# Tracking Deformable Surfaces with Optical Flow in the Presence of Self Occlusion in Monocular Image Sequences

Anna Hilsmann and Peter Eisert
Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institute Einsteinufer 37, 10587 Berlin, Germany
anna.hilsmann@hhi.fraunhofer.de

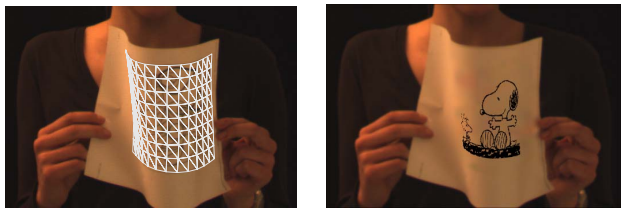
## Abstract

*In this paper, we present a direct method for deformable surface tracking in monocular image sequences. We use the optical flow constraint instead of working with distinct features. The optical flow field is regularized with a 2-dimensional mesh-based deformation model. The formulation of the deformation model contains weighted smoothing constraints defined locally on topological vertex neighborhoods.*
*2-dimensional deformation estimation in the presence of self-occlusion is a very challenging problem. Naturally, a 2-dimensional mesh folds in the presence of self-occlusion. We address this problem by weighting the smoothness constraints locally according to the occlusion of a region. Thereby, the mesh is forced to shrink instead of fold in occluded regions. Occlusion estimates are established from shrinking regions in the deformation mesh. Finding the best transformation then amounts to minimizing an error function that can be solved efficiently in a linear least squares sense.*

## 1. Introduction

The problem of capturing non-rigid motion in monocular image sequences has been addressed in many fields of application including medical imaging [9], object-based video compression [2, 10] or augmented reality [12]. We are particularly interested in images of surfaces whose deformations are difficult to describe, such as drapery of textiles or other surfaces that are bent in a way that they occlude parts of themselves.

One approach is to formulate elastic deformations in 3D [13, 15]. However, recovering 3D position and deformation from monocular surfaces is an ill-posed problem [13]. Another approach is to make use of image based deformation models in 2D. In this case, the deformation model must be able to cope with fold-overs and self occlusions. This problem has been lately addressed by Gay-Bellile *et al*. [3] who penalize a variation in the spatial warp derivative along some direction to prevent mesh-foldings and cope with self-occlusions.

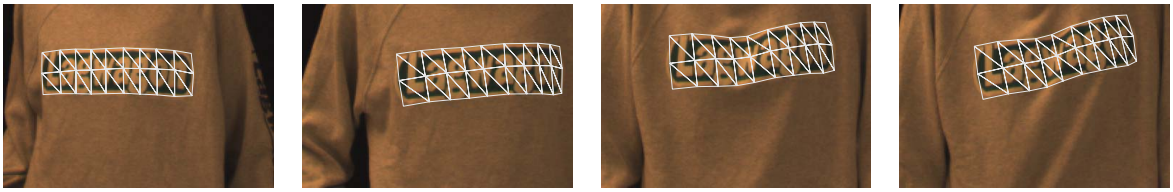


(a) (b)

Figure 1. Deformation estimation and augmentation of a deformed piece of paper under self occlusion and illumination changes.

Inspired by their work, we formulate an optical-flow-based approach to deformable surface tracking using a mesh-based deformation model together with smoothing constraints that force the mesh to shrink instead of fold in presence of self-occlusion. Fitting then amounts to minimizing one error functional consisting of two parts. The first part formulates the error given by the optical flow constraint and the motion model. The second part is a formulation of deformation smoothness. The resulting linear equation system can be solved efficiently in a least squares sense. We explicitly do not penalize deformations with a variation of the sign of partial derivatives. Instead, we allow mesh shrinking from both sides of the occlusion boundary and weight the smoothness constraints locally according to the occlusion of a region to cope with large occlusions. The contribution of this paper is a simpler registration scheme than the one presented in [3] by exploiting topological relationship of the warp. It can be efficiently solved in closed-form due to its linear least squares form.

