Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Automatic Epipolar Geometry Recovery Using Two Images

Rogério Yugo Takimoto [*,1] **André Challella das Neves** [*,2]
Thiago de Castro Martins [*,3] **Fábio Kawaoka Takase** [**,4]
Marcos de Sales Guerra Tsuzuki [*,5]

[*] *Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil.*
Mechatronics and Mechanical Systems Engineering Department
Computational Geometry Laboratory

[**] *Mind Tecnologia e Conhecimento, São Paulo, Brazil*

Abstract: The analysis and recovery of the epipolar geometry is a crucial step to perform a 3D reconstruction of a scene. This work uses two uncalibrated images as input to compute the epipolar geometry of a scene. This is done in two steps: 1. automatically feature points extraction and 2. feature points mapping determination. The feature points from the two images are automatically extracted through the SIFT algorithm. The ICP algorithm is used to compute an initial correspondence among the feature points by comparing their associated information. A novel robust mapping determination algorithm is proposed to speed up the matching process while the accuracy is maintained. The main idea is that the order in which three visible feature points in the 3D are seen must be the same independently of the camera position. The Delaunay triangulation creates coherently oriented triangles from the obtained inliers. Inliers that defined non coherently triangles are removed. The convex hull of the Delaunay triangulation is used to determined a new set of 8 points and a new set of inliers is determined. The proposed algorithm was tested and showed to be robust.

Keywords: Epipolar geometry, 3D reconstruction, affine transformation, Delaunay triangulation.

1. INTRODUCTION

The estimation of 3D information is a very important problem in computer vision. At present, there are two main approaches to accomplish this task. The first approach is based on a previous camera calibration. So that, the imaging sensor model that relates 3D object points to their 2D projections on the image is known (Ito, 1991). Calibration cannot be used in active systems due to its lack of flexibility, as optical and geometrical characteristics of the cameras might change dynamically. The second approach is based on computing either the epipolar geometry between both imaging sensors (Hartley and Zisserman, 2000) or an Euclidean reconstruction (Hartley, 1993). An application of scene reconstruction using epipolar geometry was first published by Longuet-Higgins (1981).

The human vision system can easily distinguish a 3D object from the background in an image due to some previous knowledge like colors, texture, shadow and geometric context. Thus, the objective of 3D reconstruction is to automatically extract useful information from images in a similar way as the human visual system.

Recovering epipolar geometry between uncalibrated cameras usually is realized in three steps: determination of feature points, finding reliable corresponding feature points, and estimation of the epipolar geometry. In this work, the Scale Invariant Feature Transform (SIFT) algorithm proposed by Lowe (2004) is used to determine the feature points from both images. The epipolar geometry is determined by matching feature points in both images. This is very difficult task, and it is generally accepted that incorrect matches cannot be avoided in the first stage of the process.

The well known standard RANdom SAmple Consensus (RANSAC) proposed by Fischler and Bolles (1981) does not model the matching process, it is a black box that generates several random tentative correspondences. Several robust estimation algorithms have been proposed to overcome this problem: adaptive real time RANSAC (Raguram et al., 2008), MLESAC (Torr and Zisserman, 2000), PROSAC (Chum and Matas, 2005), and others. Those algorithms try to remove mismatches created by repetitive patterns, occlusions and noise.

[1] e-mail: takimotoyugo@gmail.com. This author was supported by CNPq (grant 500801/2009-6).
[2] e-mail: andrechalella@gmail.com.
[3] e-mail: thiago@usp.br. This author was partially by FAPESP (grant 2009/14699-0).
[4] e-mail: fktakase@gmail.com.
[5] e-mail: mtsuzuki@usp.br. This author was partially supported by CNPq (grant 304258/2007-5).

In this paper, the computational cost is diminished by grouping the feature points in triangles and their topological orientation is determined. It is assumed that the orientation of three visible feature points must be the same independently of how they are seen. The triangles are consistently created using the Delaunay triangulation algorithm and the winged-edge data structure (Baumgart, 1972). The epipolar geometry is determined using at least 8 pairs of feature points. Then, the proposed algorithm will create iteratively and semi-randomly 8 pairs of consistently oriented feature points to determine the epipolar geometry, the solution is given by the candidate subset that maximizes the number of consistent points and minimizes the residual.

