P-HAF: Homography Estimation using Partial Local Affine Frames

Daniel Barath
Machine Perception Research Laboratory, MTA SZTAKI, Kende utca 13-17, Budapest, Hungary
barath.daniel@sztaki.mta.hu

Keywords: Homography, Minimal Problem, Local Affine Transformation, Stereo Vision.

Abstract: We propose an algorithm, called P-HAF, to estimate planar homographies using partially known local affine transformations. This general theory is able to exploit the affine components obtained by the commonly used partially affine covariant detectors, such as SIFT or SURF, in a real time capable way. P-HAF as a minimal solver can estimate the homography using two SIFT correspondences, moreover, it can deal with any number of point pairs as an overdetermined system. It is validated both on synthesized and publicly available datasets that exploiting all information leads to more accurate estimates and makes multi-homography estimation less ambiguous.

1 INTRODUCTION

Estimating planar correspondences is a crucial part of several vision tasks e. g. robot vision (Zhou and Li, 2006; Chen et al., 2006), camera calibration (Zhang, 2000; Ueshiba and Tomita, 2003; Chuan et al., 2003), 3D reconstruction (Zhang and Hanson, 1996; Werner and Zisserman, 2002) and augmented reality applications (Prince et al., 2002). Between two views, a planar correspondence is described by a 3×3 homography matrix which is a $P^2 \rightarrow P^2$ perspective transformation. Even though the most popular estimation techniques are based on point correspondences (Hartley and Zisserman, 2003), a homography is estimable from line (Hartley and Zisserman, 2003), region (Tanacs et al., 2014), contour (Jain and Jawahar, 2006), or affine correspondences (Köser, 2009; Chum and Matas, 2012; Barath and Hajder, 2016). Most of these algorithms include data normalization (Hartley and Zisserman, 2003), and numerical optimization to minimize the effect of the noise.

In this paper, we assume that not only the point locations but several affine components and the fundamental matrix are known.[1] Local affinities are usually represented by elliptical features. We adapt the definition used in (Chum and Matas, 2012; Barath and Hajder, 2016) where a local affine transformation is defined as the first-order Taylor-approximation of the related the homography.

Local affine transformations have become more popular in the last decade. Matas et al. (Matas et al., 2002) presented that local affinities can support stereo matching. The 3D camera pose can also be estimated using a corresponding point pair and the related affinity as it is proposed by Köser and Koch (Köser and Koch, 2008). Staying on the topic of 3D reconstruction, these transformations can facilitate the recovery of spatial point coordinates (Köser, 2009). Current 3D reconstruction pipelines exploit point correspondences as well as patches (Furukawa and Ponce, 2010; Bódis-Szomorú et al., 2014; Raposo and Barreto, 2016) to compute realistic 3D models of real-world objects. Bentolila et al. (Bentolila and Francos, 2014) proved that affine transformations put constraints on the epipoles in stereo images. Barath et al. (Barath et al., 2016b) showed that a one-to-one relationship exists between the surface normal and the local affinity.

Even though local affine transformations are useful and can significantly improve the quality of the estimation, it is time consuming to recover them – e. g. by affine covariant detectors which cannot be applied in real time. Even so, most of the detectors obtain some part of these local affinities, such as SIFT (Lowe, 1999) or SURF (Bay et al., 2006) recovering the rotational and scale components. Therefore, using solely the translation part (the point location) causes information loss. The motivation of this research is to formulate a general theory about the usage of the affine components obtained by partially affine

[1] The pre-estimation of the fundamental matrix for rigid scenes using point correspondences is a usual step in computer vision pipelines. The proposed theory is straightforward to generalize for multiple rigid motions.

Barath D.
P-HAF: Homography Estimation using Partial Local Affine Frames.
DOI: 10.5220/0006130302270235
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 227-235
ISBN: 978-989-758-227-1

covariant feature detectors. The main contributions:

1. A general theory to exploit the affine components obtained by partially affine covariant detectors which is real time capable.
2. The proposed method estimates the homography from two SIFT correspondences if the fundamental matrix is known. To our knowledge, the minimum number of required point pairs was three.

1.1 Theoretical Background

Homography is defined in this paper as a mapping $P^2 \rightarrow P^2$ which maps each vector $p_1^i = [u_1^i \quad v_1^i \quad 1]^T$ to its corresponding location $p_2^i = [u_2^i \quad v_{i,2}^i \quad 1]^T$ as $[u_2^i \quad v_2^i \quad 1]^T \sim \mathbf{H}[u_1^i \quad v_1^i \quad 1]^T$. The lower and upper indices denote the index of the current image ($\in \{1,2\}$) and the number of the feature point ($i \in [1,n]$), respectively.

