Two View Geometry Estimation by Determinant Minimization

Lorenzo Sorgi and Andrey Bushnevskiy
Technicolor Research & Innovation, Karl-Wiechert Allee 74, Hannover, Germany

Keywords: Two View Geometry, Epipolar Geometry, Perturbation Theorem, Determinant Minimization, Tetrahedron.

Abstract: Two view geometry estimation, the task of inferring the relative pose between two cameras using only the image content, is one of the fundamental and most studied problems in Computer Vision. In this paper we present a new approach for two view geometry estimation, based on the minimization of an objective function given by the overall volume of the tetrahedrons identified in 3D space by pairs of corresponding feature points. This error measure is equivalent to the determinant of a real valued square matrix, function of the point match coordinates in the camera space, and we show how to minimize it taking advantage of the Perturbation Theorem. Test performed on synthetic and real dataset confirm an increased estimation accuracy compared to the state-of-art.

1 INTRODUCTION

Given a point in one image, is it possible to constraint the position of the corresponding point in a second image? The answer to this question leads to the definition of one of fundamental theorems of the geometry of multiple views, the Epipolar Constraint, represented with a 3x3 matrix denoted as Essential Matrix E. This is a non invertible matrix of rank 2, independent of a scene structure completely constrained by the relative pose between the cameras. If a point X in 3-space is imaged as x and x' in two views, then one can show that these points satisfy the relation $\langle x', Ex \rangle = 0$, where $\langle a, b \rangle$ represents the vector inner product. This relation was first published in 1981 by Longuet-Higgins (Longuet-Higgins, 1987), who has introduced the concept of Epipolar Constraint to the computer vision community.

The first solution to the problem of Essential Matrix estimation from the image correspondences was originally proposed by Kruppa (Kruppa, 1913), where it has been shown, that given enough correspondences between two perspective views is possible to retrieve all the possible configurations of the cameras, which constitute a set of 11 solutions, among which only 10 are physically valid (Faugeras and Maybank, 1990). Most of the techniques currently used in 3D vision systems work with a closed-form high-order (13th - 10th) uni-variate polynomial equation, which encodes the solution (Nistér, 2004; Triggs, 2000; Philip, 1996). However, fifth-degree and higher-degree polynomials do not have a general solution according to the Abel-Ruffini theorem. Therefore, application of the iterative numerical routines is required, and the solution turns out to be highly unstable due to the intrinsic ill-conditioned nature of the root finding problem.

A slightly different approach, has been proposed by Batra and al. (Batra et al., 2007), where the task of Essential matrix estimation is reformulated as a constraint quadratic optimization problem, by introducing two additional constraints. In this way the authors overcome the issue of finding the root of high degree polynomials, but they have to tackle the problem in an iterative way using multiple solution seeds as starting point for the minimization step. With regards to this aspect still remains open the issue how many seed points in the solution space are required and how to sample them.

In this paper we observe, that each pair of corresponding features describes in 3D space is a tetrahedron, which has a null-volume in case of correctly estimated camera poses. Following this observation one can reformulate the two view geometry estimation problem as a minimization of the cumulative volume of the tetrahedrons defined by a set of point matches. We will show that this is equivalent to the task of minimization of the sum of the determinants of a set of square matrices, which can be solved by means of the Perturbation Theorem (Nakatsukasa, 2011).

Sorgi, L. and Bushnevskiy, A.
Two View Geometry Estimation by Determinant Minimization.
DOI: 10.5220/0005677405900594
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 3: VISAPP*, pages 590-594
ISBN: 978-989-758-175-5

2 EIGENSYSTEM PERTURBATION

In this section we introduce the perturbation theorem and describe its exploitation as solver for the matrix determinant minimization. This will provide the basic mathematical tool for the solution of the two view geometry estimation.