This paper is structured as follows. Section 2 explains the feature points detection and mapping. Section 3 explains the fundamental matrix evaluation and the proposed algorithm is in section 4. Section 5 presents some results and the conclusions are in section 6.

2. FEATURE POINTS MAPPING

The first step in the determination of the epipolar geometry is the determination of feature points and their matching. The feature points are determined using the SIFT algorithm (Lowe, 1999, 2001, 2004) that is a robust method to extract and describe feature points. This approach transforms an image into a large collection of local feature vectors, each of which is invariant to image translation, scaling and rotation, and partially invariant to illumination changes and affine or 3D projection. The steps are described as follow:

2.1 Scale-Space Extrema Detection

The feature points are detected using a cascade filtering approach that can identify locations in image scale space that are invariant with respect to image translation, scaling, rotation and are minimally affected by noise and small distortions. Lindeberg (1994) has shown that under some assumptions on scale invariance, the Gaussian kernel and its derivatives are the only possible smoothing kernels for scale space analysis. To achieve rotation invariance and a high level of efficiency, Lowe (2004) has chosen to select key locations at maxima and minima of a difference of Gaussian function applied in scale space. The scale space of an image is defined as a function, $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$, that is produced from the convolution of a variable scale Gaussian, $G(x, y, \sigma)$, with an input image, $I(x, y)$.

To identify the stable keypoint locations in the scale space, the difference-of-Gaussian (DoG) convolved with the image $D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$ will be used. The DoG will be computed from the difference of two nearby scales separated by a constant multiplicative factor k.

The DoG provides a close approximation to the scale-normalized Laplacian of Gaussian $\sigma^2\nabla^2 G$. As shown by Lindeberg (1994) and by Mikolajczyk (2002), the normalization of the Laplacian with the factor σ^2 is required for true scale invariance and the maximum and the minimum of $\sigma^2\nabla^2 G$ is the most stable image features compared to a range of other possible image functions, such as the gradient, Hessian, or Harris corner function (Mikolajczyk et al., 2005).

2.2 Keypoint Location Determination

The local maxima and minima of $D(x, y, \sigma)$ is found by comparing each sample point to its 8 neighbors in the current image and 9 neighbors in the scale above and below. A sample point is selected as a key point candidate only if it is larger than all of these 26 neighbors or smaller than all of them. After finding a keypoint candidate, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge, to be rejected.

The low contrast criteria is not sufficient to reject the keypoints because the DoG function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise (Lowe, 2004).

2.3 Orientation Assignement

Local extrema detected in DoG scale-space are called keypoints after the operations of improving positioning accuracy and eliminating low-contrast points. To determine the keypoint orientation, an orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint. The peaks in the orientation histogram correspond to the dominant directions of local gradients. The highest peak in the histogram is detected, and then any other local peak that is within 80% of the highest peak is used to also create a keypoint with that orientation. Therefore, for locations with multiple peaks of similar magnitude, there will be multiple keypoints created at the same location and scale but different orientations (Lowe, 2004).

2.4 Keypoints Descriptor

Once an image location, scale, and orientation have been assigned to each keypoint it is possible to impose a 2D coordinate system to describe the local image region and provide invariance with respect to these parameters. The next step is to compute a descriptor for the local image region that is distinct yet invariant to additional variations such as change in illumination and 3D pose.

2.5 Keypoints Mapping

After the SIFT algorithm has been applied to the images, it is possible to determine the correspondence among the keypoints. The mapping happens by comparing each keypoint descriptor that is formed from a 4×4 array of histograms with 8 orientation bins in each. Therefore, each keypoint has a feature vector with $4 \times 4 \times 8 = 128$ elements.

