Theory and Practice of Structure-from-Motion using Affine Correspondences

Carolina Raposo and João P. Barreto
Institute of Systems and Robotics, University of Coimbra
Coimbra, Portugal
{carolinaraposo,jpbar}@isr.uc.pt

Abstract

Affine Correspondences (ACs) are more informative than Point Correspondences (PCs) that are used as input in mainstream algorithms for Structure-from-Motion (SfM). Since ACs enable to estimate models from fewer correspondences, its use can dramatically reduce the number of combinations during the iterative step of sample-and-test that exists in most SfM pipelines. However, using ACs instead of PCs as input for SfM passes by fully understanding the relations between ACs and multi-view geometry, as well as by establishing practical, effective AC-based algorithms. This article is a step forward into this direction, by providing a clear account about how ACs constrain the two-view geometry, and by proposing new algorithms for plane segmentation and visual odometry that compare favourably with respect to methods relying in PCs.

1. Introduction

An Affine Correspondence (AC), denoted in this paper by $(\mathbf{x}, \mathbf{y}, \mathsf{A})$, consists in a Point Correspondence (PC) across views plus a 2×2 affine transformation A that maps image points in the neighbourhood of $\mathbf{x}$ into image points around $\mathbf{y}$ (see Fig 1). Since an affine map describes well the warp undergone by a local image patch while the camera moves, the concept of AC is broadly used for tracking and/or matching points across views [1, 14], or estimating similarity models as in [17], where local transformations are exploited for geometrical alignment. However, and despite ACs being often readily available, mainstream methods for relative pose estimation, such as [21, 8, 7], use as only input the PCs $(\mathbf{x}, \mathbf{y})$ completely disregarding the information about local affine maps.

This fact has been noticed by previous authors that conducted seminal research in jointly using PCs and affine maps for Structure-from-Motion (SfM). Perdoch *et al.* [16] and Riggi *et al.* [19] proposed to use ACs for generating additional PCs and estimate the epipolar geometry. However, these new matches are mere approximations that do not necessarily correspond to correct PCs. Koser [11] studied the relationships between ACs and homographies and advanced a single point method to compute the relative pose between a plane and a camera [10]. Recently, Bentolila and Francos proved that 1 AC puts 3 constraints on the fundamental matrix [3]. Despite these seminal works, neither the theory relating ACs with multi-view geometry is fully understood, nor exist practical algorithms such that ACs can become an effective alternative to PCs in SfM pipelines.

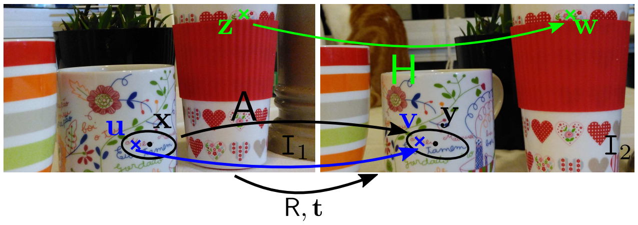


Figure 1: Images $\mathtt{I}_1$ and $\mathtt{I}_2$ are provided by two cameras related by $\mathsf{R}, \mathbf{t}$ that observe the same scene. They are used for extracting point $(\mathbf{z}, \mathbf{w})$ and affine $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ correspondences. In this configuration, since the tangent plane to the surface in points $\mathbf{x}$ and $\mathbf{z}$ is the same, both the point match and the AC are compatible with homography H.

The most obvious benefit in using ACs is that models can be estimated from fewer correspondences. Thus, and since SfM pipelines invariably comprise an iterative step of sample-and-test (*e.g.* RANSAC [6]), robustness and complexity can dramatically improve by reducing the number of possible combinations, as it happened in the past with the introduction of minimal solvers [15]. These advantages can be specially useful for applications with high combinatorics as it often arises in problems of multi-model fitting or single-model fitting with high percentages of outliers. Examples for the former are applications in plane detection [18] or multibody SfM [20], and for the latter the case of SfM in scenes dominated by deformable surfaces [13]. This article starts by investigating how ACs constrain two-view geometry deriving both new and known relations in a unified, systematic way. Also, it uses the new relations and insight to propose algorithms for detecting planes in the scene and recovering the camera relative pose using ACs. The contributions can be summarized as follows:

Characterization of the family of homographies compatible with an AC: It is well known that the affine mapping A is equal to the Jacobian of the homography H induced by the plane tangent to the 3D surface containing the image point [10, 9, 11, 3]. This paper shows that an AC is compatible with a 2-parameter family of homographies containing H, that not all correspondences are compatible with this family, and that 2 additional PCs, or 1 extra AC, define a particular instance of this family that can be estimated in closed-form. Although some of these facts have already been stated by Koser [11] and Chum *et al*. [4], we provide more clear, intuitive derivations that lead to explicit formulas, generalize results, and give new insights on the topic.

Epipolar geometry from ACs: It is shown in [3] that each AC puts 3 constraints on the parameters of the fundamental matrix, having derived these constraints for the case of the image coordinates being centred in the correspondence. This article derives these 3 constraints without requiring any change of coordinates. The advantages of these new equations are twofold: first, the constraints can be used for both essential and fundamental matrix estimation; and second, the well known 5-point and 7-point algorithms [15, 8, 7] can be applied with almost no changes to determine the epipolar geometry from ACs. In this case we simply substitute the bilinear relations arising from 5 or 7 PCs by the new constraints arising from 2 or 3 ACs, respectively.

Image plane segmentation using ACs: The article derives, for the first time, the constraints that must be verified by a PC or AC to be compatible with the 2-parameter family of homographies associated with an initial AC. These constraints have a clear geometric interpretation and can be used as a metric for segmenting correspondences according to planes present in the scene. Comparative experiments show the benefits of this direct metric with respect to sophisticated global multi-model fitting approaches [12] using the 4-point algorithm for homography estimation [8].

Visual odometry using ACs: We propose an algorithm for estimating the essential matrix from 2 ACs that is extensively tested against the 5-point method [21] in real sequences. This work provides for the first time convincing experimental evidence that ACs are a viable alternative to PCs for visual odometry and can be highly advantageous in the presence of many outliers as it happens in scenes with multiple moving objects and/or deformable surfaces.

2. Geometric Relations between ACs and Homographies

In this section, theoretical results on the relationship between ACs and homographies are presented. We start by reviewing background concepts and afterwards show how additional PCs and ACs constrain the homography.