The remainder of this paper is structured as follows. Section 2 briefly reviews the existing literature on deformable image registration and augmentation of elastic surfaces. In



(a) (b) (c) (d)

Figure 2. Tracking a logo on a shirt. (sequence: 500 frames, 25 fps)

Section 3 we propose our approach to optical flow based deformable surface tracking. Section 4 describes our approach to occlusion handling. In Section 5 we report experimental results on several data sets of deforming surfaces.

## 2. Related Work

In general, the literature distinguishes between feature-based [12], marker-based [14] and image-based [1, 7, 15] deformable surface tracking or registration methods in monocular image sequences.

Scholz and Magnor [14] use color-coded patterns and a-priori knowledge about surface connectivity for tracking of garments in monocular image sequences. Their system can cope with large external and self-occlusions. They determine shading maps by removing the color markers used for tracking and interpolating the deleted regions with a surface reconstruction method for height fields. However, the assumption of such a-priori knowledge is often difficult to achieve in most applications.

Pilet *et al*. [12] propose a feature-based real-time method for deformable object detection and tracking that uses a wide baseline matching algorithm [6] and deformable meshes. Furthermore, they estimate the irradiance of the surface using a reference image with Lambertian illumination. Our tracking approach for deformable surfaces is similar regarding the motion model but we use direct image information instead of distinct keypoints. A feature-based approach might not be suitable for registration in the presence of self-occlusion as there might not be enough feature-points to recover the correct warp.

Generally, image-based methods yield more accurate results in non-rigid deformation estimation. In [1] Bartoli and Zisserman present an optical flow based approach that uses radial basis functions to regularize the flow field. They iteratively insert new center-points for the radial basis functions based on examination of the error image after each iteration. The number of centers grows until the algorithm converges. Lim and Yang [7] also introduce a direct method for recovering non-rigid object motion with radial basis functions. They estimate point correspondences and parameters for the radial basis functions simultaneously in a stiff-to-flexible approach. Torresani *et al*. [15] describe a flow-based method that produces 3D reconstruction from single-view video by exploiting rank constraints on the tracking matrix.

Little research has been done to cope with self-occlusions in 2D deformable image registration and tracking. In [8] Lin and Liu introduce an algorithm for specific regular patterns. They establish visibility maps to deal with occluded regions. Recently, Gay-Bellile *et al*. [3] proposed a direct non-rigid registration method that detects self-occlusion as shrinking areas in the 2D warp. The warp is forced to shrink by penalizing a variation in the sign of the partial derivatives the warp along some direction. A binary decision excludes self-occluded pixels from consideration in the error function.

## 3. Optical Flow Based Deformable Surface Tracking

In order to track a deformable surface in a video sequence we exploit the optical flow constraint [5] equation along with a predefined motion model. Finding the best transformation then amounts to minimizing a quadratic error functional:

$$E = \sum_{i=1}^{n} \left( \nabla I\left(x_i, y_i\right) \cdot \mathbf{d}\left(x_i, y_i\right) + \frac{\partial I}{\partial t}\left(x_i, y_i\right) \right)^2 \quad (1)$$

where $\nabla I\left(x_i, y_i\right)$ denotes the spatial derivatives of the image $I$ at pixel position $\left[x_i, y_i\right]^T$ and $\frac{\partial I}{\partial t}\left(x_i, y_i\right)$ denotes the temporal gradient between two images. $\mathbf{d}\left(x_i, y_i\right)$ denotes the displacement vector at position $\left[x_i, y_i\right]^T$ and is defined by the motion model described below. $n$ is the number of pixels selected for contribution to the error function, i.e. pixels where the gradient is non-zero.