Homography and Fundamental Matrix. The well-known relationship

$$[e_2]_\times \mathbf{H} = \mathbf{F},$$

where

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix},$$

and $e_2 = [e_x \quad e_y]^T$ are the homography, fundamental matrix and epipole on the second image, respectively, decreases the degrees-of-freedom of the homography estimation to three as it is shown in (Barath and Hajder, 2016) in detail. As a consequence, homography $\mathbf{H}$ is determined by its last row (h_{31}, h_{32}, and h_{33}) as follows:

$$h_{1j} = e_x h_{3j} + f_{2j}, \quad h_{2j} = e_y h_{3j} + f_{1j}, \quad j \in \{1,2\}. \tag{1}$$

To form a three-point-solver, called 3PT in the further sections, Eq. 1 is substituted into the system given by $p_2 \sim \mathbf{H} p_1$. The obtained Direct Linear Transform-like (DLT) system is as

$$\begin{aligned} (u_1 e_x - u_1 u_2)h_{31} + (v_1 e_x - v_1 u_2)h_{32} + \\ (e_x - u_2)h_{33} = -u_1 f_{21} - v_1 f_{22} - f_{23}, \\ (u_1 e_y - u_1 v_2)h_{31} + (v_1 e_y - v_1 v_2)h_{32} + \\ (e_y - v_2)h_{33} = u_1 f_{11} + v_1 f_{12} + f_{13}. \end{aligned} \tag{2}$$

Note that one point pair yields only one equation since the fundamental matrix reduces the DoF of the correspondence problem to one since the point pairs have to lie on the related epipolar lines. Thus three correspondences are enough to estimate the homography.

Homography estimation using three correspondences and the fundamental matrix is well-known (Hartley and Zisserman, 2003).

Local Affine Transformation regarding to the i-th correspondence

$$\mathbf{A}^i = \begin{bmatrix} a_{11}^i & a_{12}^i & a_{13}^i \\ a_{21}^i & a_{22}^i & a_{23}^i \end{bmatrix} \tag{3}$$

is defined as the partial derivative of the related homography $\mathbf{H}$ (Köser, 2009; Chum and Matas, 2012; Molnár and Chetverikov, 2014) as follows:

$$\begin{aligned} a_{11}^i &= \frac{\partial u_2^i}{\partial u_1^i} = \frac{h_{11} - h_{31} u_2^i}{s}, \\ a_{12}^i &= \frac{\partial u_2^i}{\partial v_1^i} = \frac{h_{12} - h_{32} u_2^i}{s}, \\ a_{21}^i &= \frac{\partial v_2^i}{\partial u_1^i} = \frac{h_{21} - h_{31} v_2^i}{s}, \\ a_{22}^i &= \frac{\partial v_2^i}{\partial v_1^i} = \frac{h_{22} - h_{32} v_2^i}{s}, \end{aligned} \tag{4}$$

where $s = h_3[u_1^i, v_1^i, 1]^T$ and h_j is the j-th row of $\mathbf{H}$ ($j \in \{1,2,3\}$). These four parameters of $\mathbf{A}^i$ are responsible for horizontal and vertical scales, shear and rotation. The last column of $\mathbf{A}^i$ determines the translation as $a_{13}^i = u_2^i - u_1^i$ and $a_{23}^i = v_2^i - v_1^i$.

Homography $\mathbf{H}$ defines the correspondence between the coordinates in the first (u_1 and v_1) and second (u_2 and v_2) images as

$$u_2 = \frac{h_1^T [u_1, v_1, 1]^T}{h_3^T [u_1, v_1, 1]^T}, \quad v_2 = \frac{h_2^T [u_1, v_1, 1]^T}{h_3^T [u_1, v_1, 1]^T}.$$

2 HOMOGRAPHY ESTIMATION FROM AFFINE TRANSFORMATION

In this section, we show that the problem becomes much simpler if the epipolar geometry and local affine transformations are known.

2.1 Homography from Affinities

As it is shown in the Homography from Affine transformation and Fundamental matrix (HAF) method (Barath and Hajder, 2016) the estimation of a homography can be written in an inhomogeneous, linear form if at least one local affine transformation and the epipolar geometry is known. The coefficient matrix $\mathbf{C}$ is as follows:

$$\mathbf{C} = \begin{bmatrix} a_{11}^i u_1^i + u_2^i - e_x & a_{11}^i v_1^i & a_{11}^i \\ a_{12}^i v_1^i + u_2^i - e_x & a_{12}^i u_1^i & a_{12}^i \\ a_{21}^i u_1^i + v_2^i - e_y & a_{21}^i v_1^i & a_{21}^i \\ a_{22}^i v_1^i + v_2^i - e_y & a_{22}^i u_1^i & a_{22}^i \end{bmatrix} \tag{5}$$

The equation system can be formed as $\mathbf{C}y = d$, where vector $d = [f_{21}, f_{22}, -f_{11}, -f_{12}]$ is the inhomogeneous part while $y = [h_{31}, h_{32}, h_{33}]^T$ is the vector of the unknown parameters. The optimal solution in the least squares sense is given by $y = \mathbf{C}^\dagger d$ where $\mathbf{C}^\dagger$ is the Moore-Penrose pseudo-inverse of matrix $\mathbf{C}$. Note that augmenting this system with the formulas regarding to the point locations (Eqs. 2) leads to more robust estimation.

