Ninth Hungarian Conference on Computer Graphics and Geometry, Budapest, 2018

Structure from Motion via Affine Correspondences

Ivan Eichhardt,[1,2] and Levente Hajder[1,2]

[1] Eötvös Loránd University, Budapest, Hungary
[2] MTA SZTAKI, Budapest, Hungary

Abstract

A novel surface normal estimator is introduced using affine-invariant features extracted and tracked across multiple views. Normal estimation is robustified and integrated into our reconstruction pipeline that has increased accuracy compared to the State-of-the-Art. Parameters of the views and the obtained spatial model, including surface normals, are refined by a novel bundle adjustment-like numerical optimization. The process is an alternation with a novel robust view-dependent consistency check for surface normals, removing normals inconsistent with the multiple-view track. Our algorithms are quantitatively validated on the reverse engineering of geometrical elements such as planes, spheres, or cylinders. It is shown here that the accuracy of the estimated surface properties is appropriate for object detection. The pipeline is also tested on the reconstruction of free-form objects.

1. Introduction

One of the fundamental goals of image-based 3D computer vision[17] is to extract spatial geometry using correspondences tracked through at least two images. The reconstructed geometry may have a number of different representations: points clouds, oriented point clouds, triangulated meshes with/without texture, continuous surfaces, etc. However, frequently used reconstruction pipelines[9, 15, 2, 27] deal only with the reconstruction of dense or semi-dense point clouds. These methods include Structure from Motion (SfM) algorithms[17] for which the input are 2D coordinates of corresponding feature points in the images.

These feature points used to be detected and matched by classical algorithms such as the one proposed by Kanade-Lucas-Tomasi[35, 5], but nowadays affine-covariant feature[21, 7, 37] or region[22] detectors are frequently used due to their robustness to viewpoint changes. These detectors provide not only the locations of the features, but the shapes of those can be retrieved as well. The features are usually represented by locations and small patches composed of the neighboring pixels. The retrieved shapes determine the warping parameters of the corresponding patches between the images. The first order approximation of a warping is an affinity[24], there are techniques such as ASIFT[26] that can efficiently compute the affinity. Affine-covariant feature detectors[21, 7, 37] are invariant to translation, rotation, and scaling. Therefore, features and patches can be matched between images very accurately.

State-of-the-art 3D reconstruction methods usually resort only to the location of the region centers. The main purpose of this paper is to show that Affine Correspondences (ACs) can significantly enhance the quality of the reconstruction compared to the case when only 2D locations are considered. However, the application of ACs does not count as a novelty in computer vision. Matas *et al.*[23] showed that image rectification is possible if the affine transformation is known between two patches, then the rectification can aid further patch matching. Köser & Koch[19] proved that camera pose estimation is possible if only the affine transformation between two corresponding patches is known. Epipolar geometry of a stereo image pair can also be determined from affine transformations of multiple corresponding patches. This is possible if at least two correspondences are taken as it was demonstrated by Perdoch *et al.*[29]. Bentolila *et al.*[8] proved that three affine transformations give sufficient information to estimate the epipole in stereo images. Lakemond *et al.*[20] discussed that an affine transformation gives additional information for feature correspondence matching, useful for wide-baseline stereo reconstruction.

Theoretically, this work is inspired by the recent studyies of Molnar and Eichhardt[25] and Barath *et al.*[6]. They showed that the affine transformation between corresponding patches of a stereo image pair can be expressed using the camera parameters and the related normal vector. The main theoretical value in their works is the deduction of a general relationship between camera parameters, surface normals and spatial coordinates. Moreover, they proposed several surface normal estimators for the two-view case in[6], including an L_2-optimal one. In our paper, their work is extended to the multi-view case, with robust view-dependent geometric filtering, removing normals inconsistent with the multiple-view track.

Our research is also inspired by multi-view image-based algorithms such as Furukawa & Ponce[16] and Delaunoy & Pollefeys[11]. The former one, similarly to our work, also has a way to estimate surface normals, however, Bundle Adjustment[4] (BA) is not applied after their reconstruction, and the normal estimation is based on photometric similarity using normalized cross correlation. The latter study extends the point-based BA with a photometric error term. In this paper, we propose a complex reconstruction pipeline including surface point and normal estimation followed by robust BA.

One field of applications of accurate 3D reconstruction is Reverse Engineering[31] (RE)[†], the proposed reconstruction pipeline is validated on the RE of geometrical elements. RE algorithms are usually based on non-contact scanners such as laser or structured-light equipments, but there are cases when the object to be scanned is not available at hand, only images of it. Software to reconstruct planar surfaces using solely camera images already exist, e.g. Insight3D[1‡], however, *ours is the first study, to the best of our knowledge, that deals with the reconstruction of spheres and cylinders based on images.*

The **contributions** of our paper are as follows:

- A novel multi-view normal estimator is proposed. To the best of our knowledge, only stereo algorithms[6, 19] exist to estimate surface normals.
- A novel Bundle Adjustment (BA) algorithm is introduced that simultaneously optimizes the camera parameters, with an alternating step that removes outlying surface normals.
- It is showed that the quality of the surface points and normals resulted by the proposed AC-based reconstruction is satisfactory for object fitting algorithms. In other words, image-based reconstruction and reverse engineering can be integrated.
- The proposed algorithm can cope with arbitrary central projective cameras, not only perspective ones are considered, providing surface normals using a wide range of cameras.

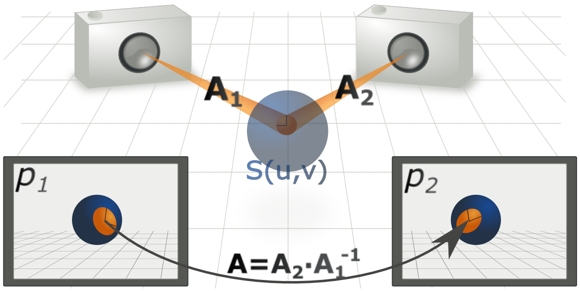

Figure 1: Illustration of cameras represented by projection functions p_i, $i = 1,2$. $\mathbf{A}_i$ is the local mapping between the surface $S(u,v)$ and its projection onto image i. Relative affine transformation between images is denoted by matrix $\mathbf{A}$.

