

# Deformable Image Alignment as a Source of Stereo Correspondences on Portraits

David C. Schneider<sup>1,2</sup>, Markus Ketter<sup>1</sup>, Anna Hilsmann<sup>1,2</sup>, Peter Eisert<sup>1,2</sup>

david.schneider@hhi.fraunhofer.de

<sup>1</sup>Fraunhofer Heinrich Hertz Institute, Berlin, Germany

<sup>2</sup>Humboldt Universität zu Berlin, Germany

## Abstract

*Mesh-based deformable image alignment (MDIA) is an algorithm that warps a template image onto a target by deforming a 2D control mesh in the image plane, using an image-based nonlinear optimization strategy. MDIA has been successfully applied to various nonrigid registration problems, deformable surface tracking and stabilization of scene-to-camera motion in video. In this paper we investigate the use of image-based MDIA for computing dense correspondences for 3D reconstruction of human heads from high resolution portrait images. Human heads are topologically simple in 3D while providing textures which are challenging to match, such as hair and skin. We find that even with a simple piecewise affine deformation model MDIA delivers excellent correspondence results. We propose a robust, piecewise optimization scheme to compute MDIA on very high resolution images. We address issues of regularization and luminance correction and discuss the role of epipolar constraints. The correspondences retrieved with our approach facilitate the estimation of camera extrinsics and yield highly detailed meshes of the head.*



Figure 1. Reconstruction result computed from correspondences obtained with mesh-based deformable image alignment. The lower row shows details of zipper, mouth and eyebrow.

## 1. Introduction

Algorithms for deformable image alignment aim at aligning two images by “warping” the template image onto the target image according to a deformation model, minimizing a measure of error between the deformed template and the target. A popular and computationally efficient class of deformation models is based on meshes. Here, each pixel of the template is unambiguously coupled to surrounding vertices of a control mesh in the image plane, causing the image to deform along with the mesh. The alignment problem reduces to finding a control mesh deformation, i.e. an offset vector for every vertex. This can be stated as an image-based energy minimization task and solved efficiently in a robust Gauss-Newton framework. The approach

has been successfully applied not only to image registration [1, 22] but also to monocular tracking of nonrigid surfaces [12, 8, 23] and, recently, to stabilization of strong camera motion in endoscopic video [17].

Similar to image registration and tracking, image-based 3D reconstruction is at its core a problem of establishing correspondences. In this paper, we investigate the use of the successful mesh-based deformable image alignment (MDIA) approach with a piecewise affine deformation model for establishing dense correspondences for the reconstruction task. Portrait images showing the full head and a part of the upper body are used as the subject for reconstruction. The reconstruction task is eased by the simple topology of the expected 3D shape and the small amount of

occlusions and depth discontinuities. It is challenging, however, with respect to the image textures to be matched: Hair, for example, is difficult due to its intricate geometry and its complex interaction with light. Skin can be problematic as it may contain very little gradient information and is prone to specularities which violate the brightness constancy assumption.

In order to recover a high degree of detail, we use a coarse-to-fine strategy with very high resolution images (18 megapixels) and very fine meshes (five pixels vertex spacing). This requires a new computational strategy for computing MDIA in a piecewise fashion: We start with a global optimization on the full image and switch to patch-wise optimization when image and mesh resolution become prohibitively large. To improve convergence we use nested iterations over image scale and regularization parameters. A simple but efficient luminance correction is used to cope with violations of the brightness constancy assumption. For reconstruction, we use three images taken with an upright triangular camera setup and a baseline of roughly 20 cm. The images of the left and right lower cameras are registered with the image of the upper central camera and correspondences over three views are used for depth computation.

In contrast to generic image registration and nonrigid tracking, the stereo matching problem has an additional constraint, namely epipolar geometry. If the camera setup is calibrated, the vertex displacements can be restricted to the epipolar lines. A straightforward way of implementing this is to rectify the image pair and only optimize for the  $X$  or  $Y$  components of the displacement vectors, depending on the direction of rectification.

