

# Multi-H: Efficient Recovery of Tangent Planes in Stereo Images

Daniel Barath<sup>1</sup>  
barath.daniel@sztaki.mta.hu

Jiri Matas<sup>2</sup>  
matas@cmp.felk.cvut.cz

Levente Hajder<sup>1</sup>  
hajder.levente@sztaki.mta.hu

<sup>1</sup> DEVA Research Laboratory,  
MTA SZTAKI, Budapest, Hungary

<sup>2</sup> Centre for Machine Perception,  
Department of Cybernetics  
Czech Technical University,  
Prague, Czech Republic

---

## Abstract

Multi-H – an efficient method for the recovery of the tangent planes of a set of point correspondences satisfying the epipolar constraint is proposed. The problem is formulated as a search for a labeling minimizing an energy that includes a data and spatial regularization terms. The number of planes is controlled by a combination of Mean-Shift and  $\alpha$ -expansion.

Experiments on the fountain-P11 3D dataset show that Multi-H provides highly accurate tangent plane estimates. It also outperforms all state-of-the-art techniques for multi-homography estimation on the publicly available AdelaideRMF dataset. Since Multi-H achieves nearly error-free performance, we introduce and make public a more challenging dataset for multi-plane fitting evaluation.

## 1 Introduction

Understanding the structure of indoor and outdoor environments is important in many applications of computer vision. Man-made objects commonly consist of planar regions, particularly in an urban environment or indoor scenes. Many algorithms, for diverse problems, exploit the information captured by planes or planar correspondences. Such problems include camera calibration [6, 25, 61], robot navigation [9, 53], augmented reality [20] and 3D reconstruction [28, 62].

This paper addresses the problem of accurate tangent plane estimation by partitioning the feature correspondences satisfying the epipolar constraint according to the similarity of their tangent planes. A plane-to-plane correspondence in two images is defined by a homography [7] which can be estimated in many ways. Methods based on point [7], line [7], conic [12, 23], local affine frame [8] or region [18] correspondences have been proposed.

Several techniques are available for the estimation of multiple homographies. The popular RANSAC paradigm has been extended to multiple plane fitting by sequential RANSAC [26, 29] and multiRANSAC [34]. However, the RANSAC strategy suffers from the low inlier ratio of each individual homography. J-Linkage [24] and the recently proposed T-Linkage [10] are based on the analysis of randomly selected clusters in the preference space which is defined by the assignment costs of data points to clusters. J-Linkage merges the initial clusters

in the order of their Jaccard distances i.e. the overlap between two sets. T-Linkage extends this approach to a continuous preference space and modifies the distance function between two clusters to the Tanimoto distance. Both algorithms decide whether a plane is significant on the basis of the number of the associated inliers.

The closest work is the PEARL algorithm of Boykov *et al.* [8]. In PEARL, the multi-model fitting problem is cleanly formulated as optimization of a global energy functional. The hypotheses are initialized by stochastic sampling. The data term of the energy functional captures the cost of a point to homography assignment. A second term introduces spatial regularization reflecting an assumption that the geometric models have non-overlapping spatial supports and that correspondences which are close are more likely to belong to the same model. A third term penalizes the number of the models.

Like PEARL, we formulate the problem as a search for energy minimizing labeling. The energy proposed here is similar: it consists of the same data and spatial regularization terms. However, in the proposed algorithm, called Multi-H, the third term of PEARL is omitted as we control the model complexity by a combination of Mean-Shift [6] and  $\alpha$ -expansion [9].

Multi-H benefits from a deterministic initialization which we show that together with a repeated use of Mean-Shift leads to results superior to PEARL. The proposed method exploits the result of Barath *et al.* [10] and estimates a homography from a single correspondence and the related affinity. Another strong point is that hard decisions whether a plane is significant or not are avoided since that depends on the application field. Small planes are beneficial *e.g.* for reconstruction, however, we introduce a significance criterion for the problem of dominant plane retrieval.

The contributions of the paper are: (i) the method for assigning point correspondences to planes according to the similarity of their tangents that leads to high-quality estimates of surface normals. Not deciding whether a plane is significant, we benefit from both weakly and strongly supported planes. (ii) It is shown that the common stochastic sampling stage of multi-homography fitting algorithms can be improved upon. The Multi-H partitioning significantly outperforms state-of-the-art multi-homography fitting techniques. (iii) We introduce new, more challenging image pairs for multi-homography estimation and make them publicly available together with the annotation<sup>1</sup>.

## 2 Multiple Homography Estimation – Multi-H

Multi-H estimates tangent plane parameters at each point correspondence by assigning them to shared planes. Its only required input is an image pair. The output of the algorithm is a set of homographies defining the tangent planes and a label for each point correspondence associating it to a homography.

