massachusetts institute of technology — artificial intelligence laboratory

Light Field Morphable Models

Chris Mario Christoudias, Louis-Philip Morency and Trevor Darrell

AI Memo 2003-010 April 2003

Abstract

Statistical shape and texture appearance models are powerful image representations, but previously had been restricted to 2D or simple 3D shapes. In this paper we present a novel 3D morphable model based on image-based rendering techniques, which can represent complex lighting conditions, structures, and surfaces. We describe how to construct a manifold of the multi-view appearance of an object class using light fields and show how to match a 2D image of an object to a point on this manifold. In turn we use the reconstructed light field to render novel views of the object. Our technique overcomes the limitations of polygon based appearance models and uses light fields that are acquired in real-time.

1. Introduction

Appearance models are a natural and powerful way of describing objects of the same class. Multidimensional morphable models [14], active appearance models [9] and their extensions have been applied to model a wide range of object appearance. The majority of these approaches represent objects in 2D and model view change by morphing between the different views of an object. Modeling a wide range of viewpoints in a single 2D appearance model is possible, but requires non-linear search [18]. Additionally, object self-occlusion introduces holes and folds in the synthesized target view which are difficult to overcome.

A polygonal based 3D morphable model was proposed by Blanz and Vetter [5]. With their approach the view is an external parameter of the model and does not need to be modeled as shape variation. However, this technique is based on a textured polygonal mesh which has difficultly representing fine structure, complex lighting conditions and non-lambertian surfaces. Due to the accuracy of the 3D surfaces needed in this approach, the face scans of each prototype subject cannot be captured in real-time and fine structure such as hair cannot be acquired.

We propose a 3D morphable model using image-based rendering rather than rendering with a polygonal mesh. We use a light field representation, which does not require any depth information to render novel views of the scene. Although a morphable model may be defined using any image-based rendering technique, we found a light field representation to be advantageous for this reason. Our approach can easily model complex scenes, lighting effects, and can be captured in real-time using camera arrays [23, 22].

With light field rendering each morphable model prototype consists of a set of sample views of the plenoptic function [1]. A reference prototype is computed and shape and texture is defined for each sample view relative to that prototype. The resulting morphable model can be matched to a 2D image of a novel object by searching over shape and texture parameters. The estimated shape and texture parameters may be used to synthesize an image-based reconstruction of the object, which in turn may be used to render novel views of the object.

In this paper we present a light field morphable model of the human head. The model was built by collecting light fields of 50 subjects using a real-time camera array [23] (see Figure 1). The light field morphable model is formally described in Section 3. In Section 4 we show how to match a light field morphable model to a single 2D image. Finally, in Section 5 we demonstrate our head model and in Section 6 provide concluding remarks and discuss future work.

2. Previous Work

Statistical models based on linear manifolds of shape and texture variation have been widely applied to the modeling, tracking, and recognition of objects with sets of 2D features [4, 11, 14]. In these methods small amounts of pose change are typically modeled implicitly as part of shape variation on the linear manifold. For representing objects with large amounts of rotation, nonlinear models have been proposed, but are complex to optimize [18].

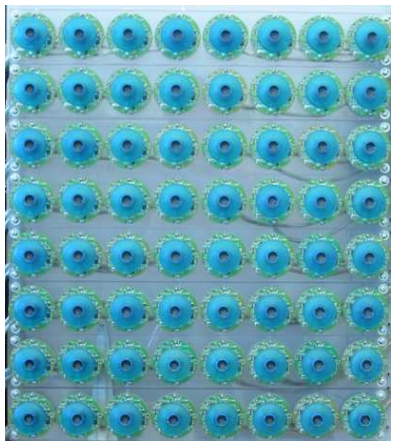


Figure 1: Light field camera array [23].

An alternative approach to capturing pose variation is to use an explicit multi-view representation which builds a PCA model at several viewpoints. This approach has been used for pure intensity models [2] as well as shape and texture models [10]. A model of inter-view variation was also recovered in [10], and missing views could be reconstructed. However, in this approach views were relatively sparse, and individual features were not matched across views; images were rendered using only texture from a single viewpoint.

