

OPTICAL FLOW BASED TRACKING AND RETEXTURING OF GARMENTS

Anna Hilsmann, Peter Eisert

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute
Einsteinufer 37, 10587 Berlin, Germany
Email: {anna.hilsmann, eisert}@hhi.fraunhofer.de

ABSTRACT

In this paper, we present a method for tracking and retexturing of garments that exploits the entire image information using the optical flow constraint instead of working with distinct features. In a hierarchical framework we refine the motion model with every level. The motion model is used to regularize the optical flow field such that finding the best transformation amounts in minimizing an error function that can be solved in a least squares sense. Knowledge about the position and deformation of the garment in 2D allows us to erase the old texture and replace it by a new one with correct deformation and shading properties without 3D reconstruction. Additionally, it provides an estimation of the irradiance such that the new texture can be illuminated realistically.

Index Terms— Optical Flow, Garment Tracking, Deformable Meshes, Virtual Clothing

1. INTRODUCTION

The problem of capturing non-rigid motion has been addressed in many fields of application including medical imaging [1], object-based video compression [2], [3] or augmented reality [4]. It has been successfully demonstrated for human motion analysis and face tracking. Automated tracking of 3D deformations of garments in monocular video sequences is challenging because the complexity of the deformation field is unknown a priori and therefore the assumed deformation model has to cope with various types of deformations.

Cloth tracking in particular has been addressed by a number of researchers. Previous work has focused on color-coded patterns [5], [6], [7] for multiview systems [8], [5], [9], [6], [7]. Some researches also proposed methods for arbitrary textures [8], [4] using feature-based approaches or for monocular sequences [4], [10]. Little research has been done in the field of optical flow based garment tracking [9], [11].

Guskov *et al.* [5] introduced a multiview real-time system based on color-coded markers with a limited size of codewords. Their system fails for large motions and the markers have to be quite large. Scholz *et al.* [6] improved upon Guskov *et. al* by using a color-code with more codewords that can cope with fast motion. It makes use of the a priori knowledge of surface connectivity and color coded pattern. In [10] they used the same color code for tracking of garments in monocular video sequences. White *et al.* [7] also used color markers and multiple synchronized cameras.

In [8] Pritchard *et al.* introduced a feature-based approach to cloth motion capturing with a calibrated stereo camera pair to acquire 3D structure. They use SIFT features to establish correspondences. Pilet *et al.* [4] proposed a feature-based real-time method

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET II (Contract IST-1-038398).

for deformable object detection and tracking that uses a wide baseline matching algorithm [12] and deformable meshes. Our approach is similar to theirs referring to the motion model but we use direct image information instead of distinct keypoints.

Torresani *et al.* [11] describe a flow-based method that produces 3D reconstruction from single-view video by exploiting rank constraints on optical flow. In [9] Scholz *et al.* presented an optical flow based multicamera system where optical flow tracking is combined with silhouette matching. In [13] we demonstrated a method to track deformations of a shirt in a pre-segmented binary video sequence using the optical flow and deformable meshes with weighted temporal and spatial smoothing constraints.

The main contribution of our tracking method is a formulation of an optical-flow-based error function together with a predefined mesh-based motion model that can be easily minimized in a linear least-squares sense. Using direct image information yields more accurate results than distinct features or markers. Furthermore, we use the estimated motion field to retexture garments in a very simple and efficient way.

The remainder of this paper is structured as follows. In Section 2 we present our optical flow based tracking algorithm. Section 3 briefly describes our method for retexturing of garments before we present our results in Section 4.

2. OPTICAL FLOW BASED GARMENT TRACKING

In order to track an object in a video sequence we exploit the optical flow constraint equation along with a predefined motion model. Finding the best transformation then amounts to minimizing a quadratic error that can be solved in a least-squares sense:

$$E = \sum_{i=1}^n \|I_x(x_i, y_i) \cdot d_x(x_i, y_i) + I_y(x_i, y_i) \cdot d_y(x_i, y_i) + I_t(x_i, y_i)\|_2 \quad (1)$$

where $I_x(x_i, y_i)$ and $I_y(x_i, y_i)$ are the spatial derivatives of the image at pixel position $[x_i, y_i]^T$ and $I_t(x_i, y_i)$ denotes the intensity change between two images and n is the number of pixels selected for contribution to the error function, i.e. pixels where the gradient is non-zero.

The optical flow equation has some constraints that have to be fulfilled. First, it is valid only for small displacements between two successive frames because it is derived assuming the image intensity to be linear. In order to cope with larger displacements we use a hierarchical framework. Second, it assumes uniform lambertian illumination. Illumination changes are taken into account by using bandpass-filtered images.