### 2.1 Point Correspondences with Local Affine Transformations

Several methods are available for the estimation of a local affine transformation at a detected point pair. We prefer to use affine-covariant feature detectors [12] since they provide point correspondences and affinities at the same time. We use MODS<sup>2</sup> [13] since it is significantly faster than ASIFT [14]. MODS provides high quality local affine transformations as well as the epipolar geometry  $F$ . The output point correspondences are consistent with fundamental

<sup>1</sup><http://web.eee.sztaki.hu/home4/node/56>

<sup>2</sup>Available at <http://cmp.felk.cvut.cz/wbs/>

matrix  $F$ . A different source of point correspondences with local affinities can be used, but the transformations must be consistent with  $F$  since Multi-H exploits this property.

Let us denote the  $i$ -th homogeneous point in  $k$ -th image with  $\mathbf{p}_k^i = [p_k^{i,x} \ p_k^{i,y} \ 1]^T$ ,  $i \in [1, N]$ ,  $k \in \{1, 2\}$ , and the related local affinity with  $\mathbf{A}_k^i$ . The transformation between the infinitely close vicinities of the two points is the one transforming the first affinity to the second as  $\mathbf{A}^i \mathbf{A}_1^i = \mathbf{A}_2^i$ . Thus  $\mathbf{A}^i = \mathbf{A}_2^i (\mathbf{A}_1^i)^{-1}$ . The elements of  $\mathbf{A}^i$  in row-major order are  $a_{11}^i$ ,  $a_{12}^i$ ,  $a_{21}^i$ , and  $a_{22}^i$ . Fig. 1 visualizes some local affine transformations using ellipses. To make



Figure 1: Corresponding local affine transformations visualized by ellipses.

the measured affinities as accurate as possible, the EG- $L_2$ -Opt correction is applied [10].

Homography  $\mathbf{H}_i$  is calculated for every affine transformation  $\mathbf{A}_i$  and the corresponding point pair by the Homography from Affine transformation and Fundamental matrix method (HAF) [10]. HAF estimates a homography from only one affine correspondence if the fundamental matrix is given by solving a system of linear, inhomogeneous equations  $\mathbf{C}\mathbf{x} = \mathbf{b}$  with coefficient matrix

$$\mathbf{C} = \begin{bmatrix} a_{11}^i p_1^{i,x} + p_2^{i,x} - e^x & a_{11}^i p_1^{i,y} & a_{11}^i \\ a_{12}^i p_1^{i,y} + p_2^{i,x} - e^x & a_{12}^i p_1^{i,x} & a_{12}^i \\ a_{21}^i p_1^{i,x} + p_2^{i,y} - e^y & a_{21}^i p_1^{i,y} & a_{21}^i \\ a_{22}^i p_1^{i,y} + p_2^{i,y} - e^y & a_{22}^i p_1^{i,x} & a_{22}^i \end{bmatrix}, \quad (1)$$

where  $\mathbf{e} = [e^x \ e^y]^T$  is the epipole on the second image. Vector  $\mathbf{b} = [f_{21} \ f_{22} \ -f_{11} \ -f_{12}]$  is the inhomogeneous part of the four equations and  $\mathbf{x} = [h_{31} \ h_{32} \ h_{33}]^T$  is the vector of the unknown parameters. The optimal solution in the least squares sense is given by  $\mathbf{x} = \mathbf{C}^\dagger \mathbf{b}$  where  $\mathbf{C}^\dagger$  is the Moore-Penrose pseudo-inverse of matrix  $\mathbf{C}$ . The homography matrix is finally calculated using its last row [10] as follows:  $h_{1j} = e^x h_{3j} + f_{2j}$ ,  $h_{2j} = e^y h_{3j} + f_{1j}$ , where  $j \in \{1, 2\}$  and  $f_{lm}, l, m \in \{1, 2, 3\}$ , are elements of the fundamental matrix  $\mathbf{F}$ .

## 2.2 Alternating Minimization

After the initialization described in the preceding section (Section 2.1), the set of homographies is improved by alternating three steps (see Alg. 1).

