Template-free Shape from Texture with Perspective Cameras

Anna Hilsmann[12]
anna.hilsmann@hhi.fraunhofer.de
David C. Schneider[12]
david.schneider@hhi.fraunhofer.de
Peter Eisert[12]
peter.eisert@hhi.fraunhofer.de

[1] Computer Vision & Graphics Group
Fraunhofer Heinrich Hertz Institute
Berlin, Germany

[2] Visual Computing Group
Humboldt University of Berlin
Berlin, Germany

Abstract

This paper formulates the Shape-from-Texture (SFT) problem of deriving the shape of an imaged surface from the distortion of its texture as a *single-plane/multiple-view* Structure-from-Motion (SFM) problem under full perspective projection. As in classical SFT formulations we approximate the surface as being piecewise planar. In contrast to many methods, our approach does not need a frontal view of the texture or the texture elements as reference, as it optimizes 3D patch positions and orientations from transformations between texture elements in the image. The reconstruction results in minimizing a large sparse linear least squares cost function based on the reprojection error, a planarity constraint and the estimated rigid motion between patches. Texture element positions in the image are estimated under the assumption of a regular texture from clustered feature points representing repeating appearances in the image. We present results obtained with synthetic data as well as real data to evaluate our method.

1 Introduction

The term Shape-from-Texture (SFT) covers a class of methods for computing the 3D shape of a textured surface from a single image by exploiting texture distortion as a cue for shape. In this paper, we present an SFT formulation, which is equivalent to a *single-plane/ multiple-view* pose estimation problem statement [17, 22] under perspective projection. As in the classical SFT setting, we assume that the texture is constructed of one or more repeating texture elements, called *texels*, and assume that these texels are small enough such that they can be modeled as planar patches. In contrast to the classical setting, we do not assume that a fronto-parallel view of the texture element is known a priori. Instead, we formulate the SFT problem akin to a Structure-from-Motion (SFM) problem, given n views of the same planar texture patch. We assume a full perspective camera model with known intrinsics and estimate the patch poses from estimated homographies between the distorted texel appearances in the image. Each homography between two arbitrary patches yields an estimate of the normal vector of one of the two patches (referred to as reference in the following) and the rigid motion between the two patches. By using each patch as reference to all other patches in turn, we get enough constraints to set up a stable linear cost function to optimize the 3D

BMVC 2011 http://dx.doi.org/10.5244/C.25.26

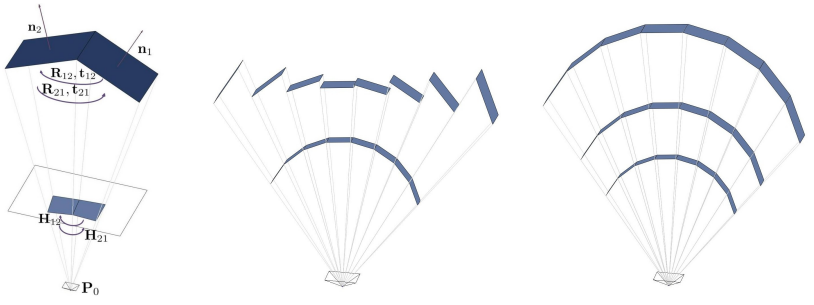
poses of the texel patches. A smooth surface is computed by regression with approximating thin-plate splines using the estimated patch centroids as data points. The final reconstruction is up to a single global scale factor.

The remainder of this paper is structured as follows. The next section briefly reviews related work and states our contribution. Section 3 explains our template-free method to reconstruct a 3D surface given its imaged texture under a full perspective camera model. In Section 4 we present results on synthetic and real data with unknown texel shape and appearance.

2 Related Work

The literature distinguishes between statistical and geometrical SFT methods. Statistical methods are often used for natural textures with statistic properties and rely on spatial frequency or density properties [6, 18, 21]. Geometric methods are generally used for deterministic textures consisting of repeated 2D geometric shapes, called *texels*. These methods reconstruct the 3D surface by exploiting the relation between 3D shape, the assumed imaging process and the texture distortion in the image. As we are interested in artificial textures, such as cloth, we will concentrate on geometric SFT approaches here. These methods differ in the assumed camera model, e.g. orthographic [4, 5], scaled orthographic [2], perspective [8]. Most methods assume a fronto-parallel appearance of the texture element given and estimate a 2D transformation, i.e. a homography or an affine transformation, between this template and the texel appearances in the image [2, 8]. This is equivalent to the assumption that a Eucledian frame attached to the reference plane is known which simplifies the deduction of the 3D patch pose from the 2D transformation. 3D position and orientation are then directly deduced from each 2D transformation to the template [2, 17]. Some approaches were proposed for orthographic camera models which estimate the orientation of a texel from affine transformations *between* the texels in the image without the need of a texel template [4, 5, 12, 14]. However, often only the surface orientation but no position in 3D space is estimated and the surface has to be deduced from a following normal integration step [10]. Often, the texel positions in the image are assumed to be known. To detect putative texels in an image without a template, methods exist that make use of repeating texel appearances [11, 16].