2.2 Affine Transformation Model

Let us denote the affine transformation related to the i-th ($i \in [1, N]$) point pair without the translation part as follows:

$$\mathbf{A}^i = \begin{bmatrix} a_{11}^i & a_{12}^i \\ a_{21}^i & a_{22}^i \end{bmatrix} = \begin{bmatrix} \cos(\alpha^i) & -\sin(\alpha^i) \\ \sin(\alpha^i) & \cos(\alpha^i) \end{bmatrix} \begin{bmatrix} s_x^i & w^i \\ 0 & s_y^i \end{bmatrix} = \begin{bmatrix} s_x^i \cos(\alpha^i) & w^i \cos(\alpha^i) - s_y^i \sin(\alpha^i) \\ s_x^i \sin(\alpha^i) & w^i \sin(\alpha^i) + s_y^i \cos(\alpha^i) \end{bmatrix} \quad (6)$$

Variables α^i, s_x^i, s_y^i, and w^i are the rotational angle, scales along x and y axes, and the shear parameters, respectively.

2.3 Homography from Partially Known Affine Transformation

In this section, it is shown that not the full local affinity is necessary for homography estimation, but their parts – obtained by e. g. SIFT or other partially affine covariant detector – can also be exploited. *In the rest of this paper the proposed method is called P-HAF as the abbreviation of Partial HAF.* Let us substitute Eqs. 6 into Eqs. 5 as

$$h_{31}\left(s_x \cos(\alpha^i) u_1^i + u_2^i - e_x\right) + h_{32} s_x^i \cos(\alpha^i) v_1^i + h_{33} s_x^i \cos(\alpha^i) = f_{21}, \quad (7)$$

$$h_{32}\left((w^i \cos(\alpha^i) - s_y^i \sin(\alpha^i)) v_1^i + u_2^i - e_x\right) + h_{31}(w^i \cos(\alpha^i) - s_y^i \sin(\alpha^i)) u_1^i + h_{33}(w^i \cos(\alpha^i) - s_y^i \sin(\alpha^i)) = f_{22}, \quad (8)$$

$$h_{31}\left(s_x^i \sin(\alpha^i) u_1^i + v_2^i - e_y\right) + h_{32} s_x^i \sin(\alpha^i) v_1^i + h_{33} s_x^i \sin(\alpha^i) = -f_{11}, \quad (9)$$

$$h_{32}\left((w^i \sin(\alpha^i) + s_y^i \cos(\alpha^i)) v_1^i + v_i^2 - e_y\right) + h_{31}(w^i \sin(\alpha^i) + s_y^i \cos(\alpha^i)) u_1^i + h_{33}(w^i \sin(\alpha^i) + s_y^i \cos(\alpha^i)) = -f_{12}. \quad (10)$$

These four equations contain the affine transformation in an easy-to-handle form. For a given part of the affinity, e. g. rotation and scale, the appropriate equations can be selected and used. After the selection, the given system is linear, inhomogeneous and can be solved as in Section 2.1.

2.4 Specialization to SIFT Features

The popular SIFT (Lowe, 1999) detector obtains rotation and scale covariant features, therefore, the proposed theory can be specialized to use SIFT. Beside the point locations the rotation and the scale is given for each feature point. After the matching process the related parts of the local affine transformation are as follows:

$$s = \frac{s_2}{s_1}, \quad \alpha = \alpha_2 - \alpha_1,$$

where s_1, s_2, α_1, and α_2 are the scales and angles on the two images, respectively. Here, we assume s as horizontal scale, thus only Eqs. 7, 9 have to be kept. Even though one SIFT correspondence yields three equations – one from the locations and two from the affinity –, the two regarding to the affine parts are linearly dependent. *As a consequence, two SIFT correspondences are enough for homography estimation* – and the system has been already overdetermined.

For $N \geq 2$ point pairs, an overdetermined, inhomogeneous, linear system is formed.

2.5 Normalization

As it is well-known, normalization of the input data is a usual and important part of homography estimation (Hartley and Zisserman, 2003) due to the numerical instability. Let us denote the normalization transformations by $\mathbf{T}_1$ and $\mathbf{T}_2$ where the normalized homography is calculated as $\hat{\mathbf{H}} = \mathbf{T}_2 \mathbf{H} \mathbf{T}_1^{-1}$. The transformation matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ are special affine transformations: they consist of translation and scale. The horizontal and vertical scales of the two transformations are denoted by l_x^k and l_y^k ($k \in \{1, 2\}$), respectively.