Consider the setup of Fig 1 where two cameras, related by a rotation R and a translation $\mathbf{t}$, observe a scene, originating images $\mathtt{I}_1$ and $\mathtt{I}_2$. Let $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ be an AC such that the patches surrounding $\mathbf{x}$ and $\mathbf{y}$ are related by a non-singular 2×2 matrix A, with

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\mathsf{T}, \mathbf{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^\mathsf{T}, \mathsf{A} = \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix}. \quad (1)$$

For a point correspondence $(\mathbf{u}, \mathbf{v})$ in the patch, it comes that

$$\mathbf{v} = \mathsf{A}\mathbf{u} + (\mathbf{y} - \mathsf{A}\mathbf{x}). \quad (2)$$

Let us also assume that the patches are related by an homography

$$\mathsf{H} \sim \begin{bmatrix} h_1 & h_4 & h_7 \\ h_2 & h_5 & h_8 \\ h_3 & h_6 & h_9 \end{bmatrix}, \quad (3)$$

such that in non-homogeneous coordinates

$$\mathbf{v} = \mathbf{f}(\mathbf{u}) = \delta_\mathbf{u}^{-1} \begin{bmatrix} h_1 u_1 + h_4 u_2 + h_7 & h_2 u_1 + h_5 u_2 + h_8 \end{bmatrix}^\mathsf{T}, \quad (4)$$

where $\delta_\mathbf{p} = h_3 p_1 + h_6 p_2 + h_9$. As first proposed by Koser *et al*. in [10, 9], approximating Eq 4 using the first-order Taylor expansion around $\mathbf{x}$ yields $\mathbf{v} = \mathbf{f}(\mathbf{x}) + \mathsf{J}_\mathbf{f}(\mathbf{x})(\mathbf{u} - \mathbf{x})$, where $\mathsf{J}_\mathbf{f}$ is the Jacobian of $\mathbf{f}$. Knowing that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$, the expression can be written as

$$\mathbf{v} = \mathsf{J}_\mathbf{f}(\mathbf{x})\mathbf{u} + (\mathbf{y} - \mathsf{J}_\mathbf{f}(\mathbf{x})\mathbf{x}). \quad (5)$$

Relating Eq 2 and 5, it can be seen that $\mathsf{A} = \mathsf{J}_\mathbf{f}(\mathbf{x})$, meaning that the affine transformation A is the Jacobian of the homography defined in point $\mathbf{x}$.

2.1. 2-Parameter Family of Homographies

The result introduced by Koser *et al*. [10, 9] allows the homography to be defined as a function of the AC. From $\mathbf{y} = \mathbf{f}(\mathbf{x})$, two constraints on the parameters of the homography are obtained. This allows (h_7, h_8) to be written as

$$\begin{bmatrix} h_7 \\ h_8 \end{bmatrix} = \begin{bmatrix} (h_3 - h_1)x_1 + (h_6 - h_4)x_2 + h_9 \\ (h_3 - h_2)x_1 + (h_6 - h_5)x_2 + h_9 \end{bmatrix}. \quad (6)$$

Replacing this result in Eq 4, the Jacobian becomes

$$\mathsf{J}_\mathbf{f}(\mathbf{x}) = \delta_\mathbf{x}^{-1} \left[\begin{bmatrix} h_1 & h_4 \\ h_2 & h_5 \end{bmatrix} - \mathbf{y} \begin{bmatrix} h_3 & h_6 \end{bmatrix} \right], \quad (7)$$

which, since $\mathsf{J}_\mathbf{f}(\mathbf{x}) = \mathsf{A}$, yields that

$$\begin{bmatrix} h_1 & h_4 \\ h_2 & h_6 \end{bmatrix} = \delta_\mathbf{x}\mathsf{A} + \mathbf{y} \begin{bmatrix} h_3 & h_6 \end{bmatrix}. \quad (8)$$

Replacing the results of Eq 6 and 8 in the homography 3, it comes that the AC $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ induces a two-parameter family of homographies H defined as

$$\mathsf{H} \sim \delta_\mathbf{x} \begin{bmatrix} \mathsf{A} & \mathbf{y} - \mathsf{A}\mathbf{x} \\ \mathbf{0} & 1 \end{bmatrix} + \begin{bmatrix} \mathbf{y} \\ 1 \end{bmatrix} \begin{bmatrix} h_3 & h_6 & h_9 - \delta_\mathbf{x} \end{bmatrix}. \quad (9)$$

Note that this is a two-parameter family since, although there are 3 unknowns, there are only 2 degrees-of-freedom (DOFs) because H is defined up to scale. Several authors [10, 9, 3] suggest to make $h_9 = 1$ in order to fix the scale factor, which has the drawback of not avoiding singular configurations. H is non-singular whenever A is full rank and $\delta_{\mathbf{x}} \neq 0$. In order to assure that H is always full rank, we introduced the following change of parameters:

$$\begin{bmatrix} g_3 \\ g_6 \\ g_9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x_1 & x_2 & 1 \end{bmatrix} \begin{bmatrix} h_3 \\ h_6 \\ h_9 \end{bmatrix}, \tag{10}$$

which leads to

$$\mathsf{H} \sim g_9 \begin{bmatrix} \mathsf{A} & \mathbf{y}-\mathsf{A}\mathbf{x} \\ \mathbf{0} & 1 \end{bmatrix} + \begin{bmatrix} \mathbf{y} \\ 1 \end{bmatrix} \begin{bmatrix} g_3 & g_6 & -x_1 g_3 - x_2 g_6 \end{bmatrix}. \tag{11}$$

Making $g_9 = 1$ fixes the scale factor and avoids singular configurations, yielding

$$\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A}) = \begin{bmatrix} \mathsf{A} + \mathbf{y}\mathbf{g}^{\mathsf{T}} & \mathbf{y} - (\mathsf{A} + \mathbf{y}\mathbf{g}^{\mathsf{T}})\mathbf{x} \\ \mathbf{g}^{\mathsf{T}} & 1 - \mathbf{g}^{\mathsf{T}}\mathbf{x} \end{bmatrix}, \tag{12}$$

with $\mathbf{g} = \begin{bmatrix} g_3 & g_6 \end{bmatrix}^{\mathsf{T}}$.

In case H is a perspectivity, there are 4 solutions that can be determined by solving two second-order equations in $\mathbf{g}$ that force the first and second columns of H to have the same norm and be orthogonal. Note that this generalizes the result by Koser and Koch [10] that requires the origin of plane coordinates to be coincident with $\mathbf{x}$.

