THREE DIMENSIONAL SCENE RECONSTRUCTION USING EPIPOLAR GEOMETRY

Rogério Yugo Takimoto, takimotoyugo@gmail.com
André Chalella Neves, andrechalella@gmail.com
Thiago de Castro Martins, thiago@usp.br
Computational Geometry Laboratory, Escola Politécnica da USP

Fábio Kawaoka Takase, fktakase@gmail.com
Mind Tecnologia e Conhecimento, São Paulo, Brazil

Marcos de Sales Guerra Tsuzuki, mtsuzuki@usp.br
Computational Geometry Laboratory, Escola Politécnica da USP

Abstract. *The analysis and recovery of the epipolar geometry is an important step to perform 3D scene reconstruction. In this work, the epipolar geometry of a scene was automatically recovered using two uncalibrated images. This is usually realized in three steps: automatically determination of feature points, feature points mapping determination and estimation of the epipolar geometry. The ICP (iterative closest point) algorithm is used to compute an initial correspondence among the feature points by comparing their associated information. In this initial mapping a proportion of these correspondences are mismatches, therefore it is necessary to refine the correspondence between the points. One approach, to eliminate mismatched points from our data, is the use of the RANSAC (RANdom SAmple Consensus) algorithm. Since the RANSAC method is nondeterministic and treats all samples equally regardless of their quality, several approaches were published to overcome the RANSAC inability in modeling the matching process. A novel robust mapping determination algorithm is proposed here 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. This technique guides the process of choosing the set of points used to determine the epipolar geometry and to reject the inconsistent matches. The results of the proposed algorithm show that the use of the Delaunay triangulation improves the accuracy of the epipolar geometry determination when the number of outliers increases.*

Keywords: *Epipolar Geometry, Three Dimensional Reconstruction, Affine Transformation.*

1. INTRODUCTION

The 3D information recover using 2D images is very useful in the 3D reconstruction process, it can be used in perception sensors of the desired environment and generally can be done through two approaches: laser scanning based and computer vision based. The computer vision based approach can be classified in two main methods. The first method depends on a previous camera calibration to determine its position relative to a reference coordinate system (Ito, 1991). In a traditional camera calibration, a calibration object is used as spatial reference object. The main advantage of this approach is that it can give better accuracy. However, the main disadvantage occurs in active systems where the optical and geometrical characteristics of the cameras might change dynamically depending on the imaging scene and camera motion. The second approach uses two uncalibrated images with epipolar geometry. The algorithm has as input exclusively the pair of images, with no other a priori information is required; and the output is the estimated fundamental matrix together with a set of interest points in correspondence. The epipolar geometry recover using two uncalibrated images must deal with the point detection and matching problem. It is possible to enumerate several interest point detectors algorithms in the computer vision area, such as the Moravec (1977) detector, SUSAN (Smith and Brady, 1995) detector, Harris (Harris and Stephens, 1988) detector and FAST (Rosten and Drummond, 2006) detector. For the point matching, information surrounded interest points are used and some point descriptors are available, such as the SIFT (Scale Invariant Feature Transform) operator (Lowe, 1999) and SURF (Speed Up Robust Features) operator (Bay *et al.*, 2008) . Comparisons between the operators showed that each operator has its own advantages and limitations and that no single algorithm has been accepted as the best choice for all applications as discussed in (Tissainayagam and Suter, 2004; Schmid *et al.*, 2000).

In this work, the epipolar geometry recovery using two uncalibrated cameras can be achieved in three steps. First, the SIFT algorithm is applied in both images to determine the keypoints and a set of initial keypoints correspondence (Lowe, 1999). Since it is expected that a proportion of these correspondences are in fact mismatches, the RANSAC (Fischler and Bolles, 1981) algorithm will be used to estimate true corresponding keypoints and the epipolar geometry. The epipolar geometry is determined using at least 7 pairs of corresponding keypoints. It is known that the RANSAC 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.

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 the metric reconstruction. Section 6 presents some results and the conclusions are in section 7.

2. KEYPOINTS MAPPING

The correspondence between keypoints is realized in two steps: initially they are determined and a matching algorithm is applied. The keypoints are determined using the SIFT algorithm (Lowe, 1999, 2001, 2004). The SIFT algorithm is a robust method used to extract and describe image keypoints. It can also extract the feature vectors around the extracted keypoints. This algorithm transforms an image in a local feature vectors invariant to image scaling, translation, rotation and partially invariant to image illumination and affine transform to 3D projection. Using this algorithm it is possible to recover the image features and get a reliable corresponding points mapping between different views of an object or scene. The steps of the SIFT algorithm are described in the following.