Lobay and Forsyth [12] presented an approach that estimates the frontal appearances of regular texture elements by clustering similar image patches. They estimate patch orientations from a sufficient number of orthographic views in an Expectation-Maximization approach similar to self-calibration in SFM. An orthographic camera is a good affine approximation of a projective camera if the distance between camera and object is large. However, for smaller distances between camera and object it deviates from the true projection. In contrast to most existing approaches, we consider the SFT problem as a *single-plane/multiple-view* SFM problem with a full perspective camera model. Our method is inspired by recent work of Varol *et al.* [19] who used homography decomposition [15, 22] to reconstruct deforming surfaces in monocular video sequences. We do not require the frontal appearance or shape of the texel given but instead estimate the 3D texel positions and orientations from homographies between the texel appearances in the image (such that a normal integration step is not needed). As we use each patch as reference in turn, we can optimize the 3D patch positions over a large sparse linear system instead of calculating them directly from transformations to a single reference. To detect the texel instances in the image, we assume that the texture is regular and extract a 2D grid lying on the deformed texture from clustered features points.



(a) (b) (c)

Figure 1: (a) Scenario with two planes, (b) local scale ambiguity, (c) global scale ambiguity

3 Shape from Texture as Structure from Motion

Notation

We denote a 3D point in Cartesian coordinates with $\mathbf{X} = [X\ Y\ Z]^T$ and its image projection with $\mathbf{x} = [x\ y]^T$. The index of a point is denoted by a superscript and the patch or plane index is denoted by a subscript, i.e. $\mathbf{X}_k^i$ is the 3D point i lying on patch k and $\mathbf{x}_k^i$ is its projected image point. Generally, $a^{(i)}$ denotes the i^{th} entry of vector $\mathbf{a}$ and $A^{(ij)}$ denotes the i^{th} row and j^{th} column of matrix $\mathbf{A}$. diag$[\mathbf{A}_1 ... \mathbf{A}_n]$ is a block diagonal matrix built from the matrices $\mathbf{A}_1 ... \mathbf{A}_n$ while diag$_n[\mathbf{A}]$ is a block diagonal matrix built from n-times matrix $\mathbf{A}$.

Plane Induced Homographies and their Decomposition

Let $\mathbf{P}_0 = \mathbf{K}\left[\mathbf{I}|\mathbf{0}\right]$ and $\mathbf{P}_1 = \mathbf{K}\left[\mathbf{R}|\mathbf{t}\right]$ be the projection matrices of two cameras with the same intrinsics $\mathbf{K}$ and the rigid motion between them described by a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$. Assume a plane with normal vector $\mathbf{n}$ and distance d from the origin is projected into both camera frames. The projections of a point on the plane into the images of $\mathbf{P}_0$ and $\mathbf{P}_1$ are then related by a homography which is given as [7]

$$\mathbf{H} = \mathbf{R} - \frac{\mathbf{t}\mathbf{n}^T}{d} \quad . \tag{1}$$

Methods for decomposing a homography into $\mathbf{R}, \mathbf{t}/d$ and $\mathbf{n}$ are detailed in [15, 22]. The decomposition results in two distinct solutions for the relative (camera) motion and the plane normal in the camera referential coordinate frame. This ambiguity can be solved if a third view of the plane is given as one solution of each decomposition yields the same normal vector for the reference plane.