Normalization of Point Pairs and Fundamental Matrix. The normalized point pairs are calculated on the first and second images as $\hat{p}_1^i = \mathbf{T}_1 p_1^i$, and $\hat{p}_2^i = \mathbf{T}_2 p_2^i$, respectively. The normalization formula for the fundamental matrix is written (Hartley and Zisserman, 2003) as

$$\hat{\mathbf{F}} = \mathbf{T}_2^{-T} \mathbf{F} \mathbf{T}_1^{-1}. \quad (11)$$

Normalization of the Affine Transformation. The normalizing transformation modifies the basic equations written in Eqs. 4. For example,

$$\left(h_{31} u_1^i + h_{32} v_1^i + h_{33}\right) \hat{a}_{11}^i = \frac{l_x^2}{l_x^1} h_{11} - \frac{l_x^2}{l_x^1} u_i^2 h_{31} \quad (12)$$

where $l_x^i = \mathbf{T}_{i,11}$, $l_y^i = \mathbf{T}_{i,22} (i \in \{1,2\})$ are the horizontal and vertical scales of the i-th normalizing transformation, respectively. The left side of Eq. 12 is the multiplication of the projective depth and first affine parameter in the normalized system. After elementary modification, it is straightforward to prove that the affine parameters are modified as

$$\hat{a}_{11}^i = \left(l_x^2/l_x^1\right) a_{11}^i, \quad \hat{a}_{12}^i = \left(l_x^2/l_y^1\right) a_{12}^i,$$
$$\hat{a}_{21}^i = \left(l_y^2/l_x^1\right) a_{21}^i, \quad \hat{a}_{22}^i = \left(l_y^2/l_y^1\right) a_{22}^i.$$

The normalized affine transformations modify Eqs. 7–10 as

$$h_{31}\left(s_x^i\left(l_x^2/l_x^1\right)\cos(\alpha_i)u_1^i + u_2^i - e_x\right) + \left(l_x^2/l_x^1\right)\left(h_{32}s_x^i\cos(\alpha_i)v_1^i + h_{33}s_x^i\cos(\alpha_i)\right) = f_{21}, \quad (13)$$

$$h_{32}\left((w_i\cos(\alpha_i) - s_y^i\sin(\alpha_i))\left(l_x^2/l_y\right)v_1^i + u_2^i - e_x\right) + \left(l_x^2/l_y^1\right)h_{31}(w_i\cos(\alpha_i) - s_y^i\sin(\alpha_i))u_1^i + \left(l_x^2/l_y^1\right)h_{33}(w_i\cos(\alpha_i) - s_y^i\sin(\alpha_i)) = f_{22}, \quad (14)$$

$$\left(l_y^2/l_x^1\right)h_{31}\left(s_x^i\sin(\alpha_i)u_1^i + v_2^i - e_y\right) + \left(l_y^2/l_x^1\right)\left(h_{32}s_x^i\sin(\alpha_i)v_1^i + h_{33}s_x^i\sin(\alpha_i)\right) = -f_{11}, \quad (15)$$

$$\left(l_y^2/l_y^1\right)h_{32}\left((w_i\sin(\alpha_i) + s_y^i\cos(\alpha_i))v_1^i + v_2^i - e_y\right) + \left(l_y^2/l_y^1\right)h_{31}(w_i\sin(\alpha_i) + s_y^i\cos(\alpha_i))u_1^i + \left(l_y^2/l_y^1\right)h_{33}(w_i\sin(\alpha_i) + s_y^i\cos(\alpha_i)) = -f_{12}. \quad (16)$$

If the system is combined with Eqs. 2 an inhomogeneous, linear system of equations is obtained. Note that the normalized correspondences and $\hat{\mathbf{F}}$ are used in Eqs. 13– 16.

2.6 Algorithmic Details

Alg. 1 shows the P-HAF algorithm specialized to SIFT features. The required input is a set of point correspondences P and the related rotation R and scale S components, for each. The output is the homography. Note that for functions with parameter $\cdots$, all the available ones are passed.

2.7 Processing Time

Time Demand of the Algorithm. The processing time of the proposed algorithm depends on the solution of the inhomogeneous, linear system which can be carried out via Moore-Penrose pseudo-inverse. On a serial processor its time complexity is $O(m^3) + O(r^3)$ where m and r are the row number of the coefficient matrix $\mathbf{A}$ and its rank, respectively.

Algorithm 1: P-HAF for SIFT points.