The match assignment can be done by computing a similarity metric between descriptors. Commonly used similarity metrics includes sum of square differences, sum of absolute differences, normalized correlation, and Mahalanobis distance metrics. Tests performed with the nearest-neighbor method (Muja and Lowe, 2009) showed that the

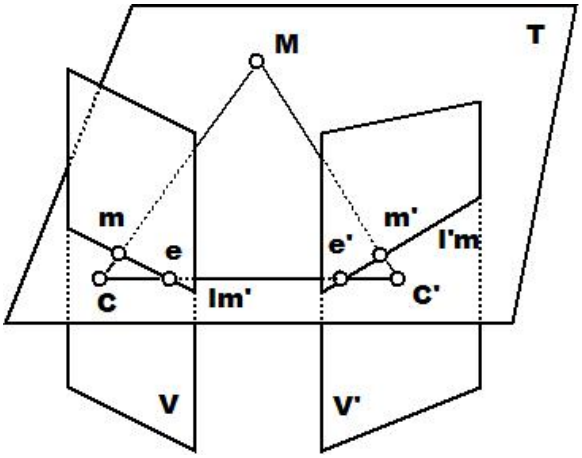


Fig. 1. Epipolar Geometry of two images. Here we have the two camera centres C and C', the 3D point M e the two projection planes V and V'.

number of corresponding points with false matches increases as the displacement between the images increases.

Therefore, the Iterative Closest Point (ICP) was used. The ICP algorithm was originally introduced to register 3D data sets by Chen and Medioni (1992) and Besl and McKay (1992), it is also used with 2D data sets to register images mainly on medical applications. The ICP algorithm iteratively performs two operations until convergence: the data matching and the transformation estimation to align the data sets. The ICP algorithm takes two data sets as input representing salient points of a reference image I_1 and a target image I_2. The goal is to compute the parameters of the transformation matrix $\boldsymbol{M}$ that best aligns the transformed points. For 2D Euclidean transformation the parameters are the rotation matrix $\boldsymbol{R}$ and the translation vector $t = (t_x, t_y)$.

3. FUNDAMENTAL MATRIX EVALUATION FROM FEATURE POINTS PAIRS

The epipolar geometry exists between any two-camera systems. Consider the case of two cameras as shown in Fig. 1. Let C and C' be the optical centers of the first and second cameras, respectively. Given a point m in the first image, its corresponding point in the second image is constrained to lie on a line called the epipolar line of m, denoted by l'_m. The line l'_m is the intersection of the plane T, defined by m, C and C' (known as the epipolar plane), with the second image plane V'. This is because image point m may correspond to an arbitrary point on the semi-line CM (M may be at infinity) and that the projection of CM on V' is the line l'_m. Furthermore, one observes that all epipolar lines of the points in the first image pass through a common point e', which is called the epipole.

If m (a point in V) and m' (a point in V') correspond to a single physical point M in space, then m, m', C and C' must lie in a single plane. This is the well-known coplanarity constraint in solving motion problems when the intrinsic parameters of the cameras are known (Longuet-Higgins, 1981).

The mathematical expression that relates corresponding points in two different images is (Luong et al., 1993):

$$m^T \cdot \mathbf{F} \cdot m' = 0 \tag{1}$$

where $\mathbf{F}$ is a 3×3 matrix called fundamental matrix, m is an image point and m' is its corresponding point in the other image. In particular, for $m = (x, y, 1)^T$ and $m' = (x', y', 1)^T$, each pair of corresponding feature points give us a linear equation.

Considering n pairs of corresponding points and denoting f the 9 elements vector made up of the entries of $\mathbf{F}$, it is possible to obtain a set of linear equations of the form:

$$\mathbf{A} \cdot f = 0 \tag{2}$$

where $\mathbf{A}$ represents a homogeneous set of equations, and f can only be determined up to scale. For a solution to exist, the matrix $\mathbf{A}$ must have rank at most 8, in this case the solution is unique and can be determined by the generator of the right null-space of $\mathbf{A}$.

The fundamental matrix can be recovered using the normalized 8 points algorithm and performs as well as the best iterative algorithms (Karlstroem and Takase, 2005). The 8 points algorithm solution involves the solution of a set of linear equations where the linear least squares minimization can be used.