2. Surface Normal Estimation.

An Affine Correspondence (AC) is a triplet $(\mathbf{A}, \mathbf{x}_1, \mathbf{x}_2)$ of a 2×2 relative affine transformation matrix $\mathbf{A}$ and the corresponding point pair $\mathbf{x}_1, \mathbf{x}_2$. $\mathbf{A}$ is a mapping between the infinitesimally small environments of $\mathbf{x}_1$ and $\mathbf{x}_2$ on the image planes. ACs can be extracted from an image pair using affine-covariant feature detectors[21, 7, 26, 37].

Let us consider $S(u,v) \in \mathbb{R}^3$, a continuously differentiable parametric surface and function $p_i : \mathbb{R}^3 \to \mathbb{R}^2$, the camera model, projecting points of S in 3D onto image 'i':

$$\mathbf{x}_i \doteq p_i\left(S\left(u_0, v_0\right)\right), \tag{1}$$

for a point $(u_0, v_0) \in \mathrm{dom}(S)$. Assume that the pose of view i is included in the projection function p_i. The Jacobian of the right hand side of Eq. (1) is obtained using the chain rule as follows:

$$\mathbf{A}_i \doteq \nabla_{u,v}\left[\mathbf{x}_i\right] = \nabla p_i\left(\mathbf{X}_0\right) \nabla S\left(u_0, v_0\right), \tag{2}$$

where $\mathbf{X}_0 = S(u_0, v_0)$ is a point of the surface. $\mathbf{A}_i$ can be interpreted as a local relative affine transformation between small environments of the surface S at the point (u_0, v_0) and its projection at the point $\mathbf{x}_i$. Remark that the size of matrices $\nabla p_i(\mathbf{X}_0)$ and $\nabla S(u_0, v_0)$ are 2×3 and 3×2. See Fig. 1 for the explanation of the parameters.

Matrix $\mathbf{A}$, the relative transformation part of ACs, can also be expressed using the Jacobians defined in Eq. (2) as follows

$$\mathbf{A}_2\mathbf{A}_1^{-1} = \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{3}$$

Two-view Surface Normal Estimation The relationship[6] of the surface normals and affine transformations are as follows:

$$\mathbf{A}_2\mathbf{A}_1^{-1} \sim \left[\mathbf{w}_{ij} \cdot \mathbf{n}\right]_{i,j} = \begin{bmatrix} \mathbf{w}_{11} \cdot \mathbf{n} & \mathbf{w}_{12} \cdot \mathbf{n} \\ \mathbf{w}_{21} \cdot \mathbf{n} & \mathbf{w}_{22} \cdot \mathbf{n} \end{bmatrix}, \tag{4}$$

[†] Reverse engineering, also called back engineering, is the processes of extracting knowledge or design information from anything man-made and re-producing it or re-producing anything based on the extracted information. Definition by Wikipedia.

[‡] Insight3D is an open-source images-based 3D modeling software.

where

$$\begin{aligned}\mathbf{w}_{ij} &\doteq \delta_j\left(\mathbf{a}_{2-j+1}^T \times \mathbf{b}_i^T\right),\\ \delta_j &= \begin{cases}1, & \text{if } (j=1)\\ -1, & \text{if } (j=2),\end{cases}\\ \begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} &= \nabla p_1(\mathbf{X}_0),\\ \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} &= \nabla p_2(\mathbf{X}_0),\\ \begin{bmatrix}\mathbf{S}_u & \mathbf{S}_v\end{bmatrix} &= \nabla S(u_0, v_0).\end{aligned}$$

Operator $\sim$ denotes equality up to a scale.

The above relation in Eq. (4) is deduced through a series of equivalent and up-to-a-scale transformations, using a property[24] of differential geometry $[\mathbf{n}]_\times \sim \left(\mathbf{S}_v\mathbf{S}_u^T - \mathbf{S}_v\mathbf{S}_u^T\right)$ with $\|\mathbf{n}\| = 1$:

$$\begin{aligned}\mathbf{A} =&\mathbf{A}_2\mathbf{A}_1^{-1} \sim \mathbf{A}_2 \operatorname{adj}(\mathbf{A}_1) =\\ &= \cdots =\\ &= \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix}\left(\mathbf{S}_v\mathbf{S}_u^T - \mathbf{S}_v\mathbf{S}_u^T\right)\begin{bmatrix}\mathbf{a}_2^T & -\mathbf{a}_1^T\end{bmatrix} \sim\\ &\sim \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix}[\mathbf{n}]_\times\begin{bmatrix}\mathbf{a}_2^T & -\mathbf{a}_1^T\end{bmatrix} =\\ &= \left[\delta_j\left(\mathbf{a}_{2-j+1}^T \times \mathbf{b}_i^T\right)\right]_{i,j} =\\ &= \left[\mathbf{w}_{ij}\cdot\mathbf{n}\right]_{i,j}. \end{aligned} \tag{5}$$

The relation between the measured relative transformation $\mathbf{A}$ and the formulation (4) is as follows:

$$\begin{aligned}a_{11} &\sim \mathbf{w}_{11}\cdot\mathbf{n},\\ a_{12} &\sim \mathbf{w}_{12}\cdot\mathbf{n},\\ a_{21} &\sim \mathbf{w}_{21}\cdot\mathbf{n},\\ a_{22} &\sim \mathbf{w}_{22}\cdot\mathbf{n}.\end{aligned} \tag{6}$$

