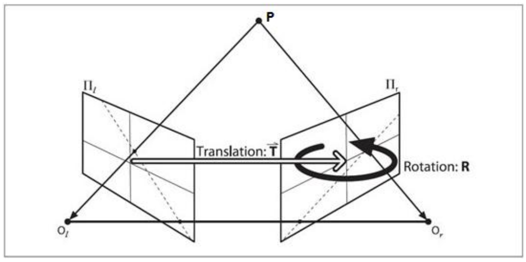

Figure 5. The relation between both stereo pair cameras

In this point the relation between the camera 3D coordinates and the so called world 3D coordinates are known as it is described by formula 4.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \cdot \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (4)$$

The elements indexed by subscript *W* are related to 3D world coordinates. This world coordinates system is common for both considered cameras. As an origin of world coordinates system is the 3D space of the first camera assumed.

It is important to realize, that due to the fact that the objects having different sizes and different scene distances can also look the same in image, so the translation only relative up to the scale can be obtained.

2.6. Image rectification

The image rectification is basically the transformation process used to project two or more images onto a common image space in such a way, that the epipolar lines become collinear and parallel to one of the axes, usually the horizontal one [13]. The advantage of this configuration is to make searching for the corresponding points in both images of stereo pair easier, as you can see in Fig. 6 [14].

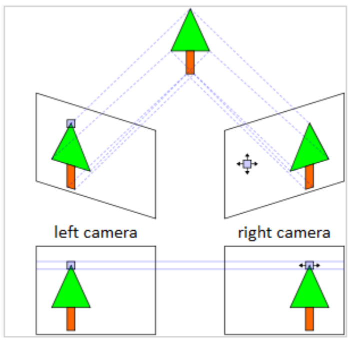

Figure 6. The rectification of camera stereo pair

In case of calibrated stereo pair it is possible to use knowledge of rotation matrix and translation vector and to find the rectification rotation matrix [15]. This matrix can be used to proper rotate both cameras to achieve a rectification configuration of stereo pair.

For our purposes it is not necessary to rectify the whole images, transformation of the detected corner points is quite enough.

2.7. Image disparity

Under the assumption of calibrated and rectified stereo pair, the disparity by definition measures the displacement of a point between two images in pixels [16]. The situation is illustrated in Fig. 7.

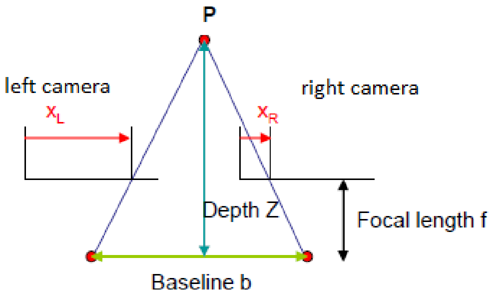

Figure 7. The situation in rectified stereo pair

The symbols x_L and x_R denote the horizontal coordinates of common point in image planes of both cameras.
Mathematically this configuration can be described by the following formula, where the relation between the disparity value d and the depth of a point Z or baseline B is attached.

$$Z = \frac{f \cdot B}{d} \tag{5}$$

The sign f stands for focal length in pixels and B is a baseline or in the other worlds the looking displacement between both cameras in SI units. It is obvious, that the points placed farther from the camera have a lower disparity value and vice versa.
It is important to notice, that due to the mentioned fact the object having different sizes and scene distances can look similarly in image, we need to know some intrinsic prior information like focal length f and point depth Z in rectified 3D coordinate system to be able to determine the real displacement in physical units.

3. EXPERIMENTAL TESTS

For the purpose to demonstrate the possibility of mutual camera distance estimation, the camera stereo pair configurations using calibrated camera were prepared. All particular presented steps were followed to determine the position displacement of the same camera. The both camera positions, which were mutually translated and rotated, give us the stereo pair configuration.
The camera calibration process is illustratively shown in Fig. 8. and the example of the camera stereo pair preparation is in Fig. 9.

Figure 8. The camera calibration

Figure 9. The stereo pair configuration

The detected points correspondingly matched are illustrated in Fig. 10.

Figure 10. The matched corner points in image stereo pair

For the easier detection of corner points the chessboard pattern was used. These points were used to determine the fundamental matrix (12 points chosen in total, 4 points in every chessboard) and during the next steps both sets of points were rectified. 9 of point pairs, 3 from every chessboard, were further investigated to find the displacement through disparity computation. Since the prior information is necessary, the real scene depth Z value in *cm* unit considers the rectified coordinate system was known for all these points.
To make the analysis more robust, three image resolutions were processed: 2560×1920, 1280×960 and 640×480 pixel size. Also the multiple mutual camera distances and image pairs were tested, specifically the 7 pairs for each 20, 30, 40 cm, 5 pair for 50, 60, 70 cm and 3 image pairs for 80, 90 and 100 cm.

4. EXPERIMENTAL RESULTS

The first results are related to procedure stability investigation. As this whole displacement estimation process is a set of theoretic mathematical algorithms and approaches, it could easily give us physically not possible results. That’s why we decided to use our own filter to recognize if the results are acceptable or not. The first criterion was the fact, that all the corresponding points (usually 144) triangulated using the found matrix R and the vector t had to have the positive depth Z coordinate. If they hadn’t, the tested stereo pair was denied. Another criterion was the difference between the found and the real camera displacement. Because the real camera shift as the ground truth was known and we got 9 results of the displacement from 9 points investigated through disparity, in case that at least one result was off more than 50% in the contrary to real one, again the tested stereo pair was denied.
These stability results are listed in Table 1. The number of all tested image stereo pairs for every resolution was 45, so according to the results, the described procedure has a success rate more than 50%.