**(1) Mean-Shift.** Fig. 2 shows that after initialization some of the homographies estimated from a single correspondence coincide with a surface tangent plane (columns one and two) and some do not (columns three and four). In each column of Fig. 2, the correspondence

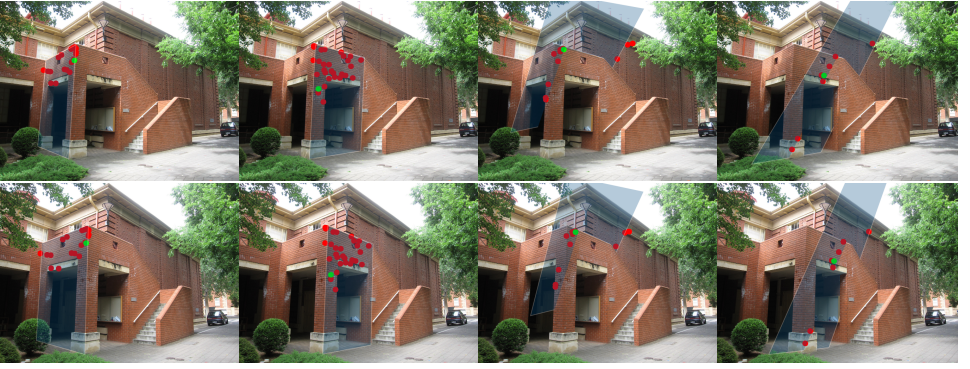


Figure 2: The images (top, bottom) of the johnsona pair. Blue shaded quadrangles visualise homographies coinciding (columns 1 and 2) and not coinciding (3 and 4) with a surface tangent plane. The correspondence initializing the homography is marked green. The red points are inliers obtained by thresholding the re-projection error at 3.0 pixels.

---

### Algorithm 1 The Multi-H Algorithm.

---

**Input:**  $I_1, I_2$  – images;  $P, A, F := \text{MODS}(I_1, I_2)$  [□]

$P$  – point correspondences;  $A$  – affine transformations;  $F$  – fundamental matrix

**Output:**  $\mathcal{H}$  – obtained homographies;  $L$  – obtained labeling

- 1:  $\mathcal{H}^0 := \text{HAF}(P, A, F)$  [□] ▷ Initialization with point-wise homographies
  - 2:  $i := 0$ ;
  - 3: **repeat** ▷ Alternating Minimization
  - 4:    $i := i + 1$ ;
  - 5:    $\mathcal{H}^i := \text{MeanShift}(\mathcal{H}^{i-1})$  ▷ Default  $\varepsilon = 2.7$
  - 6:    $L^i := \alpha\text{-expansion}(P, \mathcal{H}^i)$  ▷ Default  $\lambda = 0.5, \gamma = 0.005$
  - 7:    $\mathcal{H}^i := \text{LSQHomographyRefinement}(P, A, L^i, F)$
  - 8: **until** Convergence ▷ if  $\mathcal{H}^i = \mathcal{H}^{i-1} \wedge L^i = L^{i-1}$
  - 9:  $\mathcal{H} := \mathcal{H}^i; L := L^i$
- 

initializing the homography is marked green, and its  $\varepsilon$ -inliers are in red, with threshold  $\varepsilon = 3.0$  pixels. The tangent planes are visualized by blue quadrangles.

We assume that tangent plane homographies are shared by a number of points and their parameters emerge as modes in the homography space. Since we do not know the number of tangent planes in the scene, the mode-seeking Mean-Shift [□] algorithm is adopted. The projection of the  $i$ -th homography in the constructed  $6D$  homography space is

$$\mathbf{v}^i = \begin{bmatrix} w_1^{i,x} & w_1^{i,y} & w_2^{i,x} & w_2^{i,y} & w_3^{i,x} & w_3^{i,y} \end{bmatrix}, \quad (2)$$

where

$$\mathbf{w}_1^i = \frac{\mathbf{H}^i \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T}{H_{33}^i}, \quad \mathbf{w}_2^i = \frac{\mathbf{H}^i \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T}{H_{13}^i + H_{33}^i}, \quad \mathbf{w}_3^i = \frac{\mathbf{H}^i \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T}{H_{23}^i + H_{33}^i}.$$

The denominator of each  $\mathbf{w}^i$  is the projective depth of the transformed point in the numerator. Each vector  $\mathbf{v}^i$  determines a homography which can be recovered from three

points  $[0 \ 0 \ 1]^T$ ,  $[1 \ 0 \ 1]^T$ ,  $[0 \ 1 \ 1]^T$  and their projections if the fundamental matrix is known  $[\mathbf{H}, \mathbf{Q}]$ . Even though there are several possible representations for a homography (e.g. using its elements, projecting four points, etc.), we prefer to use a low-dimensional one – the processing time of Mean-Shift highly depends on the dimension of the problem. Since each coordinate pair  $[v_k^i \ v_{k+1}^i]$ ,  $k \in \{1, 3, 5\}$ , is a given point projected by  $\mathbf{H}^i$  the distance function  $d$  is chosen as the mean Euclidian distance between the three coordinate pairs where  $v_k^i$  is the  $k$ -th coordinate of vector  $v^i$ . The distance between the  $i$ -th and  $j$ -th feature vectors is defined as

$$d(\mathbf{v}^i, \mathbf{v}^j) = \frac{1}{3} \sum_{k=1}^3 \|[v_{2(k-1)+1}^i \ v_{2(k-1)+2}^i]^T - [v_{2(k-1)+1}^j \ v_{2(k-1)+2}^j]^T\|_2.$$