Input: P – points on the first and second images
R, S – rotation and scale for each point pair
$\mathbf{F}$ – fundamental matrix
Output: $\mathbf{H}$ – homography

1: $P, R, S :=$ Normalization(...); ▷ Sec. 2.5
2: $\mathbf{C}, b :=$ BuildCoefficientMatrix(...); ▷ Eqs. 7, 9
3: $x := \mathbf{C}^\dagger b$; ▷ † is the Moore-Penrose pseudo-inverse
4: $\mathbf{H} :=$ HomographyFromFundMat(x, $\mathbf{F}$); ▷ Eq. 1

Remark that it is reduced to $O(m) + O(r^3)$ in parallel computing (Courrieu, 2008). Therefore, P-HAF is computable in a few milliseconds (see Table 2).

Time Demand of RANSAC. Augmenting RANSAC (Fischler and Bolles, 1981) or other robust statistics (Maronna et al., 2006) with P-HAF significantly reduces the iteration number, thus higher processing speed is achieved. Table 1 reports the required iteration number (Hartley and Zisserman, 2003) of RANSAC to converge using different minimal methods (columns) as engine. Rows show the ratio of the outliers.

Table 1: Required iteration number of RANSAC augmented with minimal methods (columns) with 95% probability on different outlier levels (rows).

	# of required points		
outl.	**2**	3	4
50%	**11**	23	47
80%	**74**	373	1871

It can be seen that using two points leads to significantly less iterations, thus speeding up the process, especially for high outlier ratio.

3 EXPERIMENTAL RESULTS

The aim of this section is to show that the proposed theory works both on synthetic and real world data. All algorithms ended with a numerical refinement stage using Levenberg-Marquardt optimization technique (Moré, 1978) to minimize re-projection error.

Table 2: The processing time (in milliseconds) of normalized P-HAF – including normalization – implemented in Matlab and C++. The first row shows the time of P-HAF applied to a minimal subset – two correspondences. The second one reports the mean time on all pairs of the AdelaideRMF and Multi-H datasets. On average, P-HAF is applied to 27 SIFT point pairs as an overdetermined system.

	Matlab (ms)	C++ (ms)
2 points	0.336	0.005
N points	1.106	0.012

The competitor methods are the Direct Linear Transformation (DLT) and Three Point Method (3PT) applied to normalized data.

3.1 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 – using uniform distribution – on a plane represented by function $S_c(u,v) = [u \quad v \quad 60]^T$, $(u,v \in [-20,20])$. Both cameras point towards the origin. Their common focal length and principal point are 600 and $[300 \quad 300]^T$, respectively. Fundamental matrix $\mathbf{F}$ is computed from projection matrices $\mathbf{P}_1$, and $\mathbf{P}_2$ (Hartley and Zisserman, 2003).

A plane passing through the origin is generated with random orientation and sampled in 50 different locations – these points are projected onto cameras $\mathbf{P}_1$ and $\mathbf{P}_2$. Zero-mean Gaussian-noise is added to the point coordinates. The local affinity related to each point pair is calculated from the plane parameters (Barath and Hajder, 2016) and the noisy point locations, then decomposed into the form

$$\mathbf{A} = \begin{bmatrix} s_x \cos(\alpha_i) & w\cos(\alpha) - s_{i,y}\sin(\alpha) \\ s_x \sin(\alpha_i) & w\sin(\alpha) + s_{i,y}\cos(\alpha) \end{bmatrix},$$

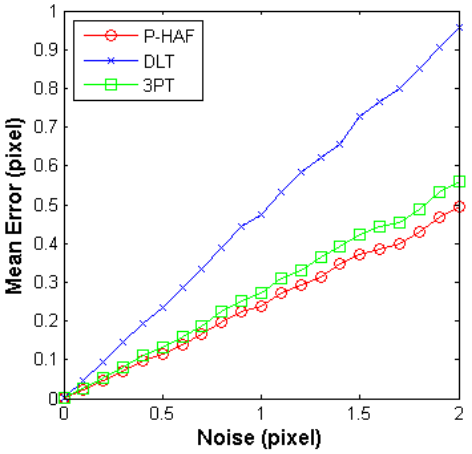
and angle α, scale s_x are kept. Tests are repeated 500 times on every noise level.

Fig. 1(a) and Fig. 1(b) visualize the mean and median errors of the normalized DLT, 3PT and P-HAF methods plotted as the function of the σ value of the zero-mean Gaussian-noise. P-HAF achieves the lowest mean and median errors. Fig. 1(c) shows the effect of the normalization. Even though the difference is not significant, the normalized algorithm is the most accurate estimator.