In the uncalibrated case, the most rigorous strategy would be to include the camera parameters in the optimization and keep the vertex displacement consistent with the last estimate of the cameras in the iterative optimization process. This approach ultimately boils down to an intensity-based bundle adjustment (e.g. [19], section 3.4. and references therein). However, bundle adjustment is known to require a good starting point for both cameras and geometry in order to converge. Therefore it is typically used as the last stage in 3D reconstruction [11]. In this paper, however, we are interested in using nonrigid image alignment to establish correspondences from the beginning. The most simple approach to do this is to ignore the epipolar constraint and use the freeform deformation model as it is used in generic registration and tracking. If the image alignment is of sufficient quality, the correspondences can be used to compute the extrinsic parameters of the setup. This strategy is similar to a feature-based approach to 3D (e.g. using SIFT, SURF, etc.), where feature extraction is agnostic of epipolar geometry and the correspondences are used later to retrieve the cameras. In contrast to a feature-based strategy, the correspondences generated with the im-

age alignment approach are dense.

While the epipolar-agnostic approach is an over-parametrization, it is a challenging benchmark for the non-rigid alignment algorithm with respect to correspondence quality. As the focus of this paper is on the capabilities of MDIA, all results shown were generated with the extrinsically uncalibrated, epipolar-agnostic approach and all camera extrinsics were computed from the correspondences established by the image alignment. We find that, despite of the simple piecewise affine deformation model we use, correspondence quality is excellent when MDIA is used with the proposed computational scheme.

The paper is organized as follows. After reviewing related work in the following section, we briefly recapitulate the MDIA formalism in section 3 and discuss the proposed computational scheme in detail in section 4. Finally, results and a conclusion are given in section 5.

## 2. Related work

Our work is closely related to several lines of research which we can only sketch briefly in the following.

**Mesh-based deformable image alignment**, as used in this work, has been applied to various registration problems, e.g. [1, 23, 22]. It has been extended by several authors to nonrigid surface tracking in monocular video sequences, where the warp is estimated for each frame pair [12, 8, 23]. In [17], MDIA is used to stabilize scene-to-camera motion in endoscopic video. Different ways of coupling the location of a pixel to the locations of its surrounding vertices are used, for example radial basis functions in [8] and barycentric coordinates in [12]. Barycentric coordinate warps amount to a piecewise affine transformation of the image and are employed in this work as well. Some authors deal with violations of the brightness constancy assumption; for example, [12] estimates a multiplicative photometric term along with the deformation. Our luminance correction model is similar but decoupled from warp estimation (section 4).

An extensive discussion of **3D reconstruction**, even if restricted to the literature specialized on the human head or face, is beyond scope. Of all aspects of 3D reconstruction, that of finding correspondences is most relevant to this work. Correspondence search can be classified into active and passive, as well as into calibrated and uncalibrated methods. Active methods modify the scene to ease correspondence detection, e.g. by structured light projections. Passive methods on the other hand rely entirely on scene features and require a sufficiently textured surface. With high resolution cameras, the fine details of human skin can be exploited for impressive reconstructions [2]. The other line of classification, calibrated vs. uncalibrated, relates to whether knowledge of scene geometry is used. The vast majority of stereo approaches, generic ones such as PMVS

[7] as well as specialized methods for the face [2, 4], use an extrinsically calibrated setup in order to restrict the search for correspondences to epipolar lines. Volumetric and hull based methods, too, rely on extrinsic camera parameters for ray computation. Uncalibrated correspondence search is often associated with matching feature descriptors like SIFT, SURF, Harris corners, etc. The downside of feature matching for reconstructing the human head is the sparsity of the recovered 3D points. Another class of algorithms that deal with uncalibrated images use statistical semantic models to restrict the space of possible 3D shapes [3, 6]. These morphable models are extremely successful on the face, but fail to represent small-scale individual details due to their statistical nature. Also, to the best knowledge of the authors no morphable model covering the complete head including hair exists.

Finally, **optical flow methods**, which share assumptions and strategies with MDIA, recover dense image correspondences as well. Optical flow is often investigated in a variational setting, e.g. [16, 20]. Some approaches, however, also employ triangle meshes, for example the discrete algorithm by [9] or the finite element approach of [5]. Optical flow and MDIA algorithms often share the brightness constancy assumption, the violation of which severely hampers results. Both techniques may involve similar smoothness terms in their models, such as a Laplacian in the classic variational approach of [13]. Also, several computational techniques for improving robustness and convergence are used with both methods. Examples are coarse-to-fine warping schemes which have been theoretically justified in [16], or the use of robust error functions such as Huber’s [14] in the data term [21, 8].

### 3. Mesh-based deformable image alignment

In the following we briefly summarize our approach to MDIA, which largely follows the methodology used in deformable surface tracking, e.g. [12, 1].

