Accurate Closed-form Estimation of Local Affine Transformations Consistent with the Epipolar Geometry

Daniel Barath[1]
barath.daniel@sztaki.mta.hu
Jiri Matas[2]
matas@cmp.felk.cvut.cz
Levente Hajder[1]
hajder.levente@sztaki.mta.hu

[1] DEVA Research Laboratory
MTA SZTAKI, Budapest, Hungary

[2] Centre for Machine Perception,
Department of Cybernetics
Czech Technical University,
Prague, Czech Republic

Abstract

For a pair of images satisfying the epipolar constraint, a method for accurate estimation of local affine transformations is proposed. The method returns the local affine transformation consistent with the epipolar geometry that is closest in the least squares sense to the initial estimate provided by an affine-covariant detector. The minimized L_2 norm of the affine matrix elements is found in closed-form. We show that the used norm has an intuitive geometric interpretation.

The method, with negligible computational requirements, is validated on publicly available benchmarking datasets and on synthetic data. The accuracy of the local affine transformations is improved for all detectors and all image pairs. Implicitly, precision of the tested feature detectors was compared. The Hessian-Affine detector combined with ASIFT view synthesis was the most accurate.

1 Introduction

The paper addresses the problem of precise estimation of local affine transformations in rigid 3D scenes[1]. Computer vision problems addressed by exploiting local features, e.g. structure-from-motion, commonly rely on point-to-point correspondences. Using the full local affine transformation has only become more popular in the last decade. Matas et al. [16] showed that local affine transformations facilitate two-view matching. Köser and Koch [12] proved that the 3D camera pose estimation is possible if the corresponding affinity and location of only one patch is given. Köser [11] showed that 3D points can be precisely triangulated from local affinities. Bentolila *et al.* [5] proved that affine transformations give constraints for estimating the epipoles in the images. Current 3D reconstruction pipelines use point correspondences as well as patches [6, 7, 24] in order to compute realistic 3D models of real-world objects. If the epipolar geometry is known, a homography can be estimated from a single local affinity [1]. Barath *et al.* [2] showed that there is a one-to-one relationship between the surface normal and the local affinity.

[1] The generalization to multiple rigid motions each satisfying a different epipolar constraint is straightforward.

Pages 11.1-11.12
DOI: https://dx.doi.org/10.5244/C.30.11

The main goal of the paper is to show how to optimally correct local affine transformations between two frames, in the least squares sense, if the fundamental matrix F is known. The fundamental matrix can either be estimated from the local affine transformations [5, 24] to be refined or from point-to-point correspondences [8]. In calibrated set-ups, F is available.

The refinement of the translation part has been solved by Hartley and Sturm [9] who exploit the fact that point locations have to satisfy the epipolar geometry: if a point is given in the first image, its correspondence in the second frame must lie on its epipolar line [10]. The closest, in the least squares sense, locations are computed as the roots of a polynomial of degree 6. The method proposed in this paper can be seen as an extension of the Hartley and Sturm method as we consider the full local affinity and present two additional constraints induced by the epipolar geometry.

Local affine transformations are commonly provided by three types of affine-covariant detectors. The first group, including MSER [15], estimates full local affine transformations directly. The second group optimizes the initial estimates – both Harris-Affine [17] and Hessian-Affine [18] perform the so-called Baumberg iteration [3] in order to obtain high-quality affinities. Finally, some methods generate synthesized views related by affine transformations and feature detectors are applied to these images. By combining the estimates of the detector with the transformation related to the current synthetic view, a local affinity is given for each point correspondence. The most frequently used combined view synthesizer and feature detector is the Affine SIFT (ASIFT) [23]. However, affine version of commonly used detectors like SURF [4], ORB [25], BRISK [13], etc. can easily been constructed using the synthesizer part of ASIFT. Matching On Demand with view Synthesis [20] (MODS) is a recently proposed method that obtains a mixture of MSER, ORB and Hessian-Affine points and does as little view-synthesizing as required to detect a predefined number of point pairs.

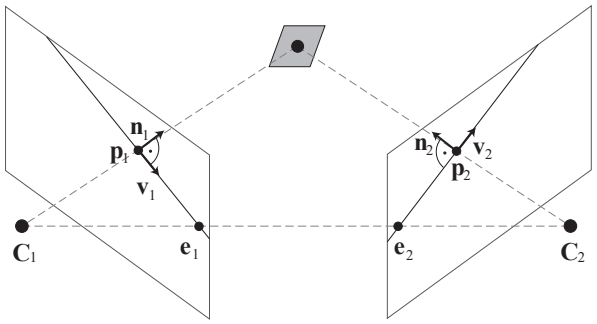
The contributions of the paper are: the introduction of two novel constraints for local affine transformations which make them consistent with the epipolar geometry (EG), and the algorithm to estimate an EG-L_2-Optimal (EG-L_2-Opt) affine transformation in the least squares (LSQ) sense by enforcing the proposed constraints. It is also proven that the LSQ optimization of the parameters has geometric and algebraic interpretations. We show experimentally that the EG-L_2-Opt procedure improves the accuracy of the output of all affine-covariant feature detector. As a side-effect, we determine the accuracy of affine-covariant feature detectors using ground truth data.