On each pyramid level we present our model M as a triangulated

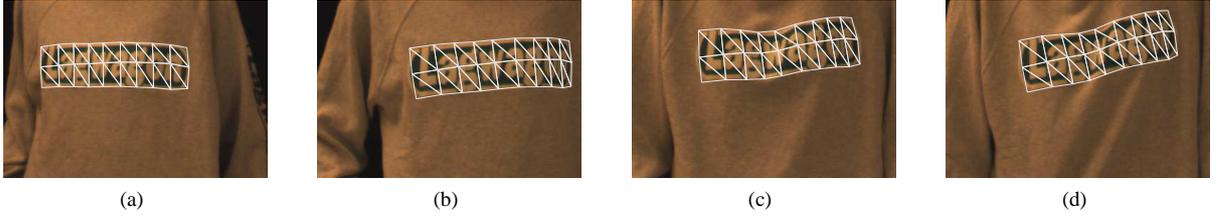


Fig. 1. Tracking a logo on a shirt. (sequence: 500 frames, 25 fps)

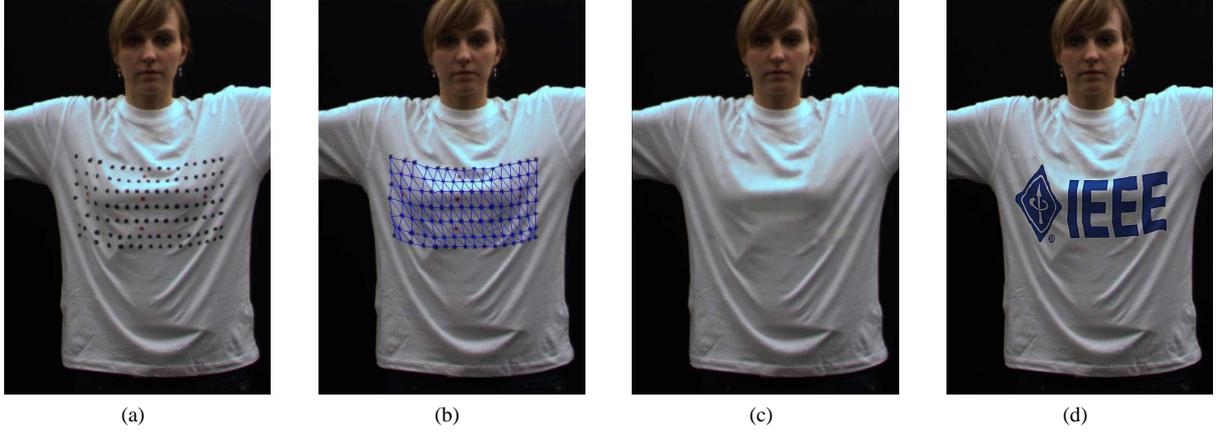


Fig. 2. Input image (a), position and deformation of the deformable mesh (b), removed dots (c), retexturing with correct deformation and shading of the new texture

regular 2D mesh with m vertices \mathbf{v}_k , ($k = 1..m$). The position of each vertex \mathbf{v}_k is given by its image coordinates $\mathbf{p}_k = [x_k, y_k]^T$. Each pixel $\mathbf{p}_i = [x_i, y_i]^T$ in the image can be represented by its barycentric coordinates of its enclosing triangle t .

$$[x_i, y_i]^T = \sum_{\substack{j=1 \\ \mathbf{v}_{jt} \in \mathbf{v}_k}}^3 B_{jt}(x_i, y_i) \cdot [v_{jtx}, v_{jty}]^T \quad (2)$$

$$\sum_{j=1}^3 B_{jt}(x_i, y_i) = 1, 0 \leq B_{jt} \leq 1$$

where $B_{jt}(x_i, y_i)$, ($j = 1, 2, 3$) are the three barycentric coordinates of pixel $[x_i, y_i]^T$ computed on the original mesh and \mathbf{v}_{jt} are the three vertices of the enclosing triangle t of the mesh consisting of vertices \mathbf{v}_k . With a mesh deformation, $[x_i, y_i]$ is mapped onto $[x'_i, y'_i]$. Thus, we are looking for a deformation of the mesh, i.e. a displacement of each vertex \mathbf{v}_k to \mathbf{v}'_k such that the barycentric coordinates of $[x'_i, y'_i]^T$ are those of $[x_i, y_i]^T$.

