HOMOGRAPHY ESTIMATION USING LOCAL AFFINE FRAMES

Sagy Zino and Joseph M. Francos

Ben Gurion University of the Negev
Department of Electrical and Computer Engineering
sagyz@ee.bgu.ac.il, francos@ee.bgu.ac.il

ABSTRACT

The homography between pairs of images is typically computed from the correspondence of key features such as points, lines, conics and other geometric entities, each contributing information to fix the required 8 degrees of freedom (DOF). In the past years there have been attempts to use conics as correspondence features as a minimum of two correspondence pairs are required. However, the resulting methods either place restrictions on the problem (such as pure camera rotation, known calibration) or result in iterative non-linear solutions or in over-parameterized linear problems. Throughout this paper we propose a simple, direct linear transformation (DLT) like solution to the problem of homography estimation using local affine frames, without any restrictions on the physical model. We also provide approximated statistical analysis of the proposed algorithm and a comparison with the performance of the DLT algorithm. In addition, this method can be easily adapted to similar problems which employ a DLT-like algorithm.

Index Terms— Homography estimation, Local affine frames, Perturbation analysis

1. INTRODUCTION

The planar homography is a collinearity preserving mapping between points on projected planes. The homography maps a coordinates vector $\mathbf{x} = (x, y)^T$ to another coordinates vector $\mathbf{x}' = \mathbf{h}(\mathbf{x}) = (x', y')^T$. With the use of homogeneous coordinates, this transformation becomes linear and is represented as a non-singular 3×3 matrix, usually denoted by H. In this manner, a simpler relation is established between the coordinate vectors as $\mathbf{x}' = \lambda \mathrm{H}\mathbf{x}$.

Homographies play an important role in the geometry of multiple views [1], [2], [3] as their estimation is the basic building block in solving more complex computer vision tasks such as camera calibration and distortion correction [4], 3D reconstruction, stereo vision [5], mosaicing [6] and pose estimation [7], [8]. In real life applications, the homography is typically estimated from noisy correspondences, making linear estimation methods ineffective, either due to large re-projection errors of the results, or due to the large amount of correspondences required, which are not always available. Typically, linear estimators are used as the basic hypothesis in a more robust RANSAC [9] algorithm.

Our approach is a direct derivation similar to that of the DLT algorithm [1] but instead of using four points to fix the required 8 DOF of the homography, we redefine the information extracted from the local affine frames as derivatives of the homography at a given point, allowing us to fix 6 DOF for each correspondence pair. Therefore, as few as two pairs are required to estimate the homography.

The rest of the paper is organized as follows. Background and prior works are reviewed in Sec. 2. The proposed method is described in Sec. 3. Statistical performance analysis of the proposed algorithm and comparison with the DLT algorithm is given in Sec. 4. Examples are given in Sec. 5. Conclusions are drawn in Sec. 6.

2. BACKGROUND AND PRIOR WORK

The most closely related works to the one presented in this paper are those using conics. Conics are second degree curves, of which, ellipses are most commonly used to complete the task at hand. In homogeneous coordinates, a general conic can be expressed as

$$\mathbf{x}^T \mathrm{C}\mathbf{x} = 0, \tag{1}$$

where C is a real symmetric 3×3 matrix whose entries are the conic coefficients. Assuming we have a pair of corresponding conics, C and C', these conics satisfy the projective relation

$$\mathrm{C}' = \lambda \mathrm{H}^{-T} \mathrm{C} \mathrm{H}, \tag{2}$$

where λ is a scaling constant. A conic correspondence provides up to 5 constraints on the homography, therefore, two correspondences are sufficient to determine the required 8 DOF.

In [10] a general conic is described as a circle undergoing a series of pure similarity, affine and projective transformations $\mathrm{H}_s, \mathrm{H}_a, \mathrm{H}_p$, respectively. Writing those relations explicitly, a set of 14 non-linear equations in 14 variables is obtained. This set is further reduced to a univariate polynomial equation of degree 8 which has at most 8 solutions for H_p.