(2) **The  $\alpha$ -expansion  $[\mathbf{Q}]$**  step minimizes the following energy:

$$E(L) = \frac{1}{\lambda} E_d(L) + \lambda E_s(L), \quad (3)$$

where  $L$  is the current labeling,  $E_d(L)$  and  $E_s(L)$  the data and smoothness terms;  $\lambda$  controls their balance. The data term is defined as

$$E_d(L) = \sum_{i=1}^N \|\mathbf{p}_2^i - \frac{\mathcal{H}^{l_i} \mathbf{p}_1^i}{\mathcal{H}_{31}^{l_i} p_1^{i,x} + \mathcal{H}_{32}^{l_i} p_1^{i,y} + \mathcal{H}_{33}^{l_i}}\|_2, \quad (4)$$

where  $\mathcal{H}^{l_i}$  is the homography associated with label  $l_i \in L$  of the  $i$ -th correspondence. The second term,  $E_s$ , reflects the assumption that neighboring points are more likely to belong to the same homography.  $E_s$  is equal to the number of neighboring points that are labeled differently:

$$E_s(L) = \sum_{i=1}^N \sum_{j=1}^N \mathcal{A}_{ij} [l_i \neq l_j], \quad (5)$$

where  $N$  is the number of correspondences, the Iverson bracket  $[\cdot]$  is equal to one if the condition inside holds and zero otherwise, and the elements of the adjacency matrix  $\mathcal{A}_{ij}$  are equal to 1 if correspondences  $i$ -th and  $j$ -th are spatial neighbors, 0 otherwise. The correspondences are considered to be neighbors if their distance in a  $4D$  concatenated coordinate space – the vector associated with a correspondence is  $[p_1^x \ p_1^y \ p_2^x \ p_2^y]^T$  – is below  $\gamma$ , a control parameter. Matrix  $\mathcal{A}$  is calculated efficiently using FLANN, the Fast Library for Approximate Nearest Neighbors  $[\mathbf{Q}]$ .

The energy cannot increase in this step due to the nature of the  $\alpha$ -expansion algorithm. A point is assigned to no plane if its distance from the closest one is greater than  $3\epsilon$  which is an empirically set threshold.

(3) **The Least-Squares Homography Refinement** runs the HAF method  $[\mathbf{H}]$  on the correspondences associated with each homography by the current labeling. The number of the homographies is unchanged. The energy decreases or remains the same since  $E_d$  is the sum of the re-projection errors which are minimized.  $E_s$  is unchanged since the labeling does not change.

**Convergence** is reached when both the number of the clusters and the energy remain unchanged in two iterations. As the first stage does not increase the number of clusters, the other stages decrease the energy, and the set of labeling is finite, convergence is ensured. In experiments reported in Section 3, Alg. 1 converged no later than after eight iterations.

	R	PEARL	QP-MF	FLOSS	ARJMC	SA-RCM	J-Lnkg	T-Lnkg	Multi-H
johnsonna	4	4.02	18.50	4.16	6.48	5.90	5.07	4.02	<b>2.41</b>
johnsonnb	7	18.18	24.65	18.18	21.49	17.95	18.33	18.17	<b>4.46</b>
ladysymon	2	5.49	18.14	5.91	5.91	7.17	9.25	5.06	<b>0.00</b>
neem	3	5.39	31.95	5.39	8.81	5.81	3.73	3.73	<b>0.00</b>
oldclassicswing	2	1.58	13.72	1.85	1.85	2.11	0.27	0.26	<b>0.00</b>
sene	2	0.80	14.00	0.80	0.80	0.80	0.84	0.40	<b>0.00</b>
mean		5.91	20.16	6.05	7.56	6.62	6.25	5.30	<b>1.19</b>
median		4.71	18.32	4.78	6.20	5.86	4.40	3.87	<b>0.00</b>

Table 1: Misclassification error (%) for the two-view plane segmentation. The selected image pairs are a subset – the same as used in [10] – of the 19 pairs of AdelaideRMF dataset. The number of the ground truth planes is denoted with  $R$ .