2.2. Using PCs to Constrain the Homography

The two-parameter family of homographies in Eq 12 can be further constrained by using PCs. Consider an additional match $(\mathbf{z}, \mathbf{w})$, as depicted in Fig 1, which is used to determine the homography $\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A})$ by estimating $\mathbf{g}$. Although at first $(\mathbf{z}, \mathbf{w})$ may appear to provide two constraints in H that should suffice to uniquely determine $\mathbf{g}$, this is not the case as proven next.

Assume that $(\mathbf{z}, \mathbf{w})$ is compatible with H such that

$$k \begin{bmatrix} \mathbf{w}^{\mathsf{T}} & 1 \end{bmatrix}^{\mathsf{T}} = \mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A}) \begin{bmatrix} \mathbf{z}^{\mathsf{T}} & 1 \end{bmatrix}^{\mathsf{T}}, \tag{13}$$

with k being a scale factor. From Eq 12, it comes in a straightforward manner that $k = \mathbf{g}^{\mathsf{T}}(\mathbf{z} - \mathbf{x}) + 1$, which, by replacing in Eq 13, yields

$$(\mathbf{w} - \mathbf{y})(\mathbf{z} - \mathbf{x})^{\mathsf{T}}\mathbf{g} = \mathsf{A}(\mathbf{z} - \mathbf{x}) - (\mathbf{w} - \mathbf{y}). \tag{14}$$

Two important facts arise from Eq 14. The first is that since $(\mathbf{w}-\mathbf{y})(\mathbf{z}-\mathbf{x})^{\mathsf{T}}$ is a rank-1 matrix, it can be concluded that a point match $(\mathbf{z}, \mathbf{w})$ only puts one constraint in $\mathbf{g}$. Thus, two point correspondences are required to fully constrain $\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A})$. The second fact is that the span S of matrix $(\mathbf{w}-\mathbf{y})(\mathbf{z}-\mathbf{x})^{\mathsf{T}}$ is 1-dimensional, with $\mathsf{S} = \lambda(\mathbf{w}-\mathbf{y}), \forall \lambda \in \mathbb{R}$, which means that H is compatible with $(\mathbf{z}, \mathbf{w})$ iff $\mathsf{A}(\mathbf{z} - \mathbf{x}) - (\mathbf{w} - \mathbf{y}) \in \mathsf{S}$. In other words, $(\mathbf{z}, \mathbf{w})$ is compatible with the homography H iff the following holds

$$(\mathbf{w} - \mathbf{y})^{\mathsf{T}}\mathsf{P}\mathsf{A}(\mathbf{z} - \mathbf{x}) = 0, \text{ with } \mathsf{P} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{15}$$

Note that the geometric meaning of Eq 15 is that vectors $\mathbf{c}_1 = \mathsf{A}(\mathbf{z} - \mathbf{x})$ and $\mathbf{c}_2 = \begin{bmatrix} w_2 - y_2 & -(w_1 - y_1) \end{bmatrix}^{\mathsf{T}}$ must be orthogonal. This finding allows using the angle between $\mathbf{c}_1$ and $\mathbf{c}_2$ as an error metric for checking compatibility between a PC and an homography induced by an AC. PCs not verifying Eq 15 cannot lie in a plane that contains the 3D point that gives rise to the AC. Although the condition is necessary but not sufficient to assure coplanarity, it can be used as an error metric for plane segmentation or tracking tasks. Section 4.1 validates the practical usefulness of the metric in a planar segmentation experiment.

2.3. Estimating the Homography using ACs

Instead of using PCs, an additional AC can be applied to further constrain the homography $\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A})$. Let $(\mathbf{z}, \mathbf{w}, \mathsf{B})$ be an extra AC that lies in the same plane as $(\mathbf{x}, \mathbf{y}, \mathsf{A})$, or that corresponds to the same plane tangent to the surface in the point of correspondence. This implies that there must be a choice of parameters $\mathbf{g}, \mathbf{m}$ such that $\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A}) = k\mathsf{H}(\mathbf{m}; \mathbf{z}, \mathbf{w}, \mathsf{B})$, where k is a scale factor.

Considering the homography as represented in Eq 12, the relations $k{=}1{+}\mathbf{g}^{\mathsf{T}}(\mathbf{z}{-}\mathbf{x})$ and $\mathbf{g}{=}k\mathbf{m}$ are obtained. Replacing in Eq 12, and making $\mathsf{M}{=}\left((\mathbf{z}{-}\mathbf{x})^{\mathsf{T}}\mathbf{g}\right)\mathsf{B}$, it comes

$$k\mathsf{H}(\mathbf{m}; \mathsf{B}, \mathbf{z}, \mathbf{w}) = \begin{bmatrix} \mathsf{B}{+}\mathsf{M}{+}\mathbf{w}\mathbf{g}^{\mathsf{T}} & \mathbf{w}{-}\mathsf{B}\mathbf{z}{-}\mathsf{M}\mathbf{z}{-}\mathbf{g}^{\mathsf{T}}\mathbf{x}\mathbf{w} \\ \mathbf{g}^{\mathsf{T}} & 1 - \mathbf{g}^{\mathsf{T}}\mathbf{x} \end{bmatrix}. \tag{16}$$

Since it is known that $\mathsf{H}(\mathbf{g}; \mathbf{x}, \mathbf{y}, \mathsf{A}){=}k\mathsf{H}(\mathbf{m}; \mathbf{z}, \mathbf{w}, \mathsf{B})$, the following system of six equations is obtained

$$\begin{array}{lcl} \mathsf{A}{-}\mathsf{B}{-}(\mathbf{w}{-}\mathbf{y})\mathbf{g}^{\mathsf{T}}{-}\mathsf{M} & = & 0 \\ \mathbf{y}{-}\mathsf{A}\mathbf{x}{-}(\mathbf{w}{-}\mathsf{B}\mathbf{z}) + (\mathbf{w}{-}\mathbf{y})\mathbf{x}^{\mathsf{T}}\mathbf{g}{+}\mathsf{M}\mathbf{z} & = & \mathbf{0} \end{array}. \tag{17}$$

This allows the computation of the 2 unknown terms of the homography, $\mathbf{g}$, from linear least squares. Therefore, replacing in Eq 12 or 16, H becomes fully determined.