2.1 Scale-Space Extrema Detection

The keypoints are detected using a cascade filtering approach that uses efficient algorithms to identify candidate locations that are then examined in further detail. The invariance to changes in scale is determined by searching for stable features across all possible scales (Lowe, 2004). The scale-space is constructed using a Gaussian function (Lindeberg, 1994) $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$ where $*$ is the convolution operator in x and y with the Gaussian function

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma} e^{\frac{-x^2+y^2}{2\sigma^2}}.$$

To identify stable keypoints in the scale-space, the difference-of-Gaussian (DoG) convolved with the image $D(x, y, \sigma)$ is used. The DoG is 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 *et al.* (2005), 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 Keypoints Localization

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 only if it is larger than all of these 26 neighbors or smaller than all of them. After finding a keypoint candidate by comparing a pixel to its neighbors, 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 has a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise. Since a poorly defined peak in the DoG function has a large principal curvature across the edge but a small curvature in the perpendicular direction, the objective in this step is to eliminate the keypoints that have high ratio between the principal curvatures (Lowe, 2004).

2.3 Keypoints Orientation Assignment

In this step, a consistent orientation is assigned to each keypoint based on local image properties. This orientation is recovered from the gradient magnitude and orientation. The scale of the keypoint is used to select the Gaussian smoothed image, $L(x, y)$ with the closest scale, so that all computations are performed in a scale invariant manner. For each image

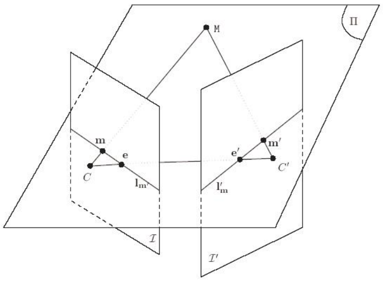


Figure 1. Epipolar geometry of two images. There are two camera centres C and C', the 3D point M and the two projection planes V and V'.

sample, $L(x, y)$, at this scale, the gradient magnitude $m(x, y) = \sqrt{L_x^2 + L_y^2}$ and the orientation

$$\theta(x, y) = \arctan\left(\frac{L_x}{L_y}\right)$$

where $L_x = L(x+1, y) - L(x-1, y)$ and $L_y = L(x, y+1) - L(x, y-1)$, are precomputed. 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 create a keypoint with that orientation. Therefore, for locations with multiple peaks of similar magnitude, there are 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. A traditional approach to compute the keypoint descriptor is to sample the local image intensities around the keypoint at appropriate scale and matching with a normalized correlation measure. This has limitations as it is highly sensitive to changes in the image that cause misregistration of samples such as in affine or 3D pose variations and random noise interference. A better approach is to allow the gradient at a particular orientation and spatial frequency to shift locations over a small field rather than being precisely localized allowing for matching and recognition of 3D objects from a range of viewpoints. In this approach, it is necessary to sample the image gradients and orientations around the keypoint location by applying the scale of the keypoint to determine the level of Gaussian blur for the image. Orientation invariance is achieved by rotating the coordinates of the descriptor and the gradient orientation relative to the keypoint orientations.

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 number of false matches increase as the displacement between the images increases. Therefore the ICP was used, it is another approach for solving the correspondence problem, although 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 $\mathbf{M}$ that best aligns the transformed points. For 2D Euclidean transformation the parameters are the rotation angle φ and the translation vector $t = (t_x, t_y)$.

3. FUNDAMENTAL MATRIX EVALUATION FROM CORRESPONDING KEYPOINTS

The techniques described above provide an initial approximation for the keypoints mapping and must be refined. The epipolar geometry (Hartley and Zisserman, 2000) provides useful information to identify errors in the mapping and refine the initial keypoints mapping. The geometric entities involved in the epipolar geometry are the epipoles, the epipolar plane and the epipolar line. The focal point of the camera is the camera center. Each camera center projects onto a distinct point into the other camera's image plane. These two image points are called epipoles. The epipolar plane is the plane that contains the camera centers and one 3D point. The epipolar line is the intersection of an epipolar plane with the image plane. Note that all epipolar lines pass through the epipole and the projected point on the image plane. With this geometry, any point M in the 3D space forms with the camera centers a plane that intercepts the two images in a line that necessarily passes through the epipoles. Fig. 1 shows the epipolar geometry components. The mathematical expression that relates corresponding points in two different images is given by $m^T \cdot \mathbf{F} \cdot m' = 0$ where $\mathbf{F}$ is a 3×3 matrix called fundamental matrix, m is an image point and m' is the corresponding point in the other image (Luong *et al.*, 1993). In particular, for $m = (x, y, 1)^T$ and $m' = (x', y', 1)^T$, each pair of corresponding points gives a linear equation in terms of elements of $\mathbf{F} = [f_{ij}]$