The above considerations are equivalent to the assumption of a fixed camera $\mathbf{P}_0$ and a *moving* plane. In the SFT setting, the image contains n views of the same planar patch or, equivalently, n identical planar patches, i.e. the texture elements, under rigid motion (see figure 1(a) for an illustration with two planes). Without loss of generality we can assume that the camera is given by $\mathbf{P}_0 = \mathbf{K}\left[\mathbf{I}|\mathbf{0}\right]$ and that the intrinsic camera matrix $\mathbf{K}$ is known. For simplicity and w.l.o.g. we set it to the identity in the following. Assume we take an arbitrary patch k as reference patch and estimate the homographies $\mathbf{H}_{kl}$ between its projected image points to the image points of all other patches $l = 1 \ldots n$ (for the moment assume that these image points are known, see section 4 for how the image positions can be estimated for regular textures). The decomposition of each homography $\mathbf{H}_{kl}$ yields an estimate of the (scaled) rigid motion

from the reference patch k to patch l and of the reference patch normal[1]:

$$\mathbf{H}_{kl} \Rightarrow \mathbf{R}_{kl},\ \mathbf{t}_{kl}/d_k,\ \mathbf{n}_{k_l}$$

where $\mathbf{n}_{k_l}$ denotes the l^{th} estimate of $\mathbf{n}_k$. Note that the translation vector contains a scale ambiguity. In the following, we write $\mathbf{t}'_{kl} = \mathbf{t}_{kl}/d_k$ for the scaled translation vector. Given the estimates of the patch normal and the (scaled) rigid motion between the patches, the 3D points $\mathbf{X}^i_k,\ i = 1\ldots m$ of each patch $k = 1...n$ can now be reconstructed as follows.

Single Patch Reconstruction

To estimate the 3D position of a single point i on patch k we minimize a cost function based on the reprojection error:

$$\begin{aligned}\hat{\mathbf{X}}^i_k &= \arg\min_{\mathbf{x}^i_k} \left[\left\| \mathbf{P}_0 \begin{bmatrix} \mathbf{X}^i_k \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{x}^i_k \\ 1 \end{bmatrix} \right\|^2 + \sum_{l=1, l\neq k}^{n} \left\| \mathbf{P}_0 \begin{bmatrix} \tilde{\mathbf{X}}^i_{kl} \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{x}^i_l \\ 1 \end{bmatrix} \right\|^2 \right] \\ \tilde{\mathbf{X}}^i_{kl} &= \mathbf{R}_{kl} \cdot \mathbf{X}^i_k + \mathbf{t}'_{kl}\end{aligned} \tag{2}$$

where $\tilde{\mathbf{X}}^i_{kl}$ denotes the point $\mathbf{X}^i_k$ after applying the estimated rigid transformation onto its corresponding point on patch l. As $\mathbf{P}_0$ is given in its canonical form, this can be rewritten as

$$\hat{\mathbf{X}}^i_k = \arg\min_{\mathbf{X}^i_k} \sum_{l=1}^{n} \left\| [\mathbf{R}_{kl}|\mathbf{t}'_{kl}] \cdot \begin{bmatrix} \mathbf{X}^i_k \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{x}^i_l \\ 1 \end{bmatrix} \right\|^2 . \tag{3}$$

with $\mathbf{R}_{kk} = \mathbf{I}$ and $\mathbf{t}_{kk} = \mathbf{0}$. For a single point $\mathbf{X}^i_k$ this results in a linear equation system

$$\hat{\mathbf{X}}^i_k = \arg\min_{\mathbf{X}^i_k} \left\| \mathbf{B}^i_k \cdot \mathbf{X}^i_k - \mathbf{b}^i_k \right\|^2 \tag{4}$$

where $\mathbf{B}^i_k$ is a $2n \times 3$ matrix and $\mathbf{b}^i_k$ is a $2n \times 1$ vector built by concatenating

$$\begin{bmatrix} R^{(11)}_{kl} - x^i_l R^{(31)}_{kl} & R^{(12)}_{kl} - x^i_l R^{(32)}_{kl} & R^{(13)}_{kl} - x^i_l R^{(33)}_{kl} \\ R^{(21)}_{kl} - y^i_l R^{(31)}_{kl} & R^{(22)}_{kl} - y^i_l R^{(32)}_{kl} & R^{(23)}_{kl} - y^i_l R^{(33)}_{kl} \end{bmatrix}, \quad \begin{bmatrix} -\mathbf{t}'^{(1)}_{kl} + x^i_l \mathbf{t}'^{(3)}_{kl} \\ -\mathbf{t}'^{(2)}_{kl} + y^i_l \mathbf{t}'^{(3)}_{kl} \end{bmatrix},\ l = 1...n \tag{5}$$