2 EG-L_2-Optimal Local Affine Transformation

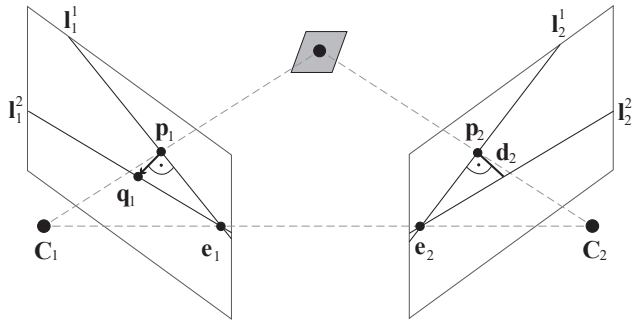
First, we discuss how to estimate an affine transformation at each corresponding point pair. Next, the compatibility constraints between an affine transformation and the fundamental matrix are presented. Finally, the computation of the EG-L_2-Opt transformation is discussed.

Local affine transformation. It is an open question how to get a good quality affine transformation related to each point pair in a real-world environment. We propose to use affine-covariant feature detectors [18] which obtain both the point locations and the affine transformations at the same time. Possibilities include ASIFT [23], MODS [20], Harris-Affine [19], Hessian-Affine [19], etc. These feature detectors provide an affine transformation for every i-th point $\mathbf{p}_k^i = [x_k^i \quad y_k^i]^T$ ($i \in [1,n]$) on the k-th image ($k \in 1,2$) as $\mathbf{A}_k^i$. The transformation $\mathbf{A}^i$ mapping $\mathbf{A}_1^i$ into $\mathbf{A}_2^i$ is obtained as

$$\mathbf{A}^i = \mathbf{A}_2^i (\mathbf{A}_1^i)^{-1}. \tag{1}$$



(a) The compatibility constraint for orientation states that $\mathbf{A}\mathbf{v}_1 || \mathbf{v}_2$ which is equivalent to $\mathbf{A}^{-T}\mathbf{n}_1 || \mathbf{n}_2$.



(b) The compatibility constraint for scale states that the ratio of $||\mathbf{p}_1 - \mathbf{q}_1||_2$ and d_2 determines the scale of the related local affine transformation perpendicular to the epipolar line.

Figure 1: EG-Consistency compatibility constraints for orientation and scale. Matrix $\mathbf{A}$ is the affine transformation, vectors $\mathbf{v}_k$ and n_k are the direction and normal of epipolar line on which point $\mathbf{p}_k$ lie in the k-th image ($k \in \{1,2\}$).

Affine compatibility – Translation. The last column of matrix $\mathbf{A}$ is responsible for the translation between the related point pair. It is shown by Hartley and Sturm [9] that it can be refined in an optimal way in the least squares sense. Their method minimizes the Euclidean distance between the original and refined positions. Then the resulting point locations are fully consistent with the epipolar geometry.

Affine compatibility – Orientation. Affine transformation $\mathbf{A}$ is considered as its left 2×2 submatrix in the following sections.

Suppose that the fundamental matrix $\mathbf{F}$ and an affine transformation $\mathbf{A}$ related to the corresponding point pair $\mathbf{p}_1$ and $\mathbf{p}_2$ are given. It is trivial that $\mathbf{A}$ is compatible with $\mathbf{F}$ only if it transforms the direction $\mathbf{v}_1$ of the related epipolar line $\mathbf{l}_1$ (on which $\mathbf{p}_1$ lies) on the first image to that of the second one $\mathbf{v}_2$. This means that $\mathbf{A}\mathbf{v}_1 \parallel \mathbf{v}_2$. It is well-known in computer graphics [26] that the direction of the normal after affine transformation is obtained as $\mathbf{A}^{-T}\mathbf{n}_1$. Therefore, formula $\mathbf{A}\mathbf{v}_1 \parallel \mathbf{v}_2$ is equivalent to

$$\mathbf{A}^{-\mathbf{T}}\mathbf{n}_1 = \beta \mathbf{n}_2, \tag{2}$$

where $\mathbf{n}_k$ and β are the normal of the k-th epipolar line ($k \in \{1,2\}$) and the scale between vectors $\mathbf{A}^{-T}\mathbf{n}_1$ and $\mathbf{n}_2$, respectively. This is visualized in Fig. 1(a).

Affine compatibility – Scale. It is shown in this section how scale β between vectors $A^{-T}\mathbf{n}_1$ and $\mathbf{n}_2$ is determined by the epipolar geometry.

