Affine Transformation from Fundamental Matrix and Two Directions

Nghia Le Minh[a] and Levente Hajder[b]
Department of Algorithms and Their Applications, Eötvös Loránd University, Pázmány Péter stny. 1/C, H-1117, Budapest, Hungary

Keywords: Affine Transformation, Epipolar Geometry, Surface Normal Reconstruction, 3D Reconstruction.

Abstract: Researchers have recently shown that affine transformations between corresponding patches of two images can be applied for 3D reconstruction, including the reconstruction of surface normals. However, the accurate estimation of affine transformations between image patches is very challenging. This paper mainly proposes a novel method to estimate affine transformations from two directions if epipolar geometry of the image pair is known. A reconstruction pipeline is also proposed here in short. As side effects, two proofs are also given. The first one is to determine the relationship between affine transformations and the fundamental matrix, the second one shows how optimal surface normal estimation can be obtained via the roots of a cubic polynomial. A visual debugger is also proposed to validate the estimated surface normals in real images.

1 INTRODUCTION

Stereo vision has been intensively researched for many decades in computer vision (Hartley and Zisserman, 2003). Classical approaches assume that there are point correspondences in two images, and then the 3D geometry of the scene and the camera parameters can be reconstructed. However, it is preferred if the cameras are calibrated, i.e. the case when the intrinsic camera parameters are estimated e.g. by the well-known chessboard-based method of Zhang (Zhang, 2000).

Recently, researchers have started to process the affine transformations between corresponding image patches, not only the corresponding point locations. A local affine transformation represents the warp between the infinitely close area around the corresponding point pairs. It can be applied for homography estimation (Barath et al., 2016), surface normal reconstruction (Köser and Koch, 2008; Barath et al., 2015); recovery of epipoles (Bentolila and Francos, 2014), camera pose estimation (Köser, 2009) as well as structure-from-motion pipelines (Raposo and Barreto, 2016; Hajder and Eichhardt, 2017).

The input of these algorithms are local affine transformations. There are many implementations for detecting the local affinities (Mikolajczyk et al., 2005). Maybe the most effective ones are affine-covariant feature detectors such as Affine-SIFT (Morel and Yu, 2009) or Hessian-Affine (Mikolajczyk and Schmid, 2002).

The goal of this paper is to show that local affine transformations can be estimated from directions around point correspondences. The proposed algorithm is theoretically based on the work of Barath et al. (Barath et al., 2017). They showed that a fundamental matrix gives two constraints for an affine transformation. Their result is exactly the same as the one published in the work of Raposo and Barreto (Raposo and Barreto, 2016), however, their proof has geometric meaning. It is demonstrated here that the other two degrees of freedom (DoF) can be determined by two corresponding directions of the images.

Contribution. The main theoretical contribution of the paper is four-fold: (i) First, the proof of (Barath et al., 2017) is reformulated for the sake of easier understanding. (ii) Then it is shown that an affine transformation can be estimated from two corresponding directions in the images if the fundamental matrix of the stereo setup is known. A linear method is introduced here. (iii) It is shown that the surface normal can be estimated if the intrinsic camera parameters are known. Barath et al. (Barath et al., 2015) showed that the optimal solution can be given via a quartic polynomial, we correct it here and prove that that polynomial is cubic. (iv) It is also demonstrated in the experiments that affine transformations can be calculated from optical flow, and these transformations can be inserted into the reconstruction pipeline.

[a] https://orcid.org/0000-0001-7690-6245
[b] https://orcid.org/0000-0001-9716-9176

Minh, N. and Hajder, L.
Affine Transformation from Fundamental Matrix and Two Directions.
DOI: 10.5220/0009154408190826
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 4: VISAPP*, pages 819-826
ISBN: 978-989-758-402-2; ISSN: 2184-4321

2 AFFINE TRANSFORMATIONS AND EPIPOLAR GEOMETRY

Given two patches, the centers are $\mathbf{p}_1$ and $\mathbf{p}_2$, the affine transformation between the shapes is

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}. \quad (1)$$

The epipolar geometry is represented by the fundamental matrix $\mathbf{F}$. As it is well-known in computer vision, it can be estimated from at least seven point correspondences (Zhang, 1998).

