Fusion of 3D Model and Uncalibrated Stereo Reconstruction

Jan Klecka, Karel Horak

Department of Control and Instrumentation, Brno University of Technology, Brno, 61600, Czech Republic
(klecka, horak)@feec.vutbr.cz

Abstract. Paper is focused on fusion of topological 3D model and color pictures of the same scene. It is describing method designed for such fusion, based on registration of uncalibrated stereo reconstruction to 3D model. This registration removes the reconstruction ambiguity, thereby makes possible acquiring projection matrices of source cameras. Projection matrices are then used for mapping 3D model into images and coloring its points, by this process visibility has to be check to covered points will not be colored. Real data experiment has been realized and the results are presented at the end of the paper.

Keywords: Data fusion, stereo reconstruction, data registration

1 Introduction

Fusion of an uncolored 3D data and color images is a process which results into creation of colored 3D model of the scene is interesting area for research today, because many different 3D sensors has become publicly available in past few years and data from this type of sensors may not contain enough information for further processing or comfortable observing. Our research in this area is specified on minimal a priory information solution. Examples of possible use of our method can be more robust localization in SLAM algorithms or just an easy virtualization of scene.

Standard way to approach this problem of this type assumes fixed set camera-3D sensor is available in order to calibration can be find before usage. The calibration is usually acquired using easy-to-detect object, with well-defined size. The object is placed, generally for a multiple times, in fields of view of both sensors from which 3D sensor coordinates system to camera coordinates system correspondences are find. These correspondences are then used to express mathematic formula which represent relations between their coordinates systems – the calibration. With this knowledge any point from 3D sensor scan can be mapped into camera image.

adfa, p. 1, 2011.

2 Algorithm outline

As proposed before, we wanted to develop method which lowers the standard ways a priory information assumption, specifically the need of calibration. Our method consist of three steps: Firstly the scene is reconstructed up to projective ambiguity from picture pair. Secondly the uncalibrated reconstruction is registered to 3D scan of the scene, which removes the ambiguity and allows to obtain projection matrices of both pictures. Finally these matrices are used to projection points of 3D scan into input pictures in order to acquire information about their color.

An experiment on real data has been performed to verify correctness of proposed algorithm. The 3D model for the experiment consist of 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.

Fig. 1. Experiment input data: Hand-held camera images (left and center) 3D scan of scene (right)

3 Uncalibrated stereo reconstruction

Every picture taken by any 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 viewpoints. 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.

3.1 Fundamental matrix

Fundamental matrix is the algebraic representation of the epipolar geometry [1]. If $\mathbf{x} = \lambda(x \quad y \quad 1)^T$ 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 defined up to scale and $\det(\mathbf{F}) = 0$, any point $\mathbf{x}$ in first image defines 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. Our experiment has been realized using following method: Firstly correspondence point has been searched using SIFT [2] 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.

3.2 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 $\mathbf{D}(\mathbf{x}) = \mathbf{x}' - \mathbf{x}$.

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 semi block matching algorithm described in [3] has been used. Rectified images and the grayscale representation of resulting disparity map is presented in the Fig. 2.

Fig. 2. Rectified images (left and middle), disparity map (right)

3.3 Triangulation

Triangulation in stereo reconstruction can be presented as a process of searching space point $\mathbf{X}$ from its projections in two different images. These projections can be found as 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 our case, we use any canonical pair instead which fits to fundamental matrix, namely:

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

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 nonzero 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 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 $\mathbf{P}$.

4 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 we using this concept 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 we decided to use idea of ICP (Iterative Closest Point) algorithm described in [4]. 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 $\mathbf{H}_0$ of searched projection transformation $\mathbf{H}$ has been got through several handpick reconstruction to 3D model correspondences. Transformation is then derived from these correspondences using homogenous method of least squares.

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 $\mathbf{H}_i$ is calculated from correspondence subset and is added to so far found transformation.

In our implementation, six correspondences have been handpicked 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.