Suppose that corresponding homogeneous point pair $\mathbf{p}_1 = [x_1 \quad y_1 \quad 1]^T$ and $\mathbf{p}_2 = [x_2 \quad y_2 \quad 1]$ are given. Let $\mathbf{n}_1 = [n_1^x \quad n_1^y]^T$ and $\mathbf{n}_2 = [n_2^x \quad n_2^y]^T$ be the normal directions of epipolar lines $\mathbf{l}_1^1 = \mathbf{F}^T\mathbf{p}_2 = [l_1^{1,a} \quad l_1^{1,b} \quad l_1^{1,c}]^T$ and $\mathbf{l}_2^1 = \mathbf{F}\mathbf{p}_1 = [l_2^{1,a} \quad l_2^{1,b} \quad l_2^{1,c}]^T$, respectively. Then the task is to define how the affine transformation $\mathbf{A}$ transforms the length of $\mathbf{n}_1$. In order to determine this scale factor let us introduce a new point as $\mathbf{q}_1 = \mathbf{p}_1 + \gamma\mathbf{n}_1$, where γ is an arbitrary scalar value. This new point determines an epipolar line $\mathbf{l}_2^2 = [l_2^{2,a} \quad l_2^{2,b} \quad l_2^{2,c}]^T$ on the second image as follows: $\mathbf{l}_2^2 = \mathbf{F}q_1 = \mathbf{F}(\mathbf{p}_1 + \gamma\mathbf{n}_1)$. Then scale β is given as the ratio of distances $d_1 = ||\mathbf{p}_1 - \mathbf{q}_1||_2$ and d_2 where d_2 is the distance between line $\mathbf{l}_2^2$ and point $\mathbf{p}_2$. The problem is visualized in Fig. 1(b) in detail. The calculation of d_2 is written by Eq. 3.

$$d_2 = \frac{|(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)x_2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)y_2 + l_2^{1,c} + f_{31} n_1^x + f_{32} n_1^y|}{\sqrt{(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2}} \tag{3}$$

It is known that point $\mathbf{p}_2$ lies on $\mathbf{l}_2^1$, which can be written as $l_2^{1,a}x_2 + l_2^{1,b}y_2 + l_2^{1,c} = 0$. This fact reduces Eq. 3 to Eq. 4.

$$d_2 = \frac{|(\gamma f_{11} n_1^x + \gamma f_{12} n_1^y)x_2 + (\gamma f_{21} n_1^x + \gamma f_{22} n_1^y)y_2 + f_{31} n_1^x + f_{32} n_1^y|}{\sqrt{(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2}} \tag{4}$$

In order to determine β, the observed point $\mathbf{q}_1$ has to be moved infinitely close to $\mathbf{p}_1$ ($\gamma \to 0$). This is written by Eq. 5.

$$\beta^2 = \lim_{\gamma \to 0} \frac{\gamma^2}{d_2^2} = \lim_{\gamma \to 0} \frac{((l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2)}{|(f_{11} n_1^x + f_{12} n_1^y)x_2 + (f_{21} n_1^x + f_{22} n_1^y)y_2 + f_{31} n_1^x + f_{32} n_1^y|^2} \tag{5}$$

After elementary modifications the final formula for scale β is given by Eq. 6.

$$\beta = \frac{\sqrt{l_2^{1,a} l_2^{1,a} + l_2^{1,b} l_2^{1,b}}}{|s_1 x_2 + s_2 y_2 + s_3|}, \quad s_i = f_{i1} n_1^x + f_{i2} n_1^y, \quad i \in \{1,2,3\}. \tag{6}$$

The EG-L_2-Opt affinity. Suppose that an observed affine transformation $\mathbf{A}'$ is given. Then let us denote that by

$$\mathbf{A}' = \begin{bmatrix} a_1' & a_2' \\ a_3' & a_4' \end{bmatrix}. \tag{7}$$

The task is to find an $\mathbf{A}$ where

$$||\mathbf{A} - \mathbf{A}'||_2^2 \tag{8}$$

is minimal and $\mathbf{A}^{-T}\mathbf{n}_1 = \beta \mathbf{n}_2$ (Eq. 2). In order to avoid inversion, it can be reformulated as $\mathbf{n}_1 = \beta \mathbf{A}^T \mathbf{n}_2$. Note that the validity of L_2 norm is discussed later in Section. 3.

