Real-Time 3D Face Tracking with Mutual Information and Active Contours

Giorgio Panin and Alois Knoll

Technical University of Munich, Informatics Faculty,
Boltzmannstrasse 3, 85748 Munich, Germany
{panin,knoll}@in.tum.de

Abstract. We present a markerless real-time, model-based 3D face tracking methodology. The system combines two robust and complimentary op-timization-based strategies, namely active contours and mutual information template matching, in order to obtain real-time performances for full 6dof tracking. First, robust head contour estimation is realized by means of the Contracting Curve Density algorithm, effectively employing local color statistics separation for contour shape optimization. Afterwards, the 3D face template is robustly matched to the underlying image, through fast mutual information optimization. Off-line model building is done using a fast modeling procedure, providing a unique appearance model for each user. Re-initialization criteria are employed in order to obtain a complete and autonomous tracking system.

Keywords: Real-time Face Tracking, Nonlinear Optimization, 3D Template Matching, Mutual Information, Active Contours, Local Statistics Contour Matching, 3D Face Modeling.

1 Introduction

Real-time 3D face tracking is an important problem in computer vision. Several approaches have been proposed and developed with different model definitions, concerning shape, degrees of freedom, use of multiple appearance/shading templates, use of natural facial features, etc.; a careful choice of model complexity is a critical issue for real-time tracking, often forcing to resort to approximate solutions, in terms of the output provided to the end user or to subsequent processing modules.

The focus of this paper concerns fast and reliable $6dof$ face pose estimation and tracking in real-time, based on a hierarchical integration of two robust high-level visual modalities. In order to motivate our approach, we first consider here related state-of-the-art methodologies, from the available literature on the subject.

The system proposed in [1] employs a robust multi-layer fusion of different visual cues in the hierarchical framework named IFA [2], proceeding from coarse to accurate visual modalities, and providing the result from the top-level tracker as output; in this work, simple template and feature point models are used at the high levels.

G. Bebis et al. (Eds.): ISVC 2007, Part I, LNCS 4841, pp. 1–12, 2007.

Other approaches, using natural face features detection and tracking in a monocular setting, have been presented in [3][4]. In paticular, in [3] a 3D face model is fitted by matching features across subsequent frames, with an approach combining RANSAC [5] and Particle Filters [6] under frame-to-frame epipolar constraints. [4] combines on-line and off-line information by matching local features and optimizing a robust least-squares global cost function. Although this approach can provide a better stability, precision and speed, due to the use of local optimization techniques, the joint use of online and offline information may pose additional choices, concerning the overall cost function parameters, and the models required for tracking.

Template-based approaches also show to be well-suited for face tracking tasks, concerning both precision and robustness issues. In [7], shape and appearance parameters are optimized at the same time under a 2D piece-wise affine deformation model; although the 3D pose of the face is not directly provided by the estimation algorithm, it can be subsequently estimated from the set of 2D parameters, at the price of quite complex computations involving a Kalman-filter based methodology [8]. The approach [9], closer to ours, directly employs a full 3D shape and appearance template, with multiple appearance models and robust nonlinear least-squares optimization. The 3D shape of the face is used together with multiple texture models of the user under different light conditions; these informations are provided off-line.

Generally speaking, the need for multiple light/appearance models can constitute a major drawback of most template-based approaches, forcing the user to provide off-line several reference images for the textured model. This motivation leads to consider more general similarity functions for template matching, in order to provide robustness of the system while using a few, or even a single appearance model, during all of the tracking task.

Motivated by the previous considerations, in this paper we propose a novel model-based approach fusing two complimentary visual modalities, namely active contour head tracking and face template matching; both modules are based on fast, local optimization of robust cost functions, and at the same time require off-line a minimal set of modeling operations. The paper is organized as follows: Section 2 describes the framework for providing the textured template. Section 3 states the overall tracking algorithm. Section 4 deals with the solution to head contour fitting, employing a fast implementation of the CCD algorithm, while Section 5 describes robust template matching optimizing mutual information similarity index. Finally, Section 6 provides experimental results obtained from a complete implementation of the proposed system.

2 Face Template Modeling

Off-line, given two photos of the subject to be tracked in front and profile views (Fig. 1), we adapt a generic 3D head model by employing part of the model-fitting methodology given in [10], where a few pre-selected feature lines (e.g. eyes, lips, cheek contour) are marked onto the photos, and a least-squares fitting of the mesh is performed.