We denote by  $K$  the number of vertices in the mesh and assume that the vertices are indexed in the range  $\mathcal{V} = \{1 \dots K\}$  so they can be identified by their indices. The topology of our control mesh is illustrated in figure 2. Similarly, we assume the image pixels to be indexed in range  $\mathcal{P} = \{1 \dots N\}$ . We denote vertex coordinates of the undeformed mesh by  $[u_V \ v_V]^T$ ,  $V \in \mathcal{V}$  and pixel coordinates as  $[x_P \ y_P]^T$ ,  $P \in \mathcal{P}$ .

Let  $T_P \in \mathcal{V}^3$  be the surrounding triangle of pixel  $P \in \mathcal{P}$  and be  $c_P^{(1)}$ ,  $c_P^{(2)}$ ,  $c_P^{(3)}$  the barycentric coordinates of  $P$  with respect to  $T_P$  in the *undeformed* control mesh. For each pixel that lies under the mesh we define a sparse row vector

$\mathbf{b}_P^T = [b_P^{(1)} \dots b_P^{(K)}]$  of length  $K$  as follows:

$$b_P^{(V)} = \begin{cases} c_P^{(i)} & \text{if } V \text{ is the } i\text{th vertex of triangle } T_P \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Thus  $\mathbf{b}_P$  contains the three barycentric coordinates of  $P$  with respect to its surrounding triangle, distributed according to the ordering of the vertices. Finally, we define the matrix of barycentric coordinates of all pixels as

$$\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_N]^T. \quad (2)$$

We can now formulate the transformation induced by a mesh deformation compactly. Denote by  $\mathbf{X}$  the matrix of all pixel coordinates and by  $\mathbf{D}$  a matrix of vertex displacements, i.e. a control mesh deformation:

$$\mathbf{X} = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} \Delta u_1 & \Delta v_1 \\ \vdots & \vdots \\ \Delta u_K & \Delta v_K \end{bmatrix}$$

Then the pixel coordinates  $\mathbf{X}^*$  after the deformation are

$$\mathbf{X}^* = \mathbf{X} + \mathbf{B}\mathbf{D} \quad (3)$$

The inverse problem, i.e. retrieving  $\mathbf{D}$  from an image pair, can be stated as an instance of the more general parametric warp estimation problem. Denote by  $\mathcal{I}$ ,  $\mathcal{K}$  a pair of single-channel images which we regard as mappings of coordinates to intensities. Let  $\mathcal{W} : (\mathbb{R}^2, \mathbb{R}^K) \rightarrow \mathbb{R}^2$  be a parametric warp, i.e. a mapping of image coordinates depending on some parameter vector  $\boldsymbol{\theta}$ , which is the quantity to be estimated. Define the pixel-wise residual as the intensity difference under the warp:

$$r(\boldsymbol{\theta}) = \mathcal{I}([x \ y]) - \mathcal{K}(\mathcal{W}([x \ y], \boldsymbol{\theta})) \quad (4)$$

Estimating  $\boldsymbol{\theta}$  amounts to minimizing the energy

$$\arg \min_{\boldsymbol{\theta}} \sum_{i=1}^N \rho(r_i(\boldsymbol{\theta})) + \mathcal{R}(\boldsymbol{\theta}) \quad (5)$$

where  $\rho$  is a norm-like function such as the Huber function [14], which was used for all results in this paper.  $\mathcal{R}(\boldsymbol{\theta})$  is a placeholder for smoothness terms which will be addressed below. Note that equations (4) and (5) are only valid under the brightness constancy assumption.

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 [15]. This requires the Jacobian of the energy function, whose rows are given by

$$\nabla r_i = -\nabla \mathcal{K}^T|_{\mathcal{W}([x_i \ y_i], \boldsymbol{\theta})} \cdot \mathbf{J}_{\mathcal{W}}. \quad (6)$$

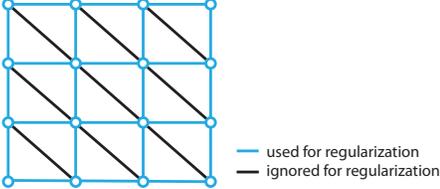


Figure 2. Mesh topology and subgraph used for regularization.