Scale β can be calculated as it is proposed in the previous section (Eq. 6). Therefore, condition

$$\mathbf{n}_1 - \beta \mathbf{A}^T \mathbf{n}_2 = 0 \tag{9}$$

is linear in the parameters of the affine transformation $\mathbf{A}$. Eq. 9 yields one equation for each coordinate (x and y) as follows:

$$n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3 = 0, \quad n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4 = 0. \tag{10}$$

Let us introduce a cost function J applying the constraints defined in Eqs. 8, 10. Using Lagrange multipliers, the cost function is as follows:

$$J(\mathbf{A}, \lambda_1, \lambda_2) = \frac{1}{2} \sum_{i=1}^{4} (a_i - a_i')^2 +$$
$$\lambda_1 (n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3) + \lambda_2 (n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4), \tag{11}$$

where λ_1 and λ_2 are the Lagrange multipliers. Eq. 8 yields non-negative values. Therefore, the optimal solution is given by the partial derivatives of J:

$$\frac{\partial J}{\partial a_1} = a_1 - a_1' - \beta n_2^x \lambda_1 = 0, \quad \frac{\partial J}{\partial a_2} = a_2 - a_2' - \beta n_2^x \lambda_2 = 0,$$
$$\frac{\partial J}{\partial a_3} = a_3 - a_3' - \beta n_2^y \lambda_1 = 0, \quad \frac{\partial J}{\partial a_4} = a_4 - a_4' - \beta n_2^y \lambda_2 = 0,$$
$$\frac{\partial J}{\partial \lambda_1} = n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3 = 0, \quad \frac{\partial J}{\partial \lambda_2} = n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4 = 0.$$

This is an inhomogeneous, linear system of equations which can be written in form $\mathbf{Cx} = \mathbf{b}$, where $\mathbf{x} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & \lambda_1 & \lambda_2 \end{bmatrix}^T$, $\mathbf{b} = \begin{bmatrix} a'_1 & a'_2 & a'_3 & a'_4 & -n_1^x & -n_1^y \end{bmatrix}^T$, and $\mathbf{C}$ are the vector of the unknown parameters, inhomogeneous part, and coefficient matrix, respectively. $\mathbf{C}$ is as follows:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & -\beta n_2^x & 0 \\ 0 & 1 & 0 & 0 & 0 & -\beta n_2^x \\ 0 & 0 & 1 & 0 & -\beta n_2^y & 0 \\ 0 & 0 & 0 & 1 & 0 & -\beta n_2^y \\ -\beta n_2^x & 0 & -\beta n_2^y & 0 & 0 & 0 \\ 0 & -\beta n_2^x & 0 & -\beta n_2^y & 0 & 0 \end{bmatrix}.$$

The solution is $\mathbf{x} = \mathbf{C}^{-1}\mathbf{b}$. See Alg. 1 for the pseudo-code of the proposed algorithm.

Algorithm 1 EG-L_2-Optimal Affine Transformation

procedure CORRECTAFFINETRANSFORMATION
  Input:
  $\mathbf{F}$ – fundamental matrix.
  $\mathbf{p}_1$, $\mathbf{p}_2$ – corresponding point pair.
  $\mathbf{A}'$ – measured affine transformation.
  Output:
  $\mathbf{A}$ – optimally refined affine transformation.
  Algorithm:
  $\mathbf{l}_1 := \mathbf{F}^T\mathbf{p}_2$; $\mathbf{l}_2 := \mathbf{F}\mathbf{p}_1$; $\mathbf{n}_1 := [l_1^a; l_1^b]/|[l_1^a; l_1^b]|_2$; $\mathbf{n}_2 := [l_2^a; l_2^b]/|[l_2^a; l_2^b]|_2$;
  $s_1 := f_{11}n_1^x + f_{12}n_1^y$; $s_2 := f_{21}n_1^x + f_{22}n_1^y$; $s_3 := f_{31}n_1^x + f_{32}n_1^y$;
  $\beta := (1/|s_1x_2 + s_2y_2 + s_3|)\sqrt{l_2^a l_2^a + l_2^b l_2^b}$;
  $\mathbf{C} := eye(6,6)$; $\mathbf{C}_{55} := 0$; $\mathbf{C}_{66} := 0$;
  $\mathbf{C}_{15} := -\beta n_2^x$; $\mathbf{C}_{26} := -\beta n_2^x$; $\mathbf{C}_{35} := -\beta n_2^y$; $\mathbf{C}_{46} := -\beta n_2^y$;
  $\mathbf{C}_{51} := -\beta n_2^x$; $\mathbf{C}_{62} := -\beta n_2^x$; $\mathbf{C}_{53} := -\beta n_2^y$; $\mathbf{C}_{64} := -\beta n_2^y$;
  $\mathbf{b} := [a'_1; a'_2; a'_3; a'_4; -n_1^x; -n_1^y]$;
  $\mathbf{x} := \mathbf{C}^{-1}b$;
  $\mathbf{A} := [x_1, x_2; x_3, x_4]$;

3 Is LSQ Minimization of the Affine Parameters Correct?

It is shown in this section that the minimization of the Frobenious-norm has both algebraic and geometric interpretations for local affine transformations.