This process can be extended to calculate the entire homography between two views using two conic correspondences.

In [11] a general method of constructing a linear set of equations in the entries of H is derived. It is shown that two conics $\mathrm{C}_i, \mathrm{C}_j$ and their corresponding counterparts $\mathrm{C}'_i, \mathrm{C}'_j$ can be used along with a normalization scheme to transform the conic relation into a linear relation of the form

$$\mathrm{C}'^{-1}_i \mathrm{C}'_j \mathrm{H} = \mathrm{H} \mathrm{C}_i^{-1} \mathrm{C}_j. \tag{3}$$

Thus, three or more conic correspondences are enough to uniquely determine the homography. In the minimal case of two conic correspondences, the authors show that there exists a four-fold ambiguity of solutions and an additional procedure is required to determine the exact solution.

Another related work is that of Köser and Koch. In [12] they address the problem of pose estimation using a method called "spatial resectioning" from a single differential feature. Although the required transformation in this problem is a 2D perspecitivity (which is determined by 6 DOF), the authors introduce the relation between the global behavior of the function that is representing the homography to its local behavior as a local affine frame, using Taylor series expansion (a detailed explanation is given in Section 3). This relation is further exploited in this paper to provide full homography estimation.

3. HOMOGRAPHY ESTIMATION USING LOCAL AFFINE FRAMES

Having established correspondence and affine normalization parameters between a pair of patches, we now turn to redefine this information in terms that are suited to homographies.

Consider a pair of corresponding patches $\{\Omega_i, \Omega'_i\}$ satisfying the affine relation

$$\mathbf{x}' = \mathrm{B}_i \mathbf{x} + \mathbf{t}_i, \tag{4}$$

for each $\mathbf{x} \in \Omega_i, \mathbf{x}' \in \Omega'_i$, where B_i is a 2×2 linear transformation matrix and $\mathbf{t}_i$ is a translation vector. In the small area of these patches this affine relation approximates, with a fair accuracy, the local behavior of a homography. Therefore, it is possible to refer to the centers of the patches, $\mathbf{x}_i$ and $\mathbf{x}'_i$ as a pair of points having additional directional information. More specifically, by differentiating eq. (4) with respect to $\mathbf{x}$ we obtain this information in terms of the derivatives

$$\frac{\mathrm{d}\mathbf{x}'}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}_i} = \left[\frac{\partial \mathbf{x}'}{\partial x}\Big|_{\mathbf{x}_i} \quad \frac{\partial \mathbf{x}'}{\partial y}\Big|_{\mathbf{x}_i} \right] = \mathrm{B}_i. \tag{5}$$

On the other hand, differentiating the Taylor series expansion of each of the entries in $\mathbf{h}(\mathbf{x})$ and evaluating it at a non-singular point $\mathbf{x}_i$ (a non-ideal point, in projective geometry terms) we obtain

$$\frac{\mathrm{d}\mathbf{h}(\mathbf{x})}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}_i} = \frac{\mathrm{d}\mathbf{x}'}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}_i}, \tag{6}$$

Thus, a relation between the affine transformation of local affine features and the derivatives of the homography is established by substituting (5) into (6)

$$\frac{\mathrm{d}\mathbf{h}(\mathbf{x})}{\mathrm{d}\mathbf{x}}\Big|_{\mathbf{x}_i} = \mathrm{B}_i. \tag{7}$$

This relation allows us to substitute the measured local affine information as the point value of the homography derivative, giving us a new type of constraint on the homography.