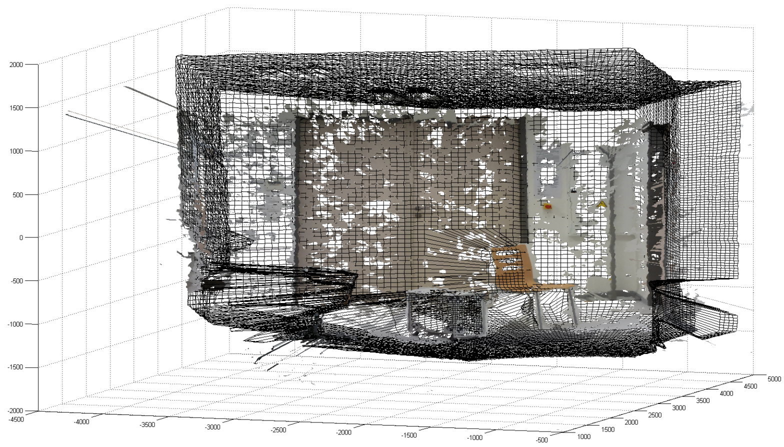

Fig. 3. Registration after 15 iterations

5 Coloring 3D model

Registration stereo reconstruction to 3D model is informational equivalent of a priory camera pair calibration, so relations between images and 3D model coordinates systems are known now and the last problem, which we dealt in this section, is how this knowledge can be used to merge information in images and 3D model. However, because following approach is greatly dependent on format of 3D model it is multiple ways to solve this problem. We describe method of our implementation used in our experiment. This part of algorithm is dividable to three parts: Firstly we acquire 'true' projection matrices. Secondly every point of 3D model is check if it's visible on images. And finally to every point which is visible at least in one image is assigned color.

5.1 Projection matrices

For proper explaining process of recovering 'true' projection matrices let us define: canonical projection matrix pair $\mathbf{P}, \mathbf{P}'$ used to obtain uncalibrated reconstruction $\mathbf{X}$ from images $\mathbf{x}, \mathbf{x}'$, projection transformation $\mathbf{H}$ acquired by registration part, 'true' stereo reconstruction $\mathbf{X}_t = \mathbf{H}\mathbf{X}$ and 'true' projection matrix pair $\mathbf{P}_t, \mathbf{P}_t'$.

Relation between 'true' and canonical projection matrices is then mathematically defined as:

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

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

We verified correctness of acquired matrices by projecting every point of 3D model into both images, thereby we can assigned color to this points by interpolation and after that visually check alignment. One of this way colored model can be seen in Fig. 4 – cause of illustration purposes presented model has been generated from 10 times denser 3D model (obtained by linear interpolation).

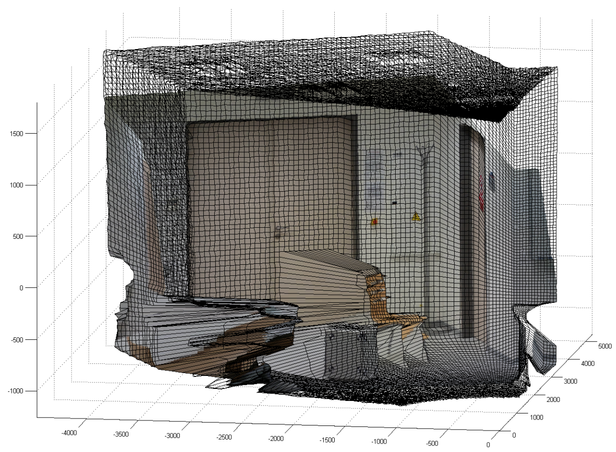

Fig. 4. Texturized 3D model using one image

5.2 Visibility check

Because 3D sensor and source camera of input images have generally different fields of view, 3D model contains some points which should not be projected into image. This points can be categorized into two groups: points which are projected outside the borders of image and points which would be projected into image but from camera viewpoint are covered behind some closer surface.