Matrix $\mathbf{A}$ without the translation is a 2×2 linear transformation, therefore, it is determined by two points. (The projection of the origin remains the same.) Let us choose points

$\begin{bmatrix}1 & 0\end{bmatrix}^T$ and $\begin{bmatrix}0 & 1\end{bmatrix}^T$. Then the minimizing formula for the former one is as follows:

$$\left\|\mathbf{A}\begin{bmatrix}1\\0\end{bmatrix}-\mathbf{A}'\begin{bmatrix}1\\0\end{bmatrix}\right\|_2^2=\left\|(\mathbf{A}-\mathbf{A}')\begin{bmatrix}1\\0\end{bmatrix}\right\|_2^2=$$
$$\left\|\begin{bmatrix}a_1-a_1' & a_2-a_2'\\ a_3-a_3' & a_4-a_4'\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right\|_2^2=\left\|\begin{bmatrix}a_1-a_1'\\ a_3-a_3'\end{bmatrix}\right\|_2^2=$$
$$(a_1-a_1')^2+(a_3-a_3')^2=0. \tag{12}$$

The minimization for the second point is fairly similar as

$$\left\|\mathbf{A}\begin{bmatrix}0\\1\end{bmatrix}-\mathbf{A}'\begin{bmatrix}0\\1\end{bmatrix}\right\|_2^2=\left\|\begin{bmatrix}a_2-a_2'\\ a_4-a_4'\end{bmatrix}\right\|_2^2=(a_2-a_2')^2+(a_4-a_4')^2=0. \tag{13}$$

By combining both Eqs. 12, 13 the Frobenious-norm of difference matrix $\mathbf{A}-\mathbf{A}'$ is obtained. As a consequence, *minimizing the Frobenious-norm of the difference matrix is equivalent to the optimization of its effect on points.* Therefore, the squared differences of the parameters have both algebraic and geometric interpretations.

4 Experimental Results

First, we show how to get ground truth affine transformations. Then we test the proposed theory on both synthesized and real-world data.

4.1 Affine Transformation from Homography

Local affine transformation $\mathbf{A}$ can be derived from the parameters of the homography [21]. The last column of the affine transformation $\mathbf{A}$ determines the translation. Suppose that homography $\mathbf{H}$ is given. The correspondence between homogeneous points $\mathbf{p}_1=[x_1 \quad y_1 \quad 1]^T$ and $\mathbf{p}_2=[x_2 \quad y_2 \quad 1]^T$ is written as $\mathbf{H}\mathbf{p}_1\sim\mathbf{p}_2$. The linear part (left 2×2 submatrix) of the affine parameters can be written as the partial derivatives of this perspective transformation:

$$a_{1j}=\frac{h_{1j}-h_{3j}x_2}{s} \quad a_{2j}=\frac{h_{2j}-h_{3j}y_2}{s} \quad j\in\{1,2\}, \tag{14}$$

where $s=\mathbf{h}_3^T\mathbf{p}_1$[2]. This is described in dept in [1]. The translation part of $\mathbf{A}$ is determined by the point locations. During the experiments, the ground truth local affine transformations are calculated using this relationship from the ground truth homographies.

4.2 Synthesized tests

For synthesized testing, two perspective cameras are generated by their projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. Their positions are randomized in the plane $Z=60$ which is parallel to plane XY. Both cameras point towards the origin. Their common focal length and principal point are 600 and $[300 \quad 300]^T$, respectively. Then 50 spatial points are generated on a random plane that passes through the origin, and the points are projected onto the cameras. The ground truth affine transformation related to each point is calculated using Eq. 14 based on the homography. Tests are repeated 500 times at every noise level.

[2]Parameter $\mathbf{h}_i^T$ is the i-th row of H.

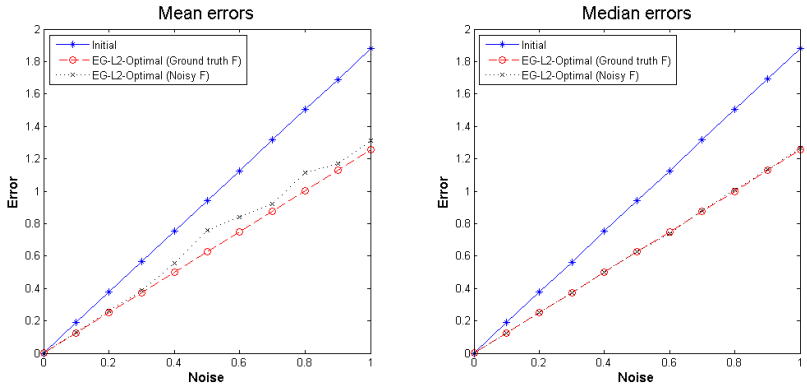


Figure 2: Error of the original and optimal affine transformations w.r.t. the noise level. The average L_2 distance from the ground truth transformation is plotted as a function of the σ value of the Gaussian noise (in pixels). The noise is added to the affine transformations and point locations. (Red Curve) The ground truth fundamental matrix is used. (Black Curve) The fundamental matrix is estimated using the noisy point correspondences by the normalized 8-point algorithm followed by a Levenberg-Marquardt optimization minimizing the symmetric epipolar error. In the median figure, the black and red curves coincide.