Next consider the basic building block of the DLT equations that corresponds to the i-th correspondence pair given in [1], assuming normalized homogeneous coordinates

$$\begin{pmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & y'_i \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -x'_i \mathbf{x}_i^T \\ -y'_i \mathbf{x}_i^T & x'_i \mathbf{x}_i^T & \mathbf{0}^T \end{pmatrix} \mathbf{h} = 0, \tag{8}$$

where $\mathbf{h}$ is vector composed of the rows of H. In order to incorporate the homography derivative information into the homography estimation equations, we need equations where the derivatives appear explicitly. However, any attempt to differentiate the homography equations results in a non-linear set of equations. Instead, we reformulate the DLT equations using the vector to skew symmetric matrix operator and Kronecker product

$$\left([\mathbf{x}']_\times \otimes \mathbf{x}^T \right) \mathbf{h} = \mathbf{0}. \tag{9}$$

The resulting system of equations is identical to (8) but in this form, it is easier to perform implicit differentiation as differentiation and the vector to skew symmetric operator commute and differentiation of the Kronecker product is identical to the product rule of differentiation. Applying implicit differentiation to (9) with respect to x and y yields

$$\left(\left[\frac{\partial \mathbf{x}'}{\partial x} \right]_\times \otimes \mathbf{x}^T + [\mathbf{x}']_\times \otimes \frac{\partial \mathbf{x}^T}{\partial x} \right) \mathbf{h} = \mathbf{0}, \tag{10}$$

$$\left(\left[\frac{\partial \mathbf{x}'}{\partial y} \right]_\times \otimes \mathbf{x}^T + [\mathbf{x}']_\times \otimes \frac{\partial \mathbf{x}^T}{\partial y} \right) \mathbf{h} = \mathbf{0}. \tag{11}$$

Stacking both Eq. (10) and Eq. (11), evaluated at a point $\mathbf{x}_i$ we obtain the basic equation block of the proposed method, to which we shall refer to as the Direct Linear Transformation - Local Affine Frames (DLT-LAF) method. The derivatives of the homogeneous coordinates are $\frac{\partial \mathbf{x}^T}{\partial x} = (1, 0, 0)$ and $\frac{\partial \mathbf{x}^T}{\partial y} = (0, 1, 0)$. It should be emphasized that $\frac{\partial \mathbf{x}'}{\partial x}\Big|_{\mathbf{x}_i}$ and $\frac{\partial \mathbf{x}'}{\partial y}\Big|_{\mathbf{x}_i}$ are measured quantities, obtained from local affine frames, and not by differentiating another set of measurements. Each set of equations (10), (11) yields 3 equations. In total, a pair of corresponding patches yields 6 linearly independent equations in H, two from spatial information and four of directional information, Therefore, two pairs of corresponding patches suffice to estimate a homography.

4. APPROXIMATED STATISTICAL ANALYSIS

In order to understand the effects of the noise on each parameter of the estimated homography, we must formulate the resulting estimate in terms of the noise free homography and the noise terms. However, such formulation is mathematically complicated as it would require extracting an exact analytic expression of the right singular vector that corresponds to the smallest singular value. Let A denote the DLT (or DLT-LAF) model matrix. In order to overcome the above difficulty, we assume that the noise induces small perturbations in the model matrix, such that the perturbed matrix $\tilde{\mathrm{A}}^T\tilde{\mathrm{A}}$ may be written as

$$\tilde{\mathrm{A}}^T\tilde{\mathrm{A}} = \mathrm{A}^T\mathrm{A} + \sum_{k=1}^{K} \mathrm{E_k}\varepsilon_k. \quad (12)$$

In this manner, second order eigenvector perturbation expansion can be used to extract the linear portion of the effects of the noise in terms of the noise free model matrix A, noise matrices E_k and noise terms ε_k.

The aim of this expansion is to obtain an approximation of the form

$$\tilde{\mathbf{h}} = \mathbf{h} + \Xi\boldsymbol{\varepsilon}, \quad (13)$$

where $\tilde{\mathbf{h}}$ is a second order approximation of the estimated homography, the true homography $\mathbf{h}$ is the eigenvector of $\mathrm{A}^T\mathrm{A}$ that corresponds to the smallest eigenvalue, Ξ is a coefficient matrix and the error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_K)^T$ which contains linear and quadratic error terms. In this form we can easily calculate the first and second order statistics of the error as

$$\mathrm{E}(\tilde{\mathbf{h}}) = \mathbf{h} + \Xi\mathrm{E}(\boldsymbol{\varepsilon}), \quad (14)$$

$$\mathrm{cov}(\tilde{\mathbf{h}}) = \Xi\mathrm{E}\left((\boldsymbol{\varepsilon} - \mathrm{E}(\boldsymbol{\varepsilon}))\,(\boldsymbol{\varepsilon} - \mathrm{E}(\boldsymbol{\varepsilon}))^T \right)\Xi^T. \quad (15)$$