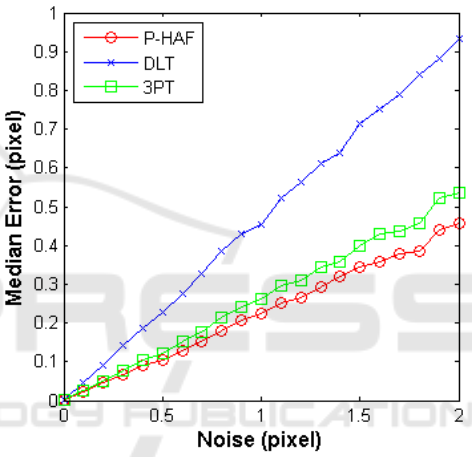
3.2 Homography Estimation

In order to test P-HAF on real world images AdelaideRMF (Wong et al., 2011) and Multi-H (Barath et al., 2016a) datasets are used. They consist of images of different sizes and point correspondences assigned to planes. Figure 2 shows four example images – the first one from each stereo pair – from the datasets. The left column is from Multi-H, pairs `boxesandbooks` and `glasscasea`, and the right one from AdelaideRMF – pairs `elderhalla` and `bonhall`. Points are painted by circles and each is assigned to a plane by color.

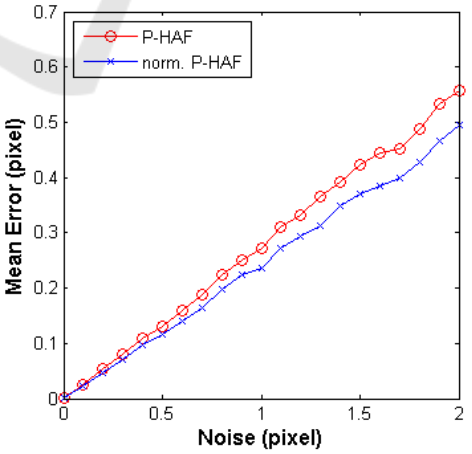
Annotations contain no information about the rotational or scale components, therefore, SIFT detector is applied to each image pair. Then the closest detected feature is paired to every annotated one. If



(a) Comparison of methods, mean error.



(b) Comparison of methods, median error.



(c) The effect of the normalization, mean error.

Figure 1: Re-projection error (vertical axis) calculated from 500 tests on each noise level. Parameter σ of the zero-mean Gaussian-noise added to the point coordinates is shown on the horizontal axis.

Table 3: The mean re-projection error (in pixels) of the methods applied to the AdelaideRMF and Multi-H datasets. Each row represents an image pair and each column consists of the re-projection errors of a method. Homographies are estimated using the 25% of the correspondences, re-projection error is computed w.r.t. all of them.

Test case	**P-HAF**	DLT	3PT
barrsmith	**27.01**	35.98	27.22
bonhall	0.97	**0.82**	0.99
bonython	**1.33**	1.35	1.35
boxesandbooks	**2.06**	8.46	2.12
elderhalla	3.26	3.50	**3.20**
elderhallb	**4.73**	5.34	5.21
glasscasea	**7.85**	26.76	9.63
glasscaseb	9.63	21.29	**7.95**
graffiti	**0.92**	1.01	0.96
hartley	1.98	**1.61**	2.01
johnssona	10.78	**10.39**	11.29
johnssonb	**5.28**	6.34	5.67
ladysymon	7.55	7.58	**7.50**
library	**4.82**	6.03	4.97
napiera	15.01	**14.48**	17.78
napierb	17.74	30.61	**17.28**
neem	**4.31**	5.44	5.64
nese	**4.35**	6.90	4.66
sene	**4.07**	7.88	4.73
unihouse	8.80	**5.38**	5.58
unionhouse	**7.01**	7.53	**7.01**
mean	**7.11**	10.22	7.27
median	**4.82**	6.90	5.58

the distance is greater than 5 pixels the point pair is omitted from the evaluation. The fundamental matrix **F** is estimated by the RANSAC eight-point technique (Hartley and Zisserman, 2003) with threshold value set to 1.0 followed by a Levenberg-Marquardt optimization minimizing symmetric epipolar distance. Every homography is estimated using the 25% of the correspondences, however, the reported re-projection errors are computed using all of them.

In Table 3, the mean re-projection errors (in pixels) are reported. Rows represent different test pairs from the AdelaideRMF and Multi-H datasets, columns show the related errors. It can be seen that the mean errors of P-HAF and 3PT are quite similar, even so, P-HAF is slightly better. The median error of P-HAF is significantly better than that of DLT and 3PT. This is expected since DLT and 3PT use a smaller part of the underlying affine transformation – the translation – while P-HAF exploits all the available information.

Figure 2: Example images from the image pairs of Multi-H (left column) and AdelaideRMF (right column) datasets. Points are marked by circles, planes are denoted by color.

