CAMERA CALIBRATION BY REGISTRATION STEREO RECONSTRUCTION TO 3D MODEL

Klečka Jan

Doctoral Degree Programme (1), FEEC BUT

E-mail: xkleck01@stud.feec.vutbr.cz

Supervised by: Horák Karel

E-mail: horak@feec.vutbr.cz

Abstract: Paper aims at unusual way to camera calibration. The main idea is that by registration of uncalibrated stereo reconstruction to 3D model of the same scene is eliminated ambiguity of the reconstruction. The reason for this is that exact metric scene reconstruction from image pair can be understate as information equivalent to calibration of the source camera pair. Described principles were verified by experiment on real data and results are presented at the end of the paper.

Keywords: EEICT, camera calibration, stereo reconstruction, data registration

1. INTRODUCTION

Camera calibration is a process used to find a set of parameters for a mathematical model of camera which then represents relation between real world coordinate system and coordinate system of calibrated camera. Standard way to deal with this problem consists basically from three parts: Firstly is necessary to have an object with points which are easy-to-find in a camera picture and simultaneously these points have well-defined relative real world position. Secondly picture of the object is need to be taken, with the examined camera, so consequently correspondences between the camera coordinates system and the real world coordinates system can be obtain from it. And finally a set of camera model parameters which fit these correspondences is found by solving set of equations.

The method described by this paper essentially differs from the standard way. At first instead of the a priory defined object is used only a 3D model of some reasonably structured scene. To this model is then registered a stereo reconstruction of captured scene, so at least two pictures is needed. And at the end is transformation obtained by registration process used to acquire projection matrices of both input pictures.

An experiment on real data has been performed to verify correctness of proposed algorithm. The 3D model for the experiment consist from multiple scans of 2D laser scanner SICK LMS 111 with measurement plane oriented vertically and rotated around vertical axis in 65° range with 0.5° resolution. Pictures have been captured by hand held camera.

Figure 1: Input data of the experiment

2. UNCALIBRATED STEREO RECONSTRUCTION

Every picture taken by a camera is a projection of in general 3D scene in to a plane, so it's obvious that due to this mechanism is one dimension (usually called depth) lost. Stereo reconstruction is process of recovering the information about lost dimension from two pictures of the same scene, taken from slightly different viewpoint. Principle of this method is based on premise that every point in image can be present as a ray in the 3D space. So if projection of a 3D point can be detected in two different images then its space position is found as an intersection of their respective rays. However for a proper determination of position and orientation of a ray in space from image coordinates it is necessary to have calibrated cameras. But, as shown in [1], even without knowledge of calibration, restraints of epipolar geometry allows to obtain some reconstruction, nevertheless only up to projective transformation ambiguity. This process consist of three steps: Firstly fundamental matrix is need to be computed. Then it is necessary to find as much correspondence points as possible – usually for this step is preferred to use disparity map. Finally fundamental matrix is used to figure out pair of projection matrices witch can be used to protectively ambiguous reconstruction of every point correspondence through triangulation.

2.1. FUNDAMENTAL MATRIX

Fundamental matrix is the algebraic representation of the epipolar geometry [1]. If $\mathbf{x}$ and $\mathbf{x}'$ represent positions of corresponding points in homogenous coordinates then fundamental matrix $\mathbf{F}$ for these two images will satisfy equation for every correspondence:

$$\mathbf{x}'^{\mathrm{T}}\mathbf{F}\mathbf{x} = \mathbf{0} \tag{1}$$

Important properties of this matrix are: $\mathbf{F}$ is rank two matrix with seven degrees of freedom – is scale invariant and $\det(\mathbf{F}) = 0$, any point $\mathbf{x}$ in first image define on second image so-call epipolar line, on which correspondence to $\mathbf{x}$ can be found, as $l' = \mathbf{F}\mathbf{x}$, all epipolar lines cross each other in one point called epipole e (or e' in second image).

