Vision, Modeling, and Visualization (2011)
Peter Eisert, Konrad Polthier, and Joachim Hornegger (Eds.)

A Global Optimization Approach to High-detail Reconstruction of the Head

David C. Schneider[1,2], Markus Kettern[1], Anna Hilsmann[1,2] and Peter Eisert[1,2]

[1]Fraunhofer Heinrich Hertz Institute, Berlin, Germany
[2]Humboldt Universität zu Berlin, Germany

Abstract

The paper presents an approach for reconstructing head-and-shoulder portraits of people from calibrated stereo images with a high level of geometric detail. In contrast to many existing systems, our reconstructions cover the full head, including hair. This is achieved using a global intensity-based optimization approach which is stated as a parametric warp estimation problem and solved in a robust Gauss-Newton framework. We formulate a computationally efficient warp function for mesh-based estimation of depth which is based on a well known image-registration approach and adapted to the problem of 3D reconstruction. We address the use of sparse correspondence estimates for initializing the optimization as well as a coarse-to-fine scheme for reconstructing without specific initialization. We discuss issues of regularization and brightness constancy violations and show various results to demonstrate the effectiveness of the approach.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction and related work

In this paper we describe an image based approach to 3D reconstruction of the human head from calibrated stereo image pairs without markers or projections. In contrast to the majority of methods in the literature, our approach is designed to reconstruct the complete head including the approximate shape of the subject's hair, as well as head-and-shoulder portraits that include some clothing. Our method does not employ a statistical model of the head and could, in principle, be used to reconstruct other types of objects. It has been optimized, however, for the human head and we have not yet evaluated it otherwise. The head is an easy subject for reconstruction with respect to topology and depth discontinuities. It is challenging, however, with respect to the textures that have to be matched: Many "materials" are either smooth (e.g. skin at low image resolutions), highly self-similar (e.g. hair or skin with pore structure at high image resolutions) or geometrically complex (again, hair).

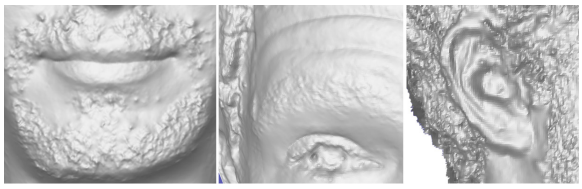

Figure 1: *Details of some reconstructions computed with the proposed approach.*

The images we use are captured with consumer-grade high resolution SLR cameras (Canon EOS 550D, Tokina 35mm fixed focal lens, 18 megapixels) under diffused studio flash lighting (Hensel). The extrinsic and intrinsic camera parameters are computed with help of a calibration object. The background is neutral and the images are masked with an automatic procedure such that the portrait appears on a perfectly white background.

We formulate our reconstruction approach as a global mesh-based nonlinear optimization problem. The following discussion of related work aims at positioning the proposed method in relation to some of the many existing approaches to 3D reconstruction. Is is not and cannot be comprehensive.

By *mesh-based* we mean that we do not estimate an independent depth at every pixel in a stereo pair as many

Figure 2: *Very high detail reconstructions (rendered depth maps), computed with the coarse-to-fine scheme without initial shape estimate. Mesh spacing is 3 pixels, other parameters as in figure 5. Fine structures become visible on the face as well as on clothing.*

depth-map techniques do [SS02], but at discrete vertices of a triangle mesh. The projection of the mesh, however, is fixed in one image of the stereo pair. In 3D the scene is assumed to be planar between the vertices. Therefore, our approach could be regarded as patch based with an *a priori* assumption of patch connectivity. While depth is estimated only at discrete vertices of a mesh, we use differential image information at every pixel in order to compute vertex depth. Our problem formulation is therewith related to triangulation based optical flow estimation techniques [GHN*10, CSM05]. It also shares several computational strategies with modern flow approaches such as a "data term plus regularizer" formulation, coarse-to-fine warping to improve convergence [PBB*06] or the use of robust error functions in the data term (e.g. [WTP*09, GBBS10]). Similar to [BBB*10, BHPS10] we rely on high image resolutions to recover facial details.