$$\begin{aligned} x' \cdot x \cdot f_{11} + x' \cdot y \cdot f_{12} + x' \cdot f_{13} + \\ y' \cdot x \cdot f_{21} + y' \cdot y \cdot f_2 + x' \cdot f_{23} + \\ x \cdot f_{31} + y \cdot f_{32} + f_{33} = 0. \end{aligned}$$

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

$$\mathbf{A} \cdot f = 0 \tag{1}$$

where

$$\mathbf{A} = \begin{pmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{pmatrix}$$

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, 2007). The 8 points algorithm solution involves the solution of a set of linear equations where the linear least squares minimization can be used. The original algorithm was introduced by Longuet-Higgins (1981) and the improvement is given by the systematic normalization of the corresponding points. In this algorithm, a simple transformation (translation and scaling) on the coordinates of the corresponding points before formulating the linear equation $\mathbf{A} \cdot f = 0$ leads to an enormous improvement in the condition of the problem and hence of the stability of the result without adding complexity to the algorithm. The normalization method consists in translating all corresponding points so that the centroid of the points coincides with the origin of image then scaling all coordinates so that the mean distance of the points to the origin is equal to $\sqrt{2}$.

The normalization method improves the results accuracy, moreover, it provides invariance with respect to arbitrary choices of the scale and coordinate origin. This is because the normalization step undoes the effect of coordinate changes, by effectively choosing a canonical coordinate frame for the measurement data (Hartley and Zisserman, 2000). Hartley and Zisserman (2000) showed that the use of the normalization method is a necessary step to recover the fundamental matrix. If matrix $\mathbf{A}$ defined in (1) has rank 8, then the solution of f is unique and determined up to scale. If matrix $\mathbf{A}$ has rank 7, it is still possible to solve the linear system by making use of the singularity constraint because $\mathbf{F}$ has 7 degrees of freedom. The most important situation is when only 7 point correspondences are known, this leads to a 7×9 matrix, which generally has rank 7.

In this case, the solution to the equations $\mathbf{A} \cdot f = 0$ is a 2D space of the form $\alpha \mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2$, where α is a scalar variable. The matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ are obtained as the matrices corresponding to the generators f_1 and f_2 of the right null-space $\mathbf{A}$. The constraint that $det\mathbf{F} = 0$ (may be written as $det(\alpha \mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2) = 0$) leads to a cubic polynomial equation in α because $\mathbf{F}_1$ and $\mathbf{F}_2$ are known. This polynomial equation has one or three real solutions in α and the complex solutions are discarded (Hartley and Zisserman, 2000). Substituting α in the equation $\mathbf{F} = \alpha \mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2$ gives one or three possible solutions for the fundamental matrix.

The automatic selection of corresponding keypoints is in fact a noisy process. Since, it is always necessary to deal with imperfect selections, consequently the fundamental matrix estimation using the 8 points algorithm can lead to inaccurate estimations. In order to eliminate outliers, or mismatched points, from our data, the RANSAC algorithm (Fischler and Bolles, 1981) is used, a general robust estimator proposed by Fischler and Bolles (1981) that fits a model to the best points from the data sets by iteratively random sampling minimal data subsets (7 points). The RANSAC algorithm used the inliers and outliers concepts. The inliers of a model are the data which approximately can be fitted to a model, while the outliers of a model are the data which cannot be fitted to this model. The RANSAC algorithm requires a threshold for determining whether points are inliers or outliers, this threshold represents the maximum algebraic distance for which a point is declared inlier. Thus, this process is repeated a number of times to estimate the model in an optimal manner.