Thus, the deformation model can be stated as [13]:

$$\mathbf{d}(x_i, y_i) = \sum_{\substack{j=1 \\ t_j \in t_k}}^3 B_j(x_i, y_i) \cdot \mathbf{d}_j(x_i, y_i) \quad (3)$$

where $\mathbf{d}_j(x_i, y_i)$ are the three vertex displacements of the enclosing triangle. Inserting the motion models into equation (1) leads to an overdetermined linear equation system that is solved in a linear least-squares sense. We incorporate additional spatial and temporal

smoothing constraints for the vertex displacements such that now the error

$$E = \sum_{i=1}^n \|I_x(x_i, y_i) \cdot d_x(x_i, y_i) + I_y(x_i, y_i) \cdot d_y(x_i, y_i) + I_t(x_i, y_i)\|_2 + \lambda \sum_{k=1}^m \|\mathbf{d}_k\|_2 + \sum_{k=1}^m w_k E_s(\mathbf{d}_k) \quad (4)$$

is minimized. The first term represents the fidelity of the transformation whereas the second term is a temporal smoothing constraint. $E_s(\mathbf{d}_k)$ is some spatial smoothing function for the vertex displacement \mathbf{d}_k between two successive frames locally weighted by w_k . We chose $E_s(\mathbf{d}_k)$ to be

$$E_s(\mathbf{d}_k) = \left| d_{kx} - \frac{1}{|N_k|} \sum_{n \in N_k} d_{nx} \right| + \left| d_{ky} - \frac{1}{|N_k|} \sum_{n \in N_k} d_{ny} \right| \quad (5)$$

where N_k denotes the set of neighbor vertices of vertex \mathbf{v}_k , i.e. all vertices \mathbf{v}_k is connected to. $E(\mathbf{d}_k)$ is a measure of vertex displacement deviation to the displacements of its neighbors. Details on the

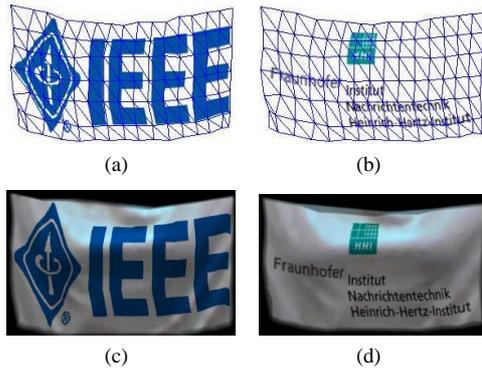


Fig. 3. Deformed texture (a,b) and correct shading of the deformed texture (c,d). The texture is correctly deformed at main wrinkles and cloth folds.

smoothing constraints can be found in [13]. The weight w_k can be e.g. adapted to the degree of ‘constrainedness’ of a vertex \mathbf{v}_k by the optical flow, e.g. the number of image points that influence its displacement in the optical flow equation. We chose w_k to be reciprocal to the norm of all linear factors of d_{kx} and d_{ky} in the linear equation system we obtain from equation (1) and (2)

$$w_k = K \cdot \frac{1}{\|\alpha_{kx}\|_2 + \|\alpha_{ky}\|_2} \quad (6)$$

where α_{kx} and α_{ky} are the vectors of linear factors of d_{kx} and d_{ky} in the linear equation system. The less pixels contribute to the vertex displacement the higher the weight w_k to the additional smoothness constraint and the more the vertex displacement is forced to equal the surrounding displacements. The factor K controls the trade-off between mesh-smoothness and mapping accuracy. To account for inaccuracies from downsampling and filtering in the image pyramid we use large values for K in levels with low resolutions and small values in levels with high resolutions, i.e. in lower levels we estimate a smooth mesh transformation representing a global transformation and admit more detailed deformations in higher image levels.

3. RETEXTURING

The tracking method described in the previous section can be applied to arbitrary surfaces that are rich enough in detail to exploit optical flow constraints. We applied to method to a shirt with a dot-pattern that is known a priori, as depicted in Fig. 2(a). This allows us to establish a shading map by erasing all dots and interpolating the deleted regions while preserving shading effects of wrinkles and fold overs. As we know the position of each dot from the tracking process, we can easily detect and remove all dots on the shirt. The dot positions known from the tracking process are dilated with a disk-shaped morphological structure element in order to remove the intensity transitions at the boundary.

As the shirt is white, we can interpolate the deleted regions in RGB color space and use the resulting image as a shading map, that is applied to the new texture by multiplication. Interpolating the RGB channels instead of working with the luminance allows us not only to recover wrinkles and folds but also lighting effects and color shifts (see Fig. 3).

For interpolation we use thin-plate splines with control points in a bounding box around each detected dot. Thin-plate splines are a

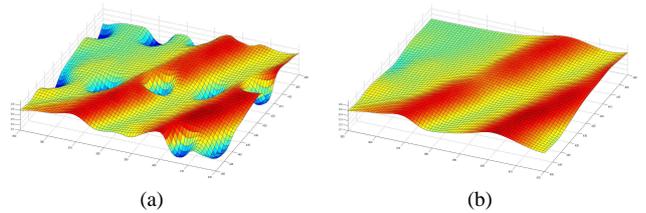


Fig. 4. Representation of intensities as height fields before (a) and after (b) dot removal and interpolation. The interpolation result yields a smooth shading map and preserves main folds and wrinkles

fundamental solution to the biharmonic function [14] meaning that they possess minimum bending energy and yield a smooth and visually pleasing result. Fig. 4 depicts a region of the shirt represented as a height field before (4(a)) and after (4(b)) the region of the dots has been interpolated. The resulting shading map is smooth while main wrinkles are preserved.