Our approach is *global* as, for a pair of views, we solve for the unknown depth of all vertices simultaneously. This is, for example, kin to graph-cut methods [SS02] for depth-map computation or the expectation maximization approach of [SFG04]. We use, however, a classic continuous optimization algorithm, namely a robust variant of Gauss-Newton [MN98]. Local approaches, on the other hand, first reconstruct or optimize over parts of the scene, e.g. feature points or oriented patches [FP08]. These are combined later using meshing techniques such as [KBH06]. We see the primary advantage of the global approach in the fact that smoothness assumptions can easily be built into the reconstruction process in the form of regularization energies.

The warp estimation framework we use to formulate our method has been widely used for 2D problems dealing with non-rigid registration. Applications include image registration (e.g. [BZ04,ZL09,ZGH09]), 2D tracking of deformable

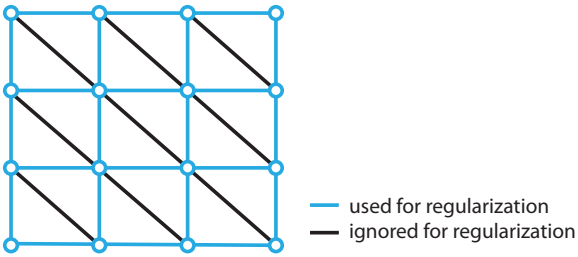

Figure 3: *Topology of the mesh used for optimization. For regularization, diagonal edges are ignored.*

surfaces (e.g. [GBBS10, ZL09, HSE10]) and compensation of camera jitter in video [SHE11]. Given the fact that registration, tracking and 3D reconstruction are ultimately based on the problem of establishing correspondences between images it is no surprise that this framework is suitable for 3D reconstruction. We have addressed this also in [SKHE11], where we explored the use of an epipolar-agnostic warping strategy for depth estimation.

While most applications of the framework rely on a brightness constancy assumption we address a term for handling luminance inconsistencies in section 2.2. Another approach to handling brightness constancy violations is described in [HE09].

2. An optimization approach to reconstruction

We formulate the 3D reconstruction problem in an intensity-based parametric warp estimation framework. In section 2.1 we describe this framework generically. The brightness constancy assumption is discussed in section 2.2. In sections 2.3 and **??** the specific warps for 3D reconstruction are intro-

duced and regularization is addressed in section 2.4. In section 2.5 we cover initialization and computational issues and discuss results.

2.1. Warp estimation framework

Let $\mathcal{I}$, $\mathcal{K}$ denote a pair of single-channel images which we regard as mappings of coordinates to intensities. For the application discussed in this paper $\mathcal{I}$ and $\mathcal{K}$ are a pair of stereo images. Let $\mathcal{W}:\left(\mathbb{R}^2,\mathbb{R}^K\right)\rightarrow\mathbb{R}^2$ be a parametric warp, i.e. a mapping of image coordinates depending on a parameter vector $\theta\in\mathbb{R}^K$, which is the quantity to be estimated. Two different warp functions for stereo reconstruction are described in the following sections. We define the pixel-wise residual or error $\mathcal{E}([x\,y],\theta)$ as the intensity difference between the images under the warp:

$$\mathcal{E}([x\,y],\theta)=\mathcal{I}([x\,y])-\mathcal{K}(\mathcal{W}([x\,y],\theta)) \tag{1}$$

Estimating θ then amounts to solving

$$\arg\min_{\theta}\sum_x\sum_y\rho(\mathcal{E}([x\,y],\theta))+\mathcal{R}(\theta) \tag{2}$$

where ρ is a norm-like function which serves as a robust error metric in the presence of luminance outliers. For all results in the paper the Huber function [Hub81] was used. $\mathcal{R}(\theta)$ is a placeholder for regularization (or smoothness) terms which will be addressed in section 2.4.