	J-Lnkg	T-Lnkg	RPA	SA-RCM	Grdy-RansaCov	ILP-RansaCov	Multi-H
mean	25.50	24.66	17.20	28.30	26.85	12.91	<b>4.40</b>
median	24.48	24.53	17.78	29.40	28.77	12.34	<b>2.41</b>

Table 2: Two-view plane segmentation. Mean and median misclassification error (%) on the 19 image pairs of the AdelaideRMF dataset.

## 3 Experimental Results

### 3.1 Comparison with Multi-homography Fitting Techniques

In this section, Multi-H is tested on the problem of significant plane retrieval. and it outperforms the state-of-the-art multi-homography fitting techniques.

**Determination of significant planes.** To determine whether a detected plane is or is not significant without strict restrictions on the minimum number of inliers, the following algorithm is introduced. (1) First, planes with less than four inliers are removed. (2) The homographies are re-computed using the standard normalized 4-point algorithm [10] followed by a numerical refinement stage minimizing the re-projection error by Levenberg-Marquardt optimization. (3) The compatibility constraint [10] for a homography and a fundamental matrix:  $\mathbf{H}^T \mathbf{F} + \mathbf{F}^T \mathbf{H} = 0$  is imposed by removing  $\mathbf{H}_i$  for which  $\|\mathbf{H}_i^T \mathbf{F} + \mathbf{F}^T \mathbf{H}_i\|_F > \theta$ . After extensive experimentation we set  $\theta = 1.0$ .

Multi-H is tested as in [10] on the AdelaideRMF dataset. For each image pair in the dataset, a set of dominant planes and point pairs on them are provided. However, affine transformations for the point pairs are not available. Thus as many correspondences and affinities as possible are obtained by MODS [10]. Then the closest match for every annotated AdelaideRMF correspondence is found among the MODS correspondences. These correspondences with the local affine transformations are the input of Multi-H.

The misclassification error (ME) is calculated as follows. First, the mapping between the ground truth  $l_{GT} \in L_{GT}$  and Multi-H  $l \in L$  labels is established. We use an iterative method, always assigning the Multi-H output homography with the highest set overlap of correspondences. The assigned Multi-H homography and ground truth one maximizing the overlap are then removed from further consideration. Note that if the assignment is not optimal, the reported misclassification errors of Multi-H are over-estimated. ME is the ratio of the number of different labels  $\sum_{i=1}^N [\mathbb{1}_{l_{GT}^i \neq l^i}]$  and the number of ground truth correspondences  $N$ .<sup>3</sup>

<sup>3</sup>The code for the ME calculation is available at <http://web.eee.sztaki.hu/home4/node/56>



Figure 3: Resulting partitioning of Multi-H on the AdelaideRMF dataset. Planes are denoted by colour. There are a few misclassified points (on the top-left and top-middle images around the edges). They are denoted by small, filled, black circles. Best viewed in colour.

Multi-H is compared with T-Linkage [10], ARJMC [19], PEaRL [8], QP-MF [30], FLoSS [9], J-Linkage [24] and SA-RCM [20] in Experiment 1 (see Table 1). Every algorithm, including Multi-H, has been tuned separately on each image pair. We prefer reporting results for a setting fixed for the whole dataset, and we do that at the end of this section, but to allow comparison with the literature we followed the per-image-parameter-setting methodology. Table 1 shows that Multi-H obtains the lowest mean and median misclassification errors on the six test image pairs evaluated in the literature [10]. Fig. 3 shows the Multi-H points color-coded by the homography they were assigned to.

Table 2 shows the mean and median misclassification errors on all 19 image pairs of the AdelaideRMF dataset. The competitor methods are T-Linkage [10], J-Linkage [24], RPA [10], SA-RCM [20], Greedy-RansaCov [10] and ILP-RansaCov [10]. Multi-H significantly outperforms all published methods. Note the significant difference in the mean and median misclassification rates obtained on the six selected image, which are commonly published (Table 1), and on the full dataset.

Even though this dataset is the most frequently used one in the multi-plane fitting literature, it consists of easy scenes where the planes are perpendicular or far from each other. In order to test the accuracy of Multi-H, we created a more challenging dataset. Examples of the new images are visualized in Fig. 4. On these images, point correspondences are detected by MODS [15] and each is manually annotated to the containing plane. Finally, outliers, i.e. non-corresponding point pairs, are added to the data. For every image pair, the first image is the ground truth and the second one is the obtained planar partitioning. Outliers are visualised by black dots on the ground truth images. Pair 4(a) is from the well-known graffiti test sequence<sup>4</sup>. Two slightly different planes present in these images. The lower plane is closer to the camera than the upper one, however, the difference is very small. Even so, Multi-H accurately distinguishes the two planes and achieves a low misclassification error of 1.19%. Image pairs 4(b) and 4(c) are a cabinet with books and a staircase viewed from above. The last two images (4(d)) visualize a room with some boxes and planar-like objects.