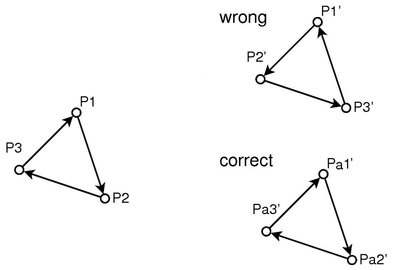


Figure 2. Given that triangle $\Delta = [P_1, P_2, P_3]$ is in the first image, and that triangles $\Delta = [P'_1, P'_2, P'_3]$ and $\Delta = [Pa'_1, Pa'_2, Pa'_3]$ are in the second image. Consider two possible correspondences between the feature points from both images: $C_1 = [P_1 \leftrightarrow P'_1, P_2 \leftrightarrow P'_2, P_3 \leftrightarrow P'_3]$ and $C_2 = [P_1 \leftrightarrow Pa'_1, P_2 \leftrightarrow Pa'_{a2}, P_3 \leftrightarrow Pa'_3]$. The correspondence defined by C_1 is wrong because both triangles have incoherently orientations. On other hand, C_2 is correct.

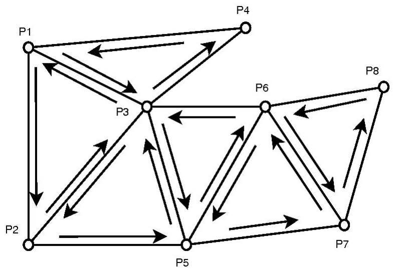


Figure 3. A plane model represents 8 feature points selected from the first image. All triangles are coherently oriented. The feature points were selected according to the crescent index order.

The epipolar geometry is used to refine the process of guided matching around the epipolar line. This geometric constraint restricts the search area and provides a weaker similarity threshold.

4. RANSAC IMPROVEMENT

The RANSAC algorithm has proven very successful for robust estimation, but having defined the robust negative log likelihood function as the quantity to be minimized it becomes apparent that RANSAC can be improved on (Torr and Zisserman, 2000). As mentioned before, several approaches were published to overcome the RANSAC inability in modeling the matching process. In this work, an algorithm to guide the points choosing process is used to decrease the numbers os false inliers, that is the numbers of outliers detected as inliers. It will be 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. 13). 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 (Mantyla, 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.

In the third step, it is create the Delaunay triangulation (Mark de Berg and Overmars, 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, that is the triangle orientation must be the same 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. Usually a threshold that leads an inlier to be incorrectly rejected only 5% of the time is usually chosen (Hartley and Zisserman, 2000). The inlier rate

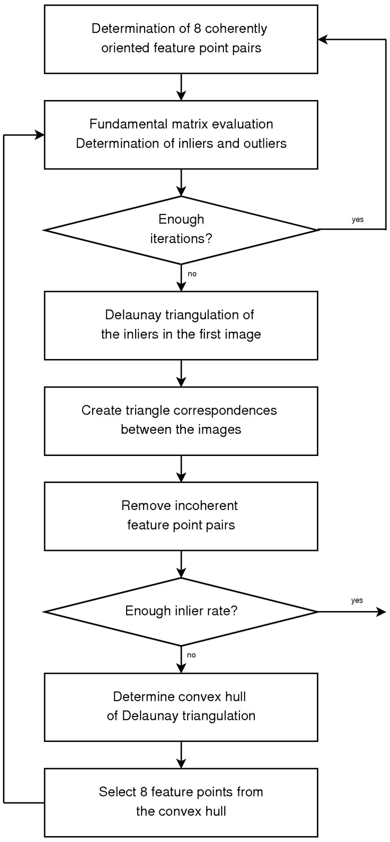

Determination of 8 coherently oriented feature point pairs
Fundamental matrix evaluation
Determination of inliers and outliers
Enough iterations?
yes
no
Delaunay triangulation of the inliers in the first image
Create triangle correspondences between the images
Remove incoherent feature point pairs
Enough inlier rate?
yes
no
Determine convex hull of Delaunay triangulation
Select 8 feature points from the convex hull

Figure 4. The proposed algorithm.

defined by $p_{in} = \frac{n_{in}}{N}$ 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 value then the algorithm is stopped. Other stop criteria similar to the one proposed in PROSAC (Chum and Matas, 2005) or MLESAC (Torr and Zisserman, 2000) can be used here. A popular stopping criterion in a RANSAC like algorithm is

$$I_t = \frac{log(1-p)}{log(1-p_{in}^s)} \approx \frac{log(1-p)}{p_{in}^s}$$

where s is the size of the random sample, I_t is the number of iterations, p_{in} is the inlier rate, and p is the required probability (Fischler and Bolles, 1981; Torr, 1995). 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.

