Nonlinear Light-Field Shape-and-Texture Appearance Manifolds

C. Mario Christoudias Louis-Philippe Morency Trevor Darrell
Vision Interfaces Group, MIT CSAIL http://groups.csail.mit.edu/vision/vip

Abstract

Statistical shape-and-texture appearance models use image morphing to define a rich, compact representation of object appearance. They are useful in a variety of applications including object recognition, tracking and segmentation. These techniques, however, have been limited to objects with Lambertian surface reflectance, simple geometry and topology. In this work we present new shape-and-texture appearance models that overcome these limitations. In the first part of our work we develop a 4D shape-and-texture appearance model, built using light-fields. This model is capable of representing objects with complex surface reflectance and geometry. We demonstrate our light-field appearance model using 50 light-fields of the human head captured from a real-time camera array. Next, we present a non-parametric appearance model of the shape and texture of objects whose appearance manifolds exhibit a complex topology, e.g. have holes. We demonstrate this model using 2D mouth images of speaking people. In our experiments we evaluate the performance of each method and provide a comparison with conventional, linear single- and multi-view deformable models.

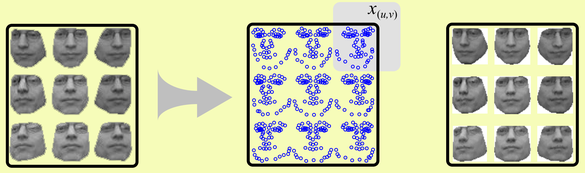
4D Representation of Appearance

Light Field $L(u,v,s,t)$

Easily models
- complex lighting
- non-lambertian surfaces
- fine structure

No 3D Geometry Required

Light Field Shape and Texture



Prototype Light Field | Light Field Shape $X = \{x_{(u,v)} \mid (u,v) \in L\}$ | Light Field Texture $G(u,v,s,t)$

Light Field Appearance Manifold (L_m)

$$L_m(p) = G_m(p) \circ D(X_r, X_m(p)), \quad D_m: R^4 \rightarrow R^4$$

4D Deformation Field

Light Field Warping Function

$$p^T = (c^T \mid t^T \mid u^T)$$

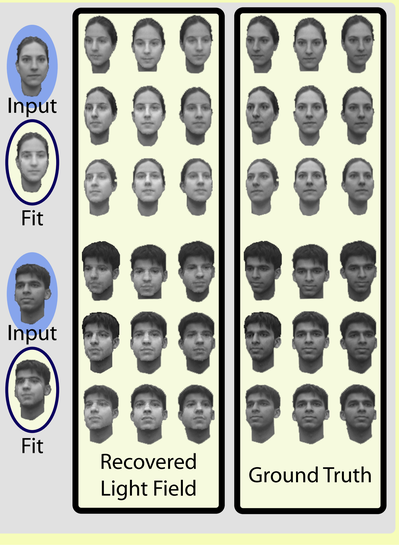
$$X_m(p) = \mathbf{S}_t(\overline{X} + \mathbf{Q}_s c)$$
$$G_m(p) = \mathbf{T}_u(\overline{G} + \mathbf{Q}_g c)$$

Reference Shape

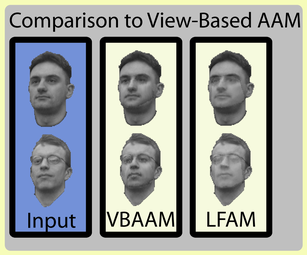
$$X_r = \mathbf{M}_\alpha \overline{X}$$