Computing of fundamental matrix can be done by several ways. On experimental data has been applied following method: Firstly correspondence point has been searched using SIFT [3] feature detector and descriptor. Then outliers in found correspondence has been removed by RANSAC algorithm which periodically compute $\hat{\mathbf{F}}$ using minimal seven point algorithm. And at last final $\mathbf{F}$ has been computed by linear optimization algorithm using all inliers followed by zeroing minimal singular value of linear criterion optimal matrix.

2.2. DISPARITY MAP

A disparity map can be presented as result of very dense correspondence search, usually so dense that disparity map has the same resolution as source images. If $\mathbf{x}$ and $\mathbf{x}'$ again represent positions of corresponding points, then related point in disparity map can be described as $\mathbf{D}(\mathbf{x}) = \mathbf{x}' - \mathbf{x}$.

Figure 2: Rectified images (left and middle), disparity map (right)

As mentioned before fundamental matrix constrains a correspondence for any $\mathbf{x}$ to be found on epipolar line l' and to make disparity map computation and representation more simple it is usual to rectify input images so their epipolar lines will become parallel to each other and to one of the axis (generally to the x axis). Then data in disparity map can be scalar because it represents difference only in one dimension.

To compute disparity map experiment data has been rectified and then algorithm described in [4] has been used. Rectified images and the grayscale representation of resulting disparity map is presented in the Fig 2.

2.3. TRIANGULATION

Triangulation in stereo reconstruction can be presented as a process of searching space point $\mathbf{X}$ from relative orientation of two rays which are respective to correspondence pair of image points $\mathbf{x}$ and $\mathbf{x}'$.

$$\mathbf{x} = \mathbf{P}\mathbf{X} \qquad \mathbf{x}' = \mathbf{P}'\mathbf{X} \tag{2}$$

Because true projection matrices are unknown in this case, it's instead used any canonical pair which fits to fundamental matrix, namely:

$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] \qquad \mathbf{P}' = \left[[\mathbf{e}']_x \mathbf{F} + \mathbf{e}'\mathbf{v}^{\mathrm{T}} \mid \lambda\mathbf{e}'\right] \tag{3}$$

Where $[\mathbf{e}']_x$ is skew symmetric matrix which satisfy $[\mathbf{e}']_x \mathbf{a} = \mathbf{e}' \times \mathbf{a}$, $\mathbf{v}$ is arbitrary vector and λ non-zero scalar. The experiment has been done with $\mathbf{v} = \mathbf{0}$ and $\lambda = 1$.

When projection matrices are defined, there is several way how to solve triangulation problem in this task. The experiment has been done using linear homogenous method:

$$\begin{bmatrix} x\mathbf{p}^{3\mathrm{T}} - \mathbf{p}^{1\mathrm{T}} \\ y\mathbf{p}^{3\mathrm{T}} - \mathbf{p}^{2\mathrm{T}} \\ x'\mathbf{p}'^{3\mathrm{T}} - \mathbf{p}'^{1\mathrm{T}} \\ y'\mathbf{p}'^{3\mathrm{T}} - \mathbf{p}'^{1\mathrm{T}} \end{bmatrix} \mathbf{X} = \mathbf{0} \tag{4}$$

Where $\mathbf{p}^{i\mathrm{T}}$ is row of $\mathbf{P}$ and $\mathbf{x} = (x, y)$.

3. DATA REGISTRATION

A purpose of data registration is to transform two data sets in such way that they will become spatially consistent. In described algorithm has been this concept used to registration the uncalibrated stereo reconstruction to 3D model of reconstructed scene because of the projective ambiguity is removed from such reconstruction. For realization of this registration has been used concept of ICP (Iterative Closest Point) algorithm described in [2]. However due to the fact that the ICP is a numerical method a guess of the solution has to be done at first.

