Modeling of Real 3D Object using Photographs

Kateřina Dařílková
Faculty of Mathematics, Physics
and Informatics
Comenius University
Mlynská Dolina
SK 84248, Bratislava, Slovakia
darilkova@sccg.sk

ABSTRACT

The goal of this work is to create several functions for processing of two input images of one object to uncover a geometry of the scene. There are well-known techniques how to compute a fundamental matrix and reconstruct 3D coordinates. Several techniques were tested to find the method that is fast and rather precise. The functions are implemented in the environment of Borland C++ Builder. Calibrated cameras and start in a situation when several points are marked correctly in both pictures are assumed.

Keywords

two view geometry, epipolar geometry, fundamental matrix, eight point method, algebraic error minimization.

1. INTRODUCTION

Computer graphics aims to model virtual reality with high realism. Several problems arise when real world objects are modeled. A model of the real object should meet some metric constraints, which are sometimes difficult or too complex to learn exactly. It is more suitable for most of situations to estimate the approximation of values. If two images of the object are available, the reconstruction of selected object points is possible, with precision up to scale, by computing some geometry around. The theory behind is called a multiple-view geometry, specifically two-view geometry.

2. BACKGROUND

The basis of theory of two-views can be dated back to year 1855. In this year french mathematician Chasles formed the problem of recovering the epipolar geometry from a seven-point correspondence. Eight years later task was solved by Hesse and in the year 1981 the original eight-point algorithm for the computation of essential matrix was introduced by Longuet-Higgins. The problem of fundamental matrix estimation is studied quite extensively from that time. A nonlinear minimization approach to estimate the essential matrix [Weng] and a distance minimization approach to compute the fundamental matrix [Luong] were described. In practice, the geometric (nonlinear) minimization approach is more reliable but computationally more expensive. Current methods improve the linear methods or accelerate the geometric minimization approach.

WSCG SHORT papers proceedings, ISBN 80-903100-9-5

WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.

3. EPIPOLAR GEOMETRY

Epipolar geometry is a natural projective geometry between two views of the scene. It is usually a picture from a camera. The type and the properties of projection are given by construction and adjustment of the camera. The epipolar geometry does not depend on a structure of the scene. It is derived from intrinsic parameters of cameras and their mutual position. Taking pictures of the scene by the camera is an arbitrary projective transformation of a 3D scene in world coordinates to a 2D image. The 3*4 *matrix of projection* P represents it.

Most of usually used cameras can be approximated by a pinhole camera model. Let us define the co-ordinate system by a position and orientation of the first camera. The matrix **P** can be simplified in this coordinate system as $\mathbf{P} = \mathbf{K}\,[\mathbf{I}_{3x3} \mid \mathbf{0}]$ where **K** is a 3*3 *calibration matrix* of the camera. Using general world coordinate system the general matrix of projection **P** has form $\mathbf{P} = \mathbf{K}\,\mathbf{R}\,[\,\mathbf{I}\ |\mathbf{t}]$. The difference is just the orientation and translation of camera towards the world coordinate system origin. **R** is a matrix of rotation and **t** is a vector of translation, called together *extrinsic parameters of camera*.

4. FUNDAMENTAL MATRIX

The epipolar geometry arises from two images. Our goal is to find an equation, which describes the relationship between the pictures. Let us find the relation which binds an image point from the first image **x**= (x, y) with an image point from the second image **x'** = (x', y') holding a constraint that they both are projections of some point **X** from the scene. Such a matrix is called the *fundamental matrix* **F**. It is the algebraic representation of the epipolar geometry. For any pair of valid points **x**, **x'** exists a 3x3 fundamental matrix **F** for which $\mathbf{x}'^{T}\,\mathbf{F}\,\mathbf{x} = 0$. **F** is of rank 2. To make the work with matrices easier, let us set the coordinate system defined by the first camera as the world coordinate system. The first camera projection matrix is then $\mathbf{P} = \mathbf{K}\,[\mathbf{I}_{3x3}\,|\mathbf{0}]$ and the second camera matrix is $\mathbf{P}' = \mathbf{K}'\,[\mathbf{R}\,|\,\mathbf{t}]$, where **R** and **t** describe a rotation and translation of the second camera towards the first one and **K** and **K'** are the calibration matrices of cameras.