Fig. 2 shows the mean (left) and median (right) distances of the original noisy transformations and that of the optimal ones w.r.t. the ground truth data. Zero-mean Gaussian noise is added to the elements of the affine transformations and point locations. The error (vertical axis) is the mean of the L_2-norms of the difference matrices of the obtained and ground truth data. The horizontal axis shows the σ value of the noise.

The red curve shows the error if the ground truth fundamental matrix is used. For the black curve, the fundamental matrix is estimated using the noisy point locations by the normalized 8-point algorithm followed by Levenberg-Marquardt optimization minimizing the symmetric epipolar error. The refined transformations are closer to the ground truth matrices than the original ones. There is no significant difference between the median and mean plots and between results obtained on the ground truth and the estimated fundamental matrix.

The processing time of the proposed method is negligible since it consists of a few operations. It is calculated in C++ in around 0.04 milliseconds per point on a 2.3 GHz PC.

4.3 Tests on Real Data

The proposed theory is tested on the annotated AdelaideRMF dataset[3] and on image pairs "graffiti"[4], "stairs" and "glasscasea" (see Fig. 3). In the last three pairs, we manually marked point correspondences and assigned them to planes. The ground truth homographies are computed using the annotated point correspondences[5].

Several affine-covariant feature detectors are run on all image pairs. The following affine-covariant detectors are applied: AAKAZE, ABRISK, AORB, ASIFT, ASURF, AHessian-Affine[6], MODS[7], MSER, Harris-Affine and Hessian-Affine[8].

[3] Available at http://cs.adelaide.edu.au/~hwong/doku.php?id=data
[4] Available at http://www.robots.ox.ac.uk/~vgg/research/affine/
[5] Available at http://web.eee.sztaki.hu/home4/node/56
[6] ASIFT is downloaded from http://www.ipol.im/pub/art/2011/my-asift. The "A-forms" of AKAZE, BRISK, ORB, SIFT, SURF, Hessian-Affine are obtained by replacing SIFT in the view-synthesizer.
[7] MODS is downloaded from http://cmp.felk.cvut.cz/wbs
[8] MSER, Har-Aff, and Hes-Aff downloaded from http://www.robots.ox.ac.uk/~vgg/research/

Detector		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	mean	median
AAKAZE	Observed	0.26	0.30	0.17	0.30	0.26	0.18	0.25	0.62	0.38	0.30	0.26
	EG-L_2-Opt	0.21	0.22	0.12	0.19	0.19	0.14	0.16	0.54	0.26	0.23	0.19
ABRISK	Observed	0.28	0.33	0.27	0.38	0.28	0.30	0.28	1.31	0.31	0.42	0.30
	EG-L_2-Opt	0.21	0.25	0.19	0.24	0.22	0.18	0.18	0.50	0.20	0.24	0.21
AHES-AFF	Observed	0.19	0.23	0.18	0.20	0.14	0.17	0.21	0.24	0.22	0.20	0.20
	EG-L_2-Opt	**0.14**	0.17	0.11	**0.13**	**0.09**	0.11	0.13	**0.14**	**0.15**	**0.13**	**0.13**
AORB	Observed	0.34	0.34	0.15	0.45	0.23	0.24	0.27	-	0.28	0.29	0.28
	EG-L_2-Opt	0.27	0.28	0.10	0.29	0.17	0.18	0.18	-	0.20	0.20	0.19
ASIFT	Observed	0.27	0.28	0.27	0.26	0.21	0.22	0.27	0.23	0.29	0.26	0.27
	EG-L_2-Opt	0.20	0.21	0.15	0.17	0.14	0.17	0.16	0.17	0.18	0.17	0.17
ASURF	Observed	0.23	0.27	0.17	0.30	0.22	0.17	0.25	0.26	0.27	0.24	0.25
	EG-L_2-Opt	0.18	0.20	0.11	0.21	0.16	0.12	0.17	0.18	0.19	0.18	0.18
HAR-AFF	Observed	0.24	0.25	0.15	0.24	0.16	0.27	0.20	0.38	0.28	0.24	0.24
	EG-L_2-Opt	0.18	0.18	**0.09**	0.19	0.12	0.19	0.13	0.35	0.17	0.16	0.18
HES-AFF	Observed	0.24	0.22	0.20	0.22	0.13	0.20	0.19	-	0.24	0.21	0.21
	EG-L_2-Opt	0.17	**0.16**	0.10	0.17	**0.09**	**0.09**	**0.12**	-	**0.15**	**0.13**	0.14
MODS	Observed	0.29	0.40	0.23	0.31	0.26	0.25	0.61	0.24	0.47	0.34	0.29
	EG-L_2-Opt	0.20	0.25	0.13	0.22	0.19	0.17	0.42	0.19	0.32	0.23	0.20
MSER	Observed	0.42	0.69	0.46	0.34	0.29	0.31	0.42	0.51	0.34	0.42	0.42
	EG-L_2-Opt	0.24	0.32	0.23	0.25	0.20	0.22	0.25	0.31	0.21	0.25	0.24

