Chapter 10
Feature Tracking for Mesh-Based Performance Capture

In this chapter, we propose our second performance capture framework. First, a robust method to track 3D trajectories of features on a moving subject recorded with multiple cameras is described. Thereafter, by combining the 3D trajectories with a mesh deformation scheme, the performance of a moving actor is captured and the high-quality scanned model can be directly animated such that it mimics the subject’s motion.

As described in Chapter 8, the generation of realistic and lifelike animated characters from captured real-world motion sequences is still a hard and time-consuming task. In the previous chapter, we presented our first performance capture system, that uses an optical flow-based 3D correspondence estimation method to capture the motion and dynamic shape of the moving actor. Although achieving good results, as mentioned in Sect. 9.2, one of the main limitations of the previous approach is its lack of robustness when the motion sequence is fast or complex.

In this chapter, we present an alternative solution, that improves the performance of the previous technique, by combining a flow-based and an image-feature based method. Furthermore, we divide the problem into two steps: first, image features in 3D space are robustly identified and tracked. Afterwards, using the 3D trajectories of the features as constraints in a Laplacian-based tracking scheme, Sect. 3.2.2, the scanned model is realistically animated over time. The simple and robust algorithm proposed here creates shape deformations for the scanned model without specification of explicit material parameters, and it works even for arbitrary moving subjects or people wearing wide and loose apparel.

The main contributions of this chapter is

• a simple and robust method to automatically identify important features and track their 3D trajectories on arbitrary subjects from multi-view video [1],
• and an efficient Laplacian-based tracking scheme that uses only a handful of feature trajectories to realistically animate a scanned model of the recorded subject [1].

springerlink.com © Springer-Verlag Berlin Heidelberg 2010
10.1 Overview

An overview of our second performance capture approach is shown in Fig. 10.1. Our system expects a multi-view video (MVV) sequence as input that shows the subject moving arbitrarily. After acquiring the sequence in our studio, Chapter 4, silhouette images are calculated via color-based background subtraction (Sect. 2.3.1), and we use the synchronized video streams to extract and track important features in 3D space over time, Sect. 10.2.

Our hybrid 3D point tracking framework jointly uses two techniques to estimate the 3D trajectories of the features from unmodified multi-view video streams. First, features in the images are identified using the Scale Invariant Feature Transform (SIFT), Sect. 2.3.4. Furthermore, SIFT is able to match a feature to its corresponding one from a different camera viewpoint. This allows us to generate a set of pairwise pixel correspondences between different camera views for each time step of input video. Unfortunately, tracking the features over time using only local descriptors is not robust if the human subject is wearing sparsely textured clothing. Therefore, we use a robust dense optical flow method as an additional step to track the features for each camera view separately in order to fill the gaps in the SIFT tracking. By

Fig. 10.1 Overview of our second performance capture framework: given a multi-view video sequence showing a moving subject, our method automatically identifies features and tracks their 3D trajectories. By applying the captured trajectories to a static laser-scan of the subject, we are able to realistically animate the scanned model making it move the same way as its real-world counterpart in the video streams.