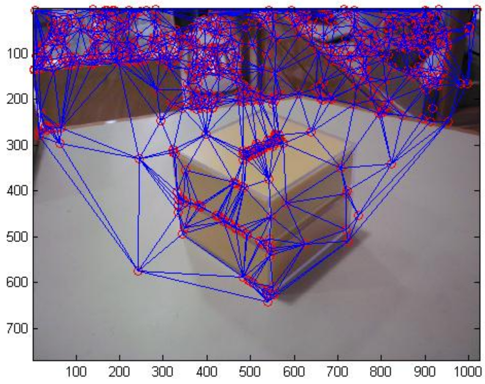
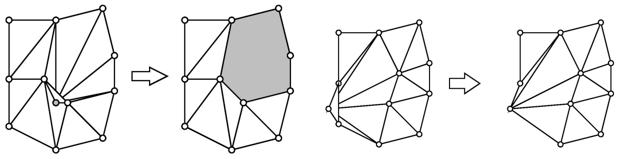


Figure 5. Delaunay triangulation created with the inliers determined in the first image.



(a) (b)

Figure 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.

5. METRIC RECONSTRUCTION

One of the most significant properties of the fundamental matrix **F** is that it may be used to determine the camera matrices of two views. The reconstruction searches the camera matrices **P** and **P′**, as well as the 3D points M_i such that $m_i = \mathbf{P} \cdot M_i$ and $m'_i = \mathbf{P}' \cdot M_i$ for a set of correspondents points between two images $m_i \leftrightarrow m'_i$, where the camera calibration $(\mathbf{P}, \mathbf{P}')$, the position and the orientation of the 3D points M_i are known. Even two images obtained by the same camera can have different camera matrices.

If there are sufficiently many point correspondences to allow the fundamental matrix to be computed uniquely, then the scene may be reconstructed up to a projective ambiguity. This is a very significant result, and one of the major achievements of the uncalibrated approach. The ambiguity in the reconstruction may be reduced if additional information is supplied about the cameras or scene. The correspondence condition $x^T \cdot \mathbf{F} \cdot m' = 0$ is a projective relationship, depending only on projective coordinates in the image. Thus, the image relationship is projectively invariant: under a projective transformation of the image coordinates $\widehat{m} = \mathbf{H} \cdot m$, $\widehat{m'} = \mathbf{H}' \cdot m'$, with $\widehat{\mathbf{F}'} = \mathbf{H}'^{-T} \cdot \mathbf{F} \cdot \mathbf{H}^{-1}$ the corresponding rank 2 fundamental matrix.

Similarly, **F** only depends on projective properties of the cameras **P** and **P′**. The camera matrix relates 3D measurements to 2D measurements and so depends on both the image coordinate frame and the choice of world coordinate frame, but the fundamental matrix **F** is unchanged by a 3D projective transformation. More precisely, if **H** is a 4×4 matrix representing a projective 3D transformation, then the fundamental matrices corresponding to the pairs of camera matrices $(\mathbf{P}; \mathbf{P}')$ and $(\mathbf{P} \cdot \mathbf{H}; \mathbf{P}' \cdot \mathbf{H})$ are the same. Thus, although a pair of camera matrices $(\mathbf{P}; \mathbf{P}')$ uniquely determine a fundamental matrix **F**, the converse is not true (Hartley and Zisserman, 2000). A pair of camera matrices **P** and **P′** corresponding to the fundamental matrix **F** can be computed using $\mathbf{P} = [\ \mathbf{I} \mid 0]$ and $\mathbf{P}' = [[e']_x \cdot \mathbf{F} \mid e']$ where **I** is the 3×3 identity matrix, 0 is a null vector and e' is the epipole such that $e'^T \cdot \mathbf{F} = 0$. One of the simplest methods to obtain the original 3D coordinates of a point projected on two images is the linear triangulation method. From the camera matrices **P** and **P′** and a pair of correspondents points $m_i \leftrightarrow m'_i$ that satisfy the epipolar constraint, $m^T \cdot \mathbf{F} \cdot m' = 0$. This constraint may be interpreted geometrically in terms of the rays in space corresponding to the two image points. This means that the two rays back-projected from image points m and m'; lie in a common epipolar plane, that is, a plane passing through the two camera centers. Since the two rays lie in a plane, they will intersect in some point. This point M projects via the two cameras to the points m and m' in the two images.

