Dynamic Facial Textures and Synthesized Body Movement for Near-Videorealistic Speech with Gesture

Richard Kennaway, Vincent Jennings, Barry Theobald, and Andrew Bangham
School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.
{jrk, vjj, bjt, ab}@cmp.uea.ac.uk

Dynamic textures for facial animation

• Static textures enhance the realism of static geometry.
• Dynamic textures enhance the realism of moving geometry.

Synthesis codebook built from training video

• Record training video:
  • one speaker, constant lighting, head-mounted camera for constant pose
  • video contains 279 sentences, containing 6315 different triphone sequences.
  • Build face model
  • Hand-label the significant points in a selection of images
  • Use principal components analysis (PCA) to build point distribution model (PDM)
  • Use PCA to build shape-free appearance model (SFAM); parameterises the variation of the texture map.
• Use models to track face in training video
  • gives equivalent model parameters for each frame.
• Build synthesis codebook
  • a continuous trajectory for each sentence passing through the model parameters for each frame.

Speech to face animation synthesis

Given a new phoneme sequence:
• extract sub-trajectories from original based on phonetic context
• concatenate sub-trajectories and apply to PDM and SFAM

Gesture notation for body animation

HamNoSys: avatar-independent transcription notation for sign languages, developed at the University of Hamburg.

Automatic translation to animation data (joint rotations) for a specific avatar, using a description of the avatar’s static body geometry.
Movement trajectories precomputed from a simplified control model, for several different sets of parameters.

Body animation data can be generated at up to 1000 frames/second = 2.5% of the time budget for 25 fps animation.

Further information and publications:
Visual speech synthesis: http://www.sys.uea.ac.uk/~bjt/
Synthetic animation for sign language: http://www.visicast.sys.uea.ac.uk/