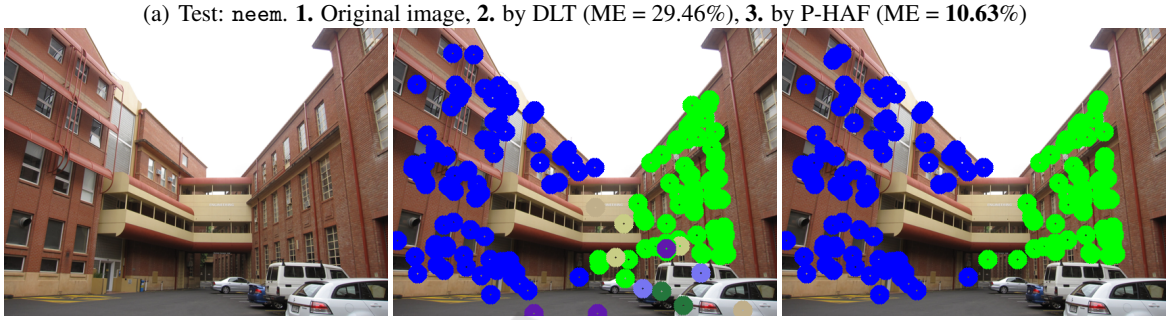
3.3 Multiple Homography Fitting

One of the main advantage of P-HAF is the required minimal point number as it is able to estimate a homography from only two SIFT correspondences. DLT needs four and 3PT three of those. Most of the robust model fitting techniques, e. g. RANSAC, are based on minimum subsets consisting of the minimum number of data to estimate a given model. Using as few data as possible makes the estimation faster, less ambiguous, and possibly more accurate.

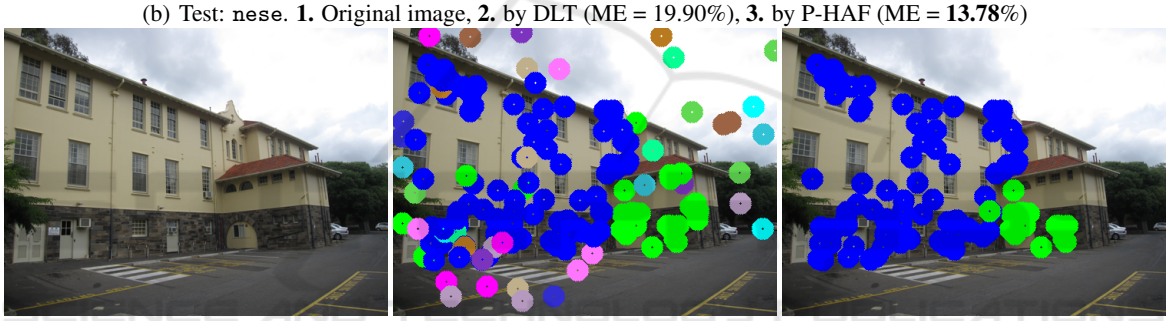
In this section, a multi-model fitting technique, PEARL (Isack and Boykov, 2012), is augmented with different model initialization methods: normalized DLT and P-HAF. The same datasets are used as in the previous experiments, AdelaideRMF and Multi-H. AdelaideRMF mainly consists of buildings while Multi-H smaller planar objects.

Fig. 3 shows the results of multi-homography fitting. Each row consists of the first image of a selected test pair. The left column shows the original image and the other ones report the obtained planar labellings obtained by PEARL with different hypothesis generation techniques: normalized DLT (middle) or P-HAF (right). The same parameters are used for all the tests and the same amount of hypothesizes are generated. The reported misclassification error (ME) is the ratio of the points assigned to wrong plane in percentage. It can be seen that *PEARL augmented with P-HAF is significantly more accurate then the one using normalized DLT.*

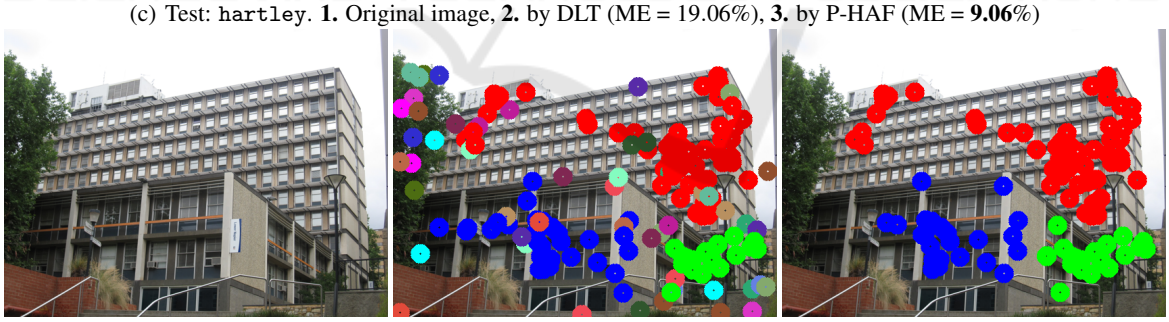
(a) Test: `neem`. **1.** Original image, **2.** by DLT (ME = 29.46%), **3.** by P-HAF (ME = **10.63**%)



