Performance Driven Face Animation via Non-rigid 3D Tracking

Wei Zhang[1] Qiang Wang[2] Xiaoou Tang[1,3]

[1]Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China

[2]Visual Computing Group, Microsoft Research Asia, Beijing, China

[3]Multimedia Lab, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

{dylan, xtang}@ie.cuhk.edu.hk qiangwa@microsoft.com

ABSTRACT

In this demo, a performance driven 3D face animation system is proposed. The proposed system consists of two key components: a robust non-rigid 3D tracking module and a MPEG4 compliant facial animation module. Firstly, the facial motion is tracked from source videos which contain both the rigid 3D head motion (6 DOF) and the non-rigid expression variations. Afterward, the tracked facial motion is parameterized via estimating a set of MPEG4 facial animation parameters(FAP). As the final step, these FAP values are transferred to the MPEG4-compliant face model for the animation purpose. The proposed tracking and animation system has a strong generalization ability and can be used in the indoor environment with no additional assumptions.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms, Experimentation

Keywords

3D Face Tracking, Performance Driven Animation

1. INTRODUCTION

Facial animation is very useful in the applications of 3D games, interactive human computer softwares, and movies, for the human face can express a wide gamut of emotions and expressions that can vary widely both in intensity and meaning[4, 5]. Since the pioneering work of [5], much has been done to increase the realism of face animation.

The approaches on performance driven facial animation have been widely studied in the community of computer vision and computer graphics during the past decades for the reduction of manually crafted animation and the potential of producing more realistic facial animation. However, the traditional approach [1] relies on the 2D feature tracking to extract the source control, which limits the application scope only to the near frontal face.

To develop a facial animation system in live scenarios, we present a novel 3D facial animation system where animation of the target model is controlled by an unconstrained animator in front of a web-camera. Different from previous approaches where the animation source is controlled by tracking 2D features, we capture the non-rigid 3D facial motion by the state-of-art 3D face tracker [6]. Moreover, it also has no limitation on the motion of the tracking target. Therefore, it guarantees the proposed animation system to work in a wide range of scenarios.

MM'09, October 19–24, 2009, Beijing, China.
ACM 978-1-60558-608-3/09/10.

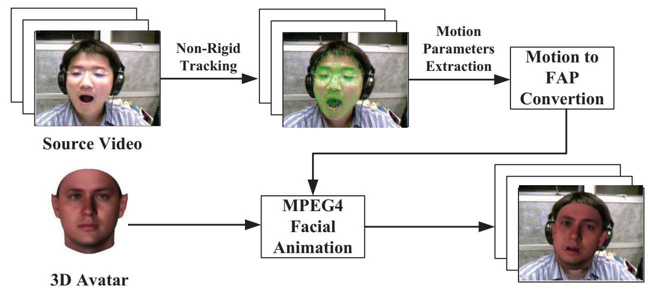

Figure 1: Flow diagram of the animation algorithm

2. ANIMATION SYSTEM DESCRIPTION

Given a video sequence as input, the proposed system is used to drive the animation of a 3D model. Fig. 1 illustrates the flow diagram of the whole animation algorithm. Firstly, non-rigid motion of the source video is captured by a non-rigid 3D face tracker. Then, the extracted motion vectors are used to calculate expression parameters which are from the definition of MPEG-4 FAP. Finally, expression parameters are transferred to the target model for animation purpose. The proposed tracking and animation system has a real-time performance. Meanwhile, the only requirement for the whole system is a neutral face input at the first frame.

2.1 Non-rigid 3D Tracking

Given a collection of image observations $\{\mathbb{I}\}$, non-rigid 3D face tracking at frame t is formulated as the Maximum A Posterior (MAP) estimation problem in Bayesian framework where the tracking state consists of non-rigid 3D shape X_t and rigid pose Q_t:

$$\{\hat{X}_t, \hat{Q}_t\} = \arg\max_{X_t, Q_t} p(X_t, Q_t|\{\mathbb{I}\}) \propto p(\{\mathbb{I}\}|X_t, Q_t) \cdot p(X_t). \quad (1)$$