Fundamental matrix and projections

The fundamental matrix depends only on the mutual position of the two cameras and their calibration. For the pair of canonic camera matrices $\mathbf{P} = [\mathbf{I}\ |\ \mathbf{0}]$, $\mathbf{P}' = [\mathbf{R}\ |\ \mathbf{t}]$ corresponding to a fundamental matrix **F** equations as follows are valid:

$$\mathbf{P} = [\mathbf{I}\ |\ \mathbf{0}],\quad \mathbf{P}' = [[\mathbf{e}']_x\,\mathbf{F} + \mathbf{e}'\mathbf{v}^{T}\ |\ d\mathbf{e}'],$$

where **e** is an epipole (a projection of the first camera centre in the second image), $[\mathbf{e}]_x$ is a cross product, a skew-symmetric matrix, **v** is any 3-vector and d is a non-zero scalar. This result can be used only in cases when calibration matrices of both cameras are equal to identity. This can be achieved by normalization. The fundamental matrix corresponding to the normalized cameras is called *essential matrix* **E** with the form $\mathbf{E} = [\mathbf{t}]_x\,\mathbf{R}$. It is singular and two of its singular values are equal. There are only two possible factorizations of **E** (ignoring signs) to a skew-symmetric matrix and a rotation matrix. Taking into consideration direction of the translation from the first to second camera, there are four possible choices for the second camera matrix **P'**. The reconstructed point X is in front of both cameras only in one of these four solutions.

Another way how to separate the projection matrices is given by [Faugeras, Luong].

5. FUNDAMENTAL MATRIX ESTIMATION

The problem of searching for projection matrices has changed to seeking for the fundamental matrix. There exist several methods how to find it. The robust ones search for the number of corresponding features in two images and start from this statistically huge set. Our input comprises several corresponding points from two images and a calibration matrix of both cameras (usually the same). Methods used in these situations can be divided into several groups: linear algorithm, algebraic minimization algorithm, distance minimization. Until now two of them were tested.

The basic linear 8-point algorithm

The best approximation of fundamental matrix is searched. The equation which defines the fundamental matrix **F** is $\mathbf{x}'^{T}\,\mathbf{F}\,\mathbf{x} = 0$ where **x** and **x'** is a pair of the matching points in the first and the second image. Their projective coordinates are $\mathbf{x}=(x, y, 1)^{T}$, $\mathbf{x}'=(x', y', 1)^{T}$. Each point match results in one linear equation in the unknown entries of **F.** For more points matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ (i = 1… n) linear equations can be stacked up into a matrix. The solution can be found by the least-square algorithm using singular value decomposition of **A**. The normalization (simple translation and scaling) of input data is very useful to make the algorithm stable. An important property of fundamental matrix is the singularity. This method, in general, does not produce matrix **F** of rank 2.

The algebraic minimization algorithm

The remaining problem is how to guarantee singularity of the constructed fundamental matrix. One possible solution is to construct the singular matrix as a product $\mathbf{F} = \mathbf{M}\,[\mathbf{e}]_x$ where M is non-singular matrix and $[\mathbf{e}]_x$ is any skew-symmetric matrix, with **e** corresponding to the epipole in the first image. To guarantee the fundamental matrix properties in such matrix **F**, a constraint on **F** is added. Matrix **F** can be computed from the image point correspondences and known epipole **e** by minimization. [Hartley, Zisserman] The estimation inaccuracy can be evaluated by an algebraic error ε. It describes a transformation which maps the estimate of the epipole $\mathbf{e}_i$ to the algebraic error $\boldsymbol{\varepsilon}_i : \mathbf{R}^3 \rightarrow \mathbf{R}^8$. The exact epipole is unknown, in reality. We acquire it's estimate using iterative methods. The Levenberg – Marquardt iterative method can be used [Numerical Recipes], [Pollefeys]. An estimation of the fundamental matrix $\mathbf{F}_0$ is calculated using different methods (the 8-point linear algorithm) to get the zero approximation of the epipole $\mathbf{e}_0$ (a right null vector of matrix $\mathbf{F}_0$). Each iteration aims to change $\mathbf{e}_i$ so that the value $||\boldsymbol{\varepsilon}_i||$ is minimized.