To remove the common scale ambiguity we divide these up-to-a-scale equations in all possible combinations:

$$\begin{aligned}&\frac{a_{11}}{a_{12}} = \frac{\mathbf{w}_{11}\cdot\mathbf{n}}{\mathbf{w}_{12}\cdot\mathbf{n}}, \frac{a_{11}}{a_{21}} = \frac{\mathbf{w}_{11}\cdot\mathbf{n}}{\mathbf{w}_{21}\cdot\mathbf{n}}, \frac{a_{11}}{a_{22}} = \frac{\mathbf{w}_{11}\cdot\mathbf{n}}{\mathbf{w}_{22}\cdot\mathbf{n}},\\ &\frac{a_{12}}{a_{21}} = \frac{\mathbf{w}_{12}\cdot\mathbf{n}}{\mathbf{w}_{21}\cdot\mathbf{n}}, \frac{a_{12}}{a_{22}} = \frac{\mathbf{w}_{12}\cdot\mathbf{n}}{\mathbf{w}_{22}\cdot\mathbf{n}}, \frac{a_{21}}{a_{22}} = \frac{\mathbf{w}_{21}\cdot\mathbf{n}}{\mathbf{w}_{22}\cdot\mathbf{n}}.\end{aligned} \tag{7}$$

The surface normal $\mathbf{n}$ can be estimated by solving the following homogeneous system of linear equations:

$$\begin{bmatrix}a_{11}\mathbf{w}_{12} - a_{12}\mathbf{w}_{11}\\ a_{11}\mathbf{w}_{21} - a_{21}\mathbf{w}_{11}\\ a_{11}\mathbf{w}_{22} - a_{22}\mathbf{w}_{11}\\ a_{12}\mathbf{w}_{21} - a_{21}\mathbf{w}_{12}\\ a_{12}\mathbf{w}_{22} - a_{22}\mathbf{w}_{12}\\ a_{21}\mathbf{w}_{22} - a_{22}\mathbf{w}_{21}\end{bmatrix}\mathbf{n} = \mathbf{0}, \text{ s.t. } \|\mathbf{n}\| = 1. \tag{8}$$

3. Proposed Reconstruction Pipeline

In this section, we describe our novel reconstruction pipeline that provides a sparse oriented point cloud as a reconstruction from photos shot from several views.

Our approach to surface normal estimation is a novel *multiple-view* extension of a previous work[6], combined with *a robust approach* to estimate surface normals consistent with all the views available for the observed tangent plane. The reconstruction is finalized by a bundle-adjustment-like numerical method, for the integrated *refinement* of all projection parameters, 3D positions and *surface normals*. Our approach is able to estimate normals of surfaces viewed by *arbitrary central-projective cameras*.

Multiple-view Surface Normal Estimation The two-view surface normal estimator (see Sec. 2) is extended to multiple views and arbitrary central projective cameras: if more than two images are given, multiple ACs may be established between pairs of views that multiplies the number of equations. The surface normal is the solution of the following problem:

$$\begin{bmatrix}a_{11}^{(1)}\mathbf{w}_{12}^{(1)} - a_{12}^{(1)}\mathbf{w}_{11}^{(1)}\\ a_{11}^{(1)}\mathbf{w}_{21}^{(1)} - a_{21}^{(1)}\mathbf{w}_{11}^{(1)}\\ a_{11}^{(1)}\mathbf{w}_{22}^{(1)} - a_{22}^{(1)}\mathbf{w}_{11}^{(1)}\\ a_{12}^{(1)}\mathbf{w}_{21}^{(1)} - a_{21}^{(1)}\mathbf{w}_{12}^{(1)}\\ a_{12}^{(1)}\mathbf{w}_{22}^{(1)} - a_{22}^{(1)}\mathbf{w}_{12}^{(1)}\\ a_{21}^{(1)}\mathbf{w}_{22}^{(1)} - a_{22}^{(1)}\mathbf{w}_{21}^{(1)}\\ \vdots\\ a_{11}^{(k)}\mathbf{w}_{12}^{(k)} - a_{12}^{(k)}\mathbf{w}_{11}^{(k)}\\ a_{11}^{(k)}\mathbf{w}_{21}^{(k)} - a_{21}^{(k)}\mathbf{w}_{11}^{(k)}\\ a_{11}^{(k)}\mathbf{w}_{22}^{(k)} - a_{22}^{(k)}\mathbf{w}_{11}^{(k)}\\ a_{12}^{(k)}\mathbf{w}_{21}^{(k)} - a_{21}^{(k)}\mathbf{w}_{12}^{(k)}\\ a_{12}^{(k)}\mathbf{w}_{22}^{(k)} - a_{22}^{(k)}\mathbf{w}_{12}^{(k)}\\ a_{21}^{(k)}\mathbf{w}_{22}^{(k)} - a_{22}^{(k)}\mathbf{w}_{21}^{(k)}\end{bmatrix}\mathbf{n} = \mathbf{0}, \text{ s.t. } \|\mathbf{n}\| = 1, \tag{9}$$

where $(1)\ldots(k)$ are indices of AC-s (*i.e.*, pairs of views).

Eliminating Dependence on Triangulation Considering *central-projective* views, $\mathbf{X}_0$ can be replaced by $p_i^{-1}(\mathbf{x}_i)$, that is the direction vector of the ray projecting $\mathbf{X}_0$ to the 2D image point $\mathbf{x}_i$. In this case, dependence on prior triangulation of the 3D point $\mathbf{X}_0$, with a possible source of error vanishes, as the equivalent ($=$) and up-to-scale ($\sim$) transformations in Eq. (5) still hold. In Eq. (4) $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{b}_1$ and $\mathbf{b}_2$, thus $\mathbf{w}_{ij}$ are redefined as follows:

$$\begin{aligned}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} &\doteq \nabla p_1\left(p_1^{-1}(\mathbf{x}_1)\right),\\ \begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} &\doteq \nabla p_2\left(p_2^{-1}(\mathbf{x}_2)\right),\end{aligned} \tag{10}$$

since the statement $\nabla p_i(\mathbf{X}_0) \sim \nabla p_i\left(p_i^{-1}(\mathbf{x}_i)\right)$ is valid for all central projective cameras.