for all estimated rigid motions $[\mathbf{R}_{kl}|\mathbf{t}'_{kl}], l = 1\ldots n$. Now let $\mathbf{X}_k$ denote the concatenated column vector of all m points $\mathbf{X}^1_k \ldots \mathbf{X}^m_k$ on patch k. $\mathbf{X}_k$ is then estimated by

$$\hat{\mathbf{X}}_k = \arg\min_{\mathbf{X}_k} \left\| \mathbf{B}_k \cdot \mathbf{X}_k - \mathbf{b}_k \right\|^2 \tag{6}$$

with a block diagonal matrix $\mathbf{B}_k = \mathrm{diag}[\mathbf{B}^1_k \ldots \mathbf{B}^m_k]$ and a column vector $\mathbf{b}_k$ constructed by concatenating $\mathbf{b}^1_k \ldots \mathbf{b}^m_k$.

To include also the estimated normal information and to force the reconstructed points to be planar we add an energy term that forces the reconstructed points to lie on a plane with the estimated normal $\mathbf{n}_{k_l}$ by additionally minimizing the plane equation

$$\left\| \mathbf{n}_{k_l}{}^T \cdot \mathbf{X}^i_k + d_k \right\|^2 \tag{7}$$

[1]Note that each decomposition of the homographies $\mathbf{H}_{kl}, l = 1...n$ yields two solutions. Ideally, one of these solutions yields a consistent normal vector $\mathbf{n}_k$ for the reference plane k for all $l = 1...n$. In practice, we determine the solutions with the most consistent estimation of the normal vector $\mathbf{n}_k$ using a Dijkstra algorithm with a cost function based on the angles between the estimated normals.

for each point $i = 1 \dots m$ on patch k and each estimate $\mathbf{n}_{k_l}, l = 1 \dots n$ of the patch normal, such that the above equation system is augmented by

$$\hat{\mathbf{X}}_k = \arg\min_{\mathbf{X}_k} \left\| \begin{bmatrix} \mathbf{B}_k \\ \mathbf{N}_k \end{bmatrix} \cdot \mathbf{X}_k - \begin{bmatrix} \mathbf{b}_k \\ -\mathbf{d}_k \end{bmatrix} \right\|^2 \tag{8}$$

where $\mathbf{N}_k = \text{diag}_m[[\mathbf{n}_{k_1} \dots \mathbf{n}_{k_n}]^T]$ and $\mathbf{d}_k = [d_k \dots d_k]^T$ is built by concatenating d_k $n \cdot m$-times. Note that by using $\mathbf{t}'_{kl} = \mathbf{t}_{kl}/d_k$ in vector $\mathbf{b}_k$ we assume $d_k = 1$. Therefore, we set $d_k = 1 \; \forall k$.

Multiple Patch Reconstruction

Until now we have addressed the reconstruction of one single arbitrary reference patch from homographies to all other patches in the image. We can now use each patch $k = 1 \dots n$ as a reference and build matrices $\mathbf{B}_k$ and $\mathbf{N}_k$ as well as the vectors $\mathbf{b}_k$ and $\mathbf{d}_k$ for all patches. Let $\mathbf{X}$ denote a vector constructed by vertically concatenating $\mathbf{X}_1 \dots \mathbf{X}_n$, i.e. a vector of all m points on all n patches. $\mathbf{X}$ can now be estimated up to scale by:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\| \begin{bmatrix} \mathbf{B} \\ \mathbf{N} \end{bmatrix} \cdot \mathbf{X} - \begin{bmatrix} \mathbf{b} \\ -\mathbf{d} \end{bmatrix} \right\|^2 \tag{9}$$

with $\mathbf{B} = \text{diag}[\mathbf{B}_1 \dots \mathbf{B}_n]$, $\mathbf{N} = \text{diag}[\mathbf{N}_1 \dots \mathbf{N}_n]$ and $\mathbf{b}$ and $\mathbf{d}$ are column vectors built by concatenating $\mathbf{b}_1 \dots \mathbf{b}_n$, and $\mathbf{d}_1 \dots \mathbf{d}_n$.

The reconstruction of all patches using equation (9) is up to scale for each single patch as we have not included any dependencies between the patches so far (see illustration in figure 1(b)). The scale ambiguity will be solved up to one single global scale in two different ways, presented in the following and compared in the next section.