6. IMPLEMENTATION

The aim of our implementation is to determine a sufficiently correct method to obtain the fundamental matrix, which is fast for available data. It is assumed, that points are assigned by operator (manually).

The 8-point normalized algorithm is fast and easy to implement method. Usually, it offers quite precise results. It is very suitable as the first step for iterative methods. If higher precision is required, the algebraic error minimization method is recommended. Distance minimization method using Sampson error is appropriate as an alternative algorithm.

Methods were implemented using a Borland C++ Builder application. For calculation of a singular decomposition of matrix and inverse matrix a suitable library was chosen (Open Computer Vision Library [OpenCV]). The reconstructed points and predefined faces can be visualized in a 3D scene generated by the OpenGL library.

The images used were not made by wide-angle lens. Information about a barrel distortion is available for some of them. The elimination of deformation did not produce increase of results correctness.

Used data

A couple of images with different accuracy and resolution were used to test implemented methods. Synthetic data, pictures from tutorial of PhotoModeler [PhotoMod] and self-made pictures done by a standard camera were used. Examples of used scenes are on figure 1.

Figure 1: Scenes
"Bench" (Res. 280x1024, Foc.l. 6.97mm),
"Boxes" (Res. 2272x1704, Foc.l. 7.19mm),
"Car" (Res. 2267x1520, Foc.l. 30.75mm).

7. TESTS AND RESULTS

Error computation

To evaluate precision of acquired fundamental matrix residual error is calculated. The tests performed show the dependency of the error value on the increasing number of corresponding points. It is important to evaluate the error over a wider group of matched points, not just for the point correspondences used to compute **F** (first 8 points). This is shown in a shape of graph curves. The comparison of errors of two tested methods are displayed in figure 2.

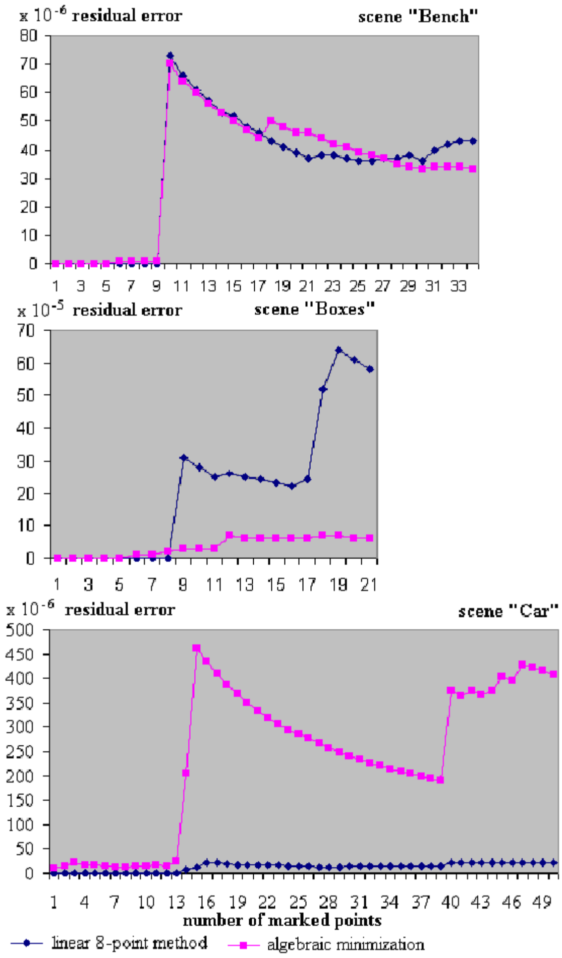

Figure 2: Graphs of methods residual error.

The graphs show the impossibility to decide positively, which method is more accurate, considering the residual error. In scenes, which are considered less stable (i.e. where the used matched points are almost coplanar) linear method works better. In scenes defined with higher precision algebraic minimization method is more accurate (e.g. in scene "Car" the result is much better).

The fundamental matrix estimate used as initial step in the first iteration is the essential part for iteration method. If the initialization is too deflected, the method diverges or converges seemingly.

Synthetic data

An idealized scene was created to understand the methods and their convergence better (Figure 3).