For arbitrary norm-like functions this energy can be minimized with a robust Gauss-Newton scheme that differs only slightly from the standard least squares case; details can be found, for example, in [MN98]. This requires the Jacobian of the energy function, whose rows are given by

$$\nabla r_i=-\nabla\mathcal{K}^T\Big|_{\mathcal{W}([x_i\,y_i],\theta)}\cdot\mathbf{J}_{\mathcal{W}}. \tag{3}$$

$\nabla\mathcal{K}^T$ is the image gradient which is evaluated at the warped coordinates. $\mathbf{J}_{\mathcal{W}}$ is the Jacobian of the warp to be optimized.

2.2. On brightness constancy

The above formulation of warp estimation relies on a brightness constancy assumption, i.e. the assumption that differences between $\mathcal{I}$ and $\mathcal{K}$ can be explained by pixels of constant brightness moving according the warp. Empirically, this proves to be wrong for head reconstruction due to non-Lambertian properties of the materials, even if the scene is carefully lit from multiple sides with studio flashes and diffusers. We therefore use a luminance correction term which is motivated as follows. We assume that in the residual $\mathcal{E}$ the non-Lambertian effects are of relatively large scale (or low frequency, but see below for a caveat) while the effects of misalignment due to an imperfect estimate of the warp parameter θ are of relatively small scale (or high frequency). However, we want to allow the areas affected by non-Lambertian effects to have sharp borders; this is required, for example, at borders between different materials in the scene. Therefore the luminance correction term $\mathcal{L}$ must reduce non-Lambertian effects in the residual but avoid to eliminate misalignment effects as these are what drives the optimization in the first place. These requirements are satisfied, for example, by the median filtered residual:

$$\mathcal{L}([x\,y],\theta)=\underset{\substack{i=-K\ldots K\\ j=-K\ldots K}}{\text{median}}\,\mathcal{E}([x\,y],\theta) \tag{4}$$

An alternative to the median filter which we will evaluate in the future is the bilateral filter.

The luminance-corrected residual is then:

$$\mathcal{E}_{\mathcal{L}}([x\,y],\theta)=\mathcal{I}([x\,y])-\mathcal{L}([x\,y],\theta)-\mathcal{K}(\mathcal{W}([x\,y],\theta)) \tag{5}$$

This is the error which we actually use for reconstruction.

2.3. A warp for 3D reconstruction

As stated in the introduction, we describe the 3D geometry of the scene by a piecewise planar triangle mesh. We assume that the projection of the mesh is regular and fixed in the first image of the stereo pair. The fixed mesh projection is illustrated in figure 3. This corresponds well with the above formulation of the warp estimation problem where the warp maps each pixel coordinate in the first image to one in the second.

The relation between two perspective camera images of a planar mesh triangle is a homography. For sufficiently large grazing angles the homography can be approximated with an affine transformation, which is a common strategy in stereo algorithms. This is a significant reduction of computational complexity as an affine transformation can be expressed as a linear function of its parameters with an adequate parametrization (see below). The simplifies the computation of the derivative in equation 3. In contrast, a homography requires a division by the homogeneous coordinate component which results in a more complicated derivative.

Mesh-based piecewise affine warp functions have been widely used for 2D vision and graphics applications, especially in conjunction with the warp estimation framework described above; references are given in the related works section. In the following we first recapitulate the piecewise affine mesh warp used, for example in [SKHE11, SHE11, HSE10], and then adjust it to the 3D reconstruction task.

Let K be the number of vertices in the mesh and assume that the vertices are indexed in the range $\mathcal{V}=\{1\ldots K\}$ so they can be identified by their indices. We denote the vertex coordinates of the undeformed mesh by $[u_V\,v_V]^T$, $V\in\mathcal{V}$ and define two vectors containing all vertex coordinates:

$$\begin{aligned}\mathbf{u}^T&=\left[u_1\ldots u_K\right]\\ \mathbf{v}^T&=\left[v_1\ldots v_K\right]\end{aligned}$$