4. PROPOSED ALGORITHM

As mentioned before, several approaches were published to overcome the RANSAC inability in modeling the matching process. In this work, it is considered that some topological characteristics must be preserved in both images. The order of three feature points that are present in both images must be coherent (see Fig. 2). It is relevant to mention that three collinear feature points can not be used in the determination of the fundamental matrix.

Before explaining the proposed algorithm, we will present the plane model (Mäntylä, 1988) that can represent coherently oriented polygons in the plane (see Fig. 3). Plane models are used to represent B-Rep solid models in the 3D space and Voronoi diagrams in the 2D space. Plane models can be constructively created using Euler Operators and can be represented by the winged edge data structure.

The proposed algorithm is shown in Fig. 4. In the first step, 8 correspondent feature points are selected and a planar graph is created. Initially, three coherently oriented feature point pairs are selected. Next, one feature point pair is selected at a time and a new triangle is added to the structure. If the new triangle pair is not coherently oriented then a new feature point pair is picked. Fig. 3 shows the created structure in the first image.

In the second step, the fundamental matrix is evaluated and the inliers and outliers are identified. The inliers are the data which approximately can be fitted to the epipolar geometry model, while the outliers are the data which cannot be fitted. If the maximum number of iterations has been reached, a new set of 8 correspondent feature points is selected.

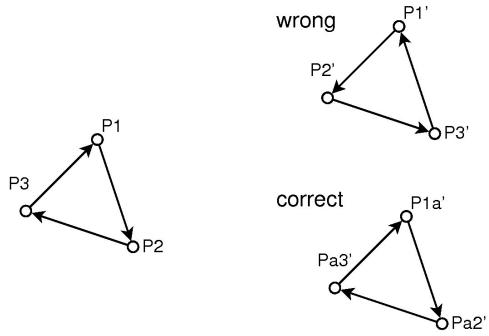


Fig. 2. Given that triangle $\Delta = [P_1, P_2, P_3]$ is in the first image, and that triangles $\Delta = [P'_1, P'_2, P'_3]$ and $\Delta = [Pa'_1, Pa'_2, Pa'_3]$ are in the second image. Consider two possible correspondences between the feature points from both images: $C_1 = [P_1 \leftrightarrow P'_1, P_2 \leftrightarrow P'_2, P_3 \leftrightarrow P'_3]$ and $C_2 = [P_1 \leftrightarrow Pa'_1, P_2 \leftrightarrow Pa'_2, P_3 \leftrightarrow Pa'_3]$. The correspondence defined by C_1 is wrong because both triangles have incoherently orientations. On other hand, C_2 is correct.

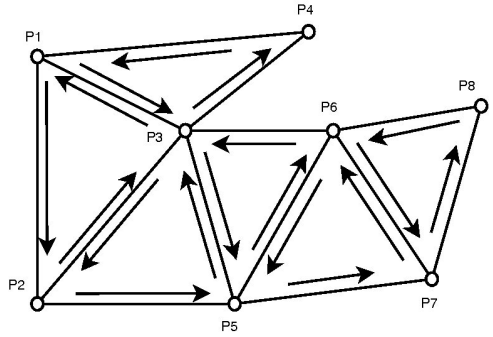


Fig. 3. A plane model represents 8 feature points selected from the first image. All triangles are coherently oriented. The feature points were selected according to the crescent index order.

In the third step, it is create the Delaunay triangulation (de Berg et al., 2008) in the first image using the determined inliers (see Fig. 5). This way, it is possible to check all inliers with a minimum set of non intersecting coherently oriented triangles in a structure similar to the one shown in Fig. 3. Now, it is possible to verify if all corresponding triangles are coherently oriented in both images. As all triangles are coherently oriented in the first image, it is necessary to exclusively check the orientation in the second image. Incoherently triangles are marked, and incoherent feature points are removed. Fig. 6 shows two possible situations and how they are processed.