While this general approach has already been investigated in [13], the assumption that the quadratic error terms in $\boldsymbol{\varepsilon}$ are negligible leads to the false conclusion that the mean of the approximated estimate is the actual homography. However, in practice the right singular vectors are extracted from the perturbed matrix $\tilde{\mathrm{A}}^T\tilde{\mathrm{A}}$, a form which ensures the existence of quadratic noise terms with a non-zero mean.

4.1. DLT and DLT-LAF methods perturbation analysis

We begin the analysis by inspecting the effects of small perturbations on the DLT model matrix A. Consider a correspondence pair $\{\mathbf{x}_i, \mathbf{x}'_i\}$, corrupted with additive noises $\boldsymbol{\varepsilon}_{\mathbf{x}_i} = (\varepsilon_{i,x}, \varepsilon_{i,y}, 0)^T$ and $\boldsymbol{\varepsilon}_{\mathbf{x}'_i} = (\varepsilon_{i,x'}, \varepsilon_{i,y'}, 0)^T$ in $\mathbf{x}_i$ and $\mathbf{x}'_i$, respectively. Substituting into (9), we obtain the noise corrupted DLT block

$$\left([\mathbf{x}'_i + \boldsymbol{\varepsilon}_{\mathbf{x}'_i}]_\times \otimes (\mathbf{x}_i^T + \boldsymbol{\varepsilon}_{\mathbf{x}_i}^T) \right)\mathbf{h} = \mathbf{0}. \quad (16)$$

Since Kronecker product is a distributive operator and the vector to skew symmetric matrix is a linear operator we may expand the left hand side of (16), which we denote by $\tilde{\mathrm{A}}_i$, to yield

$$[\mathbf{x}'_i]_\times \otimes \mathbf{x}^T + [\mathbf{x}'_i]_\times \otimes \boldsymbol{\varepsilon}_{\mathbf{x}_i}^T + [\boldsymbol{\varepsilon}_{\mathbf{x}'_i}]_\times \otimes \mathbf{x}^T + [\boldsymbol{\varepsilon}_{\mathbf{x}'_i}]_\times \otimes \boldsymbol{\varepsilon}_{\mathbf{x}_i}^T. \quad (17)$$

The first term is the noise free DLT block A_i. We further expand the remaining terms as linear combinations of the noise terms (by rearranging the terms in Eq. (16)) to obtain the perturbed DLT block $\tilde{\mathrm{A}}_i$

$$\begin{aligned}\tilde{\mathrm{A}}_i = \mathrm{A}_i &+ \varepsilon_{i,x}\mathrm{E}_{i,x} + \varepsilon_{i,y}\mathrm{E}_{i,y} + \varepsilon_{i,x'}\mathrm{E}_{i,x'} + \varepsilon_{i,y'}\mathrm{E}_{i,y'} \\ &+ \varepsilon_{i,x}\varepsilon_{i,x'}\mathrm{E}_{i,xx'} + \varepsilon_{i,y}\varepsilon_{i,x'}\mathrm{E}_{i,yx'} \\ &+ \varepsilon_{i,x}\varepsilon_{i,y'}\mathrm{E}_{i,xy'} + \varepsilon_{i,y}\varepsilon_{i,y'}\mathrm{E}_{i,yy'}. \quad (18)\end{aligned}$$

The complete perturbed model matrix $\tilde{\mathrm{A}}$ is then obtained by stacking all perturbed DLT blocks. The error blocks are padded with zero to fit the dimensions of A and their position with respect to the originating block A_i. In the final step of this analysis we obtain the required matrix $\tilde{\mathrm{A}}^T\tilde{\mathrm{A}}$. For clarity, we reassign all error terms, error blocks and indices to include all the higher order terms obtained by this multiplication to fit the notation given in Eq. (12).