Example Model Fits

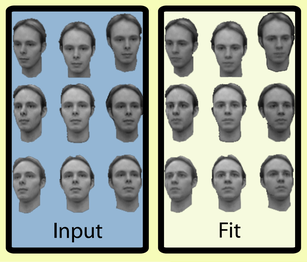
Single View Input

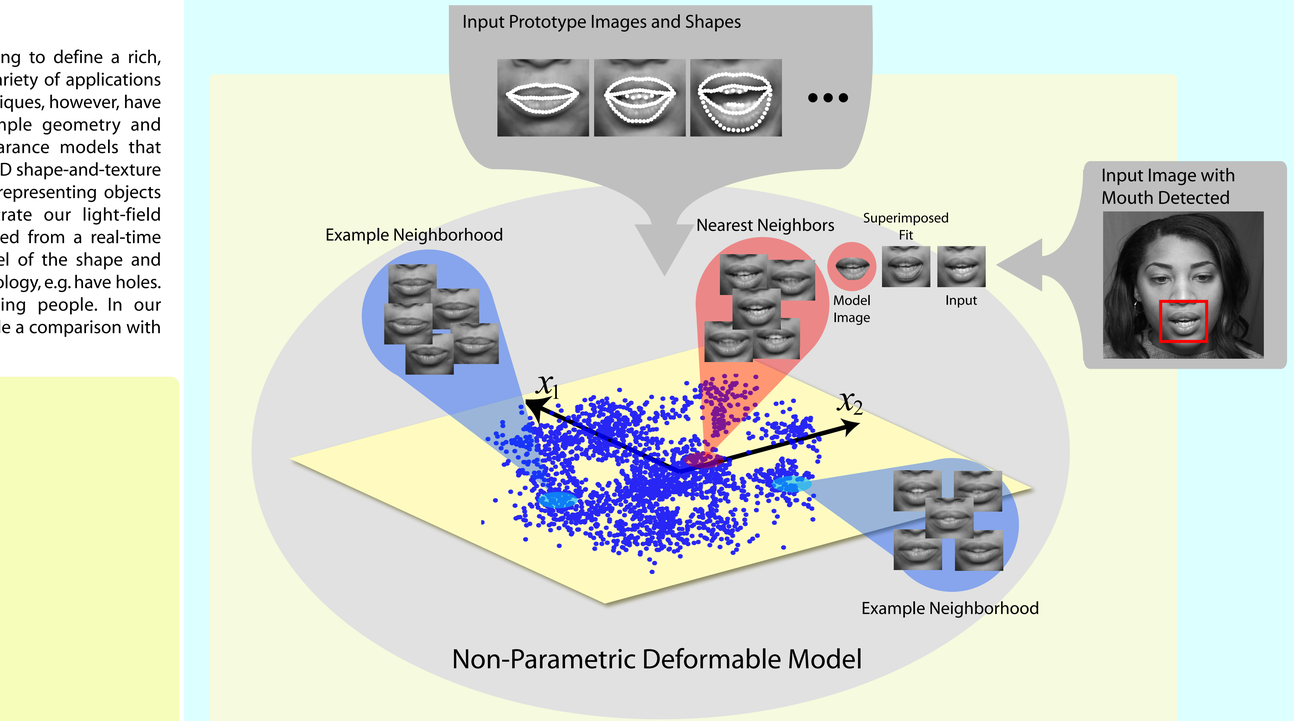


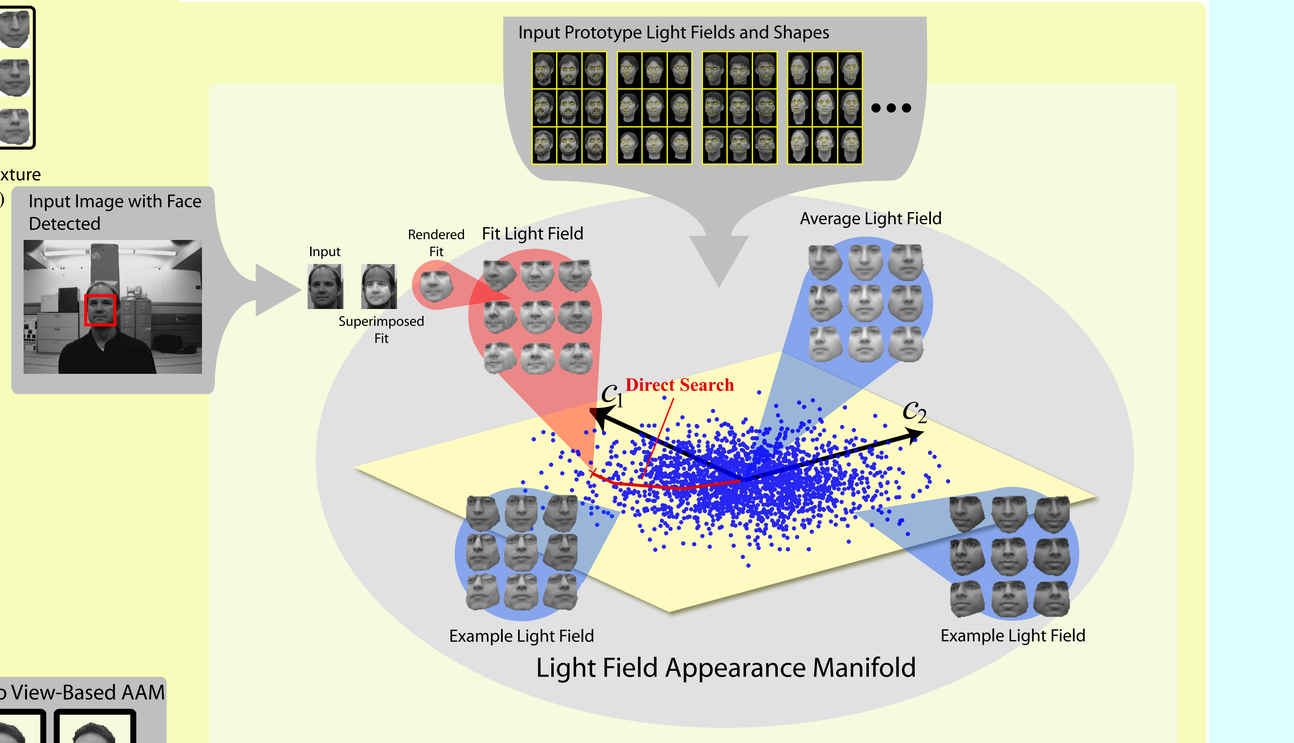
Comparison to View-Based AAM



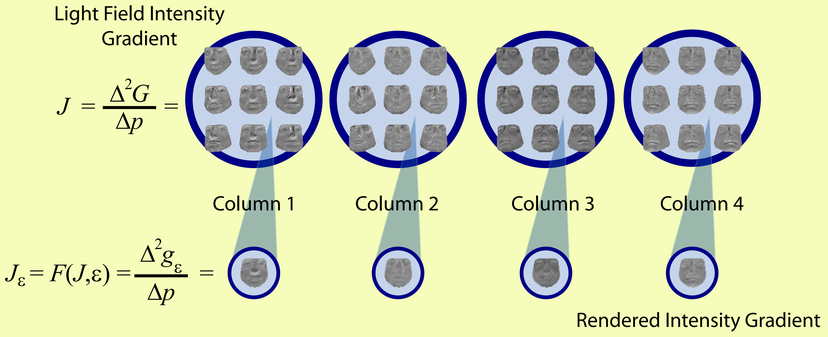
Multi-View Input







Matching to an Image (Rendering the Jacobian)



Nonlinear Appearance Manifolds: Varying Topology

Example Mouth Images

Image Morph (Mean Image)

Shape Aligned Images | Mean Image

Implicit Manifold Representation

Input | First Five Nearest Neighbors

Non-Parametric Deformable Model

Find Local Neighborhood
I: I_i, $i = 1, \ldots, k$

Analyze Input Image I