As previously observed, each AC yields 6 constraints on the parameters of the homography H. Thus, two distinct ACs yield 12 restrictions, allowing 4 constraints to be written in the terms of $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ and $(\mathbf{z}, \mathbf{w}, \mathsf{B})$ because the homography has only 8 DOFs. Using the first matrix equation of system 17 to substitute M and $(\mathbf{w}{-}\mathbf{y})\mathbf{g}^{\mathsf{T}}$ in the second, it yields two conditions as the one in Eq 15 in the terms of A and B, respectively. This is expected since, by construction, the point match of one AC is compatible with the homography induced by the other. The remaining two constraints can be obtained by determining $\mathbf{g}$ from *e.g.* the first two equations in the first matrix equation and replacing the solution in the last two. After some algebraic manipulation, this procedure yields the last matrix equation in System 18. Thus, the 4 conditions for $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ and $(\mathbf{z}, \mathbf{w}, \mathsf{B})$ to be compatible with the same homography are

$$\begin{gathered}(\mathbf{w}-\mathbf{y})^{\mathsf{T}}\mathsf{PA}(\mathbf{z}-\mathbf{x})=0\\ (\mathbf{w}-\mathbf{y})^{\mathsf{T}}\mathsf{PB}(\mathbf{z}-\mathbf{x})=0\\ \underbrace{\begin{bmatrix} s{+}a_2b_3{-}a_3b_2 & -(a_1b_3{-}a_3b_1)\\ a_2b_4{-}a_4b_2 & s{-}(a_1b_4{-}a_4b_1)\end{bmatrix}}_{\mathsf{L}}(\mathbf{w}{-}\mathbf{y})=\mathbf{0},\ \text{with}\\ s{=}\frac{[-a_2{+}b_2\quad a_1{-}b_1](\mathbf{w}{-}\mathbf{y}){-}(a_1b_2{-}a_2b_1)(x_1{-}z_1)}{(x_2{-}z_2)}\end{gathered}. \tag{18}$$

Note that, as reasoned for Eq 15, the last matrix constraint means that both vectors $\mathbf{l}_1^{\mathsf{T}}$ and $\mathbf{l}_2^{\mathsf{T}}$, corresponding to the first and second rows of the 2×2 matrix L, respectively, must be orthogonal to $(\mathbf{w}{-}\mathbf{y})$. Thus, in this case, there are 4 different angles that can be combined to provide an error metric of compatibility between two ACs and an homography. Since more information is being included, this metric should be more robust than the one computed solely from PCs. Experiments in Section 4.1 on planar segmentation confirm this hypothesis.

3. Epipolar Geometry using ACs

It has recently been shown by Bentolila and Francos [3, 2] that one AC yields 3 linear constraints on the terms of the fundamental matrix F. Their method follows a sequence of steps including: (i) coordinate shifting so that the origin is the center of an AC; (ii) estimation of the epipole location $\mathbf{e}_p$ by intersecting 3 conics computed from the 3 ACs; (iii) estimation of two PCs through line intersection for finding an homography H; (iv) computing $\mathsf{F} = [\mathbf{e}_p]_\times \mathsf{H}$. This method does not make direct use of the linear constraints since they were derived only for the AC centred in the origin, requiring many small steps to achieve an estimation of F. Also, it is not clear how it could be adapted to the calibrated case, for the estimation of the essential matrix E.

We propose a new formulation for the estimation of the epipolar geometry from ACs by deriving the linear constraints in the original coordinate system. The new method does not require any of the steps proposed in [3], being much more straightforward and easier to implement. Moreover, it can be applied both to the uncalibrated and calibrated cases.

It is known that an homography compatible with a fundamental or an essential matrix verifies the condition $\mathsf{H}^{\mathsf{T}}\mathsf{T} + \mathsf{T}^{\mathsf{T}}\mathsf{H} = 0$, for $\mathsf{T} = \mathsf{F}, \mathsf{E}$, respectively. We will derive the constraints for the case of E, with

$$\mathsf{E} = \begin{bmatrix} e_1 & e_4 & e_7\\ e_2 & e_5 & e_8\\ e_3 & e_6 & e_9\end{bmatrix}. \tag{19}$$

Consider the 2-parameter family of homographies induced by an AC $(\mathbf{x}, \mathbf{y}, \mathsf{A})$ in Eq 12. The matrix equation $\mathsf{H}^{\mathsf{T}}\mathsf{E} + \mathsf{E}^{\mathsf{T}}\mathsf{H} = 0$ yields 9 equations, 6 of which are linearly independent and can be written as

$$\underbrace{\begin{bmatrix} q_1 & q_2 & g_3 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & q_3 & q_4 & g_6 & 0 & 0 & 0\\ q_3 & q_4 & g_6 & q_1 & q_2 & g_3 & 0 & 0 & 0\\ q_5 & q_6 & \gamma & 0 & 0 & 0 & q_1 & q_2 & g_3\\ 0 & 0 & 0 & q_5 & q_6 & \gamma & q_3 & q_4 & g_6\\ 0 & 0 & 0 & 0 & 0 & 0 & q_5 & q_6 & \gamma\end{bmatrix}}_{\mathsf{N}}\mathbf{e} = \mathbf{0}, \tag{20}$$

where γ depends on the unknown $\mathbf{g}$, $\gamma{=}1{-}g_3x_1{-}g_6x_2$, $q_i, i = 1, \ldots, 6$ are defined as $q_1{=}a_1{+}g_3y_1$, $q_2{=}a_2{+}g_3y_2$, $q_3{=}a_3{+}g_6y_1$, $q_4{=}a_4{+}g_6y_2$, $q_5{=}y_1\gamma{-}a_1x_1{-}a_3x_2$ and $q_6{=}y_2\gamma{-}a_2x_1{-}a_4x_2$, and $\mathbf{e}$ is the vectorization of E by columns. Due to the sparse nature of matrix N, it is possible to combine the 6 equations in order to eliminate the two unknowns g_3, g_6. Right-multiplying the system of Eq 20 by the following matrix C

$$\mathsf{C}{=}\begin{bmatrix} x_1^2 & x_2^2 & x_1x_2 & x_1 & x_2 & 1\\ -g_6x_1^2 & -x_2(g_6x_2{-}2) & -x_1(g_6x_1{-}1) & -g_6x_1 & 1{-}g_6x_2 & -g_6\\ -x_1(g_3x_1{-}2) & -g_3x_2^2 & -x_2(g_3x_1{-}1) & 1{-}g_3x_1 & -g_3x_2 & -g_3\end{bmatrix} \tag{21}$$

yields three equations that only depend on the terms of the AC $(\mathbf{x}, \mathbf{y}, \mathsf{A})$:

$$\begin{bmatrix} x_1y_1 & x_1y_2 & x_1 & x_2y_1 & x_2y_2 & x_2 & y_1 & y_2 & 1\\ a_3x_1 & a_4x_1 & 0 & y_1{+}a_3x_2 & y_2{+}a_4x_2 & 1 & a_3 & a_4 & 0\\ y_1{+}a_1x_1 & y_2{+}a_2x_1 & 1 & a_1x_2 & a_2x_2 & 0 & a_1 & a_2 & 0\end{bmatrix}\mathbf{e}{=}\mathbf{0} \tag{22}$$

It can be seen that, as expected, the first equation corresponds to the point match. It is also important to note that the simplicity of matrix N is due to the change of variables that was performed when representing the homography (Eq 10). Moreover, solving for $\mathbf{g}$ using the first two equations in system 20 and substituting in the third, yields, after some algebraic manipulation, $\det(\mathsf{E}) = 0$.