In each image we have a measurement $m = \mathbf{P} \cdot M$, $m' = \mathbf{P}' \cdot M$, and these equations can be combined to form $\mathbf{A} \cdot M = 0$, which is linear in M. First, the homogeneous scale factor is eliminated by a cross-product to give three equations for each image point, two of them linearly independent. For the first image, $m' \cdot (\mathbf{P} \cdot M) = 0$ and

$$\begin{aligned} x(p_3^T \cdot M) - (p_1^T \cdot M) = 0 \\ y(p_3^T \cdot M) - (p_2^T \cdot M) = 0 \\ x(p_2^T \cdot M) - y(p_1^T \cdot M) = 0 \end{aligned}$$

where p_i^T are the rows of **P**. An equation of the form $\mathbf{A} \cdot X = 0$ can then be composed as

$$\mathbf{A} \cdot M = \begin{bmatrix} x(p_3^T \cdot M) - (p_1^T \cdot M) \\ y(p_3^T \cdot M) - (p_2^T \cdot M) \\ x'(p_3^{\prime T} \cdot M) - (p_1^{\prime T} \cdot M) \\ y'(p_3^{\prime T} \cdot M) - (p_2^{\prime T} \cdot M) \end{bmatrix} = 0.$$

This is a redundant set of equations and the solution is determined only up to scale. The solution is given by the unit singular vector corresponding to the smallest singular value of the singular value decomposition of **A**. This simple method is not projective-invariant and needs an additional reference to determine the scale factor and the correct solution. This can be achieved using some ground control points, that is points with known 3D locations in a Euclidean world frame.

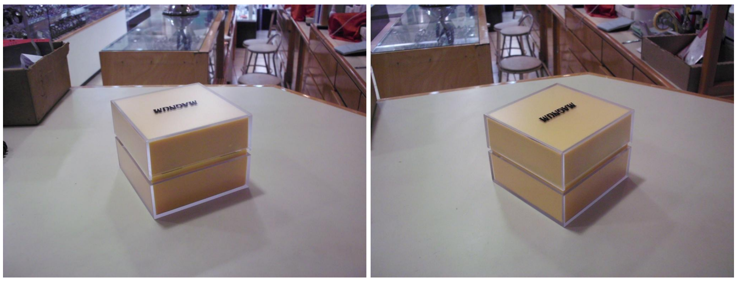


Figure 7. Images used in the epipolar geometry recover process.

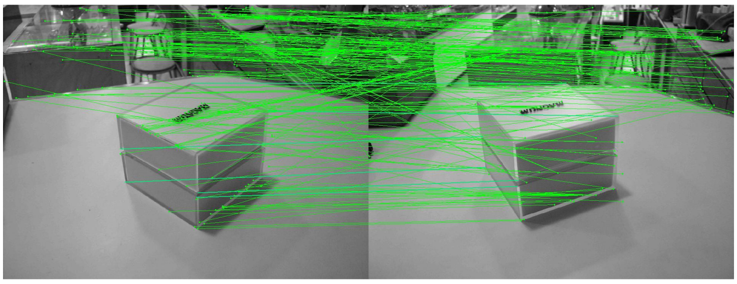


Figure 8. Correspondence lines obtained in the initial match.

6. RESULTS

The epipolar geometry automatic recover involved two algorithms implementation: the SIFT algorithm used in the point correspondence process and the RANSAC algorithm used in the correspondence points refining and the fundamental matrix recover. From the epipolar geometry recovered with this two algorithms it was possible to correctly recover the world points from an object in the scene. Fig. 7 shows the images used in the comparison, the images were captured moving the same camera in a translation movement.

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 Fig. 8. Since this initial match brings some false positives, it is necessary to refine the solution. In order to eliminate the false positives, the RANSAC algorithm is executed and the fundamental matrix is estimated to the bigger set of corresponding points. At the end of the RANSAC algorithm, it is possible to recover the epipolar geometry from the two images. Since the proposed algorithm uses the RANSAC in its core, the proposed algorithm is robust and increases the precision because it eliminates false positives that can be detected by RANSAC. Figs. 9 and 10 show the correspondents points using the original RANSAC algorithm discussed in (Hartley and Zisserman, 2000) and the proposed algorithm respectively. It is possible to observe that the original algorithm cannot completely eliminate false inliers. The proposed algorithm successfully eliminated all false inliers. Fig. 11 shows the epipolar lines recovered in the process from the fundamental matrix using the proposed algorithm. The time processing associated with the proposed algorithm decreased proportionally associated with the number of false inliers. When no false inlier is present then both algorithms have similar time processing. From the fundamental matrix obtained with the algorithm and from the knowledge of some points in the 3D world, it is possible to evaluate the 3D position. The 3D points recovered in this process generate the point clouds that can be used to model the surface of a 3D object. Figs. 12 and 13 show 3D world points of the same object from different views.