$p(X_t)$ denotes the prior constraint of 3D face deformation

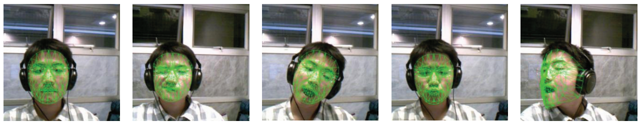

Figure 2: Nonrigid Face Tracking Results

on X_t, $p(\{\mathbb{I}\}|X_t, Q_t)$ is the likelihood distribution which describes the conditional probability of image observations $\{\mathbb{I}\}$ given the tracking state $\{X_t, Q_t\}$. The tracking algorithm consists of three key components:

1. *The prior shape model $p(X)$ characterizes the 3D shape variations from a set of training samples;*
2. *The likelihood model $p(\{\mathbb{I}\}|X, Q)$ is modeled on the 2D-3D feature correspondences. The feature set consists of the off-line trained semantic features, face silhouette features and online tracked low-level features;*
3. *The robust estimation algorithm obtains reliable results in the case of significantly noisy image observations. We make use of an hierarchical optimization strategy together with robust estimation techniques to fulfill the task.*

The generalization capability is guaranteed with the usage of deformable shape model together with the off-line trained distinctive facial features. Fig. 2 shows the results of our tracking approaches for one performer. Detailed algorithm is described in [6].

2.2 Controlling of Target Model

In order to model the expression variations from an animator, we assume that a generic expressive face shape is the sum of an identity shape and an expression deformation: $\mathbf{S}_{exp} = \mathbf{S}_{id} + \Delta\mathbf{S}$, where $\mathbf{S}_{id}$ and $\Delta\mathbf{S}$ hold respectively the face shape with neutral expression and the displacements of the vertices due to expression variations. In our work, $\mathbf{S}_{id}$ is obtained at the initialization stage when the input neutral face is registered with a deformable 3D model described in previous section. Meanwhile, $\Delta\mathbf{S}$ is expressed as a linear combination of FAPs:

$$\Delta\mathbf{S} = \sum_{i=1}^{n} \mathbf{C}_i^{fap} \alpha^i. \tag{2}$$

Given the above equation, the recovery of expression parameters α needs to minimize the difference between $\mathbf{C}_i^{fap}$ with the current motion vector $\Delta\mathbf{S}$. However, the unconstrained error minimization of linear deformation might incur unexpected distortion, as described in [2]. We take the constraint on expression parameters to a fixed region $[\alpha_l, \alpha_u]$ and the optimization problem has the following form of standard quadratic programming:

$$\begin{aligned} &min \ ||\Delta S - \sum_{i=1}^{n} C_i^{fap} \alpha^i||^2 \qquad (3) \\ &w.r.t \ \alpha_l \le \alpha \le \alpha_u, \end{aligned}$$

which can be handled efficiently with the standard numerical technique [3].

Once the expression parameters are recovered, the motion retargeting to a new face model becomes a simple task: the estimated 3D pose parameters and FAPs are transferred to the target model. With the decouple of the facial motion capture and motion retargeting, it is very easy to reuse the same source video on different models. Fig. 3 shows the typical animation results from different animators. More video results can be found at `http://mmlab.ie.cuhk.edu.hk/project.htm`.

Figure 3: Animation Results

3. CONCLUSION

We present a performance driven 3D facial animation system which consists of a real-time non-rigid 3D tracking module and a MPEG-4 compliant animation module. The proposed system allows users in front of a web-camera to control the animation of a target 3D face model. Different from previous approaches where the target face animation is driven by tracking local facial features from the source video, our animation strategy belongs to the category of motion parameterization, where the holistic 3D model's deformation is used for the calculation of high-level expression parameters calculation.

4. REFERENCES

[1] J. Chai, J. Xiao, and J. Hodgins. Vision-based control of 3d facial animation. In *Symposium on Computer Animation*, 2003.

[2] E. Chuang and C. Bregler. Performance driven facial animation using blendshape interpolation. *Computer Science Technical Report, Stanford University*, 2002.

[3] Q. P. Ekaterina, E. Kostina, and O. Kostyukova. A primal-dual active-set method for convex, 2003.

[4] P. Joshi, W. Tien, M. Desbrun, and F. Pighin. Learning controls for blend shape based realistic facial animation. In *SIGGRAPH Courses*, 2005.

[5] F. Parke. Computer generated animation of faces. In *Proceedings of ACM annual conference*, 1972.

[6] W. Zhang, Q. Wang, and X. Tang. Real Time Feature Based 3-D Deformable Face Tracking. *European Conference on Computer Vision*, 2008.