Let A and D be N-by-N symmetric real-valued matrices and $\{\lambda_i, \vec{u}_i\}_{i=1...N}$ the eigensystem of A, that is the set of eigenvalues and eigenvectors such that:

$$\begin{cases} A\vec{u}_i = \lambda_i \vec{u}_i \\ \vec{u}_i^T \vec{u}_j = \delta_{i,j} \\ A = \sum_i \lambda_i \vec{u}_i \vec{u}_i^T \end{cases}, \quad (1)$$

and $\delta_{i,j}$ is the Kronecker Delta-function

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \quad (2)$$

Considering a perturbed matrix $A' = A + \varepsilon D$, for a small ε, let $\{\lambda'_i, \vec{u'}_i\}_{i=1...N}$ be the eigensystem of A', corresponding to $\{\lambda_i, \vec{u}_i\}_{i=1...N}$. Then the following relations hold:

$$\begin{cases} \lambda'_i = \lambda_i + \varepsilon(\vec{u}_i \cdot D\vec{u}_i) + O(\varepsilon^2) \\ \vec{u}'_i = \vec{u}_i + \varepsilon \sum_{j\neq i} \frac{(\vec{u}_j \cdot D\vec{u}_i)\vec{u}_j}{\lambda_i - \lambda_j} + O(\varepsilon^2) \end{cases}. \quad (3)$$

Equations (3) are known as Perturbation Theorem.

2.1 Determinant of Perturbed Matrix

The Perturbation Theorem provides also a representation of a first order Taylor expansion of the eigensystem of a $A(x)$, denoted with $\{\lambda_i(x), \vec{u}_i(x)\}_{i=1...N}$. Let us consider a matrix function $A(x) : R \rightarrow S_{N\times N}$, where $S_{N\times N}$ is the space of N-by-N symmetric positive-definite matrices, and its first order Taylor expansion given by

$$A(x+\varepsilon) \cong A(x) + J_A(x)\varepsilon + O(\varepsilon^2), \quad (4)$$

where $J_A(x) = \frac{\partial A(x)}{\partial x}$ is the Jacobian matrix of A. Similarly the first order Taylor expansion of the corresponding eigensystem can be written as

$$\begin{cases} \lambda_i(x+\varepsilon) = \lambda_i(x) + J_{\lambda_i}(x)\varepsilon + O(\varepsilon^2) \\ \vec{u}_i(x+\varepsilon) = \vec{u}_i(x) + J_{u_i}(x)\varepsilon + O(\varepsilon^2) \end{cases}. \quad (5)$$

Equations (5) provide the eigensystem of a matrix $A(x)$ affected by a small perturbation $J_A(x)\varepsilon$, therefore the Perturbation Theorem implies that:

$$\begin{cases} J_{\lambda_i}(x) = \vec{u}_i(x) \cdot J_A(x)\vec{u}_i \\ J_{u_i}(x) = \sum_{i\neq j} \frac{(\vec{u}_j \cdot J_A(x)\vec{u}_i)\vec{u}_j}{\lambda_i(x) - \lambda_j(x)} \end{cases}. \quad (6)$$

We recall that the determinant of the matrix $A(x)$ is given by the product of its eigenvalues counted with their algebraic multiplicities. By using equations (5) one can express the determinant of the perturbed matrix as

$$detA(x+\varepsilon) = \prod_i \lambda_i(x+\varepsilon) \cong \prod_i (\lambda_i(x) + J_{\lambda_i}(x)\varepsilon) \quad (7)$$

By expanding equation (7) and neglecting the high order terms in ε, one obtains

$$detA(x+\varepsilon) = \prod_i \lambda_i(x) + \prod_i (\prod_{j\neq i} \lambda_j(x)) J_{\lambda_i}(x)\varepsilon\,. \quad (8)$$

Equation (8) provides the first order Taylor approximation of the determinant of the matrix A(x) and its derivative is given by

$$\frac{\partial detA(x)}{\partial x} \cong \prod_i (\prod_{j\neq i} \lambda_j(x)) J_{\lambda_i}(x) \quad (9)$$