Eq. (12) ensures the existence and uniqueness of a power series expansion for the eigenvector corresponding to the smallest eigenvalue of $\tilde{\mathrm{A}}^T\tilde{\mathrm{A}}$ from which we derive Eq. (13), [14]. The proof for the case of a single error term eigenvalue and eigenvector expansion can be found in [15] and is expanded to multiple error terms in [14], yielding Eq. (13). A similar analysis of the DLT-LAF method is obtained by perturbing equations (10) and (11) with both spatial and derivative noise, [14].

To verify the accuracy of the theoretical analysis we present the results of this expansion vs. the results of a simulated statistical analysis. We restrict our attention to one test case homography, given by

$$\mathrm{H} = \begin{pmatrix} -0.9527 & 3.6709 & 292.9865 \\ 2.4726 & 0.5011 & 209.3957 \\ -0.0007 & 0.0007 & 0.5463 \end{pmatrix}.$$

The simulated results were obtained by estimating the homography using the DLT-LAF method from a minimal set of two pairs of correspondences. In each trial, zero mean white Gaussian noise with varying standard deviation σ_s and zero mean white Gaussian noise with fixed standard deviation $\sigma_d = 0.2$ were added to the centers of the patches and to the derivatives of each correspondence pair. For each σ_s, 50,000 trials were conducted to determine the statistics of each entry in H. It can be seen in Fig. 1 that the resulting approximated statistics are close to the simulated statistics, making it a viable tool for comparison.

4.2. Statistical comparison between the DLT and DLT-LAF

The analysis above was used in order to provide analytic tools for comparing the performance of the DLT algorithm and the

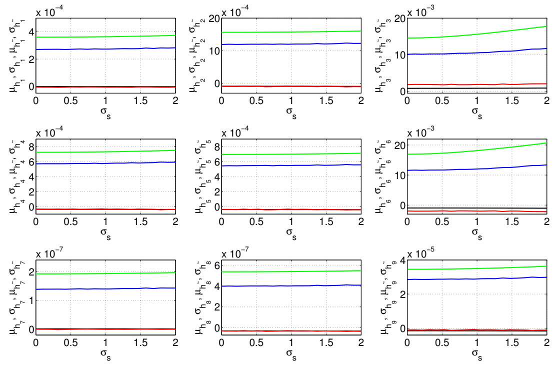

Fig. 1. Second order perturbation analysis mean and standard deviation (black and green lines, respectively) vs. simulated statistics mean and standard deviation (red and blue lines, respectively) performed on a minimal solution DLT-LAF.

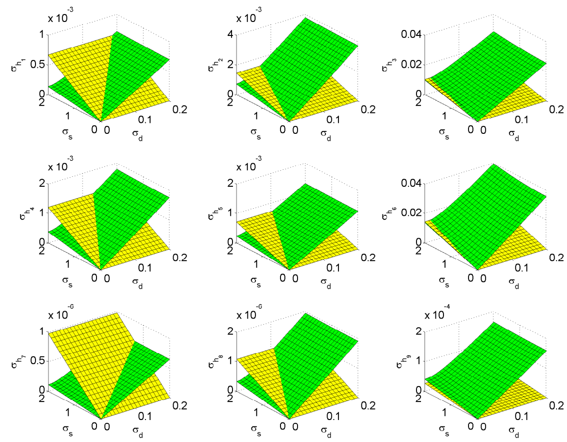

Fig. 2. Minimal solution standard deviation obtained using second order perturbation analysis. DLT (yellow) vs. DLT-LAF (green).