All the derivations obtained up to this point are valid both for the essential and fundamental matrices. Thus, since one AC provides three linear equations in the form of Eq 22, the 7-DOFs matrix F can be determined from 3 ACs and the 5-DOFs matrix E from a minimum of 2 ACs. A total of 9 and 6 equations are obtained in the uncalibrated and calibrated cases, respectively, meaning that either the 8-point or the 7-point solvers [8] can be used in the former case and the 6-point or 5-point solvers [21] in the latter. Section 5 presents experiments on the estimation of the essential matrix using 5 PCs and 2 ACs, in both rigid and non-rigid scenarios.

Note that Eq 20 can be interpreted the opposite way: suppose we know the epipolar geometry, E or F, and an AC, and wish to find the homography H compatible with both. Rewriting Eq 20 for isolating the unknown $\mathbf{g}$, a non-homogeneous system of 6 equations linear in the terms of $\mathbf{g}$

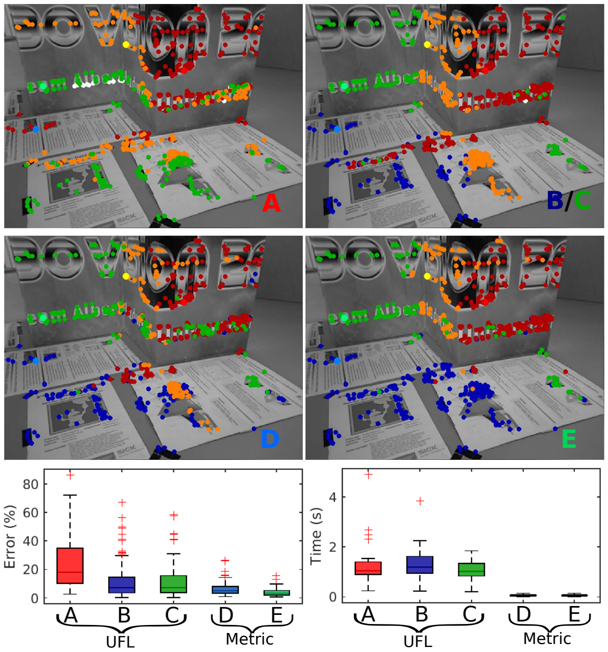


Figure 2: Experiment on planar segmentation using 5 different methods: A: 4 points UFL, B: 1 AC + 2 points UFL, C: 2 AC UFL, D: 1 AC + 2 Metric, E: 2 AC Metric. Segmentation errors and computation times are shown for each method. Planes are identified with colors and the reference AC is depicted with a lighter shade of the same color.

is obtained. Solving for $\mathbf{g}$ and substituting in Eq 12, allows for H to be fully determined.

4. Experiments in Homography Estimation

This section reports two experiments for testing the validity and usefulness of the theoretical results on homographies. The first one consists in performing planar segmentation on images that contain between 3 and 5 planes, both by formulating the problem as an Uncapacitated Facility Location (UFL)[1] problem that can be solved using message passing [12] and by using the novel error metrics proposed in Sections 2.2 and 2.3. The second experiment assesses the accuracy of homography estimation using 1 AC plus 2 PCs and 2 ACs and compares it with the 4-point linear algorithm applied in an MSAC-framework [24].

In all experiments, affine covariant features are extracted with the Hessian Laplace detector [14, 25] using the VLFeat library [26]. The affine part of the ACs, A, is refined by minimizing the photo-geometric error for increasing the estimation accuracy. The maximum number of iterations in the optimization was set to 10 in order to assure fast computation. Point normalization *a la Hartley* is always used before any linear estimation.

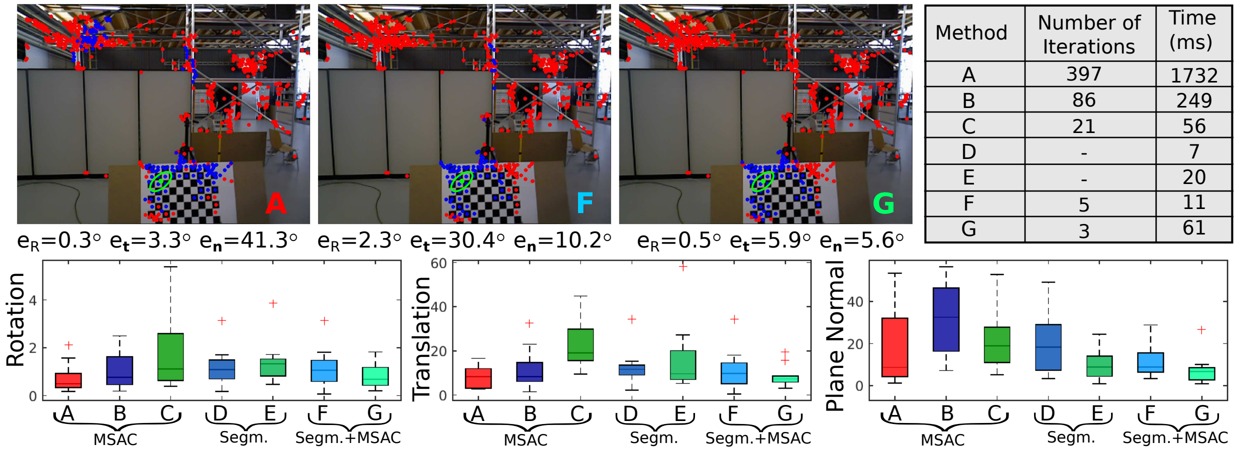
4.1. Segmentation of PCs and ACs

This experiment consists in the planar segmentation of all visible planes in 30 randomly sampled pairs of images from a sequence of the RGB-D dataset [22]. Ground truth segmentation was obtained using the method proposed in [23]. For each plane in an image pair, a reference AC was chosen by selecting the one that yielded the smallest photo-consistency error from the set of ACs belonging to that plane. Since multiple planes are being segmented simultaneously, this task is a multi-model fitting problem that can be cast as an UFL problem [12]. The goal is to assign each correspondence (client) to an homography hypothesis (facility), while simultaneously using as few hypotheses as possible. Our data cost matrix is created using the symmetric transfer error [8], and homography hypotheses are generated using the three minimal sets of 4 PCs, 1 AC plus 2 points (Section 2.2) and 2 ACs (Section 2.3). These three methods are referred to as A, B and C in Fig 2, respectively. Each homography hypothesis is generated by using the reference AC plus 3 PCs, 2 PCs or 1 AC randomly selected, for methods A, B and C, respectively. For each plane, 50 hypotheses were generated from this procedure, yielding a total of 50 × no. of planes + 1 labels per image pair due to the inclusion of the discard label.