The stereo problem is visualized in Fig. 1. The corresponding epipolar lines are as follows:

$$\mathbf{l}_2 = \mathbf{F}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix},$$
$$\mathbf{l}_1 = \mathbf{F}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}.$$

The line normals, i.e. the perpendicular direction of the lines, are computed as

$$\mathbf{n}_1 = \frac{\hat{\mathbf{l}}_1}{\sqrt{\hat{\mathbf{l}}_1^T\hat{\mathbf{l}}_1}} = \frac{\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}}{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2},$$
$$\mathbf{n}_2 = \frac{\hat{\mathbf{l}}_2}{\sqrt{\hat{\mathbf{l}}_2^T\hat{\mathbf{l}}_2}} = \frac{\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}}{\left\|\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}\right\|_2}.$$

where $\hat{\mathbf{l}}$ denotes the first two coordinates of line parameters, represented by vector $\mathbf{x}$, matrices $\tilde{\mathbf{F}}$ and $\hat{\mathbf{F}}$ consist of the first (left) two columns and (top) two rows of the fundamental matrix $\mathbf{F}$, respectively.

Points $\mathbf{p}_1$ and $\mathbf{p}_2$ lie on $\mathbf{l}_1$, and $\mathbf{l}_2$, respectively. Therefore, $\mathbf{l}_1^T\mathbf{p}_1 = 0$ and $\mathbf{l}_2^T\mathbf{p}_2 = 0$.

If another point $\mathbf{q}_1 = \mathbf{p}_1 + \varepsilon\mathbf{n_1}$ is taken in the first image, where $\mathbf{n}_1$ is the normal vector of line $\mathbf{l}_1$, then the corresponding epipolar line in the other image is

$$\mathbf{l}_2' = \mathbf{F}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \mathbf{F}\begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}.$$

The distance d of the original point $\mathbf{p}_2$ and $\mathbf{l}_2'$ is

$$d = \frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T \mathbf{l}_2'}{\sqrt{\hat{\mathbf{l}_2}'^T\hat{\mathbf{l}_2}'}},$$

where $\hat{\mathbf{l}_2'}$ is the first two coordinates of line parameters $\mathbf{l_2'}$. Thus

$$\hat{\mathbf{l}_2'} = \hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right).$$

The denominator is as follows:

$$\sqrt{\hat{\mathbf{l}_2}'^T\hat{\mathbf{l}_2}'} = \left\|\hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)\right\|_2.$$

After elementary modifications, the formula for distance d can be written as follows:

$$d = \frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\left(\mathbf{F}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \mathbf{F}\begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)}{\left\|\hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)\right\|_2} = \varepsilon\frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\mathbf{F}\begin{bmatrix} \mathbf{n}_1 \\ 0 \end{bmatrix}}{\left\|\hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)\right\|_2},$$

since $\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\mathbf{F}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} = 0$. Therefore,

$$d = \varepsilon\frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\tilde{\mathbf{F}}\mathbf{n_1}}{\left\|\hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)\right\|_2}.$$

The scale of the problem can be calculated by moving the points to infinitely close to the original position $\mathbf{p}_1$. Then,

$$s = \lim_{\varepsilon\to 0}\frac{d}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\tilde{\mathbf{F}}\mathbf{n_1}}{\left\|\hat{\mathbf{F}}\left(\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon\mathbf{n}_1 \\ 0 \end{bmatrix}\right)\right\|_2} = \frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\tilde{\mathbf{F}}\frac{\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \end{bmatrix}}{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2}}{\left\|\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}\right\|_2} = \frac{\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}^T\tilde{\mathbf{F}}\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}}{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2\left\|\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}\right\|_2} =$$
$$\frac{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2^2}{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2\left\|\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}\right\|_2} = \frac{\left\|\tilde{\mathbf{F}}^T\begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}\right\|_2}{\left\|\hat{\mathbf{F}}\begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}\right\|_2}.$$

An affinity transforms the normal of an epipolar line as $\mathbf{A}^T\mathbf{n}_2 = s\mathbf{n}_1$, where s is the scale. This scale can be eliminated by forcing the directions to be unit size.