(b) Test: `nese`. **1.** Original image, **2.** by DLT (ME = 19.90%), **3.** by P-HAF (ME = **13.78**%)



(c) Test: `hartley`. **1.** Original image, **2.** by DLT (ME = 19.06%), **3.** by P-HAF (ME = **9.06**%)



(d) Test: `napierb`. **1.** Original image, **2.** by DLT (ME = 38.22%), **3.** by P-HAF (ME = **23.17**%)

Figure 3: The results of multiple homography fitting to point correspondences. Each row is the first image of a test pair from AdelaideRMF dataset and the results of PEARL. Columns reports the obtained planar labellings of PEARL method with different hypothesis generation techniques: normalized DLT or P-HAF. The same parameters are used for all the tests and the same amount of hypothesizes are generated. The reported misclassification error (ME) is the ratio of the points assigned to wrong plane in percentage. Points are painted by circles and planes marked by color.

4 CONCLUSION

A novel minimal method is presented in this paper to improve the general point-based homography estimation by exploiting the information yielded by the commonly used feature detectors. The proposed P-HAF method is able to estimate the homography using at least two SIFT correspondences and applicable in real time. The main message of this paper is that usually there are more information about the underly-

ing homography than only the point coordinates – e. g. SIFT, SURF obtain the rotational component and the scale as well. Neglecting this information yields information loss. *We see no reasons to use the four-point algorithm instead of P-HAF* for rigid scenes if SIFT or SURF features are given.

REFERENCES

Barath, D. and Hajder, L. (2016). Novel ways to estimate homography from local affine transformations. In *In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP*, pages 432–443.

Barath, D., Hajder, L., and Matas, J. (2016a). Multi-h: Efficient recovery of tangent planes in stereo images. In *BMVC 2016, 27th British Machine Vision Conference, 19-22 September, York, England*, volume 28, page 32.

Barath, D., Molnar, J., and Hajder, L. (2016b). 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.

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

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

Bódis-Szomorú, A., Riemenschneider, H., and Gool, L. V. (2014). Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, J., Dixon, W. E., Dawson, D. M., and McIntyre, M. (2006). Homography-based visual servo tracking control of a wheeled mobile robot. *Robotics, IEEE Transactions on*, 22(2):406–415.

Chuan, Z., Long, T. D., Feng, Z., and Li, D. Z. (2003). A planar homography estimation method for camera calibration. In *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, volume 1, pages 424–429. IEEE.

Chum, O. and Matas, J. (2012). Homography estimation from correspondences of local elliptical features. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3236–3239. IEEE.

Courrieu, P. (2008). Fast computation of moore-penrose inverse matrices. *arXiv preprint arXiv:0804.4809*.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

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

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Isack, H. and Boykov, Y. (2012). Energy-based geometric multi-model fitting. *International journal of computer vision*, 97(2):123–147.

Jain, P. K. and Jawahar, C. (2006). Homography estimation from planar contours. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 877–884. IEEE.

Köser, K. (2009). *Geometric Estimation with Local Affine Frames and Free-form Surfaces*. Shaker.

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

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.

Maronna, R., Martin, D., and Yohai, V. (2006). *Robust statistics*. John Wiley & Sons, Chichester. ISBN.

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

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

Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

Prince, S. J., Xu, K., and Cheok, A. D. (2002). Augmented reality camera tracking with homographies. *Computer Graphics and Applications, IEEE*, 22(6):39–45.

Raposo, C. and Barreto, J. P. (2016). Theory and practice of structure-from-motion using affine correspondences.

Tanacs, A., Majdik, A., Molnar, J., Rai, A., and Kato, Z. (2014). Establishing correspondences between planar image patches. In *Digital Image Computing: Techniques and Applications (DICTA), 2014 International Conference on*, pages 1–7. IEEE.

Ueshiba, T. and Tomita, F. (2003). Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 966–973. IEEE.

Werner, T. and Zisserman, A. (2002). New techniques for automated architectural reconstruction from photographs. In *Computer VisionECCV 2002*, pages 541–555. Springer.

Wong, H. S., Chin, T.-J., Yu, J., and Suter, D. (2011). Dynamic and hierarchical multi-structure geometric model fitting. In *International Conference on Computer Vision (ICCV)*.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

Zhang, Z. and Hanson, A. R. (1996). 3d reconstruction based on homography mapping. *Proc. ARPA96*, pages 1007–1012.

Zhou, J. and Li, B. (2006). Robust ground plane detection with normalized homography in monocular sequences from a robot platform. In *Image Processing, 2006 IEEE International Conference on*, pages 3017–3020. IEEE.