We present our deformable motion model M as a planar triangulated regular 2D mesh with $K$ vertices $\mathbf{v_k}, (k = 1...K)$. The position of each vertex $\mathbf{v_k}$ is given by its image coordinates $\left[x_k, y_k\right]^T$. Each pixel $\mathbf{p_i} = \left[x_i, y_i\right]^T$ in the image can be represented by its barycentric coordinates of

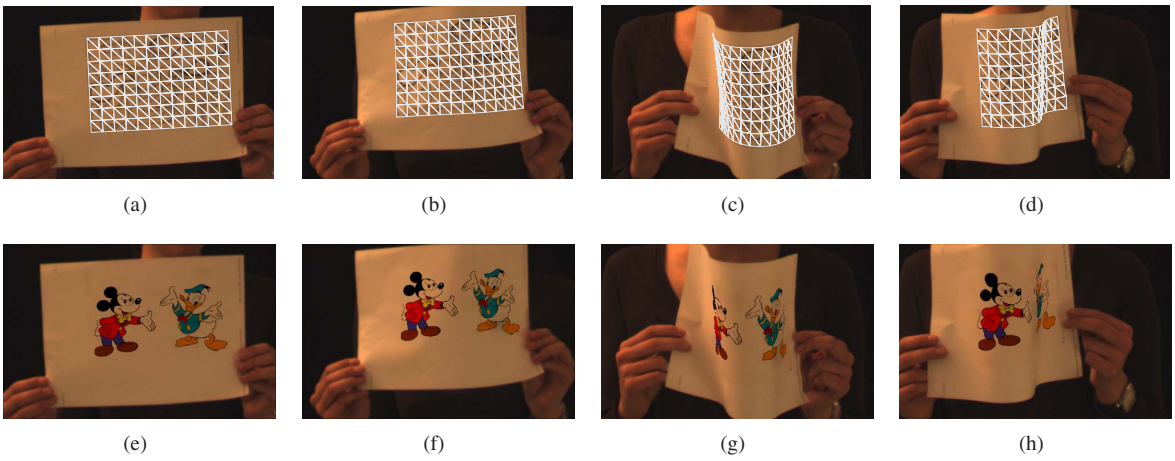


Figure 3. Tracking and augmenting a deforming piece of paper under self occlusion and illumination changes. Illumination is recovered from the deformation estimation assuming Lambertian illumination in the reference frame ([11]). The new texture is deformed and illuminated with the recovered shading map. (sequence: 500 frames, 25 fps)

its enclosing vertices.

$$[x_i, y_i]^T = \sum_{\substack{j=1 \\ \mathbf{v_j} \in \mathbf{v_k}}}^{3} B_j(x_i, y_i) \cdot \mathbf{v_j}$$
$$\sum_{j=1}^{3} B_j(x_i, y_i) = 1,\ 0 \leq B_j \leq 1 \tag{2}$$

where $B_j(x_i, y_i), (j = 1, 2, 3)$ are the three barycentric coordinates of pixel $[x_i, y_i]^T$ computed on the original mesh and $\mathbf{v_j}$ are the three vertices of the enclosing triangle of the mesh M. With a deformation of the mesh, $[x_i, y_i]^T$ is mapped onto $[x_i', y_i']^T$. Thus, we are looking for a deformation of the mesh, i.e. a displacement $\mathbf{d_k} = (d_{kx}, d_{ky})$ of each vertex $\mathbf{v_k}$ to $\mathbf{v_k'}$ such that the barycentric coordinates of $[x_i', y_i']^T$ equal those of $[x_i, y_i]^T$. Details can be found in [4]. Hence, the deformation model can be stated as:

$$\mathbf{d}(x_i, y_i) = \sum_{j=1}^{3} B_j(x_i, y_i) \cdot \mathbf{d_j}(x_i, y_i) \tag{3}$$

where $\mathbf{d_j}(x_i, y_i)$ are the three vertex displacements of the enclosing triangle. Inserting the motion model (3) into equation (1) leads to an overdetermined linear equation system. Incorporating additional smoothing constraints for the vertex displacement field yields the following error functional to be minimized:

$$E = \sum_{i=1}^{n} \left( \nabla I(x_i, y_i) \cdot \mathbf{d}(x_i, y_i) + \frac{\partial I}{\partial t}(x_i, y_i) \right)^2 + \lambda \sum_{j=1}^{m} w_j E_s(\mathbf{d_j}) \tag{4}$$