An alternative way of performing planar segmentation is by using the constraints derived in Sections 2.2 and 2.3 that must be verified for an homography induced by an AC to be compatible with a point match and another AC, respectively (methods D and E in Fig 2). In this case, explicit estimation of the homography is not required as only the constraint is used. For the case of PCs, it consists in two vectors $\mathbf{a}$ and $\mathbf{b}$ that must be orthogonal. Thus, for each point match the error is simply $e = 90 - \angle(\mathbf{a}, \mathbf{b})$. When working with ACs, a stronger error metric may be used. In this case, 4 angular constraints were derived, yielding 4 errors e which are combined by taking their weighted mean that accounts for the quality of the ACs, computed from photo-consistency. In both cases, the constraints are computed between the reference AC and all remaining correspondences. Labelling is performed by assigning correspondences that yield errors below a pre-defined threshold (1° in our experiments) to the plane and if a correspondence is assigned more that one label, the one that yielded the smallest error is chosen.

Segmentation errors and computation times are shown for each method in Fig 2. The first conclusion is that the proposed error metrics effectively segment all the planes in the scene, being much more accurate and faster than the more sophisticated UFL approach. Moreover, when formulating the segmentation task as a UFL problem, it can be seen that it is significantly better to use either the 1 AC plus 2 PCs or the 2 ACs minimal solutions than the 4-point al-

[1] A global formulation (UFL) is much more effective for multi-modal fitting than greedy generalizations of RANSAC such as [27].



Method	Number of Iterations	Time (ms)
A	397	1732
B	86	249
C	21	56
D	-	7
E	-	20
F	5	11
G	3	61

Figure 3: Results of homography estimation for a 12-image dataset from [22] using 7 different methods: A: 4 points MSAC, B: 1 AC + 2 points MSAC, C: 2 AC MSAC, D: 1 AC + 2 points Planar segmentation, E: 2 AC Planar segmentation, F: 1 AC + 2 points Planar segmentation & MSAC, G: 2 AC Planar segmentation & MSAC. Coloured boxplots were used for better visualization, where the 4-point method corresponds to red and the methods using 1 AC + 2 points and 2 ACs are shown in shades of blue and green, respectively. Rotation, translation and plane normal errors are given in degrees. For the last two, the error is the angle between the estimated and ground truth vectors. The table shows the average number of iterations and computation time of a pair of images. Inlier and outlier points and the reference AC are shown in blue, red and green.

gorithm. This is expected for two main reasons. The first is that the former solutions require less correspondences to be selected, increasing the probability of all correspondences being in the same plane. The second is that as an AC imposes 6 restrictions on the homography as opposed to 2 for a point match in method A, it is more likely that correct solutions will be chosen for methods B and C.

4.2. Structure-from-Motion

The homography associated to the checkerboard plane in Fig 3 is estimated for 12 different image pairs from dataset [22]. For each one, a reference AC on the plane was chosen as described in the previous experiment. The homography H is estimated using 7 different methods, 5 of which rely on the robust estimator MSAC [24] for selecting the inlier correspondences. Methods A, B and C in Fig 3 consist on using the reference AC and randomly selecting the rest of the required correspondences for estimating H from 4 points [8], 1 AC plus 2 points and 2 ACs, respectively. Methods F and G correspond to methods B and C with a prior planar segmentation using PCs and ACs, respectively, with the proposed metric. Finally, methods D and E are the simplest since they perform planar segmentation and estimate H from the inlier correspondences. The estimated homography is decomposed into a rotation, a translation known up to a scale factor and the plane normal. The test images contain ground truth rotation and translation. The ground truth plane equation is computed using the method presented in [23]. Presenting information on the quality of the plane normal estimation is relevant since, as observed in Fig 3, there are sets of matches that may provide homographies which originate small errors in rotation and translation but estimate the plane normal very poorly. This is very evident for method A both due to the combinatorics of the problem and because the reference AC only puts two constraints on H in this case, meaning that there are homographies that do not correspond to real scene planes that may originate minima in the cost function of MSAC.

The results in Fig 3 also show that method E performs better than method D, which is coherent with the results from the previous experiment. When combining the segmentation with an MSAC, a significant increase in the accuracy is obtained, being this the best choice of methods for the task of estimating H. Finally, the table shows that as the minimum required number of matches decreases, less iterations of MSAC are used, occurring an almost immediate convergence when applying a prior segmentation step.

5. Experiments in Epipolar Geometry

The planar segmentation and homography estimation experiments of the previous section make use of ACs extracted solely from planes in the scene. In order to assess the validity of using ACs extracted from any kind of scene, an experiment was conducted where the epipolar geometry is estimated in sequences with planes, without planes, and hybrid. Moreover, motivated by the fact that our proposed 2-AC algorithm significantly reduces the combinatorics of the essential matrix estimation problem when compared to the state-of-the-art 5-point algorithm, we evaluated the per-