DLT-LAF algorithm in cases where the solution is obtained from a minimal set. We compare the estimation error in each entry of the homography using the derived perturbation analysis. In order to illustrate the results, we again use the previously given test case H. The standard deviations of the homography entries are given in Figure 2, as functions of varying additive noise levels on both the coordinates of the patch centers and the homography derivatives. It can be seen from the first two columns of Figure 2 that the estimation error of the DLT-LAF is lower than that of the DLT for reasonably low derivative noise. However, the third column, which represents translation and homogeneous coordinate normalization is more susceptible to derivative noise. This result is consistent over many experiments we have conducted and can be explained by noting that the proposed method estimates all the entries of the homography but incorporates more "derivative" information than spatial information. Without additional weighting of the measurements, we can expect such translation errors.

5. EXAMPLES

This section describes the experiments conducted to prove the applicability of the proposed method. The images used in the experiments were taken from the Graffiti dataset. Image patches were extracted using MSER and correspondences were established using affine graph matching [16], resulting in a set of local affine frames. Having established these affine relations, a RANSAC algorithm is used to find the best possible homography with probability of 0.99. The experimental results (seen in Figure 3) show that homography estimation from local affine frames performs very well, using only the information provided by matching two local affine frames.

(a) (b)

Fig. 3. Homography estimation using local affine frames applied to the Graffiti dataset. Source images transformed and overlaid over target images. (a) #1 over #4, (b) #2 over #3.

6. CONCLUSIONS

We have proposed a simple, minimally parameterized linear method for homography estimation from correspondences of local affine frames. The proposed method provides an alternative to the unavailability of point correspondence homography estimation, in cases where as few as two local affine frames are available. The simplicity of the method makes it ideal as a basic hypothesis of the robust RANSAC estimator. Unlike other conic based methods, the proposed method does not place restrictions on the physical model and gives a unique solution.

7. REFERENCES

[1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[2] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, Cambridge, MA, USA, 1993.

[3] Q. Luong O. Faugeras and T. Papadopoulou, *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*, MIT Press, Cambridge, MA, USA, 2001.

[4] R. I. Hartley and S. Kang, "Parameter-free radial distortion correction with centre of distortion estimation," in *In ICCV*, 2005, pp. 1834–1841.

[5] C. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," Tech. Rep. MSR-TR-99-21, Microsoft Research, June 2001.

[6] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, vol. 16, pp. 22–30, 1996.

[7] W. C. Hulls W. J. Wilson and G. S. Bell, "Relative end-effector control using cartesian position based visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 12, no. 5, pp. 684–696, 1996.

[8] M. M.Y. Chang Y. K. Yu, K. H. Wong and S. H. Or, "Recursive camera-motion estimation with the trifocal tensor," *Trans. Sys. Man Cyber. Part B*, vol. 36, no. 5, pp. 1081–1090, Oct. 2006.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[10] C. V. Jawahar P. K. Mudigonda and P. J. Narayanan, "Geometric structure computation from conics," in *ICVGIP*, 2004, pp. 9–14.

[11] M. Salo J. Kannala and J. Heikkil, "Algorithms for computing a planar homography from conics in correspondence," in *Proc. the 16th British Machine Vision Conference (BMVC 2006)*, Edinburgh, UK, 2006, vol. 1, pp. 77–86.

[12] K. Kser and R. Koch, "Differential spatial resection - pose estimation using a single local image feature.," in *ECCV (4)*, David A. Forsyth, Philip H. S. Torr, and Andrew Zisserman, Eds. 2008, vol. 5305 of *Lecture Notes in Computer Science*, pp. 312–325, Springer.

[13] P. Chen and D. Suter, "Error analysis in homography estimation by first order approximation tools: A general technique," *J. Math. Imaging Vis.*, vol. 33, pp. 281–295, March 2009.

[14] S. Zino and J. M. Francos, "Homography estimation using local affine frames and second order statistics," *submitted for publication*.

[15] J. H. Wilkinson, *The algebraic eigenvalue problem*, Monographs on numerical analysis. Clarendon Press, 1988.

[16] J. Bentolila and J. M. Francos, "Affine consistency graphs for image representation and elastic matching," in *Proc. IEEE International Conference on Image Processing*, Orlando, Florida, USA, 2012.