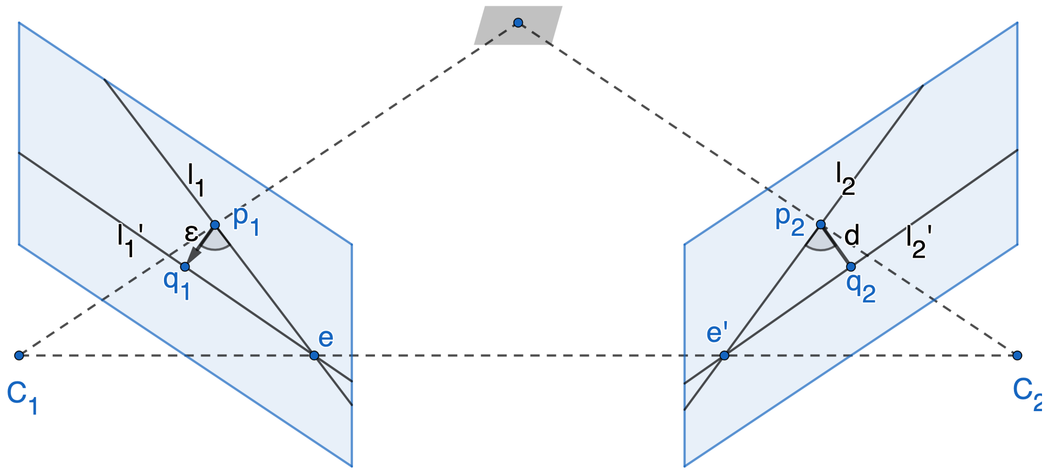

Figure 1: Basic stereo problem for scale estimation. Corresponding point pairs are $(\mathbf{p}_1, \mathbf{p}_2)$ and $(\mathbf{q}_1, \mathbf{q}_2)$; corresponding lines denoted by $(\mathbf{l}_1, \mathbf{l}_2)$ and $(\mathbf{l}'_1, \mathbf{l}'_2)$. The scale to be calculated is given by the ratio d/ε.

If the scale is substituted, and the length of the normals of epipolar lines are forced to be one, the following equation is obtained:

$$\mathbf{A}^{\mathbf{T}} \frac{\hat{\mathbf{F}} \begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix}}{\left\| \hat{\mathbf{F}} \begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} \right\|_2} = \frac{\left\| \tilde{\mathbf{F}}^T \begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix} \right\|_2}{\left\| \hat{\mathbf{F}} \begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} \right\|_2} \frac{\tilde{\mathbf{F}}^T \begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}}{\left\| \tilde{\mathbf{F}}^T \begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix} \right\|_2}.$$

The final formula that connects fundamental matrix, point locations and affine correspondences is as follows

$$\mathbf{A}^{\mathbf{T}} \hat{\mathbf{F}} \begin{bmatrix} \mathbf{p}_1 \\ 1 \end{bmatrix} = -\tilde{\mathbf{F}}^T \begin{bmatrix} \mathbf{p}_2 \\ 1 \end{bmatrix}. \tag{2}$$

The minus sign comes from the fact that the cross product matrix is a skew-symmetric one and the determinant of the camera matrices have the same sign[1]. This formula states that a 2D vector equation can be written for a valid affine transformation if the epipolar geometry, represented by the fundamental matrix, is known.

3 ESTIMATION OF AN AFFINE TRANSFORMATION IF EPIPOLAR GEOMETRY IS KNOWN

The goal of this section is to show how an affinity can be estimated for a stereo correspondence if the locations and two corresponding directions are given in the images. The locations in the images are denoted by $\mathbf{p_1} = \begin{bmatrix} u_1 & v_1 \end{bmatrix}^T$ and $\mathbf{p_2} = \begin{bmatrix} u_2 & v_2 \end{bmatrix}^T$, while the directions by $\mathbf{d}_{1i} = \begin{bmatrix} u_{1i} & v_{1i} \end{bmatrix}^T$ and $\mathbf{d}_{2i} = \begin{bmatrix} u_{2i} & v_{2i} \end{bmatrix}^T, i \in \{1, 2\}$. The affine transformation is written by a 2×2 matrix as it is defined in Equation 1.