A threshold is defined for determining whether feature point pairs are inliers or outliers, this threshold represents the maximum algebraic distance for which a pair is declared inlier. The inlier rate defined by

$$p_{in} = \frac{n_{in}}{N} \tag{3}$$

where n_{in} is the number of inliers and N is the total number of determined feature point pairs. If p_{in} is greater than a threshold then the algorithm is stopped. Other stop criteria similar to the one proposed in PROSAC (Chum

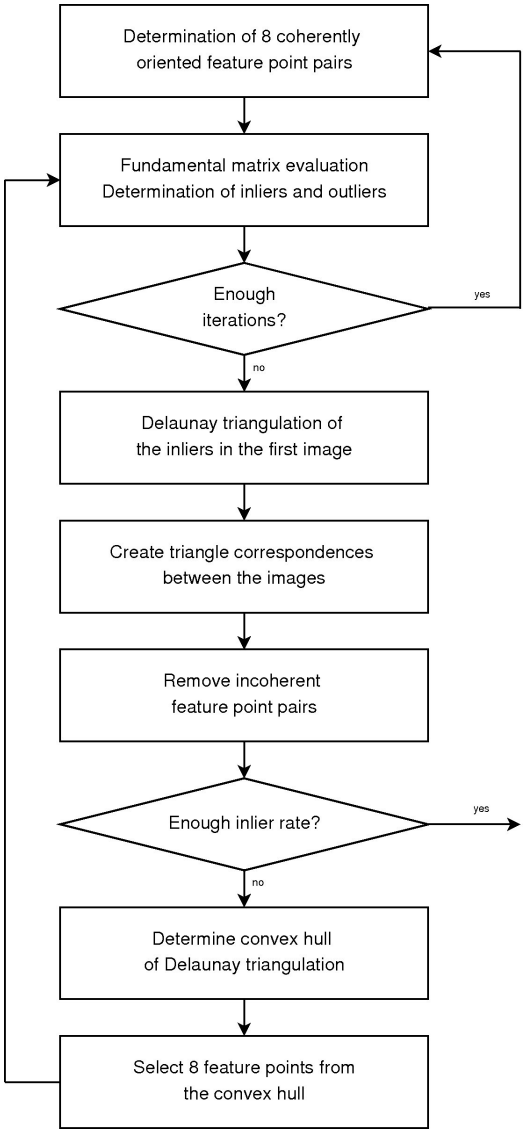


Fig. 4. The proposed algorithm.

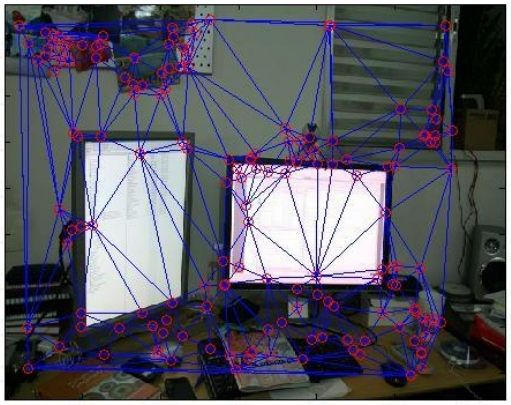


Fig. 5. Delaunay triangulation created with the inliers determined in the first image.

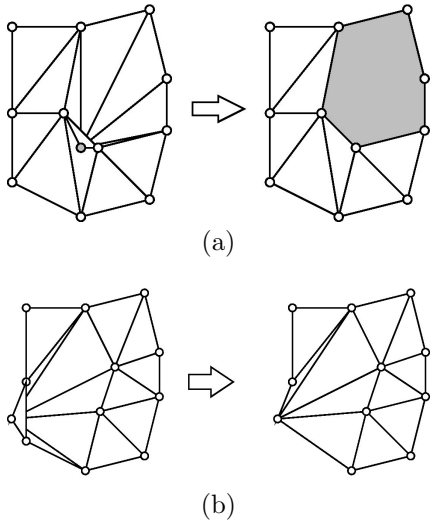


Fig. 6. It is shown two possible situations where a false inlier is present. (a) In this case the false inlier is internal to a triangle. After removing the false inlier, a polygon is created. (b) In this case the false inlier is at the boundary of the Delaunay triangulation. The inlier with less adjacent triangles is removed.

and Matas, 2005) or MLESAC (Torr and Zisserman, 2000) can be used here.