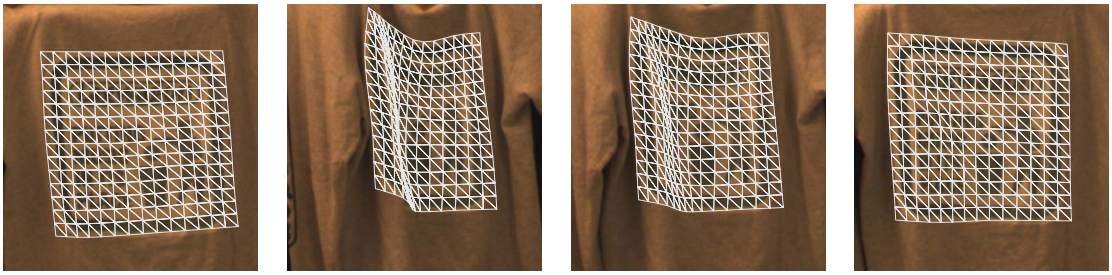
where $m$ is the number of inner vertices and $\lambda$ is the regularization parameter. The first term represents the data term of the transformation whereas the second term is a smoothing constraint for the mesh deformation field. $E_s(\mathbf{d_j})$ is a local smoothing function for the displacement $\mathbf{d_j}$ of an inner vertex $\mathbf{v_j}$ between two successive frames weighted by $w_j$. We chose $E_s(\mathbf{d_j})$ to be

$$E_s(\mathbf{d_j}) = \sum_{i=1}^{3} \left( \mathbf{d_j} - \frac{1}{D} (\mathbf{d_{ji1}} + \mathbf{d_{ji2}}) \right)^2 \tag{5}$$

where $\mathbf{d_{ji1}}$ and $\mathbf{d_{ji2}}$ denote the two neighbor vertex displacements of an inner vertex $\mathbf{v_j}$ in the topological direction $i$ in a regular triangle mesh, i.e. the horizontal, the vertical and the diagonal direction (see e.g. the regular triangle mesh in Figure 3). $E_s(\mathbf{d_j})$ is a measure of vertex displacement deviation to the displacements of its neighbors. It serves a dual purpose. Firstly, it regularizes the deformation field of the vertices. Secondly, defining it locally on the mesh allows us to weight it according to the occlusion of a region. This penalizes fold overs and forces the mesh to shrink instead of fold as this results in a smoother deformation field.

The weight $w_j$ is taken from the occlusion map described in Section 4 for each vertex. Thereby, displacements of occluded vertices are more restricted by the smoothing constraint than dis-occluded vertices and mesh foldings are prevented.

The optical flow constraint is valid only for small displacements because it assumes the image intensity to be linear between two successive frames. In order to cope with larger displacements we use a hierarchical framework.



(a) (b) (c) (d)

Figure 4. Folding and unfolding a shirt. The mesh shrinks along the occlusion boundary and recovers correctly when the shirt is unfolded. (sequence: 250 frames, 25 fps)

## 4. Self-Occlusion Handling

In order to handle large occlusions we estimate an occlusion map from shrinking areas in the deformed mesh. For each vertex $\mathbf{v_j}$ in the mesh we calculate the average distance to its vertical and horizontal neighbors and scale it by the initial vertex distance in the reference mesh:

$$D_j = \frac{1}{2\,|N_{vj}|\,D_v} \sum_{n \in N_{vj}} \|\mathbf{v_j} - \mathbf{v_n}\|_2 + \frac{1}{2\,|N_{hj}|\,D_h} \sum_{n \in N_{hj}} \|\mathbf{v_j} - \mathbf{v_n}\|_2 \tag{6}$$