An equivalent formulation can be derived also for the determinant of a non-symmetric matrices. Let us consider a D×D non-symmetric matrix $M(x)$ and its Singular Value Decomposition

$$M(x) = U \cdot \Sigma \cdot V^T \quad (10)$$

where the notation Σ is a diagonal matrix containing on the singular values $\{\sigma_j\}_{j=1,...,D}$. By applying the determinant relation $det(AB) = det(A)det(B)$ one obtains

$$detM(x) = detU \cdot det\Sigma \cdot detV^T = \prod_i \sigma_i(x) \quad (11)$$

Let us define the symmetric matrix $K(x) = M(x) \cdot M^T(x)$. The singular values of the matrix $M(x)$ are related to the eigenvalues λ_i of $K(x)$ by the expression $\lambda_i = \sigma_i^2$, therefore by differentiation we obtain

$$J_{\lambda_i} = 2\sigma_i \cdot J_{\sigma_i} \quad (12)$$

As the matrix $K(x)$ is by definition symmetric, one can apply the results of the Perturbation Theorem to compute the Jacobian J_{λ_i} of its eigenvalues and express the first order Taylor approximation of the determinant of the non-symmetric matrix $M(x)$ as

$$detM(x+\varepsilon) \cong detM(x) + \prod_i (\prod_{j\neq i} \sigma_j(x)) \frac{J_{\lambda_i}(x)}{2\sigma_i(x)} \varepsilon \quad (13)$$

2.2 Two View Geometry Estimation

Let us consider a two-view geometry model in Fig.1, where m and m' are projections of a 3D point M on the image planes of two cameras.

We assume to be working in calibrated camera condition, therefore each image point m can be mapped to

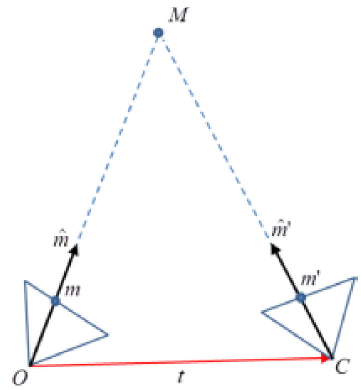


Figure 1: Two-view projection model in a calibrated camera space.

the corresponding incident vectors $\hat{m}$ by inverting the cameras projection functions (Hartley and Zisserman, 2004). Without loss of generality we can assume the projection center of the first camera to be located in the origin of the reference system, $O = [(0 \quad 0 \quad 0)]^T$. Let us also denote with $\{R,t\} \in SO(3)$ the Euclidean transformation relating two camera systems, which enables the definition of the incident vector in the second camera system as

$$\hat{m}' \sim R(M - t), \tag{14}$$

where $\sim$ denotes equality up to a non-zero scale factor.

One can easily infer that the four points $\{O, \hat{m}, t, R^T\hat{m}' + t\}$ lay on the same plane, defined by the line OC and the point M; which is also a straightforward consequence of the epipolar constraint. An alternative interpretation of the coplanarity constraint is given in 3D if one considers a solid, namely the tetrahedron, identified by the four points, (Fig.2(a)). In the ideal case the latter has null volume, however noisy image projections or incorrect 2-view geometry parameters lead to the construction of a non-zero volume (Fig.2(b)).

The volume V of a tetrahedron with vertices $\{a,b,c,d\}$ can be computed from the determinant of the matrix M_V by the relation

$$V = \frac{1}{6} det M_V \ , \tag{15}$$

where $M_V = [(a-b)\,,(b-c)\,,(c-d)\,,]$ is constructed using the vertices coordinates. Therefore by applying the previous equation 15 to the tetrahedron in (Fig.2), one can compute its volume as a fraction of the determinant of the matrix $A(R,t)$ defined as

$$A(R,t) = [-\hat{m}\,,(\hat{m}-1)\,,-R^T\hat{m}'] \tag{16}$$

where the matrix $A(R,t)$ is explicitly indicated as a function of the camera motion $\{R,t\}$.

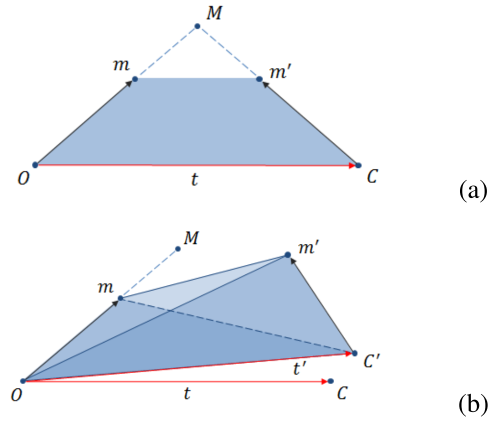


Figure 2: Tetrahedron built using corresponding feature point in case of correctly (a) and incorrectly (b) estimated camera pose.