<sup>4</sup>Available at <http://www.robots.ox.ac.uk/~vgr/research/affine/>

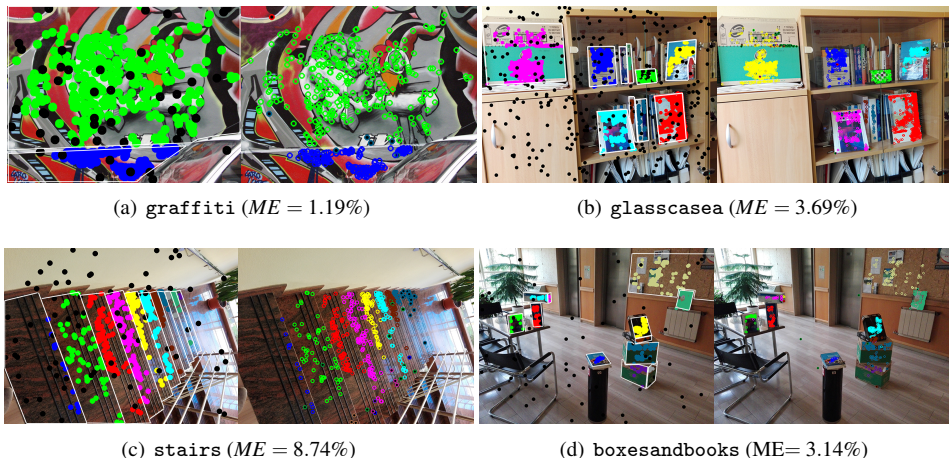


Figure 4: Four image pairs of the new dataset. Points coloured according to tangent planes, manual annotation (left) and Multi-H assignment (right). ME is the misclassification error.

	johnsa	johnsb	ladysymon	neem	old	sene	mean	median
<b>Multi-H</b>	<b>9.33</b>	<b>10.14</b>	<b>4.49</b>	<b>2.00</b>	<b>1.79</b>	<b>0.00</b>	<b>4.79</b>	<b>3.74</b>
T-Lnkg	34.28	24.04	24.67	25.65	20.66	7.63	22.82	24.36
SA-RCM	36.73	16.46	39.50	41.45	21.30	20.20	29.27	29.02
RPA	10.76	26.76	24.67	19.86	25.25	0.42	17.95	22.27

Table 3: Misclassification error (%) with a fixed parameter setup, average over 5 runs. The following abbreviations are used: johnsonna (johnsa), johnsonnb (johnsb), oldclassicswing (old).

These tests are more challenging than the ones containing buildings since the observed planar regions are very small and their orientations are in many cases similar, see e.g. the books in glasscasea.

**Proposed general configuration.** For practical point of view, it is desirable that a single setting of parameters of the method covers most common cases. Through extensive experimentation, we found that  $\lambda = 0.5$ ,  $\varepsilon = 2.7$ , and  $\gamma = 0.005$  are a robust choice. Table 3 shows the misclassification error on the AdalaideRMF dataset (average of 5 runs). The results are significantly worse than those of the separately tuned ones (see Table 1), but much better than the performance of the competitor algorithms<sup>5</sup> with a fixed set-up.

## 3.2 Evaluation of Surface Normal Accuracy

In this section, the accuracy of planes estimated by Multi-H is compared with the point-wise estimates of the affine-covariant detector. All planes returned by Multi-H are used, the significance constraint which was used in the previous section is not applied. The accuracy was assessed on the fountain-P11 dataset [22] which includes 11 images with resolution  $3072 \times 2048$ , projection and calibration camera matrices and reconstructed point clouds

<sup>5</sup>Experimental results are copied from [10]





Figure 5: Correspondence clustering into tangent planes for frames 1, 2 of the fountain-P11 set. Planes denoted by colour, estimated surface normals visualized by white line segments.

Frames	1 – 2	3 – 5	1 – 5	6 – 8	5 – 9
Affine Detector	35.7   32.7	24.9   20.3	19.0   15.8	22.5   18.6	20.0   15.4
EG- $L_2$ -Optimal	35.5   32.5	23.1   19.8	16.7   13.9	19.9   16.6	17.8   14.4
<b>Multi-H</b>	<b>14.4   9.4</b>	<b>9.0   7.5</b>	<b>7.0   5.8</b>	<b>8.8   7.3</b>	<b>7.1   5.7</b>

Table 4: Mean and median errors (in degrees) of estimated normals for selected image pairs.

with surface normals. Point correspondences of MODS [15] between selected image pairs were obtained. On average, 920 correspondences were found.