Table 1: Errors of the affine-covariant feature detectors "Observed" and their "EG-L_2-Opt" corrections. The error is the mean of the L_2-norms of the difference matrices of the obtained and ground truth affine transformations. Test pairs: (a) hartley, (b) johnsonnb, (c) neem, (d) sene, (e) oldclassicswing, (f) ladysymon (g) graffiti (h) stairs (i) glasscasea

	AAKAZE	ABRISK	AHES-AFF	AORB	ASIFT	ASURF	HAR-AFF	HES-AFF	MODS	MSER
Inliers	239	110	1420	145	2082	837	64	73	941	78
Time	81.91	81.38	89.30	86.39	81.34	84.00	4.10	3.22	52.92	0.41

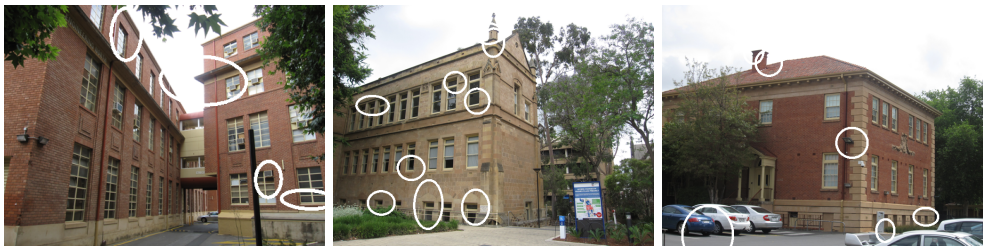
Table 2: The average number of inliers – correspondences lying on an annotated homography – for different feature detectors. Processing times in seconds on an Intel Core4Quad 2.33 GHz PC with 4 GByte memory using only a single core.[10]

Correspondences of features points obtained by matching [14] are assigned to the closest annotated homography. The distance between a point pair and a homography is defined as the re-projection error ($\mathbf{Hp}_1 \sim \mathbf{p}_2$). If a correspondence is farther from its closest homography than 1.0 px, it is discarded from the evaluation since the ground truth affine transformation for such correspondence can not be calculated. For the remaining correspondences, ground truth affine transformations are calculated using Eqs. 1. Fundamental matrices are computed by the normalized 8-point algorithm followed by a numerical refinement stage minimizing symmetric epipolar error by Levenberg-Marquardt optimization [22].

The errors are shown in Table 1. The error is the mean of the L_2-norms of the difference matrices of the obtained and ground truth data. Each column represents a test pair except the last two ones which show the mean and median errors. The corresponding odd and even rows visualize the mean error of the observed affine transformations given by each feature detector and that of the refined, EG-L_2-Opt ones. The error metric is the same as used for the synthesized tests. Every method is applied using their default parameterization. The median values show the same trend. The most important conclusion of these tests is that *the refined, EG-L_2-Opt affine transformations are always more accurate than the observed ones.*

Hessian-Affine augmented with the view-synthesizer of ASIFT (denoted by AHES-AFF)

(a) "hartley" (b) "johnsonnb" (c) "neem"



(d) "sene" (e) "oldclassicswing" (f) "ladysymon"

(g) "graffiti" (h) "stairs" (i) "glasscasea"

Figure 3: The first frames of the selected image pairs with a few local affinities each represented by an ellipse.

obtains the most accurate affine transformations (see Table 1) and provides many point correspondences as well (see Table 2). If the required number of correspondences needs not be high, Hessian-Affine without view-synthesizing might be the method of choice since it is significantly faster and its accuracy is nearly the same.

4.4 Improvements on Homography and Surface Normal Estimates

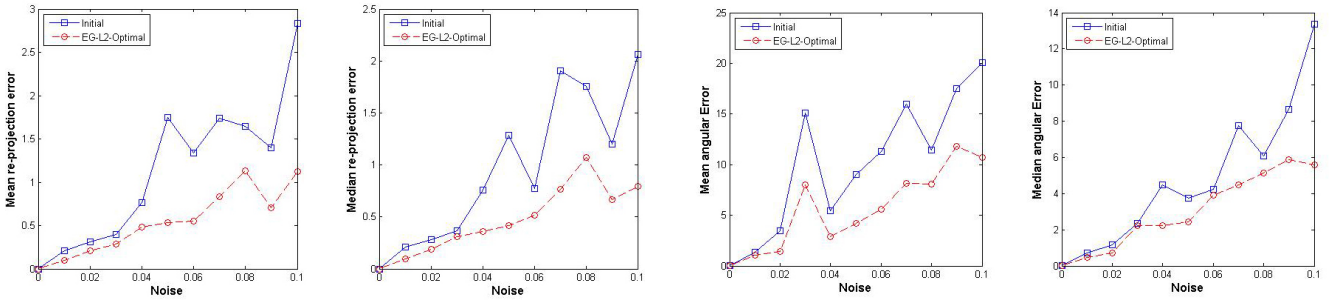
This section presents experiments showing that EG-L_2-Opt affinities lead to more accurate homography and surface normal estimates.