where $N_{vj}$ and $N_{hj}$ are the vertical and horizontal neighborhoods of vertex $\mathbf{v_j}$ and $D_v$ and $D_h$ denote the initial vertical and horizontal distances between two neighboring vertices in the regular reference mesh. By interpolating the average distances that present local mesh shrinking estimates over the entire surface we can establish an estimate of occluded regions in an occlusion map. Figure 5 shows examples of occlusion estimates for different kinds of self-occlusion. The left column shows the original images with the deformation mesh. In the middle the deformed occlusion map is shown while the right column depicts the occlusion map of the undeformed surface. The occlusion maps are used to adapt the weight $w_j$ in equation (4) for vertex $\mathbf{v_j}$ to the degree of its occlusion. Vertices in occluded regions are assigned a higher weight to the smoothness constraint than vertices in dis-occluded regions, i.e. the smaller the distance of a vertex to its neighbors the more its displacement is constrained by the surrounding displacements. Hereby, we prevent foldings because vertices in mesh shrinking regions are forced to behave like their neighbors. Additionally, when unfolding the surface the increased smoothing weight in occluded regions causes the shrunk (i.e. occluded) region of the mesh to be stretched by the vertices at the occlusion boundary whose displacements are constrained by the optical flow equation. We chose $w_j$ to be

$$w_j = \begin{cases} \frac{1}{D'_j} & \text{for } D'_j < 0.1 \\ 1, & \text{else} \end{cases} \tag{7}$$

where $D'_j$ equals $D_j$ after the vector of all $D_j$ is normalized so that the maximum of all $D_j$ is one. Hereby, we do not adapt the weight to the smoothness constraint if the mesh expands or shrinks uniformly, e.g. due to a movement toward or away from the camera.

## 5. Experimental Results

We applied our approach on several video sequences with different kinds of surfaces (paper, fabric) and different kinds of deformations and self-occlusions.

**T-Shirt Logo Sequence.** Figure 2 shows example frames of tracking a logo on a t-shirt when a person moves. The sequence contains elastic deformations that are due to stretching, small 3D rotation and extension when the person moves toward the camera. The logo on the shirt is tracked correctly in all frames.

**T-Shirt Folding Sequence.** Figure 4 depicts frames of a sequence showing the folding and unfolding of a shirt with self occlusions due to folding. The mesh shrinks along the occlusion boundary when the shirt is folded and recovers correctly when it is unfolded. Figure 5(a) - 5(c) depict the estimated occlusions for one frame of the sequence in an occlusion map of the deformed and the undeformed (i.e. the reference) surface. In Figure 5(b) the correctly estimated occlusion boundary can be seen in the occlusion map.

**Paper Sequences.** Figure 1, 3 and 6 show examples of sequences showing different kinds of paper foldings. Occlusions due to paper folding in the middle of the paper (Figure 3(d) and 6(a)) are as well registered as foldings at the surface boundary (Figure 3(c) and 6(d)). Normally, in pure 2D deformable surface tracking mesh foldings appear in the presence of self-occlusion. This behavior is undesirable as it can lead to mis-registration and e.g. augmentation of the surface fails. The bottom row of Figure 3 and Figure 1(b)