$\nabla \mathcal{K}^T$  is the image gradient which is evaluated at the warped coordinates.  $\mathbf{J}_{\mathcal{W}}$  is the Jacobian of the warp to be optimized. For piecewise affine MDIA, the warp  $\mathcal{W}$  with vertex displacements  $\mathbf{D}$  as parameters is given in equation (3). The Jacobian is simply  $\mathbf{J}_{\mathcal{W}} = \mathbf{B}$ .

Depending on the warp to be estimated, regularization terms need to be incorporated in the optimization. Our regularization consists of two components. Firstly, the discrete Laplacian of a subgraph of the control mesh is used. For any graph with  $K$  vertices, degrees  $n_1 \dots n_K$  and adjacency matrix  $\mathbf{A}$ , the uniform-weight Laplace matrix [10] is defined as

$$\mathbf{L} = \text{diag}(n_1 \dots n_K) - \mathbf{A}. \quad (7)$$

The Laplace matrix is built for a subgraph of the control mesh which is obtained by eliminating vertical edges. The regularization term used is

$$\mathcal{R}_1(\mathbf{D}) = \|\mathbf{L}\mathbf{D}\|_F^2 \quad (8)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. This is a classic regularization approach for MDIA.

The second regularization term is used only at the higher resolution levels in the image scale pyramid, which is described in more detail in the following section. It penalizes deviations of the current, high-frequency solution from the previously computed low-frequency solution  $\mathbf{D}_{\text{lf}}$ :

$$\mathcal{R}_2(\mathbf{D}) = \lambda_2 \|\mathbf{D} - \mathbf{D}_{\text{lf}}\|_F^2 \quad (9)$$

Details on computing  $\mathbf{D}_{\text{lf}}$  are given below.

## 4. Optimization scheme

We find that convergence and robustness of MDIA primarily depends on four factors: (1) the amount of image detail, (2) the relative mesh granularity, i.e. the number of pixels that contribute to each vertex in the data term, (3) the relative weight of the smoothness terms with respect to the data term, and (4) deviation from the brightness constancy assumption. Computational cost and memory usage, on the other hand, depend largely on the number  $K$  of mesh vertices, as the sparse linear system to be solved in each Gauss-Newton iteration is of size  $K \times K$ . For fine meshes and high-resolution images, memory cost is significant, despite the sparsity of the matrix.

We therefore use a two stage nested iteration approach to optimization, implementing a coarse-to-fine scheme not only with respect to image size but also to mesh granularity and smoothness. The outermost iteration loop is over image resolution which is expressed in terms of an image scaling factor  $s \in (0, 1]$  in the following. The second loop is over the weight of the Laplacian smoothness term and the innermost loop is the usual Gauss-Newton iteration. The *stages* are defined with respect to the scaling factors: On the lower scales, the mesh spans the entire image; this is the *global* stage. At a certain scale, we switch to the *local* stage, where the optimization is computed piecewise on image patches. In the following, the key ingredients of the scheme are described in detail.

**Global stage** In the global stage the mesh spans the entire image. We keep the spacing between mesh vertices constant over scale changes. Thus the mesh is rebuilt at every scale with an increasing vertex count. As image detail increases along with the scale factor, the mesh implicitly becomes more adaptive as well. It has been reported in the optical flow literature that a fine-grained stepping of scale factors leads to better convergence than the classic half-resolution Gaussian pyramid (e.g. [16]). We find this to be the case for MDIA as well and use a scale difference of 15%. At the smallest image scale ( $s = 0.05$  in our setup) the warp has to bridge the entire baseline between the camera pair. Therefore, the middle iteration loop (over the weight of the Laplacian smoothness term) has five steps at this scale while at all other scales only two iterations over regularization weights are used.

**Local stage** Due to the increasing vertex count, global computation becomes inefficient at a certain scale that depends on the available memory and computing power ( $s = 0.35$  in our setup). At this point we switch to a piecewise optimization scheme. For piecewise computation, we define a patch size which can be efficiently optimized and compute an appropriate mesh. The image is then tiled with these mesh patches with some overlap and the optimization is computed for each patch separately. Note that this is a natural starting point for parallelization. The results are joined before the next step in scale iteration. The patch size *and* the mesh are now kept constant at all scales. Therefore, the number of patches has to increase to cover the full image.