Now we consider a single pixel at coordinates $[x\,y]^T$. Let $T \in \mathcal{V}^3$ be the surrounding triangle of the pixel and let $c^{(1)}$, $c^{(2)}$, $c^{(3)}$ be its barycentric coordinates with respect to T in the undeformed control mesh. We define the sparse barycentric coordinate vector $\mathbf{b}^T = \left[b^{(1)} \dots b^{(K)}\right]$ as follows:

$$b^{(V)} = \begin{cases} c^{(i)} & \text{if } V \text{ is the } i\text{th vertex of triangle } T \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The pixel coordinates x and y can be expressed in terms of the mesh vertices:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{b}^T & \\ & \mathbf{b}^T \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \tag{7}$$

The sparsity pattern of $\mathbf{b}$ is construed to "select" the correct vertex coordinates out of $\mathbf{u}$ and $\mathbf{v}$.

Similar to the coordinates, the displacement of pixels under a warp can be expressed in terms of vertex displacements $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ with the help of barycentric coordinates. This allows the formulation of a warp parametrized by the vertex displacements:

$$\mathcal{W}\left(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \mathbf{b}^T & \\ & \mathbf{b}^T \end{bmatrix} \begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix} \tag{8}$$

Note that in general the displacement of each pixel is governed by the displacement of its three surrounding mesh vertices. This amounts to six degrees of freedom (two for each vertex). Therefore, this warp is piecewise affine in the image plane.

One of the advantages of this warp is its easy formulation (and computation) for all pixels simultaneously as a single matrix equation. Let $\mathbf{x}^T = [x_1 \dots x_N]$ and $\mathbf{y}^T = [y_1 \dots y_N]$ be vectors of all pixel coordinates considered. Further, let $\mathbf{B}$ be a matrix of barycentric coordinates:

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_N^T \end{bmatrix} \tag{9}$$

Then the warp for all coordinates can be expressed as:

$$\mathcal{W}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} \mathbf{B} & \\ & \mathbf{B} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix} \tag{10}$$

Note that the Jacobian of this warp is simply the matrix on the right hand side.

For 3D reconstruction the piecewise affine warp is an over-parametrization, as it ignores the camera calibration. In order to involve the calibration without complicating the warp, we restrict the vertex displacement to the epipolar lines as follows. For each vertex V, the direction of the constraint line is given by a vector $\begin{bmatrix} e_V^{\mathrm{X}} & e_V^{\mathrm{Y}} \end{bmatrix}$ which we assume to be of norm 1. For a displacement by d_V pixels along the line, the vertex coordinate offsets are $\Delta u_V = d_V \cdot e_V^{\mathrm{X}}$ and $\Delta v_V = d_V \cdot e_V^{\mathrm{Y}}$. With all displacement parameters d_V in a vector $\mathbf{d}^T = \begin{bmatrix} d_1 \dots d_K \end{bmatrix}$ the displacement Δx, Δy of a single pixel can be expressed as:

$$\Delta x = \mathbf{b}^T \Delta u_V = \mathbf{b}^T \operatorname{diag}\left(e_1^{\mathrm{X}} \dots e_K^{\mathrm{X}}\right) \mathbf{d} =: \mathbf{e}_{\mathrm{X}}^T \mathbf{d} \tag{11}$$

$$\Delta y = \mathbf{b}^T \Delta v_V = \mathbf{b}^T \operatorname{diag}\left(e_1^{\mathrm{Y}} \dots e_K^{\mathrm{Y}}\right) \mathbf{d} =: \mathbf{e}_{\mathrm{Y}}^T \mathbf{d} \tag{12}$$

Then we can formulate the constrained warp similar to the unconstrained one as:

$$\mathcal{W}\left(\begin{bmatrix} x \\ y \end{bmatrix}, \mathbf{d}\right) = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{\mathrm{X}}^T \\ \mathbf{e}_{\mathrm{Y}}^T \end{bmatrix} \mathbf{d} \tag{13}$$

Note a subtle difference to the unconstrained case: The epipolar constraint only makes sense if the epipolar consistent displacement (the right side of the sum above) is added to a pixel coordinate which is epipolar consistent itself. This coordinate is in general not identical with $[x\,y]^T$ and therefore denoted as $[\hat{x}\,\hat{y}]^T$. The choice of $[\hat{x}\,\hat{y}]^T$ determines which disparities are assumed when $\mathbf{d} = \mathbf{0}$ and is somewhat arbitrary. We use the projection of $[x\,y]^T$ to its epipolar line. Also note that the matrix here is not block-diagonal and the number of parameters is halved in comparison to the unconstrained warp.