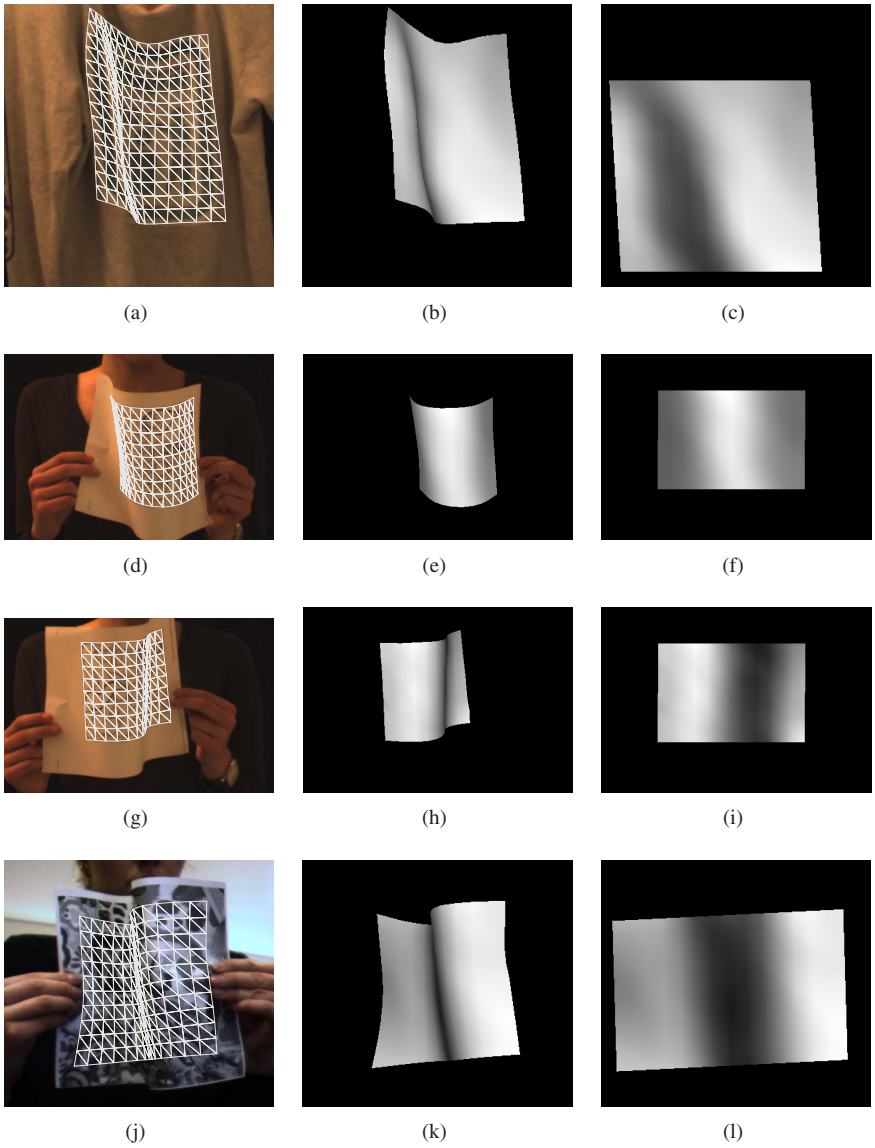


Figure 5. Examples for occlusion maps. Left column: deformation mesh mapped on the deforming fabric, middle column: occlusion map of the deformed surface, right column: occlusion map of the undeformed surface. Dark colors mark occluded regions.

show augmentation results of a paper bending sequence. Our approach for illumination estimation is similar to the one presented in [11]. Irradiance is recovered from a white normalization and the 2D deformation estimation assuming the reference frame to be illuminated uniformly. With our approach the new texture is deformed correctly in occluded regions. Figure 6 depicts two example frames from paper bending sequences with self-occlusions and compares two results with and without our occlusion handling approach, i.e. $w_j = 1$ for all vertices. The left column shows the original frame and the deformed mesh. The middle column depicts details of the deformed mesh using our approach to occlusion handling. The mesh shrinks at the occlusion boundary whereas without our approach it folds in occluded regions (see right column).

## 6. Conclusion

We presented a direct approach to deformable surface tracking in monocular image sequences taking into account the important issue of self-occluded regions. Our method is based on the optical flow constraint equation that is regularized by a mesh-based deformation model. Smoothing constraints on the deformation field are formulated locally on the mesh. Occlusion maps are estimated from lo-