[1]In other words, both image coordinate systems are left or right handed.

3.1 Estimation of an Affine Transformation

The relationship between point locations, affine transformation, and the fundamental matrix is given in Eq. 2. Substituting the coordinates, the following formula is obtained:

$$\mathbf{A}^T \tilde{\mathbf{F}} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = -\hat{\mathbf{F}}^T \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix}$$

As the multiplication of the fundamental matrix and point locations gives the epipolar line in the second image, and the transpose of the fundamental matrix and second point location yield the corresponding epipolar line in the first image, the vector-equation can be rewritten as

$$\mathbf{A}^T \begin{bmatrix} l_{2u} \\ l_{2v} \end{bmatrix} = - \begin{bmatrix} l_{1u} \\ l_{1v} \end{bmatrix},$$

where the normals of the epipolar lines are the vectors $\mathbf{n}_1 = [l_{1u} \quad l_{1v}]^T$ and $\mathbf{n}_2 = [l_{2u} \quad l_{2v}]^T$.

If the elements of the affine transformations are substituted, the following linear system of equations is given:

$$\begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix} \begin{bmatrix} l_{2u} \\ l_{2v} \end{bmatrix} = - \begin{bmatrix} l_{1u} \\ l_{1v} \end{bmatrix}. \tag{3}$$

Now the connection between the known directions is written in conjunction with the affine parameters. The directions are correctly transformed by the affine matrix, however, the lengths of the vectors are not known. This fact can be formulated as

$$\mathbf{A}\begin{bmatrix} u_{11} \\ v_{11} \end{bmatrix} = \alpha_1 \begin{bmatrix} u_{21} \\ v_{21} \end{bmatrix},$$
$$\mathbf{A}\begin{bmatrix} u_{12} \\ v_{12} \end{bmatrix} = \alpha_2 \begin{bmatrix} u_{22} \\ v_{22} \end{bmatrix},$$

where α_1 and α_2 are the unknown lengths.

It can be straightforwardly rewritten by substituting the elements of the affine transformations as

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}\begin{bmatrix} u_{11} \\ v_{11} \end{bmatrix} = \alpha_1 \begin{bmatrix} u_{21} \\ v_{21} \end{bmatrix},$$
$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}\begin{bmatrix} u_{12} \\ v_{12} \end{bmatrix} = \alpha_2 \begin{bmatrix} u_{22} \\ v_{22} \end{bmatrix}. \tag{4}$$

The final problem can be formed by merging Equations 3 and 4. The problem is a six-dimensional linear one, it is written in Equation 5. The solution is trivially obtained by multiplying the right vector by the inverse of the coefficient matrix from the left.

3.2 Affine Transformation from Optical Flow

The affine transformations can be estimated by other techniques, e.g. using an affine-invariant matcher (Yu and Morel, 2011). However, from our experience, the quality of those are not satisfactory, because the estimated transformations are highly contaminated.

We have tried another way: the affine transformations are estimated if the optical flow between the images is available. The estimation problem is an inhomogeneous linear one as it is discussed in the appendix. Therefore, the estimation is very fast, the pseudo-inverse of a matrix, corresponding to a 6D problem, has to be computed.

Problem Statement. An optical flow is given, thus the relative offset for each camera pixels are known as

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix},$$

where the vector $[x_i \quad y_i]^T$ and $[x'_i \quad y'_i]^T$ denote the pixel coordinates in the first and second images, respectively. The flow itself is represented by the offset vectors $[\Delta x_i \quad \Delta y_i]^T$.

The task is to estimate the affine transformation around the given point location $\mathbf{x}_0 = [x_0, y_0]$.'Around' means that the neighbouring pixels has to be considered. they are selected in a disk with radius R, where R is a parameter of the algorithm.