Bundle Adjustment using Affine Correspondences Let us consider all observed surface points with corresponding surface normals as the set 'Surflets'. An element of this set is a pair $\mathcal{S} = (\mathbf{X}_\mathcal{S}, \mathbf{n}_\mathcal{S})$ of a 3D point and a surface normal, has multiple-view observations constructed from ACs as follows: corresponding image points $\mathbf{x}_k \in \mathrm{Obs}_0(\mathcal{S})$ of the k-th view and relative affine transformations $\mathbf{A}_{k_1,k_2} \in \mathrm{Obs}_1(\mathcal{S})$ between the k_1-st and the k_2-nd views, $k_1 \neq k_2$.

Our novel bundle adjustment scheme minimizes the following cost, refining structure *(surface points and normals)* and motion *(intrinsic and extrinsic camera parameters)*:

$$\sum_{\mathcal{S}\in\mathrm{Surflets}} \left(\sum_{\mathbf{x}_k\in\mathrm{Obs}_0(\mathcal{S})} \mathrm{cost}^k_{\mathbf{X}_\mathcal{S}}(\mathbf{x}_k) + \lambda \sum_{\mathbf{A}_{k_1,k_2}\in\mathrm{Obs}_1(\mathcal{S})} \mathrm{cost}^{k_1,k_2}_{\mathbf{n}_\mathcal{S}}(\mathbf{A}_{k_1,k_2}) \right), \tag{11}$$

where the following cost functions based on equations (1) and (3) ensure that the reconstruction remains faithful to point observations and ACs as follows:

$$\begin{aligned} \mathrm{cost}^{k_1,k_2}_{\mathbf{n}_\mathcal{S}}(\mathbf{A}) &= \left\| \mathbf{A} - \mathbf{A}_{k_2}\mathbf{A}_{k_1}^{-1} \right\|, \\ \mathrm{cost}^k_{\mathbf{X}_\mathcal{S}}(\mathbf{x}_k) &= \left\| \mathbf{x}_k - p_k(\mathbf{X}_\mathcal{S}) \right\|. \end{aligned} \tag{12}$$

Note that if λ is set to zero in Eq. (12) the problem reduces to the original point-based bundle adjustment problem, without the additional affine correspondences. In our tests λ is always set to 1. Ceres-Solver[3] is used to solve the optimization problem. The Huber and Soft-L1 norms are applied as loss functions for $\mathrm{cost}^{k_1,k_2}_{\mathbf{n}_\mathcal{S}}$ and $\mathrm{cost}^k_{\mathbf{X}_\mathcal{S}}$, respectively.

Bundle adjustment is followed by, in an alternating scheme, a geometric outlier filtering step described below, removing surface normals inconsistent with the multiple-view track. See Fig. 2 as an overview of the successive steps in the pipeline.

Geometric Outlier Filtering This step removes all surface normals that do not fulfill multiple-view geometric requirements. Suppose that the 3D center of a tangent plane $(\mathcal{S})$ is observed from multiple views. It is clear that this surface cannot be observed 'from behind' from any of the views so the estimated surface is removed from the reconstruction if the following is satisfied:

$$\begin{aligned} &\mathbf{n}_\mathcal{S} \text{ is an outlier,} \\ \text{if } \exists \mathbf{x}_i, \mathbf{x}_j \in \mathrm{Obs}_0(\mathcal{S}), i \neq j :& \langle \mathbf{n}, \mathbf{v}_i \rangle \cdot \langle \mathbf{n}, \mathbf{v}_j \rangle < 0, \end{aligned} \tag{13}$$

where $\mathbf{v}_k$ is the direction of the ray projecting the observed 3D point on the image plane of the k-th view.

Outlier filtering is always followed by a BA-step, if more than 10 surface normals were removed in the process.

Overview of the Pipeline Our reconstruction pipeline (see Fig. 2) is the modified version of OpenMVG[27, 28], the reconstructed scene, using the proposed approach, is enhanced by surface normals, and additional steps for robustification are included. At first, we extracted Affine Correspondences using TBMR[36] and further refined them by a simple gradient-based method, similarly to[32]. Multiple-view matching resulted in sets 'Obs_0' and 'Obs_1', as described above. An incremental reconstruction pipeline[27] provides camera poses and an initial point cloud without surface normals. Our approach now proceeds by multiple-view surface normal estimation as presented in Sec. 2.

The obtained oriented point cloud and the camera parameters can be further refined by our bundle adjustment approach. Since some of the estimated surface normals may be outliers, we apply an iterative method which has two inner steps: (i) bundle adjustment and (ii) outlier filtering. The latter discards surflets not facing all of the cameras. The process is repeated until no outlying surface normals are left in the point cloud.

4. Fitting Geometrical Elements to 3D Data

This section shows how standard geometrical elements can be fitted on oriented point clouds obtained by our image-based reconstruction pipeline.

Plane. For plane fitting, only the spatial coordinates are used. Considering its implicit form, the plane is parameterized by four scalars $\mathbf{P} = [a,b,c,d]^T$. Then a spatial point $\mathbf{x}$ given in homogeneous form is on the plane if $\mathbf{P}^T\mathbf{x} = 0$. Moreover, if the plane parameters are normalized as $a^2 + b^2 + c^2 = 1$, formula $\mathbf{P}^T\mathbf{x}$ is the Euclidean distance of the point w.r.t the plane. The estimation of a plane by minimizing the plane-point distances is relatively simple. It is well-known in geometry[13] that the center of gravity $\mathbf{c}$ of spatial points $\mathbf{x} : i = 0,\ i \in [1 \ldots N]$, is the optimal choice: $\mathbf{c} = \sum_i \mathbf{x}_i / N$, where N denotes the number of points. The normal $\mathbf{n}$ of the plane can be optimally estimated as the eigenvector of matrix $\mathbf{A}^T\mathbf{A}$ corresponding to the least eigenvalue, where matrix $\mathbf{A}$ is generated as $\mathbf{A} = \sum_i (\mathbf{x}_i - \mathbf{c})(\mathbf{x}_i - \mathbf{c})^T$.