Following the previous procedure, it is straightforward to first estimate each patch up to scale using equation (9) and then solve the scale ambiguity in a second step by minimizing a cost function based on the estimated rigid motion between corresponding 3D points:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{l=1, l\neq k}^{n} \left\| (\mathbf{R}_{kl} \cdot \hat{\mathbf{X}}^i_k + \mathbf{t}'_{kl}) \cdot d_k - \hat{\mathbf{X}}^i_l \cdot d_l \right\|^2 \tag{10}$$

This results in a homogeneous system and a non-trivial solution is found by a singular value decomposition (SVD). The reconstructed points of each patch are then scaled with the estimated scale factors: $\hat{\mathbf{X}}^i_k \rightarrow d_k \cdot \hat{\mathbf{X}}^i_k$.

Alternatively, we can directly include the scale factors $\mathbf{d} = [d_1 \dots d_n]^T$ into the optimization and add an energy term based on the the estimated rigid motion between the patches to the minimization problem:

$$\sum_{l=1, l\neq k}^{n} \left\| \mathbf{R}_{kl} \cdot \mathbf{X}^i_k + \mathbf{t}'_{kl} \cdot d_k - \mathbf{X}^i_l \right\|^2 \tag{11}$$

for all points $i = 1 \dots m$ on all patches $k = 1 \dots n$. This results in the following joint homogeneous system in $\mathbf{X}$ and $\mathbf{d}$:

$$[\hat{\mathbf{X}}, \hat{\mathbf{d}}] = \arg\min_{\mathbf{X},\mathbf{d}} \left\| \begin{bmatrix} \mathbf{B} & -\tilde{\mathbf{b}} \\ \mathbf{N} & \tilde{\mathbf{I}} \\ \mathbf{M} & \tilde{\mathbf{t}}' \end{bmatrix} \cdot \begin{bmatrix} \mathbf{X} \\ \mathbf{d} \end{bmatrix} \right\|^2 \tag{12}$$

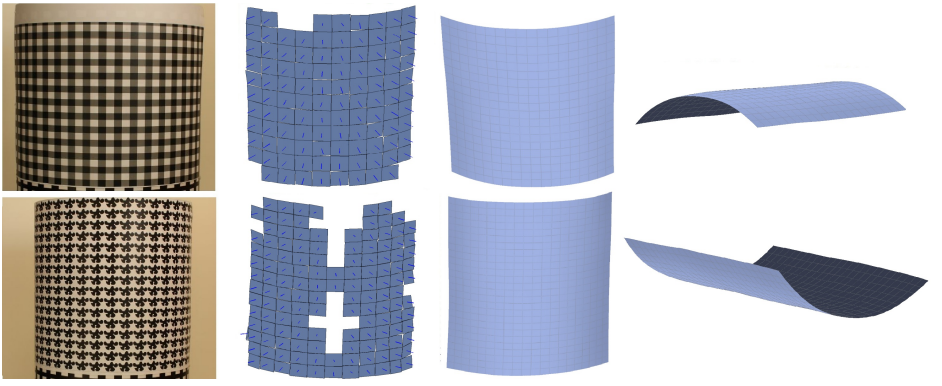


Figure 2: Original images (left), reconstructed patches with estimated normals (second left) and interpolated dense surfaces (two rightmost images). A comparison of the interpolated surface with a reconstruction from stereo correspondences showed an angular RSME of the surface normals of 4.9 degrees.

with $\tilde{\mathbf{b}} = \text{diag}[\mathbf{b}_1 \dots \mathbf{b}_n]$ and $\tilde{\mathbf{I}} = \text{diag}_n[\mathbf{1}_{nm\times 1}]$. $\mathbf{M}$ and $\tilde{\mathbf{t}}'$ describe the linear equation system resulting from equation (11) and are built as follows. For each patch k and each estimated transformation $[\mathbf{R}_{kl}|\mathbf{t}'_{kl}]$, $l = 1...n$ we build a block diagonal matrix $\tilde{\mathbf{R}}_{kl} = \text{diag}_m[\mathbf{R}_{kl}]$ and a column vector $\tilde{\mathbf{t}}'_{kl}$ by m-times repeating the translation vector $\mathbf{t}'_{kl}$. From these matrices we build $\tilde{\mathbf{R}}_k$ and $\tilde{\mathbf{t}}'_k$ by vertically concatenating $\tilde{\mathbf{R}}_{k1}...\tilde{\mathbf{R}}_{kn}$ and $\tilde{\mathbf{t}}'_{k1}...\tilde{\mathbf{t}}'_{kn}$ for each patch $k = 1 \dots n$ and finally, $\mathbf{M}$ and $\tilde{\mathbf{t}}$ are given by:

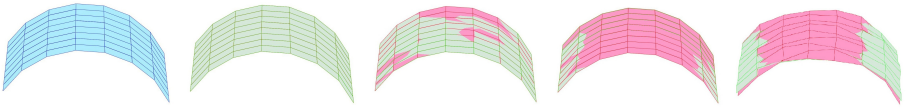
$$\mathbf{M} = \text{diag}[\tilde{\mathbf{R}}_1...\tilde{\mathbf{R}}_n] - [\mathbf{I}_{3mn}...\mathbf{I}_{3mn}]^T, \quad \tilde{\mathbf{t}}' = \text{diag}[\tilde{\mathbf{t}}'_1...\tilde{\mathbf{t}}'_n] \tag{13}$$

where $\mathbf{I}_{3mn}$ denotes the $3mn \times 3mn$ identity matrix. Equation (12) results in a homogeneous equation system and a non-trivial solution is found via an SVD. The estimated reconstruction is up to one single global scale (see figure 1(c)).

Surface Reconstruction
Until now we have assumed that the surface is piecewise planar. To get a smooth dense surface, we apply a robust surface regression method (using approximating thin-plate splines [1, 20]) that uses the patch centroids as data points (see figure 2). If a grid structure, i.e. a connectivity between the patches is known, the patch centroids can serve as new mesh vertices and a finer mesh can be constructed by inserting new vertices and interpolating their positions.

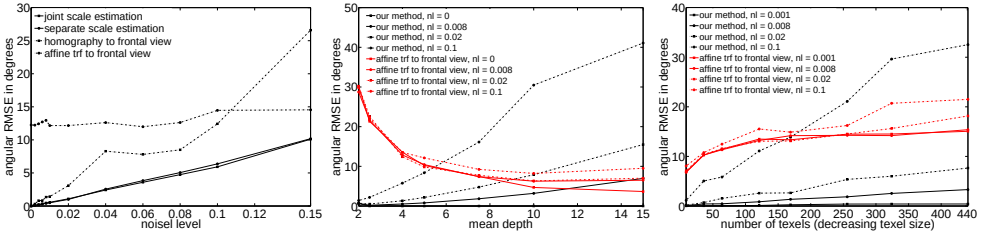
4 Experimental Results

Synthetic Data: We first applied our approach to synthetic data to assess the reconstruction quality in the presence of noise and for different viewing conditions. Following [2], we synthetically constructed a half cylinder (with radius $r = 1$) quantized into varying numbers of regular quadrilaterals and placed at different distances to the camera (we refer to the mean distance between camera and surface as *mean depth d*). An image of the surface was calculated by projecting the vertices into the image plane (using a focal length of $f = 1$). The reconstruction quality was measured using the root mean squared error (RMSE) of the surface normals. To test the stability of our approach with respect to noise, we disturbed



(a) original shape (b) $\sigma = 0\%$ (c) $\sigma = 0.002\%$ (d) $\sigma = 0.01\%$ (e) $\sigma = 0.1\%$

Figure 3: Reconstruction of a cylinder (blue) with different noise levels for joint (green) and separate (red) scale estimation ($d = 4$).



(a) (b) (c)

Figure 4: Synthetic reconstruction results of a cylinder. RMSE of the estimated normal vector angles as a function of (a) noise level at a fixed mean depth ($d = 4$) for our method compared to approaches that estimate the surface from affine transformations [2] or homographies [17] to a known frontal view of a reference patch, (b) mean scene depth and (c) number of texels for our method (with joint scale estimation, black) compared to the affine approach of [2] (red). Note that with an increasing number of texels the texel size decreases.

the image positions of the vertices with additive zero-mean Gaussian noise with a standard deviation defined as percentage of the longer image side, i.e. for a 512×1024 image a value of $\sigma = 0.1\%$ corresponds to $\sigma = 1.024$. Figure 3 shows the overlaid reconstruction results of a half cylinder using our method with joint and separate scale estimation for different noise levels. Figure 4(a) compares the angular RMSE for a fixed mean depth and a fixed number of quads over varying noise levels achieved with our approach to results of approaches that calculate the patch poses directly from affine transformations [2] or homographies [17] to a known frontal view of a template patch. Our method outperforms these two methods at this medium range ($d/f = 4$) as we do not use a single known reference patch but instead use all texel patches as references in turn and thereby can optimize over several measurements. Figure 4(b) plots the angular RMSE as a function of mean depth for a fixed number of texels and different noise levels comparing our method to [2]. With increasing distance between camera and object the degree of perspective distortion gets weaker. Hence, an affine camera better approximates the projective camera and an affine transformation more and more approximates the homography between planar patches. Following, while at shorter distances or at large distances with lower noise level our method produces more accurate reconstruction results than an affine approach, with increasing mean depth and noise level the estimation of an affine transformation tends to produce more robust results than homography estimation. The same applies for an increasing number of texels on the same surface (see figure 4(c)) as with a decreasing size of texels the viewing conditions also tend from perspective to more affine conditions.