The initial guess of searched projection transformation $\mathbf{H}_0$ has been got through several handpick reconstruction to 3D model correspondences. Transformation is then derived from these correspondences using homogenous least square method.

Then the found transformation is gradually getting precision by periodical appliance of following algorithm: Firstly the closest point in 3D model is find for every point of stereo reconstruction. Secondly some of the worst (the most distant) closest points correspondences are rejected as outliers. Then from remaining correspondences is randomly picked subset, because it turned out that full set of correspondences inliers is usually too large for effective processing. Finally transformation step $\mathbf{H}_s$ is calculated from correspondence subset and is added to so far found transformation $\mathbf{H}_i = \mathbf{H}_s \mathbf{H}_{i-1}$ where i is number of iteration.

In described experiment six correspondences has been handpick for initial guess calculation. For outliers removal 25% of most distant correspondence has been rejected. And 5000 samples has been randomly picked for transformation step calculation from inliers.

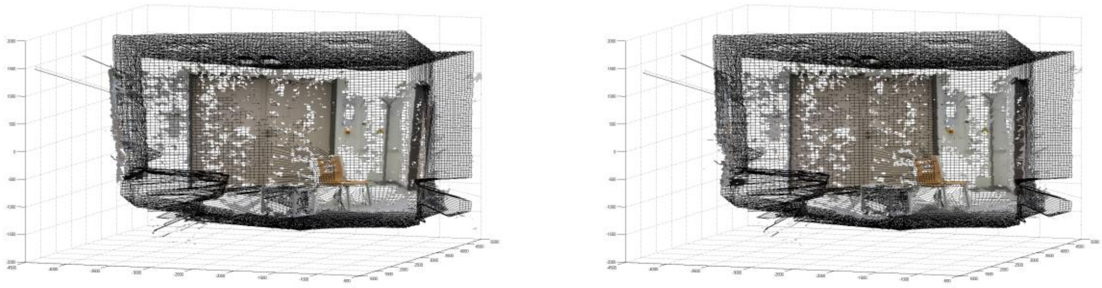

Figure 3: Initial guess of registration (left) and registration after fifteen iterations (right)

4. RESULTS

After stereo reconstruction has been registered to 3D model (in other words true stereo reconstruction represented by any point $\mathbf{X}_t = \mathbf{H}\mathbf{X}$ has been found), it is possible to use found projection transformation $\mathbf{H} = \mathbf{H}_{i_{\max}}$ to acquiring true projection matrix pair $\mathbf{P}_t, \mathbf{P}_t'$ from canonical pair $\mathbf{P}, \mathbf{P}'$ by solving following pair of equations:

$$\left.\begin{array}{r} \mathbf{x} = \mathbf{P}_t \mathbf{X}_t = \mathbf{P}\mathbf{X} \\ \mathbf{X}_t = \mathbf{H}\mathbf{X} \end{array}\right\} \mathbf{P}_t = \mathbf{P}\mathbf{H}^{-1} \tag{5}$$

Similar equation can be derive for $\mathbf{P}_t'$.

Accuracy of true projection matrix $\mathbf{P}_t$ computation is presented on fig. 4 by texturising original 3D model by projecting its points into the first image.