Sphere. Fitting sphere is a more challenging task since there is no closed-form solution when the square of the L_2-norm (Euclidean distance) is minimized. Therefore, iterative algorithms[13] can be applied for the fitting task. However, if alternative norms are introduced[30], the problem becomes simpler.

In our implementation, a simple trick is used in order to get a closed-form estimation: the center of the sphere is estimated first, then two points of the sphere are selected and connected, and a line section is obtained. The perpendicular bisector of this section is a 3D plane. If the point selection and bisector forming is repeated, the common point of these planes gives the center of the sphere. However, the measured coordinates are noisy, therefore there is no common point of all the planes. If the j-th plane is denoted by $\mathbf{P_j}$ and the circle center by $\mathbf{C}$, the latter is obtained as $\mathbf{C} = \arg\min_\mathbf{C} \sum_j \mathbf{P_j}^T\mathbf{x}$.

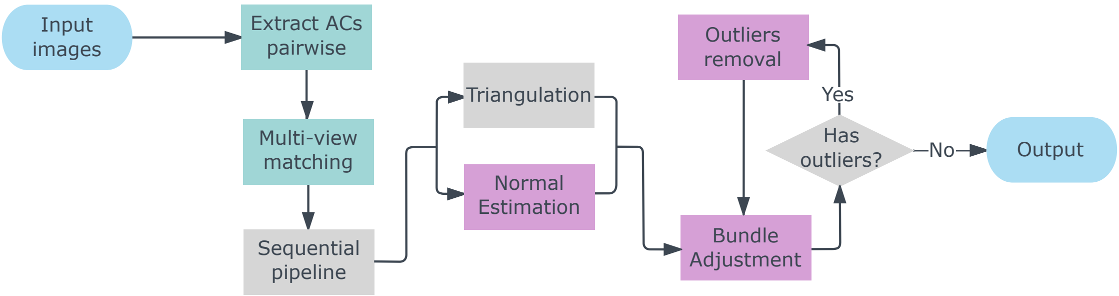

Figure 2: Reconstruction pipeline. The input is a set of photos of a scene, the output is a reconstructed point cloud with accurate normals. The central novelty of this work is highlighted in purple.

The radius of the circle is yielded as the square root of the average of the squared distances of the points and the center $\mathbf{C}$.

Cylinder. The estimation of a cylinder is a real challenge. The cylinder itself can be represented by a center point $\mathbf{C}$, the unit vector $\mathbf{w}$ representing the direction of the axis, and the radius r. The cost function of the cylinder fitting is as follows: $\sum_i \left(u_i^2+v_i^2-r^2\right)^2$, where the unit vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ form an orthonormal system, and the scalar values u_i and v_i are obtained as $u_i = \mathbf{u}_i^T(\mathbf{x}_i-\mathbf{C})$ and $v_i = \mathbf{v}_i^T(\mathbf{x}_i-\mathbf{C})$. This problem is nonlinear, therefore a closed-form solution does not exist to the best of our knowledge. However, it can be solved by alternating three steps[12]. It is assumed that the parameters of the cylinder are initialized.

1. **Radius.** It is trivial that the radius of the cylinder is yielded as the root of the mean squared of the distances between the points and the cylinder axis.
2. **Axis point.** The axis point $\mathbf{C}$ is updated as $\mathbf{C}_{new} = \mathbf{C}_{old} + k_1\mathbf{u} + k_2\mathbf{v}$, where the vectors $\mathbf{u}$, $\mathbf{v}$, and the axis form an orthonormal system. The parameters k_1 and k_2 are obtained by solving the following inhomogeneous system of linear equations:

$$2\sum_i \begin{bmatrix} u_i^2 & u_i v_i \\ u_i & v_i^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \sum_i \begin{bmatrix} \left(u_i^2+v_i^2\right)^2 u_i \\ \left(u_i^2+v_i^2\right)^2 v_i \end{bmatrix}.$$

3. **Axis direction.** It is given by a unit vector $\mathbf{w}$ represented by two parameters. The estimation of those are obtained by a simple exhaustive search.

Before running the alternation, initial values are required. If the surface normals $\mathbf{n}_i$ are known at the measured locations $\mathbf{x}_i$, then the axis $\mathbf{w}$ of the cylinder can be computed as the vector perpendicular to the normals. Thus all normal vectors are stacked in the matrix $\mathbf{N}$, and the perpendicular direction is given by the nullvector of the matrix. As the normals are noisy, the eigenvector of $\mathbf{N}^T\mathbf{N}$ corresponding to the least eigenvalue is selected as the estimation for the nullvector. The other two direction vectors $\mathbf{u}$ and $\mathbf{v}$ are given by the other two eigenvectors of matrix $\mathbf{N}^T\mathbf{N}$. The initial value for the axis point is simply initialized as the center of gravity of the points.

5. Experimental Results

The proposed reconstruction pipeline is tested on 3D reconstruction using real images. Firstly, the quality of the reconstructed point cloud and surface normals are quantitatively tested. High-quality 3D reconstruction is presented in the second part of this section.

5.1. Quantitative Comparison of Reconstructed Models

In the first test, the quality of the obtained surfaces are compared. Three test sequences are taken as it is visualized in Fig. 3: a plane, a sphere, and a cylinder. Our reconstruction pipeline is applied to compute the 3D model of the observed scenes including point clouds and corresponding normals. Then the fitting algorithms discussed in Sec. 4 are applied. First, the fitting is combined with a RANSAC[14]-like robust model selection by minimal point sampling[§] to detect the most dominant object in the scene. Object fitting is then ran only on the inliers corresponding to the dominant object. Results are visualized in Fig. 4.

The quantitative results are listed in Tab. 1. The errors are computed for both 3D positions and surface normals except for the reconstruction of the plane where the point fitting is very low and there is no significant difference between the methods. The ground truth values are provided by the fitted 3D geometric model. The angular errors are given in degrees. The least squared (LSQ), mean, and median values are calculated for both types of errors. Three surflet-based methods are compared: the PMVS algorithm[¶][16] and the proposed one with and without the BA refinement. The

[§] At least three points are required for plane fitting, four points are needed for cylinders and spheres.