This simple geometrical model allows for the formulation of the 2-view geometry estimation problem, given a set of N point correspondences between two views, as the solution of the minimization problem

$$\{\bar{R},\bar{t}\} = \underset{\{R,t\}\in SO(3)}{\arg\max} \sum_{k=1,\ldots,N} det^2(A_k(R,t)), \tag{17}$$

where $A_k(R,t)$ is the matrix build according to the equation (16) using the k-th point correspondence.

The objective function (17) is minimized using the Levenberg-Marquardt algorithm and in each iteration the normal equation is build using equations (11) and (13) to express the error contributions and their Jacobian.

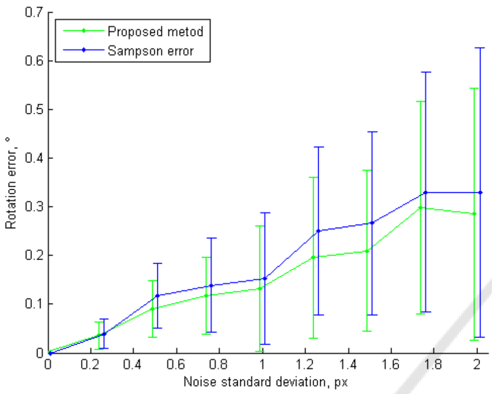
3 RESULTS

The proposed technique has been evaluated on synthetic and real data and compared with the standard approach, based on the minimization of the Sampson error (Hartley and Zisserman, 2004; Zhang, 1998).

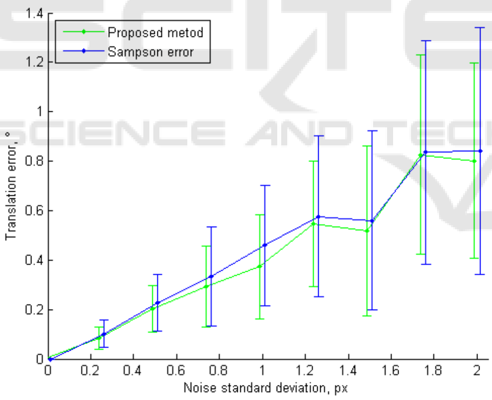
3.1 Synthetic Data

The synthetic model was comprised of a point cloud containing 100 randomly positioned 3D points and two virtual cameras, randomly located at a fixed distance with the respect to the point cloud. Each camera was modeled as a 50 horizontal field-of-view lens and 1024x768 sensor. The stereo geometry was randomly sampled from the subspace of Euclidean Transformations $\{R,t\}$, such that $\|t\| = 1$.

The image projections of the points were corrupted using a zero-mean white Gaussian noise with increasing standard deviation ranging within the interval $[0,2]$ pixels. For each level of image noise 50 random scenes were generated and the median and the standard deviation of the estimation errors were collected. The rotation and translation errors are presented by two angular measures: the angle between the estimated and real baseline vectors, and the angle associated to the difference rotation δR between the estimated $\tilde{R}$ and real R rotations, $\delta R = R^T \tilde{R}$. In both tests algorithm was initialized with the ideal stereo geometry, given by the null rotation and the unit norm x-vector $\{I_3, [1\ 0\ 0]^T\}$.



(a)



(b)

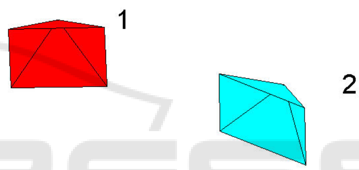
Figure 3: Synthetic results. Error for estimated camera rotation (a) and translation (b).