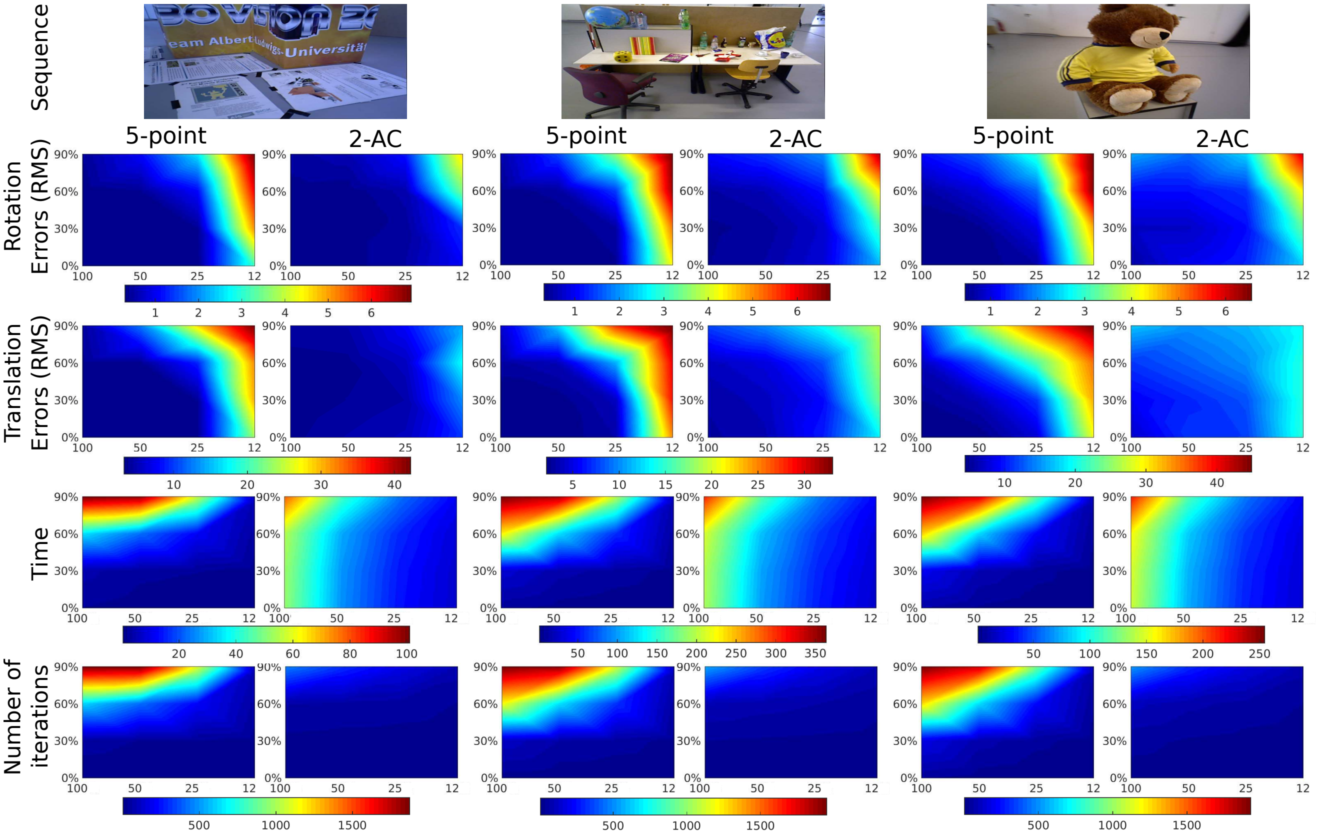


Figure 4: Experiment on the estimation of E for 3 sequences of the dataset from [22]. The number of input correspondences was reduced from 100 to 50, 25 and 12, and outliers were injected so that they constituted 0%, 30%, 60%, and 90% of the input set, originating 16 different configurations. For each sequence, color plots of the RMS rotation and translation errors are included, where the colors between the 16 error values were obtained by interpolation, for visualization purposes. Computational time for the complete sequences and average number of iterations of MSAC per image pair are shown.

formance of both methods in the presence of outliers and/or few input matches. Such conditions may occur in scenarios with very low textured surfaces where it is difficult to extract features and/or much deformation caused by the movement of pedestrians, vegetation in the wind, *etc*. Thus, in the first experiment we injected outliers and decreased the size of the data set, while in the second two real sequences dominated by large deformations were considered.

A total of 6 equations in the form of Eq 22 are obtained from 2 ACs. Our proposed method selects 5 out of the 6 equations for generating up to 10 solutions for E using the solver [21] and the remaining - that must be one corresponding to a point match - for selecting the best solution using the reprojection error. The 5-point algorithm used is the one proposed in [21]. As a final step for both methods, an iterative refinement was performed with the inlier matches.

The first experiment reports results on 3 sequences from [22], where the first only contains planes, the second contains both planar and non-planar objects and the third does not have any planar surfaces (Fig 4). The size of the data set was reduced from 100 correspondences down to 12 by randomly sampling data points from the original set, and outliers were injected by adding noise sampled from a uniform distribution of mean 0 and standard deviation 5 pixels. In Fig 4, results are shown for our proposed method and for the 5-point algorithm using as input the PCs from the extracted ACs. However, we also performed tests by using as input SIFT features and 6 PCs extracted from the 2 ACs, as proposed in [3]. These results are not shown because they compare unfavourably to the 5-point and 2-AC algorithms, respectively. For the first case, using SIFT features provides a decrease in computational time of about 20% but originates a decrease in accuracy of approximately 9% in translation and 25% in rotation. When extracting points from the ACs, besides this extra overhead, the average error also increased (3.8% in translation and 19.6% in rotation). From Fig 4 it can be seen that the 5-point algorithm is slightly superior when working with large input sets and low percentages of outliers, being, however, similar to the proposed 2-AC method for the first sequence. This implies that the quality

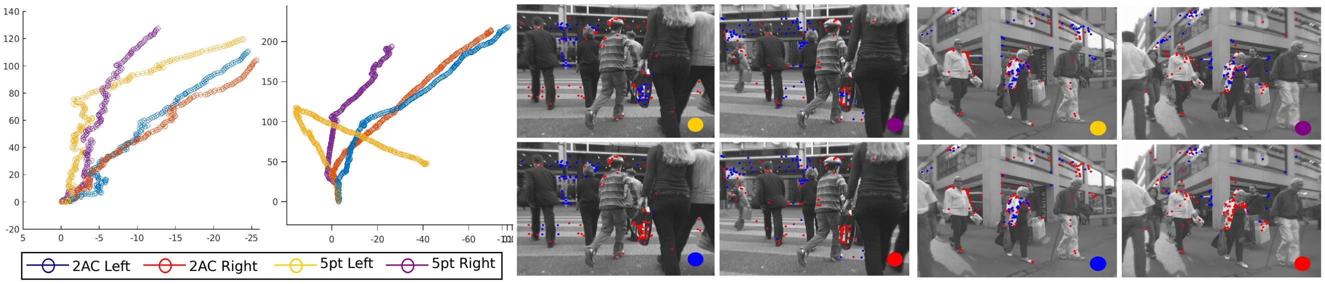


Figure 5: Experiment on the estimation of the essential matrix for a 220-image sequence (trajectories on the left) and a 600-image sequence (trajectories on the right) from the dataset presented in [5]. The estimation was performed for the left and right channels of the stereo pairs independently. The trajectories recovered for each channel, for the proposed 2-AC algorithm and the 5-point algorithm are identified with colors. For each sequence, an example showing the inlier (blue) and outlier (red) points is given. The stereo pair channel and used method are identified with a coloured circle.

of the ACs in this sequence is higher. Another important observation is that our method is significantly more resilient to outliers and small data sets. As an example, the RMS error in translation never exceeds 25° while the 5-point algorithm frequently reaches errors over 60°. Concerning the number of MSAC iterations, it becomes clear that the large decrease in the size of the minimum set of correspondences from 5 to 2 significantly favours a robust estimator. In relation to computational time, the discrepancy is not so evident due to the overhead of AC refinement. However, for large data sets with high percentages of outliers, the 2-AC method still is computationally more efficient.