The convex hull of the remaining Delaunay triangulation is determined. By using feature points preferentially from the convex hull boundary the resulting fundamental matrix maps the far away points accurately and possibly the number of inliers increases.

5. RESULTS

The epipolar geometry automatic recover involved several algorithms implementation: the SIFT used in the feature points determination, the ICP used fo define an initial correspondence between the feature from both imagens, the proposed algorithm used in the correspondence points refining and the fundamental matrix determination.

Initially the SIFT algorithm were applied in both images to obtain an initial set of corresponding points. The result of the SIFT algorithm can be viewed in the Figs. 7.(a) and 7.(b), the first image shows a correctly recognized corresponding point and the second image shows a fail in the recognition.

After the initial matching of corresponding feature points, the proposed algorithm is executed and the fundamental matrix is estimated to the bigger set of corresponding points. Fig. 8.(a) shows the outliers and the inliers at the beginning and at the end of the iteration. Through the analysis of Fig. 8.(a), it is possible to notice that the number of outliers decreases as the number of iteration increases showing that a robust matching was performed.

At the end of the proposed algorithm, it is possible to recover the epipolar geometry of the two images. Fig. 8.(b) shows the epipolar lines of the first and the second image. Although the nearest point algorithm failed to matching the corresponding points when the displacement between the two images increase, the ICP algorithm could retrieve a

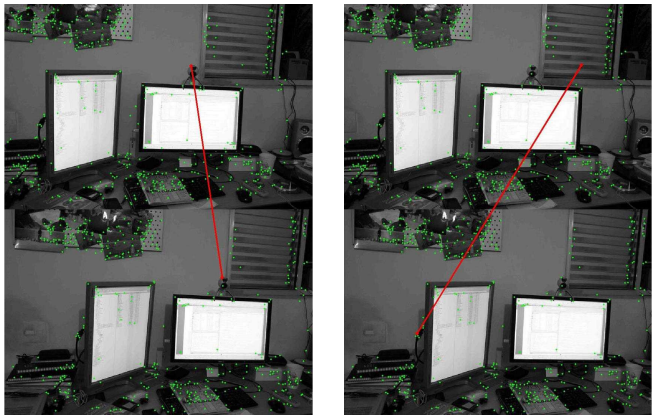


Fig. 7. (a) Images with a corresponding point recognized (initial matching) showing a correct matching. (b) Images with a mismatched corresponding point (initial matching) showing a failed matching.

Fig. 8. Inliers (yellow) and outliers (blue) at the beginning (above) and at the end of iteration (below). (b) Epipolar lines of first and second images are shown.

considerable number of corresponding points in the initial match and could recover the epipolar lines (Fig. 8.(b)).

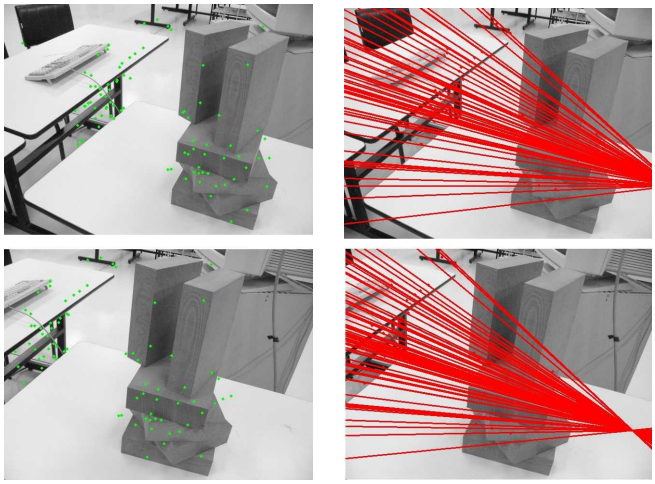
Figures 9.(a) and 10.(a) show the initial matching of corresponding feature points in two additional examples. At the end of the proposed algorithm, the epipolar geometry was recovered as can be seen in Figs. 9.(b) and 10.(b).