[¶] The implementation of PMVS included in VisualSFM library is applied. See http://ccwu.me/vsfm/.

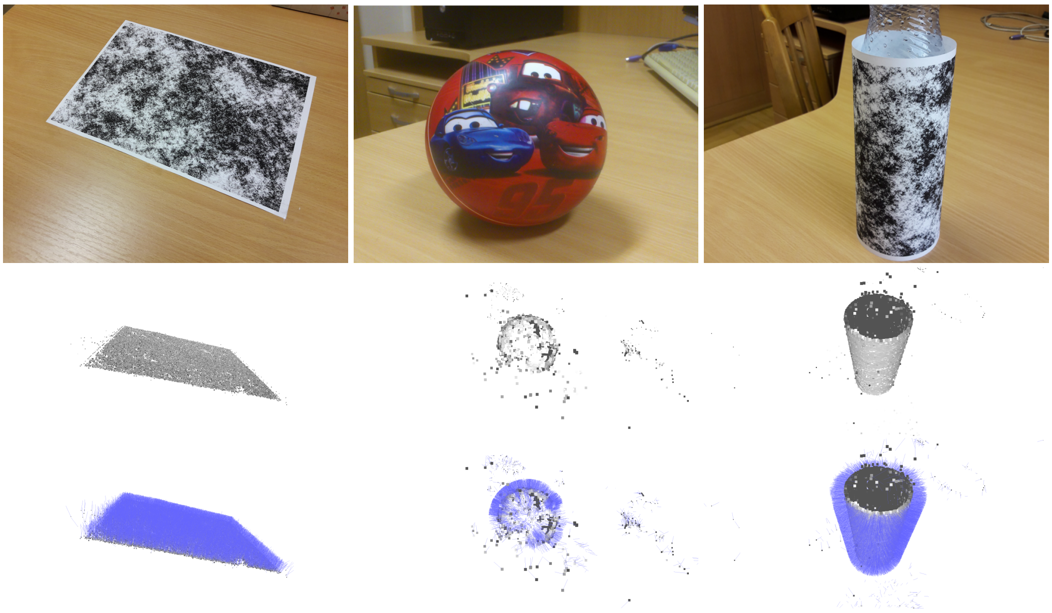

Figure 3: Test objects for quantitative comparison of surface points and normals. Top: One out of many input images used for 3D reconstruction. Middle: Reconstructed point cloud returned by proposed pipeline. Bottom: Same models with surface nor[illegible] *ved in color*.

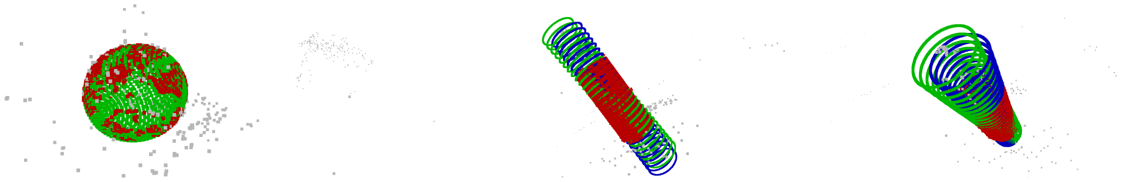

Figure 4: Reconstructed sphere (left) and two views of the cylinder (middle and right). Inliers, outliers, and fitted models are denoted by red, gray, and green, respectively. In the case of cylinder fitting, blue color denotes the initial model computed by RANSAC[14]. Inliers correspond to the RANSAC minimal model. *Best viewed in color.*

proposed pipeline outperforms the rival PMVS algorithm, with and without the additional BA step of our pipeline: the initial 3D point locations are more accurate than the result of PMVS. The difference is significant especially for the cylinder fitting: PMVS is unable to find the correct solution in this case. This example is the only one where the surface normals are required for the object fitting, the quality of the resulting normals of PMVS do not reach the desired level contrary to ours.

The proposed method and PMVS estimate surface normals at distinct points in space, however, surface normals can also be estimated by fitting tangent planes to the surrounding points. This is a standard technique in RE[31], a possible algorithm is written in Sec. 4. We used MeshLab[10] to estimate the normals given the raw point cloud. Two variants are considered: tangent planes are computed using 10 and 50 Nearest Neighboring (NN) points. The latter yields surface normals of better quality: our method computing for a distinct point in space is always outperformed by the 50 NNs-based algorithm. However, our approach outperforms the result provided by MeshLab for 10NNs for the cylinder. Moreover, the returned point locations are more accurate when the proposed method is applied. A possible future work is to estimate the normals using nearby surflets. This is out of the scope of this paper. Note that our method has the upper hand over all spatial neighborhood-based approaches for isolated points (*i.e.*, neighboring 3D points are distant in a non-uniform point cloud).

To conclude the tests, one can state that the proposed algorithm is more accurate than the rival PMVS method[16]. Image-based RE of geometrical elements is possible by applying our reconstruction pipeline. Median of the angular errors are typically between 5 and 10 degrees.

5.2. 3D Reconstruction of Real-world Objects.

Our reconstruction pipeline is qualitatively tested on images taken of real-world objects.

Table 1: Point (Pts.) and angular (Ang.) error of reconstructed surface normals for plane, sphere, and cylinder. Ground truth normals computed by robust sphere fitting based on methods described in Sec. 4. DNF: Did Not Find correct model.