The last experiment shown in Fig 5 compares the performance of the 5-point algorithm with the proposed 2-AC method in two sequences dominated by strong deformations [5]. They were acquired with calibrated stereo cameras, allowing the estimation of E individually for each channel. Results show that the trajectories obtained with our method are much more similar and smoother than the ones obtained with the 5-point algorithm, suggesting the superiority of the 2-AC approach. This can be confirmed by the examples that show how the 5-point method tends to select as inliers matches that belong to pedestrians or moving objects. From the known extrinsic calibration of the stereo cameras, it is possible to compute the estimation error between left and right channels, for each image. It was observed that similar error distributions were obtained for both methods, with the exception that the proposed method yields less outliers in the distributions. For the first sequence, using 2 ACs and 5 PCs originated 6 and 22 outliers, respectively. For the second, our method originated large errors in only 4 frames, while the 5-point algorithm failed 14 times, of which 3 were rotation errors over 30°. It is important to note that small between-channels errors are obtained whenever a method incorrectly chooses inliers in the same moving objects. The examples in Fig 5 show that this happens often for the 5-point method, leading to trajectories that are not smooth.

The experiments reported in this section confirm that using 2 ACs as opposed to 5 PCs in the estimation of E brings important benefits when dealing with scenarios where it is difficult to extract valid correspondences.

6. Conclusions

We investigate the use of ACs in homography and epipolar geometry estimation and show that both can be accomplished from as few as 2 ACs, benefiting hypothesis-and-test schemes. The geometric insights provided two new error metrics that proved to be very useful in the clustering of points in planes. Also, we derive, for the first time, the general linear constraints for an AC to be compatible with an epipolar geometry. A particular case for these constraints was determined in [3, 2] that is only valid when the coordinate system is centred in the AC. The proposed approaches are successfully applied in planar segmentation and homography estimation tasks, as well as in conventional SfM. In the latter case, our method compares very favourably with the state-of-the-art 5-point algorithm in the presence of high percentages of outliers and/or small input data sets. As future work, we intend to exploit the benefits of using ACs in multi-model fitting for other applications such as piecewise-planar reconstruction and visual odometry in the presence of non-rigid or piecewise rigid structures. We believe it is now possible to solve such problems, that have been tackled in [18] and [13] using stereo cameras, with monocular cameras, providing a significant advance in the literature.

Acknowledgements

Carolina Raposo acknowledges the Portuguese Science Foundation (FCT) for funding her PhD under grant SFRH/BD/88446/2012. The authors also thank FCT and COMPETE2020 program for generous funding through project VisArthro with reference PTDC/EEI-AUT/3024/2014.

References

[1] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004. 1

[2] J. Bentolila and J. Francos. Homography and fundamental matrix estimation from region matches using an affine error metric. *Journal of Mathematical Imaging and Vision*, 49(2):481–491, 2014. 4, 8

[3] J. Bentolila and J. M. Francos. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding*, 122:105 – 114, 2014. 1, 2, 3, 4, 7, 8

[4] O. Chum and J. Matas. Homography estimation from correspondences of local elliptical features. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3236–3239, Nov 2012. 2

[5] A. Ess, B. Leibe, K. Schindler, and L. van Gool. A mobile vision system for robust multi-person tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*. IEEE Press, June 2008. 8

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 1

[7] R. Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):580–593, Jun 1997. 1, 2

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 1, 2, 4, 5, 6

[9] K. Koser, C. Beder, and R. Koch. Conjugate rotation: Parameterization and estimation from an affine feature correspondence. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008. 2, 3

[10] K. Koser and R. Koch. Differential spatial resection - pose estimation using a single local image feature. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision ECCV 2008*, volume 5305 of *Lecture Notes in Computer Science*, pages 312–325. Springer Berlin Heidelberg, 2008. 1, 2, 3

[11] K. Kser. *Geometric estimation with local affine frames and free-form surfaces*. PhD thesis, University of Kiel, 2009. http://d-nb.info/994782322. 1, 2

[12] N. Lazic, B. J. Frey, and P. Aarabi. Solving the uncapacitated facility location problem using message passing algorithms. In Y. W. Teh and D. M. Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10)*, volume 9, pages 429–436, 2010. 2, 5

[13] M. Lourenco, D. Stoyanov, and J. P. Barreto. Visual odometry in stereo endoscopy by using pearl to handle partial scene deformation. In *Augmented Environments for Computer-Assisted Interventions*, volume 8678 of *Lecture Notes in Computer Science*, pages 33–40. Springer International Publishing, 2014. 1, 8

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005. 1, 5

[15] D. Nister. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6):756–770, June 2004. 1, 2

[16] M. Perd'och, J. Matas, and O. Chum. Epipolar geometry from two correspondences. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 215–219, 2006. 1

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, June 2007. 1

[18] C. Raposo, M. Antunes, and J. Barreto. Piecewise-planar stereoscan:structure and motion from plane primitives. In *Computer Vision ECCV 2014*, volume 8690 of *Lecture Notes in Computer Science*, pages 48–63. Springer International Publishing, 2014. 1, 8

[19] F. Riggi, M. Toews, and T. Arbel. Fundamental matrix estimation via tip - transfer of invariant parameters. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 21–24, 2006. 1

[20] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(6):983–995, June 2006. 1

[21] H. Stewénius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60:284–294, June 2006. 1, 2, 4, 7

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012. 5, 6, 7

[23] C. Taylor and A. Cowley. Parsing indoor scenes using rgb-d imagery. In *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2012. 5, 6

[24] P. H. S. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:2000, 2000. 5, 6

[25] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(3):177–280, July 2008. 5

[26] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008. 5

[27] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multiransac algorithm and its application to detect planar homographies. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–153–6, Sept 2005. 5