Multi-H partitions correspondences on the basis of their tangent planes. The partitioning is visualized in Fig. 6. A single homography is fitted using the correspondences in the same tangent plane cluster. The normals at the correspondences are calculated from the homography as the camera parameters are known. Table 4 (row 3, Multi-H) shows the mean and median angular errors of the surface normals calculated from the homographies w.r.t. ground truth data. The surface normals determined by the homographies are significantly more accurate than the estimates from the initial local affine transformations output by the detector (Table 4, first row). Normals estimated after the EG- $L_2$ -Optimal procedure [2], that improves the local affinities using constraints provided by the fundamental matrix, are significantly less accurate too (Table 4, second row).

### 3.3 Processing Time and Implementation Details

The speed of the Multi-H procedure was measured on two sets consisting of 100 and 500 correspondences. Since a randomized version of Mean-Shift was used, the algorithm ran 100 times. The mean number of iterations of Algorithm 1 was approx. 6 in both cases. The average processing times for the 100 and 500 correspondences were 0.04 and 0.80 sec. on a desktop PC with Intel Core i5-4690 CPU, 3.50 GHz using 4 cores.

Each column of Fig. 6(a) shows the processing time (in milliseconds) for an image pair. The parts of each bar visualize the time of the different algorithmic steps. The data shows that Multi-H has negligible time demand compared to the feature detection process (MODS). The bars associated with the calculation of the adjacency matrix and point-wise homographies cannot be seen since they require approx. 4 – 6 milliseconds.

Fig. 6(b) presents the processing time of the alternating minimization. It significantly drops after the first iteration, then it is constant-like. The drop is caused by the Mean-Shift that reduces the number of homographies which speeds-up the  $\alpha$ -expansion step.

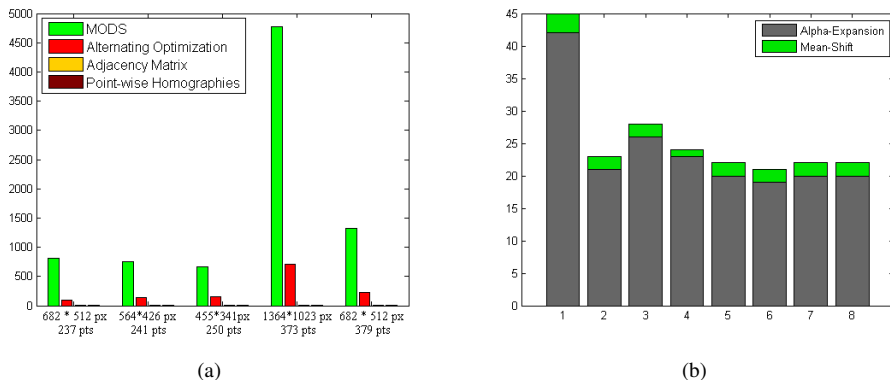


Figure 6: (a) Processing time (in milliseconds) of Multi-H applied to different image pairs. The vertical axis at each column shows the resolution of the images and the correspondence number. (b) The processing time (in milliseconds) of iterations 1-8 of the alternating minimization on the hartley pair.

Multi-H is implemented in C++. The GCOptimization<sup>6</sup> code was used for  $\alpha$ -expansion. A fast Mean-Shift implementation was downloaded from the web<sup>7</sup>.

## 4 Conclusions

The Multi-H approach for estimation of tangent planes in image pairs by partitioning feature correspondences was proposed. The method is accurate, outperforming state-of-the-art multi-homography fitting techniques for both fixed and per-image parameter setting. Experiments showed that the standard datasets are relatively easy and we therefore augmented the data with several challenging image pairs which we annotated.

In most applications, Multi-H will run significantly faster than the affine-covariant detectors providing the input. It is real-time on a standard CPU if the number of correspondences is below approx. 300. A GPU implementation of  $\alpha$ -expansion [27] will be real-time capable for significantly larger problems.

The source code and the annotated dataset are available at <http://web.eee.sztaki.hu/home4/node/56>

## Acknowledgements

This work was supported by the Hungarian National Research, Development and Innovation Office grant VKSZ 14-1-2015-0072. J. Matas was supported by the Technology Agency of the Czech Republic project TE01020415 (V3C – Visual Computing Competence Center).

