Lecture 18: Morphable Models for Video
Tony Ezzat

Description

I describe how to create, with machine learning techniques, a generative, videorealistic, speech animation module: A human subject is first recorded using a videocamera as he/she utters a pre-determined speech corpus. After processing the corpus automatically, a visual speech module is learned from the data that is capable of synthesizing the human subject, uttering entirely novel utterances that were not recorded in the original video.

The two key contributions of this work are:

1. A variant of the multi-dimensional morphable model (MMM) to synthesize new, previously unseen, mouth configurations from a small set of mouth image prototypes,

2. A trajectory synthesis technique based on regularization, which is automatically trained from the recorded video corpus and which is capable of synthesizing trajectories in MMM space corresponding to any desired utterance.

Results are presented on re-animating human subjects and celebrities. Results are also presented on a series of numerical and psychophysical experiments designed to evaluate the synthetic animations.

Suggested Reading

- Ezzat, Geiger and Poggio. **Trainable Videorealistic Speech Animation.** Proceedings of ACM SIGGRAPH 2002, San Antonio, Texas, July 2002.
- Bregler, Covell, and Slaney. **Video Rewrite: Driving Visual Speech with Audio** Proceedings of ACM SIGGRAPH, 1997 (1997)