For homography estimation the same synthetic scene is constructed as in Section 4.2: a random plane is generated and sampled at ten locations which are projected onto the cameras. The method proposed by Koeser [11] is applied to one of the ten correspondences and the related affinity. Tests are repeated 500 times for every noise level. Fig 4(a) shows that homographies calculated from the EG-L_2-Opt refined data are the most accurate ones. The error metric is the mean re-projection error (in pixels) computed for the point locations.

For surface normal estimation, the technique proposed recently by Barath *et al.* [2] is performed. They show that a one-to-one relationship exists between an affine transformation and the related surface normal and introduce normal estimators. In our tests, the same testing

[10]Information in Table 2 is not assessing the precision of affine transformation, the main topic of the paper. It complements Table 1 in providing broader characterization of detector performance.

environment is used as proposed in [2] and FNE normal estimator[11] is applied to both the initial and EG-L_2-Opt affinities. Fig. 4(b) confirms that the proposed technique makes the surface normals more accurate.



(a) (b)

Figure 4: Mean, (a) left, and median, (a) right, re-projection errors (in pixels) of the homography estimation [11] applied to the noisy and the EG-L_2-Opt refined affinities. Mean, (b) left, and median, (b) right, angular errors (in degrees) of the surface normals estimated from the initial and EG-L_2-Opt refined affinities. The errors are plotted as the function of the σ value of the isotropic 6D zero-mean Gaussian noise.

5 Conclusions

We showed how to improve the accuracy of a local affine transformation obtained by an affine-covariant feature detector by considering the epipolar constraint. The proposed algorithm is optimal in the least squares sense. Its computational cost is negligible. The proposed least squares minimization has an intuitive geometric interpretation.

The introduced EG-L_2-Opt procedure is validated on real-world image pairs. It improves the accuracy of all tested affine-covariant detectors. On average, the error of the refined affinities is reduced to about 65%. The EG-L_2-Opt affinities improve the accuracy of surface normal and homography estimates as well.

As a side-effect, the experiments quantitatively compared the precision of affine-covariant feature detectors. The Hessian-Affine detector combined with the view-synthesizer of ASIFT obtains the most accurate affinities.

The source code is available at `http://web.eee.sztaki.hu/home4/node/56`

Acknowledgement

This work was partially supported by the Hungarian National Research, Development and Innovation Office under the grant VKSZ 14-1-2015-0072. J. Matas was supported by the GACR P103/12/G084 grant.

[11] Fast normal estimator (FNE) is downloaded from `http://web.eee.sztaki.hu/home4/node/53`

References

[1] D. Barath and L. Hajder. Novel ways to estimate homography from local affine transformations. In *11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, pages 432–443, 2016.

[2] D. Barath, J. Molnar, and L. Hajder. Novel methods for estimating surface normals from affine transformations. In *Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 316–337. Springer International Publishing, 2016.

[3] A. Baumberg. Reliable feature matching across widely separated views. In *Computer Vision and Pattern Recognition*, volume 1, pages 774–781. IEEE, 2000.

[4] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[5] J. Bentolila and J. M. Francos. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding*, 122:105–114, 2014.

[6] A. Bódis-Szomorú, H. Riemenschneider, and L. Van Gool. Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels. In *CVPR*, 2014.

[7] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[8] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[9] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding: CVIU*, 68(2):146–157, 1997.

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[11] K. Köser. *Geometric Estimation with Local Affine Frames and Free-form Surfaces*. Shaker, 2009.

[12] K. Köser and R. Koch. Differential spatial resection - pose estimation using a single local image feature. In *ECCV*, pages 312–325, 2008.

[13] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*, pages 2548–2555. IEEE, 2011.

[14] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision*, ICCV, pages 1150–1157, 1999.

[15] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC*, pages 36.1–36.10, 2002.

[16] J. Matas, S. Obdrzálek, and O. Chum. Local affine frames for wide-baseline stereo. In *ICPR, Quebec, Canada, August 11-15, 2002.*, pages 363–366, 2002.

[17] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *ECCV*, pages 128–142. Springer, 2002.

[18] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[19] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[20] D. Mishkin, J. Matas, and M. Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 141:81–93, 2015.

[21] J. Molnár and D. Chetverikov. Quadratic transformation for planar mapping of implicit surfaces. *Journal of Mathematical Imaging and Vision*, 48:176–184, 2014.

[22] J. Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*. Springer.

[23] J-M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[24] C. Raposo and J. P. Barreto. Theory and practice of structure-from-motion using affine correspondences. 2016.

[25] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. IEEE, 2011.

[26] K. Turkowski. Transformations of surface normal vectors. In *Tech. Rep. 22, Apple Computer*, 1990.