Proposed Solution. The affine transformation represents the relations between corresponding neighbouring points as

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \end{bmatrix}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}\begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix}.$$

In other form:

$$\begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i & y_i & 1 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ a_5 \\ a_3 \\ a_4 \\ a_6 \end{bmatrix} = \begin{bmatrix} x'_i \\ y'_i \end{bmatrix}.$$

Left matrix is denoted by $\mathbf{C}_i$. then

$$\mathbf{C}_i \begin{bmatrix} a_1 & a_2 & a_5 & a_3 & a_4 & a_6 \end{bmatrix}^T = \mathbf{x}'_\mathbf{i}.$$

If we have N different locations, the problem is over-determined:

$$\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_N \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ a_5 \\ a_3 \\ a_4 \\ a_6 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix}$$

It should be considered for all pixels that fulfill the constraint $(\mathbf{x}_0 - \mathbf{x}_i)^T (\mathbf{x}_0 - \mathbf{x}_i) < R^2$. The problem is linear:

$$\mathbf{C}\begin{bmatrix} a_1 & a_2 & a_5 & a_3 & a_4 & a_6 \end{bmatrix}^T = \mathbf{x}',$$

where $\mathbf{C} = [\mathbf{C}_\mathbf{1}^\mathbf{T} \dots \mathbf{C}_\mathbf{N}^\mathbf{T}]^T$ and $\mathbf{x} = [\mathbf{x}_\mathbf{1}^\mathbf{T} \dots \mathbf{x}_\mathbf{N}^\mathbf{T}]^T$. The optimal solution for the affine parameters are estimated as follows:

$$\begin{bmatrix} a_1 & a_2 & a_5 & a_3 & a_4 & a_6 \end{bmatrix}^T = \left(\mathbf{C}^\mathbf{T}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{x}'.$$

Remark that optical flow for at least three locations is required. With more than three points the problem is over-determined.

$$\begin{bmatrix} l_{2u} & 0 & l_{2v} & 0 & 0 & 0 \\ 0 & l_{2u} & 0 & l_{2v} & 0 & 0 \\ u_{11} & v_{11} & 0 & 0 & -u_{21} & 0 \\ 0 & 0 & u_{11} & v_{11} & -v_{21} & 0 \\ u_{12} & v_{12} & 0 & 0 & 0 & -u_{22} \\ 0 & 0 & u_{12} & v_{12} & 0 & -v_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} -l_{1u} \\ -l_{1v} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

4 OPTIMAL SURFACE NORMAL ESTIMATION

In this section, we prove that the polynomial, defined in (Barath et al., 2015) is cubic and not quartic.

If we restrict the estimating normal into the form $n = [x, y, 1-x-y]^T$ (i.e the sum of coordinates is 1), then the roots of the mentioned polynomial are the possible values of x.

The notations come from the original paper[2]. Unfortunately, the full proof cannot be repeated here due to the page limit of the submission.

The coefficients for the polynomials are determined by the point locations in the stereo image pair and the related affine transformation.

The initial facts for our proof is that the following rules can be formed:

$$\Omega_k^1 = \Psi_k^2 = 0, \qquad \Omega_k^2 = -\Psi_k^1,$$

as it is written in page 21 of the paper (Barath et al., 2015).

Then the coefficients for the quadratic curves are as follows[3]:

$$A_2 = \sum_{k=1}^{4} \Omega_k \Omega_k^2, \qquad B_1 = \sum_{k=1}^{4} \Psi_k \Psi_k^1$$

$$C_1 = \sum_{k=1}^{4} (\Omega_k^1 \Psi_k + \Psi_k^1 \Omega_k) = \sum_{k=1}^{4} \Psi_k^1 \Omega_k$$

$$C_2 = \sum_{k=1}^{4} (\Omega_k^2 \Psi_k + \Psi_k^2 \Omega_k) = \sum_{k=1}^{4} \Omega_k^2 \Psi_k$$