Table 1. The procedures stability results

shift [cm]	*stability of procedure [pieces]*		
	2560×1920	*1280×960*	*640×480*
20	3	3	5
30	5	1	4
40	4	5	4
50	2	3	4
60	4	2	3
70	2	1	2
80	2	3	1
90	2	0	1
100	2	0	0
∑	**26** / 45	**18** / 45	**24** /45

The second comparison is about the relative errors of the found displacement. The errors were computing by the formula 6.

$$\delta_B = \frac{B' - B}{B} \cdot 100 \qquad (6)$$

The character B represents the real camera displacement (baseline) and B' is the obtained value. The values belonging to the particular shift and resolution were averaged and the results are displayed in graphs as in Fig. 11.

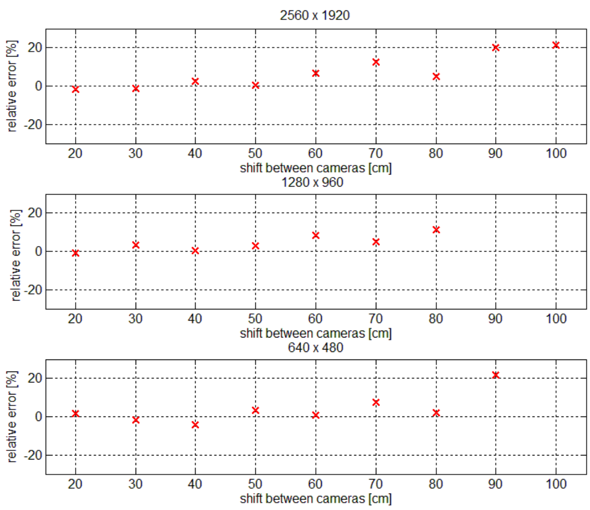

Figure 11. Averaged relative errors of found displacements

If the result is missing for specific configuration combination, it means that due to the stability failure was not obtained.

It is possible to see, that the whole procedure is more stable and produces better results if the camera shift is smaller.

5. CONCLUSION

This paper dealt with the determination of displacement between two cameras using the corresponding image stereo pair.

The whole procedure of camera shift estimation was discussed and all particular algorithm steps were explained.

To verify this approach an experimental test was performed following the considered theory. The experimental results were statistically analyzed and presented in the form of graphs and tables.

The first analysis was about the stability evaluation. Using the restricted filter to accept or dismiss the results, we found that the whole process had more than 50% of success rate. The results were listed in a table.

The second analysis dealt with relative errors of the found displacements. The corresponding results under the specific configuration were averaged and shown in a graph.

It is also important to notice, that as we found according to the obtained results, the whole described procedure gives us better results in case of stereo pair having small relative rotations and translation displacement. That's why the potential navigation using this kind of camera information turns to be also real time computation problem.

6. ACKNOWLEDGEMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0469-12.

REFERENCES

[1] D. C. Brown, "Close-Range Camera Calibration", Photogrammetric Eng. Vol. 37, pp. 855-866, 1971.
[2] G. Bradski, A. Kaehler, "Learning OpenCV", O'Reilly Media, First edition, pp. 381-384, 2008.
[3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision (ICIS), Seventh IEEE/ACIS International Conference on, pp. 91-110, 2004.
[4] A. Hast, J. Nysjo, A. Marchetti, "Optimal RANSAC – Towards a Repeatable Algorithm for Finding the Optimal Set", Journal of WSCG 21, pp. 21-30, 2013.
[5] S. Belongie, D. Kriegman, "Explanation of Homography Estimation", Department of Computer Science and Engineering, University of California, 2007.
[6] L. Sroba, R. Ravas, J. Grman, "Comparison of subpixel corner detection based on reprojection error criterion", 10th International Conference on Measurement, 2015.
[7] Z. Weixing, M. Changhua, X. Libing, L. Xincheng, "A fast and accurate algorithm for chessboard corner detection", CISP 2nd International Congress on, Image and Signal Processing, pp. 1-5, 2009.
[8] D. C. Brown, "Decentering the distortion of lenses", Photogrammetric Eng. Vol. 32, pp. 444-462, 1966.
[9] K. Riha, P. Hujka, "Epipolarni geometrie", Ustav telekomunikaci FEKT VUT v Brne, Elektrorevue, 2005.
[10] L. Jiangbo, C. Hua, L. Jian-Guang, L. Jiang, "An Epipolar Geometry-Based Fast Disparity Estimation Algorithm for Multiview Image and Video Coding", Circuit and Systems for Video Technology, IEEE Transaction on, pp. 737-750, 2007.
[11] W. Wang, H. Tsui, "A SVD decomposition of essential matrix with eight solutions for the relative positions of two perspective cameras", 15th International Conference on Pattern recognition, pp. 362-365, 2000.
[12] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, pp. 312-313, 2003.
[13] D. Oram, "Rectification for any epipolar geometry", British Mechine Vision Conference, pp. 653-662, 2001.
[14] G.Gerig, "Image Rectification (Stereo)", CS 6320 Spring, 2012.
[15] A. Fusiello, E. Trucco, A. Verri, "A compact algorithm for rectification of stereo pairs", Machine Vision and Application Volume 12, pp. 16-22, 1999.
[16] N. Navab, Ch. Unger, "Rectification and Disparity", Computer Aided Medical Procedures, Technical University of Munich.