Shape models with 3D features have the advantage that viewpoint change can be explicitly optimized while matching or rendering the model. Vetter [5] showed how a morphable model could be created from 3D range scans of human heads. This approach represented objects as simply textured 3D shapes, and relied on high-resolution range scanners to construct a model; non-lambertian and dynamic effects are difficult to capture using this framework. With some manual intervention, 3D models can be learned directly from monocular video [12, 16]; an automatic method for computing a 3D morphable model from video was shown in [6]. These methods all used polygonal mesh modes for representing and rendering shape.

Image-based models have become popular in computer graphics recently; with these approaches 3D object appearance is captured in a set of sampled views or ray bundles. Lightfield [15] and Lumigraph [13] rendering techniques provide a method for creating a new image by resampling the set of stored rays that represent an object. These techniques can accurately capture non-lambertian appearance and fine scale shape. Most recently the Unstructured Lumigraph [7] was proposed, and generalized the Lightfield/Lumigraph representation to handle arbitrary camera placement and geometric proxies.

A method to morph two lightfields was presented in [24]; this algorithm extended the classic Beier and Neely algorithm to work directly on the sampled lightfield representation and to account for self-occlusion across views. Features were manually defined, and only a morph between two (synthetically rendered) light fields was shown

in this work.

In this paper we develop the concept of a light field morphable model, in which 3 or more light fields are ”vectorized” (in the sense of [4]) and placed in correspondence. We automatically construct a light field morphable model of facial appearance from real images, and show how that model can be automatically matched to single static intensity images. Our model differs from the multi-view appearance model of [10] in that coefficients between views are explicitly linked, and that we do not model any pose variation within the shape model at a single view. Additionally we can render appearances with view-dependent texture effects, and use a densely sampled camera array. We construct our face model using images of 50 individuals captured in real-time using a 6x8 camera array.

3. Light Field Manifolds

In this section we present the light field morphable model. We first discuss light field morphing and how a morph between many light fields may be defined. Next we provide a formal description of the light field appearance manifold.

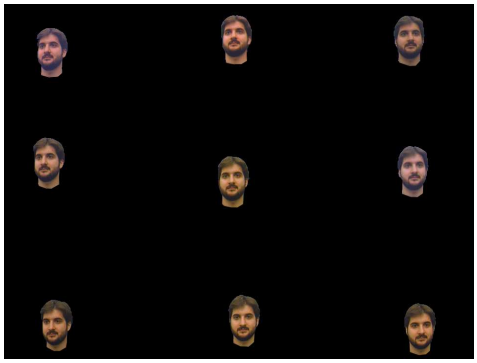
As described in [24], a light field morph is defined by the 2D deformation fields between corresponding views of two light fields. The collection of 2D deformations constitute a 3D deformation between the objects represented by each light field. The resulting morphed light field is an object that represents a smooth transition from one source object to another. We constructed a morph among more than two light fields by selecting a reference light field and then computing a deformation field between the views of each light field with that of the reference. The morphed light field is obtained by blending the aligned views of each source light field.

In a feature based framework [24] the reference may be computed as a polygonal mesh whose vertices are a weighted combination of those specified for each source light field. The background edges of the reference are defined in a similar manner. The geometry of the average light field is specified by an equal weighting of the source features. Alternatively, when the objects represented by each light field are of the same class, correspondence may be computed automatically using optical flow [3]. We have implemented both approaches, and below report results using automatic model construction with optic-flow correspondence estimation.