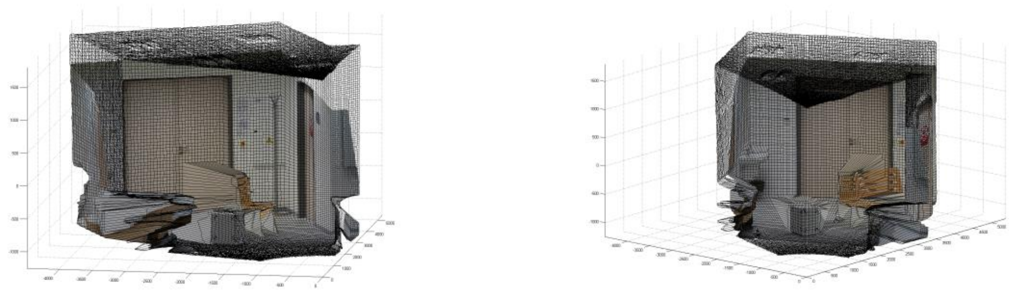

Figure 4: 3D model textured by projection into first image

For obtaining internal and external camera parameters have been true projection matrices decompose using QR decomposition into a form: $\mathbf{P}_t = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$ where $\mathbf{K}$ is upper triangular matrix, $\mathbf{R}$ orthogonal matrix of rotation transformation and $\mathbf{t}$ is a translation vector. However because standard QR decomposition decompose matrix into product of orthogonal matrix and upper triangular matrix (in this order) some modification has to be done. At first let's introduce rearrangement function which swaps matrix elements position by following rule: ${}^{i,j}a = {}^{n-j,m-i}b$ where ${}^{i,j}a$ is element of $m \times n$ sized matrix $\mathbf{A}$ on position $[i, j]$ and b is element of matrix $\mathbf{B} = \mathrm{re}(\mathbf{A})$. Secondly $\mathbf{P}_t$ has been split in to square matrix $\mathbf{P}_{KR}$ formed by its first three columns and vector $\mathbf{p}_{Kt}$ formed by its last column. Then the decomposition has been realized as:

$$\mathbf{QR}_r = \mathrm{qr}(\mathrm{re}(\mathbf{P}_{KR})) \rightarrow \begin{matrix} \mathbf{K} = \mathrm{re}(\mathbf{R}_r) \\ \mathbf{R} = \mathrm{re}(\mathbf{Q}) \\ \mathbf{t} = \mathbf{K}^{-1}\mathbf{p}_{Kt} \end{matrix} \tag{6}$$

Example of experimental data decomposition:

$$\mathbf{P}_t \Rightarrow \mathbf{K} = \begin{pmatrix} 2592 & 48 & 1448 \\ 0 & 2555 & 1395 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{R} \approx \begin{matrix} \alpha_x = -70° \\ \alpha_y = -20° \\ \alpha_z = -172° \end{matrix}, \mathbf{t} = \begin{pmatrix} -712 \\ -343 \\ 166 \end{pmatrix}$$

5. CONCLUSION

The described method shows itself to be able of acquiring camera projection matrices of stereo pair images. Useful application of this algorithm can be found in fusion 3D and 2D data. The weakest point is now in registration stereo reconstruction to 3D model. Especially correspondence detection for the first transformation guess proven to be difficult problem to be automated. Then also robustness and effectiveness of the iterative part of algorithm aren't satisfyingly high enough. Improvement of registration part will be subject of further research.

ACKNOWLEDGEMENT

The completion of this paper was made possible by grant No. FEKT-S-14-2429 - "The research of new control methods, measurement procedures and intelligent instruments in automation", and the related financial assistance was provided from the internal science fund of Brno University of Technology and also with assistance Competence Center realized by TACR (reg. number TE01020197).

REFERENCES

[1] Richard Hartley, and Andrew Zisserman. *Multiple view geometry in computer vision*, 2nd ed. Cambridge: Cambridge University Press, 2003. ISBN 05-215-4051-8.

[2] Paul J. Besl, and Neil D. McKay. (1992). "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. vol. 14, issue 2, pp. 239-256. Available from: http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/xpl/articleDetails.jsp?arnumber=121791 DOI: 10.1109/34.121791.

[3] David G. Lowe. (2004). "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* [online]. vol. 60, issue 2, pp. 91-110. Available from: http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf DOI: 10.1023/B:VISI.0000029664.99615.94

[4] Heiko Hirschmüller. (2008). "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. vol. 30, issue 2, pp. 328-341. Available from: http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/xpl/articleDetails.jsp?arnumber=4359315 DOI: 10.1109/TPAMI.2007.1166