Real Data: One advantage of the method presented in this paper in contrast to standard SFT methods is that a frontal appearance of the texel is not needed a priori for reconstruction. If a frontal appearance of the texel is not given as template, we cannot simply use standard matching techniques to detect the patches in the image. Therefore, we present here an appli-

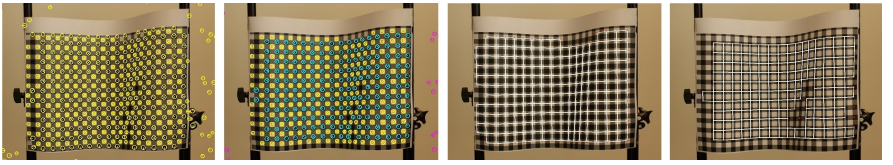


Figure 5: From left to right: detected feature points, feature clusters marked in different colors and estimated mesh models for each cluster.

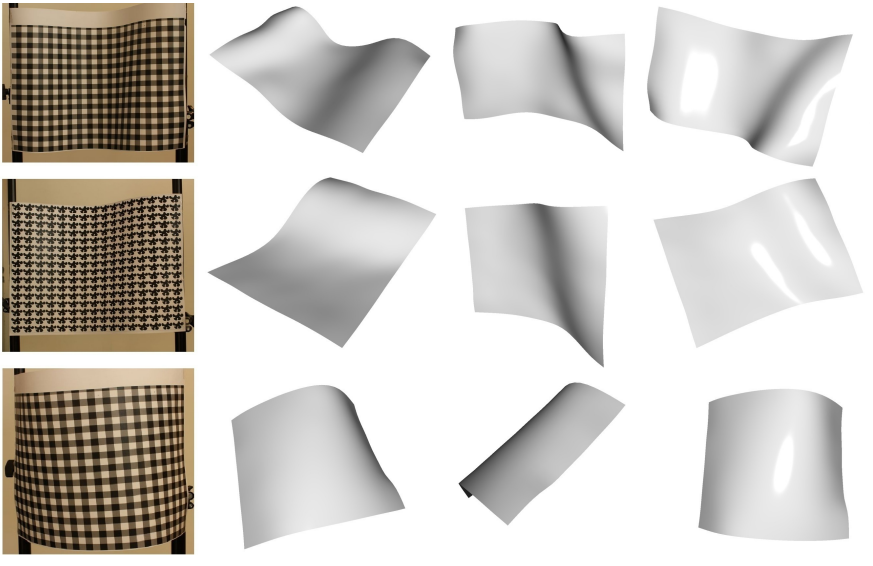


Figure 6: Input images (left) and reconstructed surfaces shown from different points of view.

cation to regular textures, i.e. texture that have been created by regular tiling with the same texel. Such textures are ubiquitous in the real world, e.g. in cloth patterns. Note that theoretically the reconstruction method presented in the previous section is not limited to regular textures and that there is no constraint on the shape of the texels. However, the synthetic experiments showed that rich and dense textures are needed to provide a trade-off between texel size and number and thus to provide enough constraints for optimization. Given an image of a deformed regular texture we extract a grid of quadrilaterals describing the texture deformation in the image projection. Our method is inspired by recent grid detection methods [16]. We generate suitable feature points on the image using SIFT and group them using unsupervised clustering [3] based on their descriptors (figure 5). For each cluster, a lattice model consistent with the geometric relationship between feature points and the assumed texture regularity is estimated. Each texel is now seen as a planar patch under perspective projection. The vertices of the grid together with additional feature points in each quadrilateral are used as input for the surface reconstruction method explained in the previous section. Figure 2 shows two reconstructed surfaces together with the estimated patch poses and normals at patch centroids. As reference, we captured these surfaces with a trifocal camera setup and used a reconstruction from stereo correspondences as ground truth. The RMSE of the estimated normals of the interpolated surface was 4.9 degrees. Figure 6 shows further reconstruction results of more complex shapes achieved with our method.