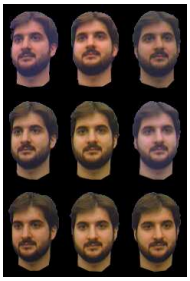
A common problem in 2D image morphing is visibility change. This phenomenon occurs when an object or part of an object is visible in the morphed image but not in one or both source images and vice versa. Visibility change yields undesirable artifacts in the morphed image known as holes and folds [8, 20]. In [24], Zhang et al. describe how to apply visibility processing, which fills holes in certain views of the morphed light field by looking in other views of the source light field where the part of the object may be visible. When constructing a morphable model one may compensate for visibility change by pre-aligning the objects of each source light field, thus eliminating the need for visibility processing.

The average light field is characterized as having the average shape and texture of a set of prototype light fields. By definition its difference in shape and texture is minimal between each of the light field prototypes and therefore it is the preferred reference light field. Using optical flow, the average light field is computed via the bootstrapping

(a)



(b)



(c)

Figure 2: Light field pre-processing. (a) 9 of 48 original views of a subject, (b) segmented views, (c) cropped views (for result display only).

algorithm outlined in [21]. This algorithm placed in the context of light fields is presented below as Algorithm 1. For efficiency we applied the algorithm independently on each view of the prototype set.

Once a reference is defined, each source light field may be described in terms of its shape and texture [14]. The shape of a light field is the collection of 2D deformation fields between it and the reference. Its texture is a light field that is geometrically aligned to the reference. The linear combination of texture and shape form an appearance manifold: given a set of light fields of the same object class, the linear combination of their texture warped by a linear combination of their shape describes a new object whose shape and texture are spanned by that of the prototype light fields. A more formal description of this manifold is provided next.

We define $L(u, v, s, t)$ to be a light field consisting of a set of sample views of the scene, parameterized by view indices (u, v) and scene radiance indices (s, t) [15]. Let $L_1, ..., L_n$ be a set of prototype light fields with reference light field L_{ref}. The shape of each light field, S_i is the mapping,

$$S_i : \mathcal{R}^4 \longrightarrow \mathcal{R}^4 \tag{1}$$

that specifies for each ray in prototype L_i the corresponding ray in the reference light field L_0. The shape vector, S_i, is formed by the concatenation of the deformation fields defined for each view of L_i.

The texture, T_i is obtained by reverse warping each L_i by S_i'. More formally,

$$T_i(u, v, s, t) = L_i(S_i'(u, v, s, t)) = L_i \circ S_i'(u, v, s, t). \tag{2}$$

In Equation 2 S_i' is the inverse deformation of S_i; it specifies the deformation from the reference to the prototype L_i. Like shape, the texture vector of a light field is the concatenation of the texture in each of its views.

Using shape and texture, the light field manifold is defined by the set of light fields L_{model} which satisfy,

$$L_{model} = \left(\sum_i b_i T_i \right) \circ \left(\sum_i c_i S_i \right) \tag{3}$$

Equation 3 is the light field morphable model. It relates each model light field to a corresponding point on the manifold parameterized by the weight vectors $\mathbf{b}$ and $\mathbf{c}$. When synthesizing a new light field from a prototypical set one specifies a point on this manifold. Likewise, when performing analysis one finds the point on the manifold corresponding to a given light field or 2D image of an object. We outline how a 2D image can be matched to the manifold in the next section. Two points on the head manifold, constructed using 50 prototype heads (see Figure 3), are shown in Figure 4. The light fields in these and subsequent figures are displayed in the cropped format illustrated in Figure 2.

4. Model Matching

Let S_i and T_i define the shape and the texture of a light field morphable model, specified by prototypes L_i, for $i = 1, ..., n$. A 2D image is matched to a point on the light

Algorithm 1 Compute Average Light Field

Let $L_1, ..., L_n$ be a set of prototype light fields.
Select an arbitrary light field L_i as the reference light field L_{ref}
repeat
  for all L_i **do**
    Compute correspondence fields S_i' between L_{ref} and L_i using optical flow.
    Backwards warp each view of L_i onto L_{ref} using S_i'.
  end for
  Compute the average over all S_i' and T_i.
  Forward warp each view of $T_{average}$ using $S_{average}$ to create $L_{average}$.
  Convergence test: is $L_{average} - L_{ref} < limit$?
  Copy $L_{average}$ to L_{ref}
until convergence