Coefficient for x^4, i.e. the highest one of the polynomial, is as follows:

$$A_2^2 B_1 - A_2 C_1 C_2 = A_2 (A_2 B_1 - C_1 C_2)$$

$$= A_2 \Big(\sum_{k=1}^{4} \Omega_k \Omega_k^2 \sum_{k=1}^{4} \Psi_k \Psi_k^1 - \sum_{k=1}^{4} \Psi_k^1 \Omega_k \sum_{k=1}^{4} \Omega_k^2 \Psi_k \Big)$$

$$= A_2 \Big(\sum_{i=1}^{4} \sum_{j=1}^{4} \Omega_i \Omega_i^2 \Psi_j \Psi_j^1 - \sum_{i=1}^{4} \sum_{j=1}^{4} \Psi_i^1 \Omega_i \Omega_j^2 \Psi_j \Big)$$

$$= A_2 \Big(\sum_{i=1}^{4} \sum_{j=1}^{4} \Omega_i \Omega_i^2 \Psi_j \Psi_j^1 - \sum_{i=1}^{4} \sum_{j=1}^{4} \Psi_i^1 \Omega_i \Omega_j^2 \Psi_j \Big)$$

$$= A_2 \Big(\sum_{i=1}^{4} \sum_{j=1}^{4} \Omega_i (-\Psi_i^1) \Psi_j (-\Omega_j^2) - \sum_{i=1}^{4} \sum_{j=1}^{4} \Psi_i^1 \Omega_i \Omega_j^2 \Psi_j \Big)$$

$$= A_2 \Big(\sum_{i=1}^{4} \sum_{j=1}^{4} \Omega_i \Psi_i^1 \Psi_j \Omega_j^2 - \sum_{i=1}^{4} \sum_{j=1}^{4} \Psi_i^1 \Omega_i \Omega_j^2 \Psi_j \Big) = 0$$

Thus, the highest coefficient vanishes, therefore the degree of the polynomial is three, in other words, the polynomial is cubic. $\square$.

5 RECONSTRUCTION PIPELINE

This section overviews the components of our reconstruction pieline.

In the first stage, point correspondences are detected using ASIFT (Yu and Morel, 2011). Then the fundamental matrix is estimated by the eight-point method (Hartley, 1995). As the cameras are precalibrated, essential matrix can also be retrieved from fundamental matrix and intrinsic camera parameters.

In order to estimate affine transformations, two corresponding directions have to be found. First, Line Segment Detector (von Gioi et al., 2012) and Line Binary Descriptor (Zhang and Koch, 2013) are applied to the image patches centered around feature points so that at least two segment pairs are matched. The directions are obtained through the normalized direction vector of those segments. Then affinities are estimated from the fundamental matrix and two directions as discussed in Section 3.

With the feature point locations and the recovered camera pose [4], one can compute a sparse 3D reconstruction. This step is carried out by applying Hartley-Sturm triangulation (Hartley and Sturm, 1997). Finally, optimal surface normals can be estimated from

[2]The proof described here cannot be understood without reading the original paper. The related part is found in the appendix of the paper (Barath et al., 2015), however, Section 2, i.e. geometric background, has to be read as well.

[3]Not all coefficients are listed, only the ones that are required to understand the proof.

[4]Camera extrinsic parameters, i.e. the relative pose, can be obtained by decomposing the essential matrix (Hartley and Zisserman, 2003)

affine transformations by the proposed modification, overviewed in Section 4, of the method of Barath et al (Barath et al., 2015).

6 EXPERIMENTS

For testing, we concentrate mainly on real-world case as the main goal of our work is to insert the surface normal estimation into a 3D reconstruction pipeline.

6.1 Synthetic Test

Synthetic test was only constructed in order to validate that the formula, given in Eq. 5, is correct. For this purpose, a simple synthetic testing environment was implemented in Octave [5]. Camera parameters as well as the scene geometry, i.e. a sphere in our environment, was randomly generated, point locations were given by projecting the points into the camera image. Ground-truth affine transformations were generating via the tangent planes of the sphere. The affine parameters can be determined by derivating the homographies, related to the tangent planes, at the corresponding point locations as it is discussed in the paper of Barath et al. (Barath et al., 2015) in detail.

Conclusion of Synthetic Test. It was successfully validated that Equation 5 is correct, the GT affine transformations were always exactly retrieved.

6.2 Visual Debugger

We have developed a tool in order to run the reconstruction pipeline and visualize the computed surface normals. We call this tool as Visual Debugger. The point correspondences as well as the directions are manually selected in order to avoid detection errors. The fundamental matrix is automatically estimated by the eight-point method (Hartley, 1995). Robustification is implemented using the standard RANSAC (Fischler and Bolles, 1981) scheme.