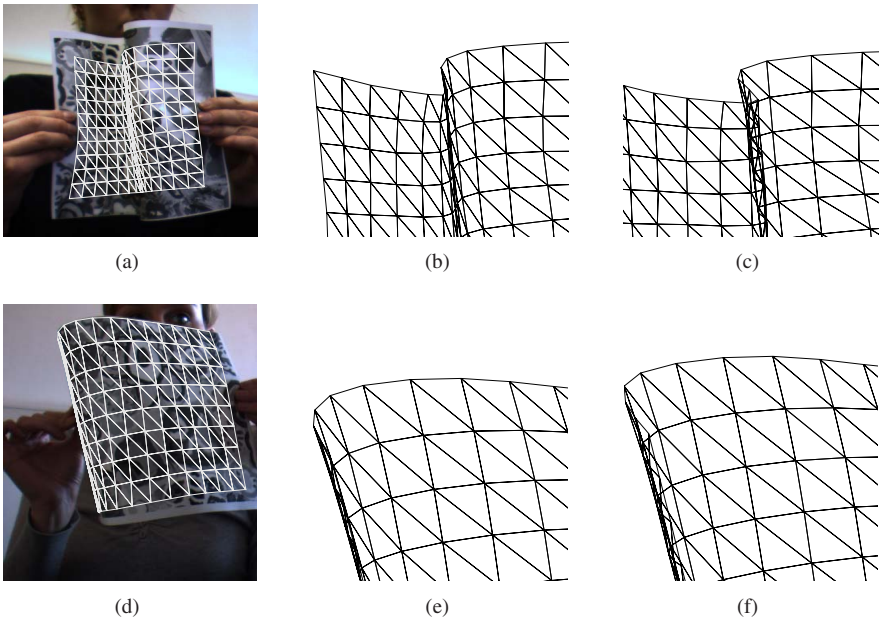



Figure 6. Occlusion handling. Left: original image, middle: detail of the deformed mesh with occlusion handling, right: detail of the deformed mesh without occlusion handling.

cal mesh shrinking properties and allow us to weight the smoothing constraints locally according to the occlusion estimate. Thereby, we prevent mesh-foldings in self-occluded regions. Experimental results on real image sequences of different kinds of deforming surfaces show that our approach successfully tracks the deforming surfaces in the presence of self-occlusions.

## References


[1] A. Bartoli and A. Zisserman. Direct estimation of non-rigid registrations. In *Proc. British Machine Vision Conf. (BMVC 2004)*, 2004. 2

[2] P. Eisert. MPEG-4 facial animation in video analysis and synthesis. *Int. Journal of Imaging Systems and Technology*, 13(5):245–250, March 2003. 1

[3] V. Gay-Bellile, A. Bartoli, and P. Sayd. Direct estimation of non-rigid registrations with image-based self-occlusion reasoning. In *Proc. Int. Conf. on Computer Vision (ICCV 2007)*, pages 1–6, 2007. 1, 2

[4] A. Hilsmann and P. Eisert. Deformable object tracking using optical flow constraints. In *Proc. 4th Int. Conf. on Visual Media Production (CVMP 2007)*, Nov. 2007. 3

[5] B. Horn and B. G. Schunck. Determining optical flow. Technical report, Cambridge, MA, USA, 1980. 2

[6] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, pages 775–781, 2005. 2

[7] J. Lim and M.-H. Yang. A direct method for modeling non-rigid motion with thin plate spline. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, volume 1, pages 1196–1202, 2005. 2

[8] W.-C. Lin and Y. Liu. Tracking dynamic near-regular textures under occlusion and rapid movements. In *Proc. European Conf. on Computer Vision (ECCV 2006)*, May 2006. 2

[9] D. Metaxas. *Physics-Based Deformable Models: Applications to Computer Vision, Graphics, and Medical Imaging*. Kluwer Academic Publishers, 1996. 1

[10] J. Ostermann. Object-oriented analysis-synthesis coding (ooasc) based on the source model of moving flexible 3D-objects. *IEEE Trans. on Image Processing*, 3(5):705–711, Jan. 1994. 1

[11] J. Pilet, V. Lepetit, and P. Fua. Augmenting deformable objects in real-time. In *Int. Symposium on Mixed and Augmented Reality*, Vienna, Austria, October 2005. 3, 5

[12] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *Int. Journal of Computer Vision*, Jan. 2007. 1, 2

[13] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for non-rigid 3–d shape recovery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(8):1481–1487, August 2007. 1

[14] V. Scholz and M. Magnor. Texture replacement of garments in monocular video sequences. In *Rendering Techniques 2006: Eurographics Symposium on Rendering*, pages 305–312, June 2006. 2

[15] L. Torresani, D. Yang, E. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2001)*, 2001. 1, 2