$$E(\mathbf{b},\mathbf{c},\mathbf{t}) = \| I \circ D(\mathbf{s}_m(\mathbf{c}), \mathbf{s}_{\text{ref}}) - \mathbf{g}_m(\mathbf{b}) \|$$

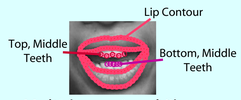
Reference Coordinate Frame
$$\mathbf{s}_{\text{ref}} = \frac{1}{k}\sum_i \mathbf{s}_i$$

Model Shape and Texture
$$\mathbf{g}_m(\mathbf{b}) = \sum_i b_i \mathbf{g}_i$$
$$\mathbf{s}_m(\mathbf{c},\mathbf{t}) = \mathbf{S}_t\left(\sum_i c_i \mathbf{s}_i\right)$$

Neighborhood Textures
$$\mathbf{g}_i = I_i \circ \mathbf{s}_{\text{ref}}$$

Experiments

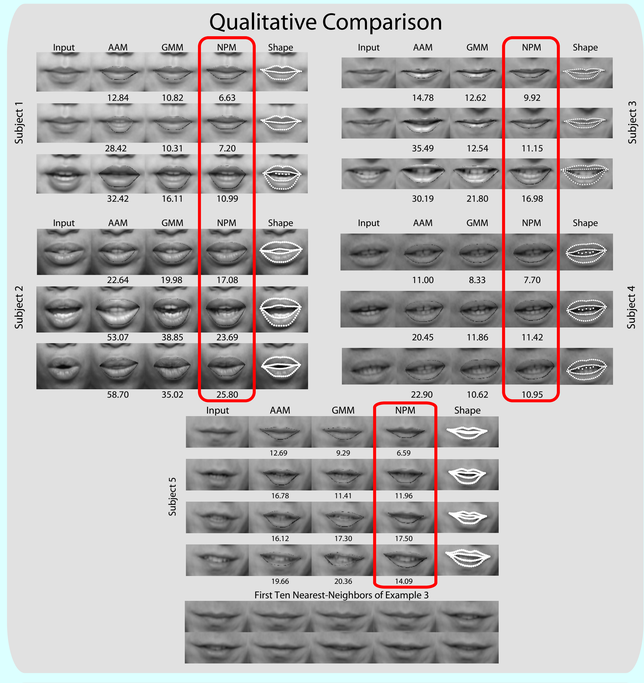
Five Subjects from AVTIMIT Database



Multidimensional Shape

Results

Qualitative Comparison



Quantitative Comparison