7. CONCLUSION

From the results it was possible to verify that the use of the SIFT and the RANSAC algorithm can be used to recover the points in the 3D world and the proposed algorithm improves the quality of the points used in the estimation of the fundamental matrix in the presence of a great number of outliers. The results showed that it is possible to identify the desired object from the generated point clouds. Moreover, the analysis of the coordinate system showed that the metric reconstruction of the object is correct and it was successfully recovered.

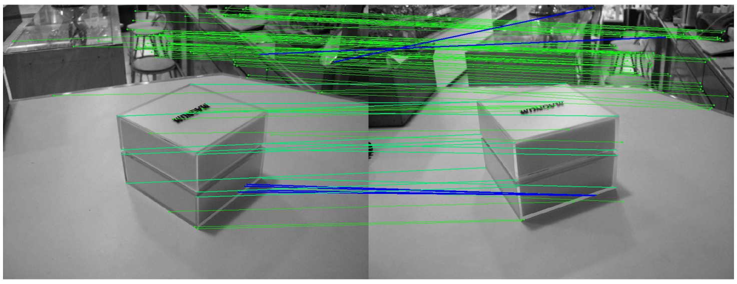


Figure 9. Correspondence lines obtained using RANSAC (false inliers detected by RANSAC in blue).

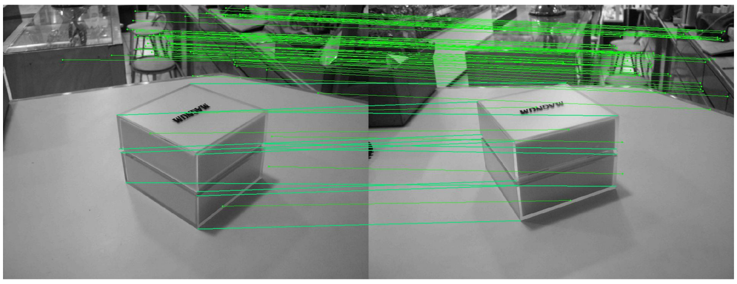


Figure 10. Correspondence lines obtained using RANSAC and the Delaunay triangulation.

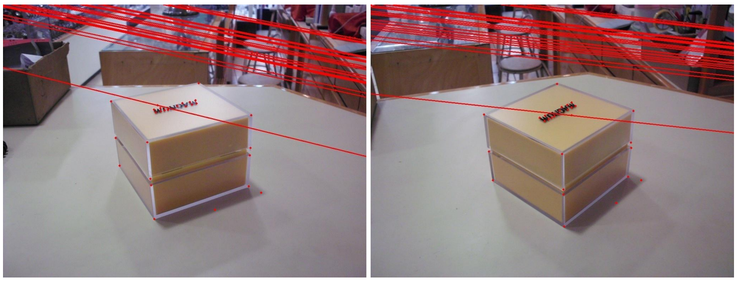


Figure 11. Epipolar lines obtained from fudamental matrix.

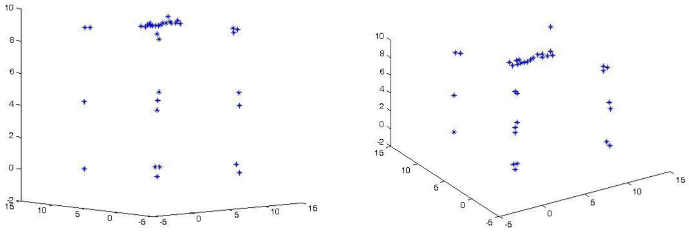


Figure 12. 3D world points of the same object.

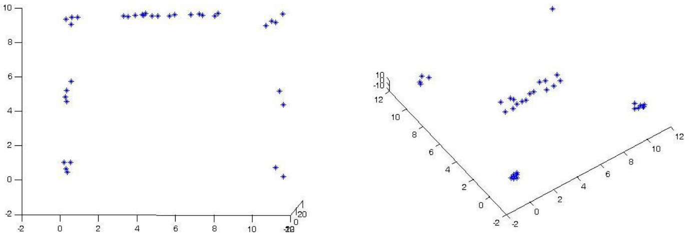


Figure 13. Side and top view of the same object.