<sup>6</sup>Available at <http://vision.csd.uwo.ca/code/>

<sup>7</sup>Available at <http://scikit-learn.org/stable/modules/clustering.html#mean-shift>

## References

- [1] D. Barath and L. Hajder. Novel ways to estimate homography from local affine transformations. In *Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, pages 432–443, 2016.
- [2] D. Barath, J. Matas, and L. Hajder. Accurate closed-form estimation of local affine transformations consistent with the epipolar geometry. In *27th British Machine Vision Conference (BMVC)*, 2016.
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence (PAMI)*, pages 1222–1239, 2001.
- [4] J. Chen, W. E. Dixon, D. M. Dawson, and M. McIntyre. Homography-based visual servo tracking control of a wheeled mobile robot. *Robotics*, pages 406–415, 2006.
- [5] Z. Chuan, T. D. Long, Z. Feng, and D. Z. Li. A planar homography estimation method for camera calibration. In *Computational Intelligence in Robotics and Automation*, pages 424–429, 2003.
- [6] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence (PAMI)*, pages 603–619, 2002.
- [7] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [8] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision (IJCV)*, pages 123–147, 2012.
- [9] N. Lazic, I. Givoni, B. Frey, and P. Aarabi. Floss: Facility location for subspace segmentation. In *International Conference on Computer Vision (ICCV)*, pages 825–832, 2009.
- [10] L. Magri and A. Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation.
- [11] L. Magri and A. Fusiello. T-linkage: a continuous relaxation of j-linkage for multi-model fitting. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3954–3961, 2014.
- [12] L. Magri and A. Fusiello. Multiple model fitting as a set coverage problem. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3326, 2016.
- [13] M. Marius and G. D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Application (VISAPP)*, pages 331–340, 2009.
- [14] 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*, pages 43–72, 2005.
- [15] D. Mishkin, J. Matas, and M. Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, pages 81–93, 2015.

- [16] J-M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, pages 438–469, 2009.
- [17] P. K. Mudigonda, C. V. Jawahar, and P. J. Narayanan. Geometric structure computation from conics. In *ICVGIP*, pages 9–14, 2004.
- [18] J. Nemeth, C. Domokos, and Z. Kato. Recovering planar homographies between 2d shapes. In *12th International Conference on Computer Vision (ICCV)*, pages 2170–2176, 2009.
- [19] T. T. Pham, T-J. Chin, J. Yu, and D. Suter. Simultaneous sampling and multi-structure fitting with adaptive reversible jump mcmc. In *Advances in Neural Information Processing Systems*, pages 540–548, 2011.
- [20] T-T. Pham, T-J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. *Pattern Analysis and Machine Intelligence (PAMI)*, pages 1658–1671, 2014.
- [21] G. Simon, A. W. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *International Symposium on Augmented Reality (ISAR)*, pages 120–128, 2000.
- [22] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.
- [23] A. Sugimoto. A linear algorithm for computing the homography from conics in correspondence. *Journal of Mathematical Imaging and Vision (JMIV)*, pages 115–130, 2000.
- [24] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *European Conference on Computer Vision (ECCV)*, pages 537–547. 2008.
- [25] T. Ueshiba and F. Tomita. Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In *International Conference on Computer Vision (ICCV)*, pages 966–973, 2003.
- [26] E. Vincent and R. Laganière. Detecting planar homographies in an image pair. In *Image and Signal Processing and Analysis (ISPA)*, pages 182–187, 2001.
- [27] V. Vineet and P. J. Narayanan. Solving multilabel mrfs using incremental  $\alpha$ -expansion on the gpus. In *Computer Vision (ACCV)*, pages 633–643. 2009.
- [28] T. Werner and A. Zisserman. New techniques for automated architectural reconstruction from photographs. In *European Conference on Computer Vision (ECCV)*, pages 541–555. 2002.
- [29] K. Yasushi and K. Hiroshi. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *British Machine Vision Conference (BMVC)*, pages 1–10, 2004.

- [30] J. Yu, T-J. Chin, and D. Suter. A global optimization approach to robust multi-model fitting. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2041–2048, 2011.
- [31] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence (PAMI)*, pages 1330–1334, 2000.
- [32] Z. Zhang and A. R. Hanson. 3d reconstruction based on homography mapping. *Proc. ARPA96*, pages 1007–1012, 1996.
- [33] J. Zhou and B. Li. Robust ground plane detection with normalized homography in monocular sequences from a robot platform. In *Image Processing*, pages 3017–3020, 2006.
- [34] M. Zuliani, C. S Kenney, and B. S. Manjunath. The multiransac algorithm and its application to detect planar homographies. In *International Conference on Image Processing (ICIP)*, pages III–153, 2005.