4. RESULTS

We apply the method described in the previous sections to a video sequence of 100 frames tracking the elastic deformations of a dot pattern on a shirt (see Fig. 5). The video is recorded at 25 frames per second with a resolution of 1024×768 pixels. Fig. 5(a)-Fig. 5(d) show example frames of the video sequence with the deformable mesh overlaid on the pattern to display tracking accuracy. From Fig. 5(d) we see that even if a point becomes invisible because of wrinkles or fold overs (left point column, third point from the bottom) the smoothing constraints preserve the mesh smoothness and the mesh does not lose track. However, the tracking method described in Section 2 is not limited to the dot pattern used here. It can also be applied to arbitrary logos on shirts that are rich enough in detail to exploit optical flow [13], [15] (see Fig. 1). In these cases other retexturing methods have to be applied [4], [15].

Fig. 3(a) and 3(b) show two close ups of deformed textures with the deformable mesh overlaid on the texture. Fig. 3(c) and 3(d) display the two deformed textures with correct shading. Our motion model correctly represents main wrinkles and cloth folds. Applying the shading maps to the new texture enhances the realistic impression even for smaller wrinkles.

5. CONCLUSION

We presented a method to track elastic deformations of garments based on optical flow and deformable meshes in single view video sequences. The system can cope with arbitrary textures that are rich enough in detail to exploit optical flow constraints. We apply the method to a shirt with a dot pattern which allows us to erase the tracked dots and establish shading maps by interpolating the deleted regions. The shading maps are applied to a new texture by multiplication in order to increase the realistic impression.

Future work will concentrate on further motion models and additional constraints that can cope with a wider range of fold overs and occlusions. The objective is to design a real-time deformable surface tracking system using faster solvers for the equation systems in the tracking and shading interpolation steps. This system can be used in Virtual Mirror scenarios.



Fig. 5. Example frames from the video with correct mesh position and deformation (a-d) and possible variations of retexturing (e-h)

6. REFERENCES

- [1] D.N. Metaxas, *Physics-Based Deformable Models: Applications to Computer Vision, Graphics, and Medical Imaging*, Kluwer Academic Publishers, 1996.
- [2] P. Eisert, “MPEG-4 facial animation in video analysis and synthesis,” *Int. Journal of Imaging Systems and Technology*, vol. 13, no. 5, pp. 245–250, March 2003.
- [3] J. Ostermann, “Object-oriented analysis-synthesis coding (ooasc) based on the source model of moving flexible 3D-objects,” *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 705–711, Jan. 1994.
- [4] P. Pilet, V. Lepetit, and P. Fua, “Fast non-rigid surface detection, registration and realistic augmentation,” *Int. Journal of Computer Vision*, Jan. 2007.
- [5] I. Guskov, S. Klivanov, and B. Bryant, “Trackable surfaces,” in *Proc. 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA 2003)*, 2003, pp. 251–257.
- [6] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor, “Garment motion capture using color-coded patterns,” *Computer Graphics Forum (Proc. Eurographics EG’05)*, vol. 24, no. 3, pp. 439–448, 2005.
- [7] R. White, K. Crane, and D. Forsyth, “Capturing and animating occluded cloth,” in *ACM Transactions on Graphics (SIGGRAPH)*, 2007.
- [8] D. Pritchard and W. Heidrich, “Cloth motion capture,” in *Eurographics*, 2003, pp. 263–271.
- [9] V. Scholz and M. Magnor, “Cloth motion from optical flow,” in *Proc. Vision, Modeling and Visualization*, 2004.
- [10] V. Scholz and M. Magnor, “Texture replacement of garments in monocular video sequences,” in *Rendering Techniques 2006: Eurographics Symposium on Rendering*, June 2006, pp. 305–312.
- [11] L. Torresani, D. Yang, E. Alexander, and C. Bregler, “Tracking and modeling non-rigid objects with rank constraints,” in *Proc. Computer Vision and Pattern Recognition (CVPR’01)*, 2001.
- [12] V. Lepetit, P. Laguerre, and P. Fua, “Randomized trees for real-time keypoint recognition,” in *Proc. Computer Vision and Pattern Recognition (CVPR’05)*, 2005, pp. 775–781.
- [13] 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.
- [14] G. Wahba, “Spline Models for Observational Data,” *Soc. Ind. Appl. Math.*, Philadelphia, 1990.
- [15] A. Hilsmann and P. Eisert, “Tracking deformable surfaces with optical flow in the presence of self occlusions in monocular image sequences,” in *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, Anchorage, Alaska, USA, June 2008.