8. ACKNOWLEDGEMENTS

RY Takimoto was supported by CNPq (Grant 500.801/2009–6), TC Martins was supported by FAPESP (Grant 2009/14699–0) and MSG Tsuzuki was partially supported by the CNPq (Grants 304.258/2007–5 and 309.570/2010–7).

9. REFERENCES

Baumgart, B.G., 1972. "Winged edge polyhedron representation." Tech Rep, Stanford Univ, Stanford, CA, USA.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L.V., 2008. "Surf: Speeded up robust features". *Comput Vis Image Und*, Vol. 110, pp. 346–359.

Besl, P. and McKay, N., 1992. "A method for registration of 3D shapes." *IEEE T Pattern Anal Mach Intell*, Vol. 14, pp. 239–256.

Chen, Y. and Medioni, G., 1992. "Object modeling by registration of multiple range images." *Image Vision Comput*, Vol. 10, pp. 145–155.

Chum, O. and Matas, J., 2005. "Matching with prosac - progressive sample consensus". In *Proc IEEE Comput Soc Conf Comput Vision Pattern Recog*. pp. 220–226.

Fischler, M.A. and Bolles, R.C., 1981. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". *CACM*, Vol. 24, pp. 381–395.

Harris, C. and Stephens, M., 1988. "A combined corner and edge detector". In *Proc 4th Alvey Vision Conf*. pp. 147–151.

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

Ito, M., 1991. "Robot vision modelling-camera modelling and camera calibration". *Adv Robotics*, Vol. 5, pp. 321–335.

Karlstroem, A., 2007. *Estimação de Posição e Quantificação de Erro Utilizando Geometria Epipolar entre Imagens*. M Eng Thesis, Escola Politécnica da Universidade de São Paulo, São Paulo.

Lindeberg, T., 1994. "Scale-space theory: A basic tool for analysing structures at different scales". *J Appl Stat*, Vol. 21, pp. 224–270.

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

Lowe, D.G., 1999. "Object recognition from local scale-invariant features". In *Proc Int Conf Comput Vision*. Corfu, Greece, pp. 1150–1157.

Lowe, D.G., 2001. "Local feature view clustering for 3D object recognition". In *Proc IEEE Conf Comput Vision Pattern Recog*. pp. 682–688.

Lowe, D.G., 2004. "Distinctive image features from scale-invariant keypoints". Tech Rep, Univ British Columbia.

Luong, Q.T., Deriche, R., Faugeras, O. and Papadopoulo, T., 1993. "On determining the fundamental matrix: Analysis of different methods and experimental results". Tech Rep 1894, INRIA.

Mantyla, M., 1988. *Introduction to Solid Modeling*. W.H. Freeman & Co., New York, NY, USA.

Mark de Berg, Otfried Cheong, M.v.K. and Overmars, M., 2008. *Computational Geometry: Algorithms and Applications*. Springer-Verlag.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L., 2005. "A comparison of affine region detectors". *Int J Comput Vision*, Vol. 65, pp. 43–73.

Moravec, H., 1977. "Towards automatic visual obstacle avoidance". In *Proc 5th Int Joint Conf Artif Intell*. pp. 584–584.

Muja, M. and Lowe, D.G., 2009. "Fast approximate nearest neighbors with automatic algorithm configuration". In *Int Conf Comput Vision Theory App*. pp. 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 *Proc 10th Eur Conf Comp Vision: Part II*. pp. 500–513.

Rosten, E. and Drummond, T., 2006. "Machine learning for high-speed corner detection". In *Eur Conf Comput Vision*. Vol. 1, pp. 430–443.

Schmid, C., Mohr, R. and Bauckhage, C., 2000. "Evaluation of interest point detectors". *Int J Comput Vision*, Vol. 37, No. 2, pp. 151–172.

Smith, S.M. and Brady, J.M., 1995. "Susan - a new approach to low level image processing". *Int J Comput Vision*, Vol. 23, pp. 45–78.

Tissainayagam, P. and Suter, D., 2004. "Assessing the performance of corner detectors for point feature tracking applications". *Image Vision Comput*, Vol. 22, pp. 663–679.

Torr, P.H.S. and Zisserman, A., 2000. "Mlesac: a new robust estimator with application to estimating image geometry". *Comput Vis Image Und*, Vol. 78, No. 1, pp. 138–156.

Torr, P., 1995. *Outlier Detection and Motion Segmentation*. Ph.D. thesis, University of Oxford.

10. Responsibility notice

The author(s) is (are) the only responsible for the printed material included in this paper