	Metrics	PMVS[16]	Ours	Ours+BA	MeshLab (10NNs)	MeshLab (50NNs)
Plane	Ang. Error (LSQ)	19.85	14.54	**13.86**	11.23	**1.98**
	Ang. Error (Mean)	13.14	9.39	**9.16**	7.43	**1.71**
	Ang. Error (Median)	6.72	5.91	**5.90**	5.07	**1.55**
Sphere	Pts Error (LSQ)	0.38 (DNF)	0.03	**0.010**	0.029	0.011
	Pts Error (Mean)	0.31 (DNF)	0.0083	**0.0076**	0.0095	0.0079
	Pts Error (Median)	0.3 (DNF)	**0.0056**	0.0062	0.0068	0.0062
	Ang. Error (LSQ)	84.1 (DNF)	19.43	**18.41**	12.50	**2.18**
	Ang. Error (Mean)	77.09 (DNF)	14.54	**13.72**	7.66	**2.36**
	Ang. Error (Median)	79.58 (DNF)	11.74	**10.83**	5.50	**1.75**
Cylinder	Pts Error (LSQ)	0.70	**0.69**	0.77	0.76	0.77
	Pts Error (Mean)	0.53	**0.51**	0.57	0.56	0.57
	Pts Error (Median)	0.42	**0.37**	0.42	0.41	0.42
	Ang. Error (LSQ)	29.76	22.48	**18.41**	22.01	**4.23**
	Ang. Error (Mean)	23.15	14.39	**13.72**	14.89	**3.22**
	Ang. Error (Median)	17.62	7.33	**5.68**	9.13	**2.60**

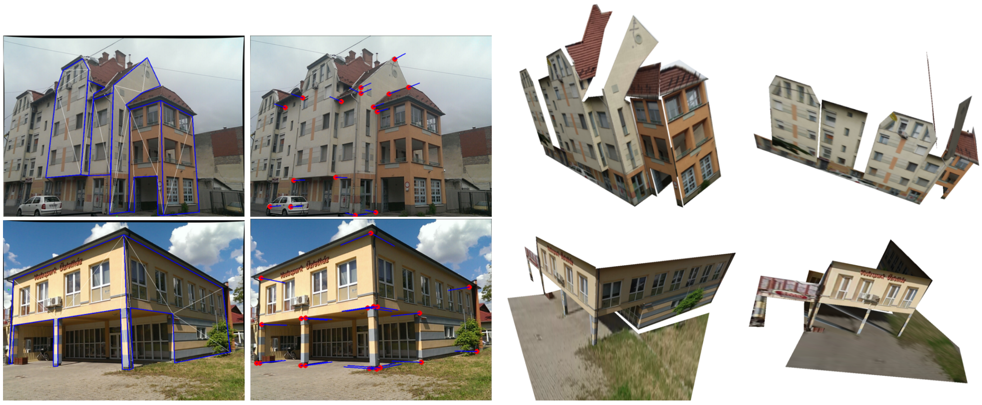

Figure 5: Reconstruction of real buildings. From left to right: selected regions in first image; regions with reconstructed normals; two different views of the reconstructed and textured 3D scene.

Reconstruction of Buildings. The first qualitative test is based on images taken of buildings. The final goal is to compute the textured 3D model of the object planes. The novel BA method is successfully applied on two test sequences of the database of the University of Szeged[34]. This database contains images and the intrinsic parameters of the cameras. For the sake of the quality, the planar regions are manually segmented in the images. Results can be seen in Fig. 5.

Free-form Surface Reconstruction. The proposed BA method is also applied to the dense 3D reconstruction of free-form surfaces as it is visualized in Figures 6 and 7. The first two examples come from the dense multi-view stereo database[33] of CVLAB[||]. The reconstruction of a painted plastic bear also demonstrates the applicability of our reconstruction pipeline as well as a reconstructed face model with surface normals in Fig. 7.

Finally, our 3D reconstruction method is qualitatively compared to PMVS of Furukawa *et al.*[16]. The Fountain dataset is reconstructed both by PMVS and our method.

[||] http://cvlabwww.epfl.ch/data/multiview/denseMVS.html

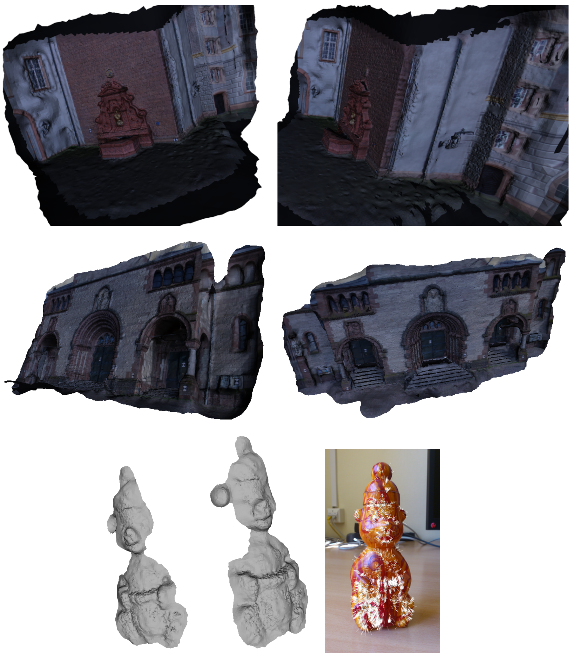

Figure 6: Reconstruction of real-world free-form objects.

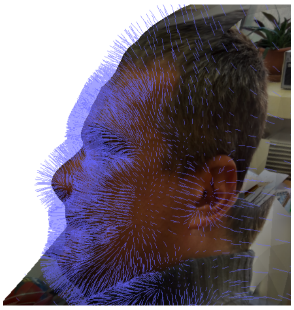

Figure 7: Reconstructed 3D face with surface normals colored by blue.

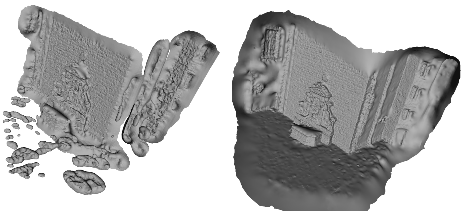

Figure 8: 3D reconstructed model obtained by Furukawa *et al.*[16] (left) and proposed pipeline (right). Out method yields a more connected surface with less holes.

Then from the 3D point cloud with surface normals the scene is obtained using the Screened Poisson surface reconstruction[18] for both methods. The comparison can be seen in Fig. 8. The proposed method extracts significantly finer details as it is visualized. As a consequence, walls and objects of the scene form a continuous surface, and the result of our method does not contain holes.