Again, the warp can be written in a single matrix equation for all pixels. We define two diagonal matrices:

$$\mathbf{E}_{\mathrm{X}} := \operatorname{diag}\left(e_1^{\mathrm{X}} \dots e_K^{\mathrm{X}}\right) \tag{14}$$

$$\mathbf{E}_{\mathrm{Y}} := \operatorname{diag}\left(e_1^{\mathrm{Y}} \dots e_K^{\mathrm{Y}}\right) \tag{15}$$

The all-pixel warp is given by:

$$\mathcal{W}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \mathbf{d}\right) = \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{bmatrix} + \begin{bmatrix} \mathbf{B}\mathbf{E}_{\mathrm{X}} \\ \mathbf{B}\mathbf{E}_{\mathrm{Y}} \end{bmatrix} \mathbf{d} \tag{16}$$

Note that this is still only an approximation to the correct 3D mesh warp described in section 2.3: While the vertices move consistently with epipolar geometry due to the constraint, the motion of the pixels in between is still described by an affine transform rather than a homography.

2.4. Regularization

The affine epipolar warp is often regularized by an energy term based on a mesh Laplacian $\mathbf{L}$ (e.g. [BZ04, HSE10]):

$$\mathcal{R}\left(\begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix}\right) = \left\| \begin{bmatrix} \mathbf{L} & \\ & \mathbf{L} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{u} \\ \Delta\mathbf{v} \end{bmatrix} \right\|^2 \tag{17}$$

For regularizing the constrained warp there are several options: Either the displacement magnitudes $\mathbf{d}$ along the epipolar lines are regularized directly. Alternatively, the components of $\mathbf{d}$ are converted back to Δu- and Δv-displacements and these are regularized in the same way as they are in the unconstrained warp. $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ can be obtained from $\mathbf{d}$ by:

$$\Delta\mathbf{u} = \mathbf{E}_{\mathrm{X}}\mathbf{d} \tag{18}$$

$$\Delta\mathbf{v} = \mathbf{E}_{\mathrm{Y}}\mathbf{d} \tag{19}$$

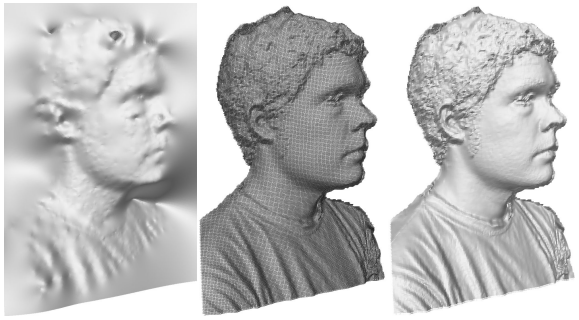

Figure 4: *Left: Initial depth estimate computed from SIFT correspondences as described in section 2.5. Center, right: Rendered low resolution reconstruction results, with and without mesh (6 megapixels, 5 pixel mesh spacing, one stereo pair). Note that the optimization was computed at a single image scale with a fixed mesh resolution, i.e. without coarse-to-fine strategies.*

Therefore the regularization energy is given by:

$$\mathcal{R}(\mathbf{d}) = \left\| \begin{bmatrix} \mathbf{L} & \\ & \mathbf{L} \end{bmatrix} \left(\begin{bmatrix} \mathbf{E}_X \\ \mathbf{E}_Y \end{bmatrix} \mathbf{d} \right) \right\|^2 = \left\| \begin{bmatrix} \mathbf{L}\mathbf{E}_X \\ \mathbf{L}\mathbf{E}_Y \end{bmatrix} \mathbf{d} \right\|^2 \quad (20)$$

Note that the matrix in the rightmost term can be precomputed. For regularization we only use horizontal and vertical edges of the mesh as shown in figure 3.