**Result propagation** The proposed optimization scheme requires us to combine solutions of different patches as well as to propagate solutions between scales and meshes that are not trivial to relate due to the different mesh resolutions. To facilitate these transformations, we use an intermediate representation of the vertex displacements as a dense vector field  $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  which is decoupled

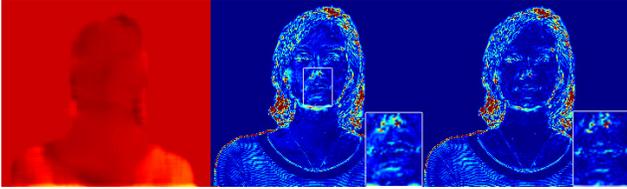


Figure 3. Luminance correction map from scale 0.07 (left), residual before (center) and after (right) correction. Note that the scale of the luminance correction map is exaggerated. As an example, the mouth region has been magnified for better visibility of the subtle but important residual differences.

from the meshes and can be evaluated at arbitrary locations. The field is implemented as a natural neighbor interpolator [18]. For a solved patch with top left corner  $\mathbf{p}$ , each vertex displacement vector  $\mathbf{u}$  at location  $\mathbf{x}$  is stored by setting  $\Psi(s^{-1}(\mathbf{x} + \mathbf{p})) \leftarrow s^{-1}\mathbf{u}$ ; i. e. the coordinate system of  $\Psi$  is that of the full image. At the next scale the interpolator is queried at the appropriate locations to initialize the displacements for the next optimization step.

**Low frequency regularization** The smoothness term  $\mathcal{R}_2$  in equation (9) regularizes displacements using a solution from a lower scale. This term is used only in the local optimization stage. The low frequency solution  $\mathbf{D}_{\text{lf}}$  used for regularization is based on the solution of the last iteration of the global scale. Note that  $\mathbf{D}_{\text{lf}}$  is not simply the solution matrix of this scale, but has to be adapted to match the location and the scale of the patch being optimized.

**Luminance correction** Violations of the brightness constancy assumption severely hamper correspondence results but are hard to avoid even with careful lighting. We therefore apply an explicit luminance correction to the target image in the global optimization stage. An additive luminance correction map is computed from the residual, i. e. the difference between the target and the template warped according to the current deformation estimate. We assume that luminance effects in the residual are of relatively large scale (or low frequency, but see below for a caveat) while the effects of misalignment are of relatively small scale (or high frequency). Luminance correction must avoid to eliminate misalignment effects as these are what drives the optimization in the first place. Also, we want to allow sharp borders for luminance effects which precludes the use of the low-pass filtered residual as an estimate for luminance effects. Instead, we use the median-filtered residual.

Formally, the luminance correction term  $\mathcal{L}([x y])$  at pixel  $[x y]$  is:

$$\mathcal{L}([x y]) = \underset{\substack{i=-K \dots K \\ j=-K \dots K}}{\text{median}} \left\{ \mathcal{K}(\mathcal{W}([x+i \ y+j], \mathbf{D})) - \mathcal{I}([x+i \ y+j]) \right\} \quad (10)$$



Figure 4. Example of an image triple used for reconstruction.



Figure 7. Points (red) and corresponding epipolar lines (green) from fundamental matrices computed from the MDIA correspondences.

$\mathcal{L}$  is upscaled and added to the target image before the next iteration of warp estimation. An example for  $\mathcal{L}$  and its effect on the residual is shown in figure 3. As we use a constant filter size ( $K = 10$ ) over all scales, the luminance map implicitly becomes more detailed on finer scales. In the piecewise stage, where the image patches are smaller and correspondence is better, a high-pass filter is used on the patches.

**Depth computation** To obtain a 3D reconstruction from the correspondences found with MDIA, the fundamental matrices between each view pair are computed using RANSAC with a tight threshold and the normalized eight point algorithm [11]. Figure 7 shows examples of epipolar lines from these fundamental matrices. For a metric reconstruction, the fundamental matrix is upgraded to a camera matrix using the intrinsic camera parameters obtained from a calibration pattern. Depths are computed by linear triangulation using correspondences between three views and the computed cameras, again following [11].

## 5. Results and conclusion

To demonstrate the quality of MDIA correspondence results, figures 5 and 6 show reconstructions of several heads, computed as described above, together with one of the respective input views for visual comparison.