6. CONCLUSION

From the results it is possible to verify that the SIFT algorithm can identify a large number of feature points. Tests showed that the performance of the ICP algorithm was better than the nearest-neighbor method in the initial matching of feature points.

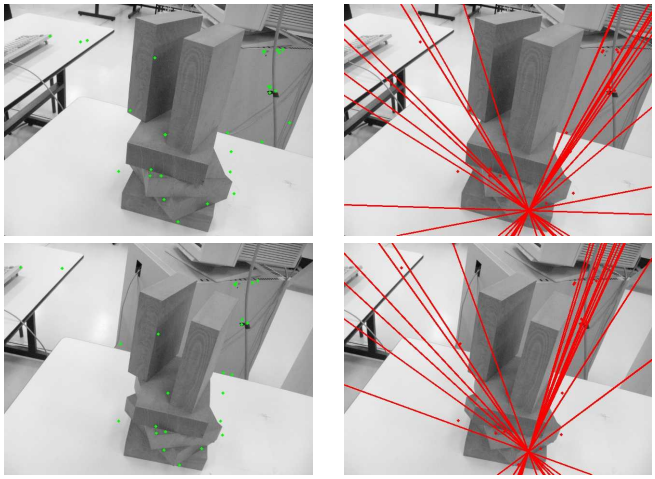
The proposed method based on triangle coherence maintenance among images showed to be effective by eliminating mismatches. The epipolar geometry was determined with fewer iterations when compared with the RANSAC algorithm.

For the future work, it is necessary to quantify the obtained results through the error analysis. Moreover, it is



(a) (b)

Fig. 9. (a) Images with a corresponding point recognized (initial matching). (b) Epipolar lines of first and second images are shown.



(a) (b)

Fig. 10. (a) Images with a corresponding point recognized (initial matching). (b) Epipolar lines of first and second images are shown.

necessary to recover the camera parameters to generate the points cloud in the space and to perform a 3D reconstruction.

REFERENCES

Baumgart, B.G. (1972). Winged edge polyhedron representation. Technical report, Stanford University, Stanford, CA, USA.

Besl, P. and McKay, N. (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 239–256.

Chen, Y. and Medioni, G. (1992). Object modeling by registration of multiple range images. *Image and Vision Computing*, 10, 145–155.

Chum, O. and Matas, J. (2005). Matching with prosac - progressive sample consensus. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, 220–226.

de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Springer-Verlag.

Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24, 381–395.

Hartley, R. (1994). I. projective reconstruction and invariants from multiple images. *PAMI*, 16, 1036–1041.

Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hartley, R.I. (1993). Euclidean reconstruction from uncalibrated views. In *Second European Workshop on Applications of Invariance in Computer Vision*, 237–257.

Ito, M. (1991). Robot vision modelling-camera modelling and camera calibration. *Advanced Robotics*, 5, 321–335.

Karlstroem, A. and Takase, F.K. (2005). Motion structure analysis and error quantification - an epipolar geometry based approach. In *International Congress of Mechanical Engineering*. Ouro Preto, Brazil.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21, 224–270.

Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293, 133–135.

Lowe, D.G. (1999). Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision*, 1150–1157. Corfu, Greece.

Lowe, D.G. (2001). Local feature view clustering for 3-d object recognition. In *Proc. 2001 IEEE Conf. Computer Vision Pattern Recognition*, 682.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computational Vision*, 60, 91–110.

Luong, Q.T., Deriche, R., Faugeras, O., and Papadopoulo, T. (1993). On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report 1894, INRIA.

Mäntylä, M. (1988). *An Introduction to Solid Modeling*. Computer Science Press.

Mikolajczyk, K. (2002). *Detection of local features invariant to affine transformations*. Ph.d. thesis, Institut National Polytechnique de Grenoble, France.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65, 43–73.

Muja, M. and Lowe, D.G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Applications*, 331–340.

Raguram, R., michael Frahm, J., and Pollefeys, M. (2008). A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In *Proceedings of the 10th European Conference on Computer Vision: Part II*.

Torr, P.H.S. and Zisserman, A. (2000). Mlesac: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1), 138–156.