A few results by the Visual Debugger are seen in Figure 2. We have tested our method on the KITTI (Geiger et al., 2012) and Malaga (Blanco et al., 2014) datasets.

6.3 Surface Normals

A fully automatic testing procedure were also carried out using the whole reconstruction pipeline. An example with the visualized normal vectors is pictured in Figure 3. The normals are differently colored: the absolute values of the coordinates of the normals are considered; if the largest absolute coordinate of a vector is x, y, and z, then yellow, white, and red color is used for drawing it, respectively.

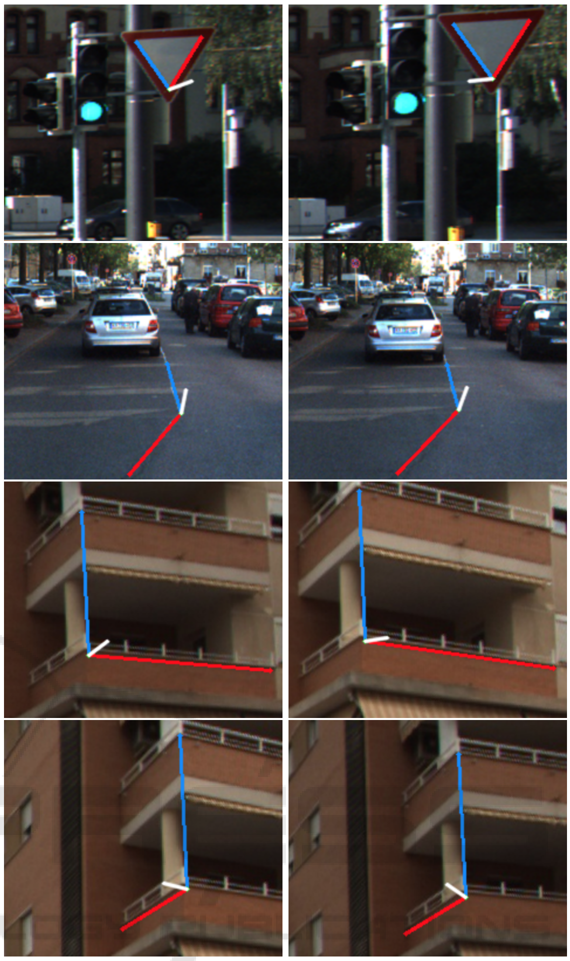

Figure 2: Results on real sequences. Each row consists of a stereo image pair. The manually selected directions are colored by red and blue. The estimated surface normals are visualized by white. *Best viewed in color*.

6.4 Surface Normal Reconstruction from Optical Flow

Although the focus of this paper is to show that affine transformations can be estimated from two corresponding directions and the fundamental matrix, we demonstrated here that there are other ways to efficiently estimate the affine transformations. In Sec. 3.2, it is overviewed how an transformation can be retrieved at a given location if the optical flow between the images is given. The surface normals can be estimated by the optimal method, proposed in Sec 4.

[5]Octave is an open-source MATLAB-clone. See http://www.octave.org.

Figure 3: Estimated surface normals visualized in an image of one of the KITTI (Geiger et al., 2012) sequences. Yellow, white and red coordinates are used for horizontal, vertical and perpendicular, to the image plane, directions. *Best viewed in color.*