The results were generated from correspondences of two

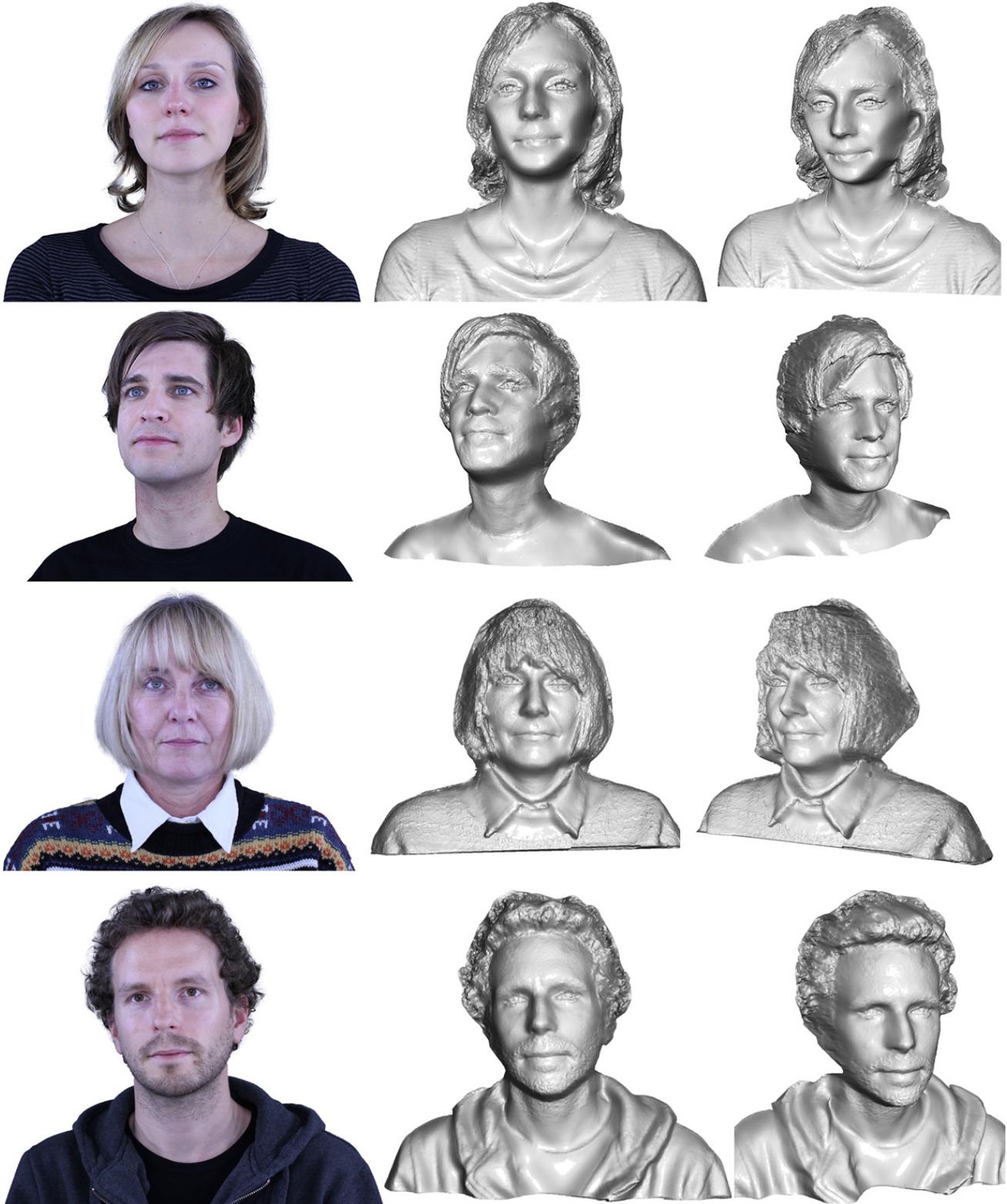


Figure 5. Reconstruction results based on MDIA correspondences from three images. Extrinsic camera parameters and depth were estimated from the computed correspondences. Note that the 3D meshes are inherently depth maps from three more or less frontal views (triangular camera setup,  $\approx 20$  cm baselines). Mesh vertex spacing is five pixels, the initial value for all displacements was 0.

