

Lecture 22: Morphable Models and Talking Faces

Description

We 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.

Two key aspects of this work are:

1. a variant of the multidimensional 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.

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)
-