

# Deformable Template Tracking in 1ms

David Joseph Tan<sup>1</sup>

tanda@in.tum.de

Stefan Holzer<sup>1</sup>

holzers@in.tum.de

Nassir Navab<sup>1</sup>

navab@in.tum.de

Slobodan Ilic<sup>2</sup>

Slobodan.Ilic@siemens.com

<sup>1</sup> CAMP

Technische Universität München

Munich, Germany

<sup>2</sup> Siemens AG

Munich, Germany

The goal of template tracking is to estimate the transformation parameters that define the motion of the planar template. Typically, the transformation parameters encode linear transformation based on homographies with 8 degrees of freedom, which allows them to track rigid motions of the template. However, when it comes to tracking surfaces that undergo non-rigid deformations, deformable transformations with higher degrees of freedom must be used.

Most works on deformable template tracking rely on feature points [1, 5, 6, 7, 8]. It follows the traditional algorithm where it detects feature points on the current frame; finds feature point correspondences between the current frame and the template; remove outliers from these correspondences; and, estimate the deformation. However, since deformable models have a higher degrees of freedom, it becomes more difficult to detect outliers. As a result, it requires a longer runtime. Therefore, this paper aims to address the problem of real-time deformable template tracking.

Instead of using tracking-by-detection approaches with feature points, our work focuses on a frame-to-frame tracking approach with a dense pixel arrangement on the template. Hence, to track the template, we use the linear predictor  $\mathbf{A}$  which establishes a linear relation between the vector of image intensity differences  $\delta\mathbf{i}$  of a template and the corresponding template transformation parameters  $\delta\boldsymbol{\mu}$ , which is written as [3]:

$$\delta\boldsymbol{\mu} = \mathbf{A}\delta\mathbf{i}. \quad (1)$$

The main benefit of using dense pixel intensities is that a lack of a large number of feature points is compensated by the dense pixel information and, thus, allows tracking of less textured surfaces such as faces.

Up to this work, linear predictors have only been used to handle linear transformations such as homographies to track planar surfaces. In this paper, we introduce a method to learn non-linear template transformations that allows us to track surfaces that undergo non-rigid deformations. These deformations are mathematically modelled using 2D Free Form Deformations (FFD) with cubic B-Splines [2, 4]. It uses control points that are uniformly arranged around the template such that the deformation of the template is modelled by the displacement of the control points. In this way, the transformation parameters in  $\delta\boldsymbol{\mu}$  is defined by the displacements of the control points.

Linear predictors are learned using a dataset of  $n_\omega$  images. Each image is a deformed version of the template, where random movements are induced on its control points. These movements correspond to the change in the parameter vector  $\delta\boldsymbol{\mu}$ . Using FFD, the location of the sample points are deformed that creates the image intensity differences  $\delta\mathbf{i}$ . Therefore, we can concatenate the vectors from  $\{(\delta\mathbf{i}_\omega, \delta\boldsymbol{\mu}_\omega)\}_{\omega=1}^{n_\omega}$  to construct the matrices  $\mathbf{Y} = [\delta\boldsymbol{\mu}_1, \delta\boldsymbol{\mu}_2, \dots, \delta\boldsymbol{\mu}_{n_\omega}]$  and  $\mathbf{H} = [\delta\mathbf{i}_1, \delta\mathbf{i}_2, \dots, \delta\mathbf{i}_{n_\omega}]$  with the relation  $\mathbf{Y} = \mathbf{A}\mathbf{H}$ , and learn the linear predictor  $\mathbf{A}$  using [3]:

$$\mathbf{A} = \mathbf{Y}\mathbf{H}^\top (\mathbf{H}\mathbf{H}^\top)^{-1}. \quad (2)$$

The simplicity of our approach allows us to track deformable surfaces at extremely high speed of approximately 1 ms per frame with a single core of the CPU, which has never been shown before.

To evaluate our algorithm, we perform an extensive analysis of our method's performance on synthetic and real sequences with different types of surface deformations. In addition, we compare our results from the real sequences to the feature-based tracking-by-detection method [5], and show that the tracking precisions are similar but our method performs 100 times faster.

Our *Supplementary Material* includes a video that shows quantitative and qualitative results to demonstrate our tracking performance under different deformations and in low-lighting condition as illustrated in Fig. 1.



Figure 1: These images are exemplary examples of tracking a template.

- [1] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 2003.
- [2] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis*, 2008.
- [3] F. Jurie and M. Dhome. Hyperplane approximation for template matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.
- [4] S. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 1997.
- [5] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision*, 2008.
- [6] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 2012.
- [7] J. Zhu and M. Lyu. Progressive finite newton approach to real-time nonrigid surface detection. In *Conference on Computer Vision and Pattern Recognition*, 2007.
- [8] J. Zhu, S. Hoi, and M. Lyu. Nonrigid shape recovery by gaussian process regression. In *Conference on Computer Vision and Pattern Recognition*, 2009.