2.5. Computational issues and results

The energy function we minimize for reconstruction is highly non-convex and hence the optimization is prone to local minima. Therefore, either an initial estimate of the shape to reconstruct is required as a starting point for the optimization, or a coarse-to-fine strategy needs to be used which smooths the error function in early iterations. We have implemented both strategies for the proposed approach. Regardless of the chosen strategy we always iterate over several, increasingly lower weights of the Laplacian smoothness term.

For the coarse to fine strategy, we iterate over both image scale and, implicitly, mesh resolution. We use a fine-grained stepping in image scale space with a scale factor difference of about 15 percent. The spacing of mesh vertices in the first view is kept fixed at, typically, five pixels. Therefore the mesh is recomputed at every scale and implicitly becomes finer in relation to the image resolution. For very fine reconstructions, as shown in figure 2, the vertex spacing is lowered further down to three pixels when the full image resolution is reached.

The meshes shown in figures 5 and 2 were computed from two stereo pairs with cameras in a triangular configuration. The left and right view were warped independently to the center center view and correspondences over all three views were used to compute the depth. The reconstruction in figure 4 was computed from a single stereo pair. Meshes shown in figures 5 and 2 were computed with the coarse-to-fine approach, i.e. without an initial shape estimate. Convergence is generally very good but failures do occur, especially on large areas with very high frequency texture detail such as smooth types of cloth and skin, e.g. in the second row of figure 5. Here, texture details only appear at very fine scales and the correspondence estimate is too far off already when these scales are reached.

Robustness and also computation speed can be improved if an initialization for the optimization is provided, i.e. an estimate of the shape to reconstruct. Excellent reconstruction results can be achieved with relatively coarse initializations, even at a fixed image scale and mesh resolution. An example is shown in figure 4. The initialization was computed from a set of sparse feature correspondences found with the SIFT algorithm [Low04]. In order to initialize the optimization, a displacement along the epipolar line must be computed from the sparse correspondences for each vertex of the optimization mesh. Let

$$[x_i\, y_i] \leftrightarrow [\hat{x}_i\, \hat{y}_i]$$

with $i = 1 \ldots N$ be a pair of corresponding feature points. For each such pair we identify its enclosing mesh triangle in the first view (where the mesh is fixed) as well as its barycentric coordinates and create a sparse row vector $\hat{\mathbf{b}}_i^T$ as described in equation (6). We then find initial vertex offsets $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ by solving the following linear system in a least squares sense:

$$\begin{bmatrix} \hat{\mathbf{b}}_1^T \\ \vdots \\ \hat{\mathbf{b}}_N^T \\ \gamma\mathbf{L} \end{bmatrix} \Delta\mathbf{u} = \begin{bmatrix} \hat{x}_1 - x_1 \\ \vdots \\ \hat{x}_N - x_N \\ \mathbf{0} \end{bmatrix} \quad (21)$$

$\mathbf{L}$ is the Laplacian of the optimization mesh and γ controls the smoothness of the result. $\Delta\mathbf{v}$ can be found analogously. Finally, the $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ offsets are mapped to the nearest epipolar lines and converted to disparities as required for the warp in equation (13). Figure 4 shows the initial shape estimate obtained from this procedure as well as the optimization result.

3. Conclusion and future work

In summary, we have described an approach to 3D reconstruction of human head and shoulder portraits which is stated in a nonlinear warp estimation framework and solved as a global optimization problem. We have formulated a computationally efficient warp function for depth estimation which is based on a widely used piecewise affine warp. We have addressed issues of luminance correction, regularization and initialization as well as coarse-to-fine strategies for reconstruction without initial depth estimate.