5 Conclusion and Future Work

We presented an SFT formulation equivalent to a plane-based SFM problem with a full perspective camera model. The surface is first approximated as being piecewise planar and positions and orientations of planar patches are optimized from estimated homographies between patch appearances in the image. A dense surface is reconstructed using a surface regression to the estimated patch centroids. In future, we will investigate Hermite regression methods that not only account for the estimated 3D positions but also the estimated normals, such as presented in [9, 13]. This will be useful at discontinuities which are currently smoothly interpolated. One advantage of our method is that it does not require a frontal view of the texture elements. Instead, we recover a 2D grid describing the deformation of a regular texture in the image from clustered feature points. To this end, we use SIFT features which are not invariant under affine or projective transformations (e.g. at strong texture deformations) such that other features will be investigated in future.

References

[1] F. L. Bookstein. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[2] T. Collins, J.-D. Durou, A. Bartoli, and P. Gurdjos. Single-View Perspective Shape-from-texture with Focal Length Estimation: A Piecewise Affine Approach. In *Proc. 3D Data Processing, Visualization and Transmission (3DPVT 2010)*, 2010.

[3] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[4] D. A. Forsyth. Shape From Texture and Integrability. In *Proc. Int. Conf. on Computer Vision (ICCV 2001)*, pages 447–452, 2001.

[5] D. A. Forsyth. Shape from Texture without Boundaries. In *Proc. Europ. Conf. on Computer Vision (ECCV 2002)*, pages 225–239, 2002.

[6] F. Galasso and J. Lasenby. Shape from Texture of Developable Surfaces via Fourier Analysis. In *Proc. of the 3rd Int. Conf. on Advances in Visual Computing*, pages 702–713, 2007.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[8] K. Ikeuchi. Shape from Regular Patterns. *Artificial Intelligence*, (1):49 – 75, 1984.

[9] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson Surface Reconstruction. In *Proc. of the fourth Eurographics Symposium on Geometry Processing (SGP 2006)*, pages 61–70, 2006.

[10] P. Kovesi. Shapelets Correlated with Surface Normals Produce Surfaces. In *Proc. Int. Conf. on Computer Vision (ICCV 2005)*, volume 2, pages 994 –1001, 2005.

[11] T. K. Leung and J. Malik. Detecting, Localizing and Grouping Repeated Scene Elements from an Image. In *Proc. Europ. Conf. on Computer Vision (ECCV 1996)*, pages 546–555, 1996.

[12] A. Lobay and D. A. Forsyth. Recovering Shape and Irradiance Maps from Rich Dense Texton Fields. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2004)*, volume 1, pages 400–406, Los Alamitos, CA, USA, 2004.

[13] I. Macedo, J. P. Gois, and L. Velho. Hermite Interpolation of Implicit Surfaces with Radial Basis Functions. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 1–8, 2009.

[14] J. Malik and R. Rosenholtz. Computing Local Surface Orientation and Shape from Texture for Curved Surfaces. *Int. Journ. Comput. Vision*, 23:149–168, June 1997.

[15] E. Malis and M. Vargas. Deeper Understanding of the Homography Decomposition for Vision-Based Control. Arobas INIRA Sophia Antipolis, Universidad der Sevilla, 2007. Technical report.

[16] M. Park, K. Brocklehurst, R. Collins, and Y. Liu. Deformed Lattice Detection in Real-World Images using Mean-Shift Belief Propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Special Issue on Probabilistic Graphical Models*, 31(1), October 2009.

[17] P. Sturm. Algorithms for Plane-Based Pose Estimation. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition, (CVPR 2000)*, pages 1010–1017, June 2000.

[18] B. J. Super and A.C. Bovik. Shape from Texture using Local Spectral Moments. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 17(4):333 –343, 1995.

[19] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-Free Monocular Reconstruction of Deformable Surfaces. In *Proc. Int. Conference on Computer Vision (ICCV 2009)*, 2009.

[20] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA, 1990.

[21] A. Witkin. Recovering Surface Shape and Orientation from Texture. *Journ. of Artificial Intelligence*, 17(1–3):17–45, 1981.

[22] Z. Zhang and A. R. Hanson. Scaled Euclidean 3D Reconstruction Based On Externally Uncalibrated Cameras. In *In IEEE Symposium on Computer Vision*, pages 37–42, 1995.