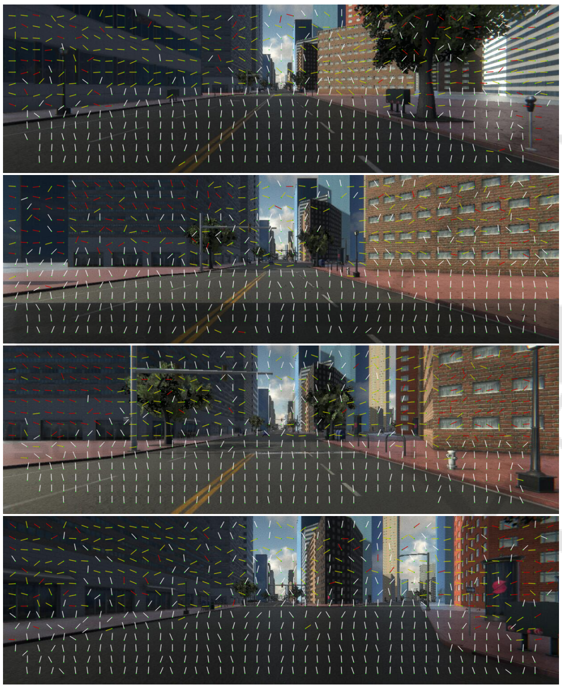

Figure 4: Estimated surface normals visualized in images generated by simulator LG-SVL (LG-, 2019). Yellow, white and red coordinates are used for horizontal, vertical and perpendicular, to the image plane, directions. Affine transformations are computed from optical flow. *Best viewed in color.*.

The image sequences were generated by the LG-SVL Simulator (LG-, 2019). The optical flows were computed by the pre-trained deep network HD3 (Yin et al., 2019). Resulting images are visualized in Figure 4, the estimated surface normals are drawn. Although affine transformations can be estimated for each pixel location, a regular sparse grid is applied for sampling due to easier interpretation.

7 CONCLUSIONS AND FUTURE WORK

In this position paper, we have presented a novel reconstruction pipeline in order to compute the surface normals of a 3D scene. It has been shown that an affine transformation can be estimated from a point and two line (direction) correspondences if the fundamental matrix is known. The surface normals themselves can be estimated via the roots of cubic polynomials.

Future Work. This paper concentrates on the theoretic aspects of the problem. More quantitative and qualitative tests are required in order to validate the practical applicability of the proposed reconstruction pipeline.

The four DoF of affine transformations means that the knowledge of epipolar geometry and two directions are enough to estimate a local affine transformation. However, there are degenerate cases when a direction and the fundamental matrix represent the same information for the estimation. These degenerate cases have also to be found and discussed. This is missing in the current form of the paper.

ACKNOWLEDGEMENTS

Nghia Le Minh was supported by the project `EFOP-3.6.3-VEKOP-16- 2017-00001`: Talent Management in Autonomous Vehicle Control Technologies, by the Hungarian Government and co-financed by the European Social Fund. Levente Hajder was supported by the project no. `ED_18-1-2019-0030`: Application domain specific highly reliable IT solutions subprogramme. It has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Thematic Excellence Programme funding scheme.

REFERENCES

(2019). LGSVL Simulator. https://www.lgsvlsimulator.com/.

Barath, D., Molnar, J., and Hajder, L. (2015). Optimal Surface Normal from Affine Transformation. In *VISAPP 2015*, pages 305–316.

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

Barath, D., Toth, T., and Hajder, L. (2017). A minimal solution for two-view focal-length estimation using two affine correspondences. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2557–2565.

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

Blanco, J.-L., Moreno, F.-A., and Gonzalez-Jimenez, J. (2014). The málaga urban dataset: High-rate stereo and lidars in a realistic urban scenario. *International Journal of Robotics Research*, 33(2):207–214.

Fischler, M. and Bolles, R. (1981). RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:358–367.

Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hajder, L. and Eichhardt, I. (2017). Computer vision meets geometric modeling: Multi-view reconstruction of surface points and normals using affine correspondences. In *2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017, Venice, Italy, October 22-29, 2017*, pages 2427–2435.

Hartley, R. I. (1995). In defence of the 8-point algorithm. *International Conference on Computer Vision*, pages 1064–1070.

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

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

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.

Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In *Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part (I)*, pages 128–142.

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

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

Raposo, C. and Barreto, J. P. (2016). Theory and practice of structure-from-motion using affine correspondences. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 5470–5478.

von Gioi, R. G., Jakubowicz, J., Morel, J., and Randall, G. (2012). LSD: a line segment detector. *IPOL Journal*, 2:35–55.

Yin, Z., Darrell, T., and Yu, F. (2019). Hierarchical discrete distribution decomposition for match density estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yu, G. and Morel, J. (2011). ASIFT: an algorithm for fully affine invariant comparison. *IPOL Journal*, 1:11–38.

Zhang, L. and Koch, R. (2013). An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *J. Visual Communication and Image Representation*, 24(7):794–805.

Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195.

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