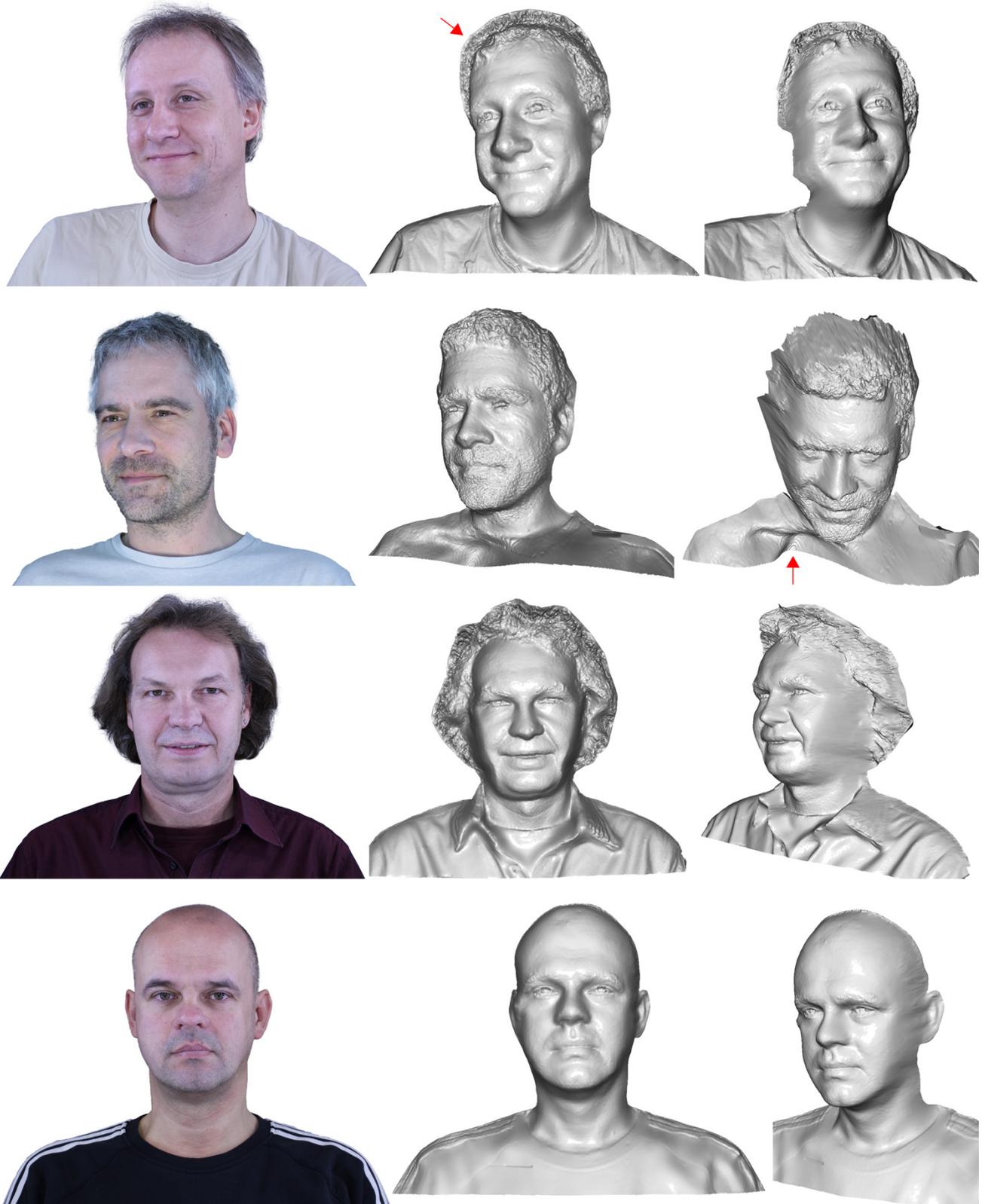


Figure 6. Additional reconstruction results; same parameters as in figure 5. In the first row, artifacts on fine strands of hair are visible. In the second row, convergence failed on the white shirt.

image pairs taken with an upright triangular camera setup (Canon EOS 550D, 35 mm fixed focal, 18 megapixels) and a baseline of roughly 20 cm. The images of the lower left and lower right camera were warped to that of the upper central camera. Although the scene was well lit from multiple sides, luminance correction proved to be crucial for the quality of the results. The images were taken with a relatively neutral background. To avoid distractions from remaining shadows and background structure, they were roughly segmented using a semi-automatic tool. An example of an input image triple is shown in figure 4. Mesh smoothness was controlled only by the smoothness term of the registration algorithm; no extra filtering was applied to the geometry. Mesh vertex spacing was five pixels. Extrinsic camera parameters and depths were computed from the MDIA correspondences. To generate the figures, all vertices on the background were removed from the meshes. For initialization, all vertex displacements were set to zero.