The results, presented in Fig.3. demonstrate, that the accuracy of the proposed method is higher than the one, based on the Sampson Error minimization. The interesting aspect is that the convergence of the minimization of the proposed error function is achieved irrespective to the proximity of the initialization point to the actual solution. This does not apply to the Sampson Error minimization approach, where a number of trials have failed to converge to the correct solution.

3.2 Real Data

For the real data test we have used a set images from two GoPro Hero 3+ action cameras, set in a stereo configuration on the planar surface Fig.4(a). The intrinsic parameters of the cameras were pre-estimated using our own calibration tool, based on the calibration approach, presented in (Kanatani, 2013). The feature points positions and their descriptors were extracted from each of two views and matched.

(a)



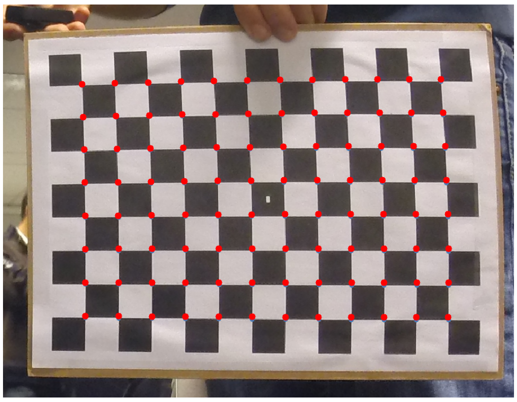
(b)

Figure 4: Real dataset test. Stereo camera configuration (a) and estimated geometry (b).

In order to minimize the influence of the outliers, only the feature correspondences, valid in both directions were used for the geometry estimation. The reconstructed geometry, presented in Fig.4(b) confirms the efficiency of the proposed method.

In order to asses the quality of the stereo geometry recovered using the proposed approach, we have designed a test using a black and white checkerboard pattern. The relative geometry of two cameras has been first estimated using the proposed method and the feature set extracted from the snapshots of the actual scene.

A set of corresponding snapshots of a checkerboard pattern then has been taken in such a way, that the grid has been simultaneously visible in both views. The grid points have been detected in each of the views, triangulated in 3D space using the estimated camera geometry and then projected back to the original views. The error then has been estimated as an average displacement between the detected and backprojected grid points (Fig.5). The resulting backprojection error with the mean $\mu = 2.9px$ and the standard deviation $\sigma = 0.67px$ confirms the accuracy of



(a)

(b)

Figure 5: Real dataset test. A crop of the black and white pattern view (a), detected (blue circle) and backprojected (red dot) points (b).

the proposed approach and suggests the possibility of its straightforward application without any subsequent refinement.

4 CONCLUSIONS

We have presented a novel approach for the two view geometry estimation, based on the tetrahedron volume minimization using a Perturbation Theorem. The evaluation using synthetic and real datasets and comparison to the standard Sampson Error minimization algorithm confirms the accuracy of the method. The approach features a major advantage, namely the geometrical nature of the objective function, which leads to an increase in the estimation accuracy and allows for a very rough initialization of the numerical iterations without the requirement of multiple seeds.

REFERENCES

Batra, D., Nabbe, B., and Hebert, M. (2007). An alternative formulation for five point relative pose problem. *Motion and Video Computing, IEEE Workshop on*, 0:21.

Faugeras, O. D. and Maybank, S. (1990). Motion from point matches: Multiple of solutions. *Int. J. Comput. Vision*, 4(3):225–246.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Kanatani, K. (2013). Calibration of ultrawide fisheye lens cameras by eigenvalue minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(4):813–822.

Kruppa, E. (1913). *Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung*. Hölder.

Longuet-Higgins, H. C. (1987). A computer algorithm for reconstructing a scene from two projections. In Fischler, M. A. and Firschein, O., editors, *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, pages 61–62. Kaufmann, Los Altos, CA.

Nakatsukasa, Y. (2011). *Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition*. PhD thesis, Davis, CA, USA. AAI3482268.

Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–777.

Philip, J. (1996). A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. *The Photogrammetric Record*, 15(88):589–599.

Triggs, B. (2000). Routines for relative pose of two calibrated cameras from 5 points. Software available from http://lear.imag.fr/people/triggs/src.

Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vision*, 27(2):161–195.