Dealing with first group is very easy – it is sufficient to check if coordinates of projected point are inside image borders, but dealing with second group desire more complex approach: Firstly surfaces have to be defined, because we had no information which scanned points form surface we assumed that every neighboring points creating surface. Then we created depth map by interpolating distance from center of projection of projected 3D points into pixels of image. During this process we segmenting points by checking if respective pixels aren't occupied by closer surface.

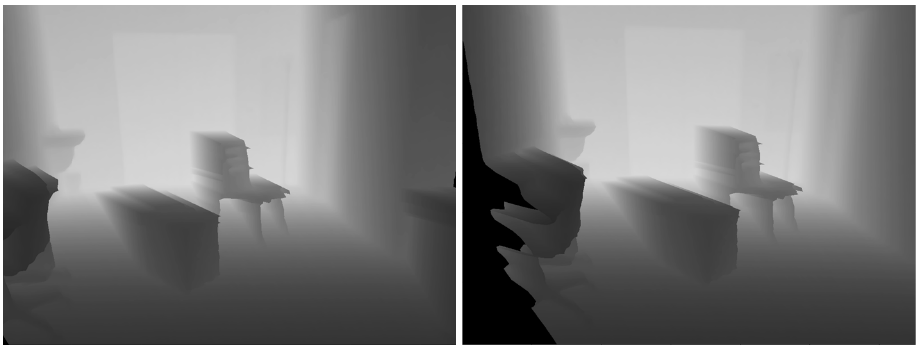

Fig. 5. Depth maps of input images

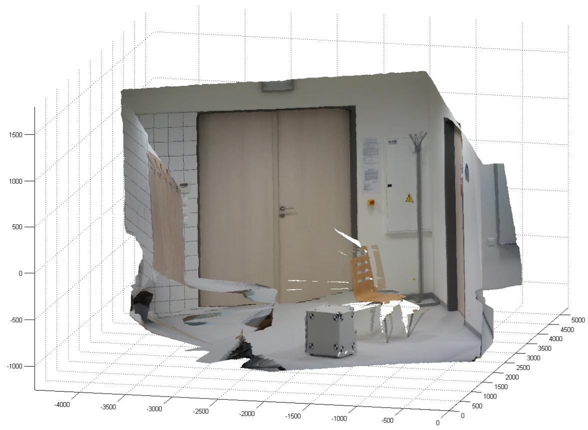

Fig. 6. The resulting fusion of input data

5.3 Color fusion

After visibility check 3D points can be classified into three groups: Points which are not visible in any image, points which are visible only in one image and points which are visible in both images. Dealing with first two groups is very strait forward – to the first group is not assigned any color and to the second group is assigned color from respective image. However, to the points from the last group can be assigned color from both images. Generally it is possible to use this fact for noise reduction through the use of appropriate combination colors from both images. Because in our experiment we dealt with static scene we assigning to this group mean from two respective colors. Results of this process can be seen in Fig. 6.

6 Conclusion

In this paper we describe method for coloring 3D model without need of a priory calibration. Main idea of our method lies in registration stereo reconstruction to colored model. From experiments result we can see that method algorithm is in general correct but alignment of data 3D profile and assigned texture is not perfect (focus e.g. on the chair in Fig. 6). The misalignment is cause by imperfect registration and by nonlinear distortion of the image source camera, which is not compensated in any step. Subject of our further research will be improvement of registration part of algorithm, especially initial guess part, because for usefulness in applications like e.g. SLAM should this part be automated.

Acknowledgments. 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. David G. Lowe.: Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision. vol. 60, issue 2, pp. 91-110 (2004).
3. Heiko Hirschmüller.: Stereo processing by semiglobal matching and mutual information. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 30, issue 2, pp. 328-341 (2008).
4. Paul J. Besl, and Neil D. McKay.: A method for registration of 3-D shapes In: IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 14, issue 2, pp. 239-256 (1992)