While MDIA converged to very good results on the face for all tested subjects, convergence on people’s clothing sometimes failed due to lack of texture. Figure 6 (second row) shows a failure of convergence on a white shirt. If sufficient texture is available, clothing can be reconstructed in similarly high detail as the face. Results on hair are often surprisingly good, but not as consistently as on the face. Problematic are thin or scattered strands (e.g. figure 6, first row) and strong depth discontinuities.

Note that all meshes shown are inherently depth maps from roughly frontal views as shown in figure 4. For a more elaborate reconstruction, depth maps from other views should be computed and fusion techniques should be employed to obtain the final mesh. This has not been implemented at the time of writing. We believe, however, that the degree of detail in our results sufficiently illustrates the very high quality of correspondences that can be achieved with MDIA and the proposed computational scheme.

While freeform MDIA as an epipolar-agnostic approach has significantly more degrees of freedom than approaches exploiting estimates of epipolar relations, we have shown that it is capable of retrieving excellent dense correspondences, providing highly detailed 3D reconstructions of the human head. These results can be achieved without requiring an initialization close the solution and without any hint on epipolar geometry. We proposed a computational scheme for solving freeform MDIA robustly on very high resolution images, addressing regularization, luminance correction for handling brightness constancy violations and piecewise optimization to cope with limitations of memory and processing power.

## References

[1] A. Bartoli and A. Zisserman. Direct Estimation of Non-Rigid Registrations. In *British Machine Vision Conference*, London, 2004. 48,

49, 50

[2] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-Quality Single-Shot Capture of Facial Geometry. *ACM Transactions on Graphics*, 29(3), 2010. 49, 50

[3] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *ACM SIGGRAPH*, 1999. 50

[4] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High Resolution Passive Facial Performance Capture. *ACM Transactions on Graphics*, 29(3), 2010. 50

[5] J. Condell, B. Scotney, and P. Morrow. Adaptive Grid Refinement Procedures for Efficient Optical Flow Computation. *International Journal of Computer Vision*, 61(1):31–54, 2005. 50

[6] M. Dimitrijevic, S. Ilic, and P. Fua. Accurate Face Models from Uncalibrated and Ill-Lit Video Sequences. In *Computer Vision and Pattern Recognition*, Washington, DC, 2004. 50

[7] Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:1–14, 2008. 50

[8] V. Gay-Bellile, A. Bartoli, and P. Sayd. Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 32(1):87–104, 2010. 48, 49, 50

[9] B. Glocker, T. Heibel, N. Navab, P. Kohli, and C. Rother. Triangle-Flow: Optical Flow with Triangulation-Based Higher-Order Likelihoods. In *European Conference on Computer Vision*, Heraklion, 2010. 50

[10] L. Grady and J. Polimeni. *Discrete Calculus*. Springer, 2010. 51

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003. 49, 52

[12] A. Hilsmann and P. Eisert. 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. 48, 49, 50

[13] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 50

[14] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981. 50

[15] P. McCullagh and J. Nelder. *Generalized Linear Models*. Chapman & Hall, 1998. 50

[16] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly Accurate Optic Flow Computation with Theoretically Justified Warping. *International Journal of Computer Vision*, 67:141–158, 2006. 50, 51

[17] D. C. Schneider, A. Hilsmann, and P. Eisert. Warp-based Motion Compensation for Endoscopic Kymography. In *Eurographics*, Llandudno, 2011. 48, 49

[18] R. Sibson. *Interpreting Multivariate Data*, chapter A Brief Description of Natural Neighbor Interpolation, pages 21–36. John Wiley, 1981. 52

[19] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle Adjustment - A Modern Synthesis. In *Vision Algorithms: Theory and Practice*. Springer, 2000. 49

[20] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and Motion-adaptive Regularization for High Accuracy Optic Flow. In *International Conference on Computer Vision*, Kyoto, 2009. 50

[21] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber-L1 Optical Flow. In *British Machine Vision Conference*, London, 2009. 50

[22] J. Zhu, L. V. Gool, and S. Hoi. Unsupervised Face Alignment by Robust Nonrigid Mapping. In *International Conference on Computer Vision*, 2009. 48, 49

[23] J. Zhu and M. Lyu. A Fast 2D Shape Recovery Approach by Fusing Features and Appearance. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 31(7):1210–1224, 2009. 48, 49