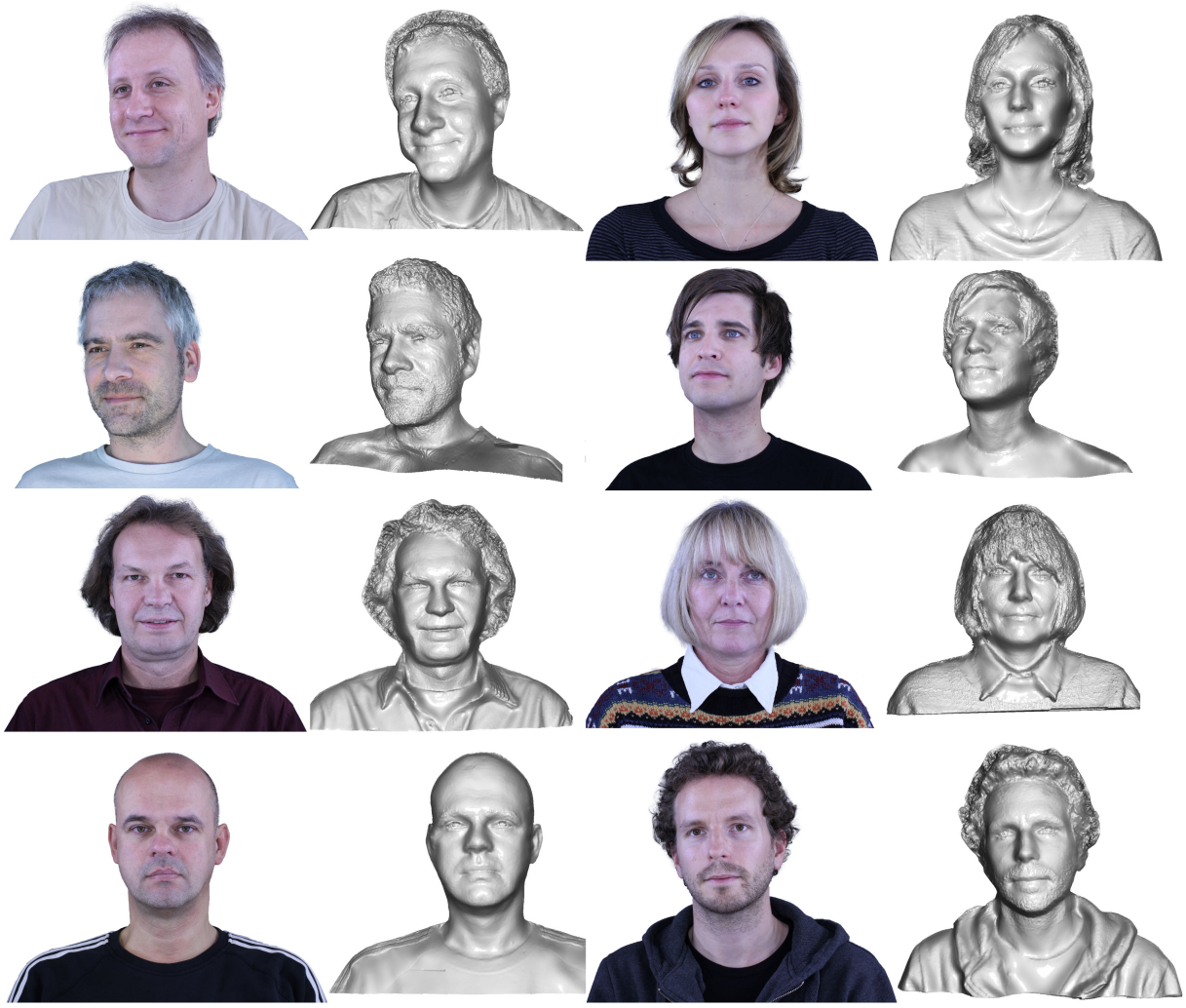

Figure 5: *Reconstruction results (rendered depth maps) with one source image. Reconstructions were computed with the coarse-to-fine scheme without initial depth estimate from two stereo pairs (triangular camera setup, left and right view warped to the center). Mesh spacing is 5 pixels, final image resolution is 18 megapixels. In the second row convergence failed on the plain cloth of the shirts as the structure of the texture only appears at very high levels in scale space.*

Regarding future work, we want to fuse meshes from multiple camera pairs in order to reconstruct a 180 degree model of the person. We think that depth map fusion can be formulated naturally in our optimization framework as a warp estimation problem over a set of images. We also will investigate strategies to further improve the geometric quality of our results, especially on hair. Finally, we plan to evaluate the performance of our method on subjects other than human heads.