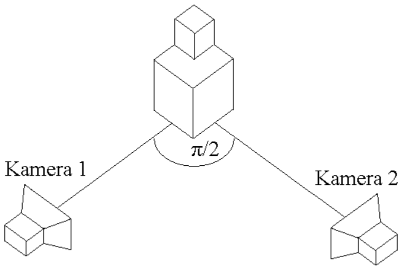

Figure 3: Synthetic scene.

The residual error was quantified here for the linear method and algebraic minimization too. Moreover, exact fundamental matrix derived from the known geometry of the scene was applied (Figure 4).

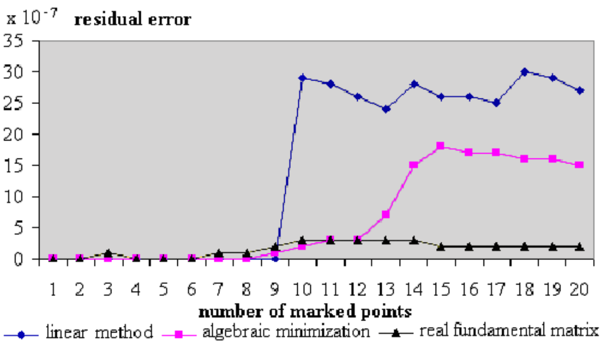

Figure 4:Graph of error for synthetic scene.

The algebraic minimization seems more suitable in this situation. All the error values are 10 times lower than in the real scenes. It is caused by much more precise selection of matching points.

Visual results

An indirect proof of the calculation accuracy is a visualization of the results as a graphical depicting of the reconstructed 3D position of the points in a virtual world. Such a visualization shows how much the reconstruction fits, that is, how the fundamental matrix fits. An experienced operator uncovers in visualization which pair of corresponding points is set incorrectly or improperly.

8. CONCLUSION

This work compares two methods for acquisition of the fundamental matrix by 8 points marked in two pictures of a scene. Linear method and the method of algebraic minimization were exploited. The presented comparisons show similarity of the methods results. Well-defined scenes have significantly better results with the fundamental matrix improved by algebraic minimization starting from a matrix given by the linear method. If the scene is described by improper set of points, the linear method is more suitable. It is possible to recognize this case by a monitoring of several properties of computation and of partial results. The choice of another set of marked points can be more efficient step to get the accurate fundamental matrix.

Standard camera suffices to take two pictures of any scene and to reconstruct its geometry. Some of camera parameters are required to be known (published usually by the producer). Much more importance is given to right choice of pictures. The user ought to arrange the scene to be heterogenous enough, to take pictures suitable for selection of 8 non-coplanar points placed in the scene uniformly. Another important condition is the precise determination of marked points in images.

9. ACKNOWLEDGEMENTS

This work is the part of project supported by APVT grant 20- 025502.

10. REFERENCES

[Faugeras, Luong] Faugeras, O., Luong, Q.-T. (2001). The geometry of multiple images. Massachusetts Institute of Technology, The MIT Press Cambridge, London, ISBN 0-262-06220-8.

[Hartley, Zisserman] Hartley, R., Zisseman, A. (2000). Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, United Kingdom, ISBN 0 521 62304 9.

[Luong] Luong, Q.-T. (1992). Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus Grande autonomie des Systemes robotiques. PhD thesis, Université de Paris-Sud, Centre d'Orsay.

[Numerical Recipes] Numerical recipes in C: the art of scientific computing, Cambridge University Press, ISBN 0-521-43108-5

[OpenCV] Open Computer Vision Library, Copyright © 2000, Intel Corporation, www.intel.com/ research /mrl/research/opencv

[PhotoMod] PhotoModeler Pro 5.0 Demo, Copyright 1993 – 2002 Eos Systems Inc. http://www.photomodeler.com/ddlp.html, (8.9.04)

[Pollefeys] Pollefeys, M.,(2002),Visual 3D Modeling from Images, Tutorial Notes, University of North Carolina - Chapel Hill, USA, www.cs.unc.edu/ ~marc/tutorial/node160.html

[Weng] Weng, j., Ahuja, N, Huang,T. (1993). Optimal motion and structure estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(9):864-884.