6. Conclusions and Future Work

Two novel algorithms are presented in this paper: (i) a closed-form multiple-view surface normal estimator and a (ii) bundle adjustment-like numerical refinement scheme, with a robust multi-view outlier filtering step. Both approaches are based on ACs detected in image pairs of a multi-view set. The proposed estimator, to the best of our knowledge, is the first multiple-view method for computing surface normal using ACs. It is validated that the accuracy of the resulting oriented point cloud is satisfactory for reverse engineering even if the normals are estimated based on distinct points in space.

A possible future work is to enhance the reconstruction accuracy by considering the spatial coherence of the surflets.

Acknowledgement.

Supported by the ÚNKP-17-3 New National Excellence Program of the Ministry of Human Capacities.

References

1. Insight3D - opensource image-based 3D modeling software. http://insight3d.sourceforge.net/. 2

2. S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Commun. ACM*, 54(10):105–112, 2011. 1

3. S. Agarwal, K. Mierle, and Others. Ceres Solver. `http://ceres-solver.org`. 4

4. B. Triggs and P. McLauchlan and R. I. Hartley and A. Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000. 2

5. S. Baker and I. Matthews. Lucas-Kanade 20 Years On: A Unifying Framework: Part 1. Technical Report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2002. 1

6. D. Barath, J. Molnar, and L. Hajder. Novel methods for estimating surface normals from affine transformations. In *Computer Vision, Imaging and Computer Graphics Theory and Applications, Selected and Revised Papers*, pages 316–337. Springer International Publishing, 2015. 1, 2, 3

7. H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 1, 2

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

9. M. Bujnak, Z. Kukelova, and T. Pajdla. 3d reconstruction from image collections with a single known focal length. In *ICCV*, pages 1803–1810, 2009. 1

10. P. Cignoni, M. Corsini, and G. Ranzuglia. MeshLab: an Open-Source 3D Mesh Processing System. *ERCIM News*, (73):45–46, April 2008. 6

11. A. Delaunoy and M. Pollefeys. Photometric Bundle Adjustment for Dense Multi-view 3D Modeling. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 1486–1493, 2014. 2

12. D. Eberly. Fitting 3D Data with a Cylinder. `http://www.geometrictools.com/Documentation/CylinderFitting.pdf`. Online; accessed 11 April 2017. 5

13. D. Eberly. Least Squares Fitting on Data. `http://www.geometrictools.com/Documentation/LeastSquaresFitting.pdf`. Online; accessed 12 April 2017. 4

14. M. Fischler and R. Bolles. RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:358–367, 1981. 5, 6

15. J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys. Building rome on a cloudless day. In *Proceedings of the 11th European Conference on Computer Vision*, pages 368–381, 2010. 1

16. Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. 2, 5, 6, 7, 8

17. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. 1

18. M. Kazhdan and H. Hoppe. Screened Poisson Surface Reconstruction. *ACM Trans. Graph.*, 32(3):29:1–29:13, 2013. 8

19. K. Köser and R. Koch. Differential Spatial Resection - Pose Estimation Using a Single Local Image Feature. In *Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV*, pages 312–325, 2008. 1, 2

20. R. Lakemond, S. Sridharan, and C. Fookes. Wide baseline correspondence extraction beyond local features. *IET Computer Vision*, 5(4):222–231, 2014. 1

21. D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 1, 2

22. J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proceedings of the British Machine Vision Conference 2002, BMVC 2002, Cardiff, UK, 2-5 September 2002*, 2002. 1

23. J. Matas, S. Obdrzálek, and O. Chum. Local Affine Frames for Wide-Baseline Stereo. In *16th International Conference on Pattern Recognition, ICPR 2002, Quebec, Canada, August 11-15, 2002.*, pages 363–366, 2002. 1

24. J. Molnár and D. Chetverikov. Quadratic Transformation for Planar Mapping of Implicit Surfaces. *Journal of Mathematical Imaging and Vision*, 48:176–184, 2014. 1, 3

25. J. Molnár and I. Eichhardt. A differential geometry approach to camera-independent image correspondence. *Computer Vision and Image Understanding*, 2018. 1

26. J.-M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009. 1, 2

27. P. Moulon, P. Monasse, and R. Marlet. Adaptive structure from motion with a contrario model estimation. In *Asian Conference on Computer Vision*, pages 257–270. Springer, 2012. 1, 4

28. P. Moulon, P. Monasse, R. Marlet, and Others. OpenMVG. `https://github.com/openMVG/openMVG`. 4

29. M. Perdoch, J. Matas, and O. Chum. Epipolar Geometry from Two Correspondences. In *18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China*, pages 215–219, 2006. 1

30. V. Pratt. Direct Least-squares Fitting of Algebraic Surfaces. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, pages 145–152, 1987. 4

31. V. Raja and K. J. Fernandes. *Reverse Engineering: An Industrial Perspective*. Springer, 2007. 2, 6

32. C. Raposo, M. Antunes, and J. P. Barreto. Piecewise-Planar StereoScan: Structure and Motion from Plane Primitives. In *European Conference on Computer Vision*, pages 48–63, 2014. 4

33. 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 *IEEE Conference on Computer Vision and Pattern Recognition, 2008.*, pages 1–8. IEEE, 2008. 7

34. A. Tanács, A. Majdik, L. Hajder, J. Molnár, Z. Sánta, and Z. Kato. Collaborative mobile 3d reconstruction of urban scenes. In *Computer Vision - ACCV 2014 Workshops - Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part III*, pages 486–501, 2014. 7

35. Tomasi, C. and Shi, J. Good Features to Track. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 593–600, 1994. 1

36. Y. Xu, P. Monasse, T. Géraud, and L. Najman. Tree-based morse regions: A topological approach to local feature detection. *IEEE Transactions on Image Processing*, 23(12):5612–5625, 2014. 4

37. G. Yu and J.-M. Morel. ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line*, 2011, 2011. 1, 2