References

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M.: High-Quality Single-Shot Capture of Facial Geometry. *ACM Transactions on Graphics 29*, 3 (2010). 2

[BHPS10] BRADLEY D., HEIDRICH W., POPA T., SHEFFER A.: High Resolution Passive Facial Performance Capture. *ACM Transactions on Graphics 29*, 3 (2010). 2

[BZ04] BARTOLI A., ZISSERMAN A.: Direct Estimation of Non-Rigid Registrations. In *British Machine Vision Conference* (London, 2004). 2, 4

[CSM05] CONDELL J., SCOTNEY B., MORROW P.: Adaptive Grid Refinement Procedures for Efficient Optical Flow Computation. *International Journal of Computer Vision 61*, 1 (2005), 31–54. 2

[FP08] FURUKAWA Y., PONCE J.: Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 1* (2008), 1–14. 2

[GBBS10] GAY-BELLILE V., BARTOLI A., SAYD P.: Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning. *IEEE Transactions On Pattern Analysis And Machine Intelligence 32*, 1 (2010), 87–104. 2

[GHN*10] GLOCKER B., HEIBEL T., NAVAB N., KOHLI P., ROTHER C.: TriangleFlow: Optical Flow with Triangulation-Based Higher-Order Likelihoods. In *European Conference on Computer Vision* (Heraklion, 2010). 2

[HE09] HILSMANN A., EISERT P.: Joint Estimation of Deformable Motion and Photometric Parameters in Single View Video. In *ICCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment* (Kyoto, 2009). 2

[HSE10] HILSMANN A., SCHNEIDER D. C., EISERT P.: Realistic cloth augmentation in single view video under occlusions. *Comput. Graph. 34* (October 2010), 567–574. 2, 3, 4

[Hub81] HUBER P.: *Robust Statistics*. John Wiley & Sons, 1981. 3

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing* (Aire-la-Ville, Switzerland, Switzerland, 2006), SGP '06, Eurographics Association, pp. 61–70. 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60, 2* (2004), 91–110. 5

[MN98] MCCULLAGH P., NELDER J.: *Generalized Linear Models*. Chapman & Hall, 1998. 2, 3

[PBB*06] PAPENBERG N., BRUHN A., BROX T., DIDAS S., WEICKERT J.: Highly Accurate Optic Flow Computation with Theoretically Justified Warping. *International Journal of Computer Vision 67* (2006), 141–158. 2

[SFG04] STRECHA C., FRANSENS R., GOOL L. V.: Wide-baseline stereo from multiple views: A probabilistic account. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 1* (2004), 552–559. 2

[SHE11] SCHNEIDER D. C., HILSMANN A., EISERT P.: Warp-based Motion Compensation for Endoscopic Kymography. In *Eurographics 2011* (Llandudno, 2011). 2, 3

[SKHE11] SCHNEIDER D. C., KETTERN M., HILSMANN A., EISERT P.: Deformable image alignment as a source of stereo correspondences on portraits. In *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA)* (Colorado Springs, 2011). 2, 3

[SS02] SCHARSTEIN D., SZELISKI R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision 47* (April 2002), 7–42. 2

[WTP*09] WERLBERGER M., TROBIN W., POCK T., WEDEL A., CREMERS D., BISCHOF H.: Anisotropic Huber-L1 Optical Flow. In *British Machine Vision Conference* (London, 2009). 2

[ZGH09] ZHU J., GOOL L. V., HOI S.: Unsupervised Face Alignment by Robust Nonrigid Mapping. In *International Conference on Computer Vision* (2009). 2

[ZL09] ZHU J., LYU M.: A Fast 2D Shape Recovery Approach by Fusing Features and Appearance. *IEEE Transactions On Pattern Analysis And Machine Intelligence 31*, 7 (2009), 1210–1224. 2