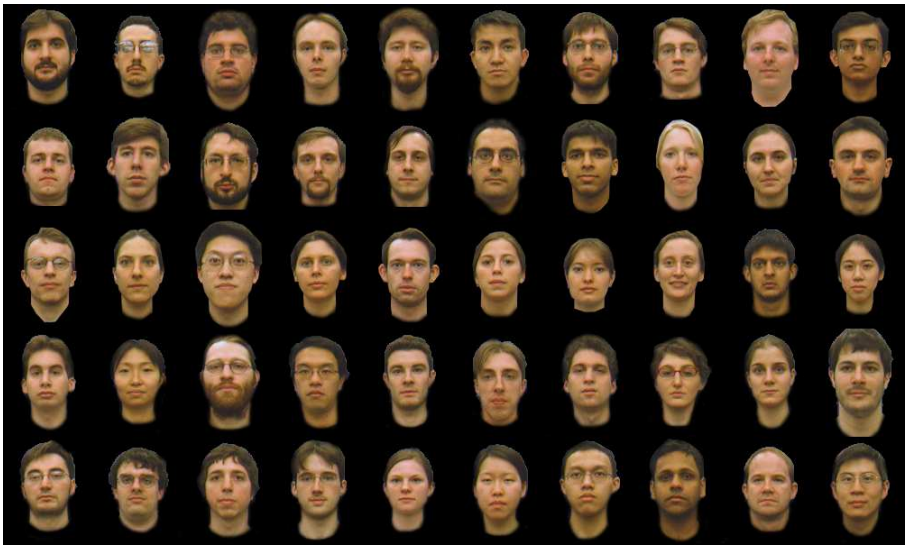


Figure 3: Frontal views of the 50 subjects used to build a light field morphable model of the human head.

field manifold by minimizing the non-linear objective function:

$$E(\mathbf{b}, \mathbf{c}, \epsilon) = \frac{1}{2} \sum_{x,y} \left(I_{novel} - F\left(L_{model}, \epsilon \right) \right)^2 \tag{4}$$

where, L_{model} is as specified in Equation 3 and F is a function that renders pose ϵ of the model light field [15, 7].

The objective function in Equation 4 compares the novel 2D image to the corresponding view in L_{model} in a common coordinate frame. Unlike [14], this is done in the coordinate frame of the novel image. This is because the shape and texture of the morphable model are computed only at the discrete views of the light field prototypes and are not defined in general at any continuous view. Thus at every iteration of the minimization, given the weight vectors $\mathbf{b}$ and $\mathbf{c}$, a morphed light field is synthesized and the estimated pose ϵ is used to render the view corresponding to that of the novel 2D image.

We match a novel image to the light field morphable model defined in Equation 3 by estimating the pose using the average light field and then solving a bilinear system in shape and texture constructed using optical flow. More specifically, given a novel 2D image, the mean squared error between this image and every discrete view of the average light field is computed and the view exhibiting smallest error with pose $\tilde{\epsilon}$, is used as an approximation to the pose of the image. Optical flow is then computed between the novel image and the selected view of the average light field resulting in S_{novel}, the shape of the novel image. Its texture T_{novel} is found by warping the novel image into the coordinate system of the reference via the deformation field defined by S_{novel}. Using the computed shape and texture we then find $\mathbf{b}$ and $\mathbf{c}$ by solving

$$\begin{aligned} T_{novel} &= \sum_i b_{\tilde{\epsilon},i} T_{\tilde{\epsilon},i} \\ S_{novel} &= \sum_i c_{\tilde{\epsilon},i} S_{\tilde{\epsilon},i} \end{aligned} \tag{5}$$

where $T_{\tilde{\epsilon},i}$ and $S_{\tilde{\epsilon},i}$ are the shape and texture vectors at the discrete pose $\tilde{\epsilon}$ and $\mathbf{b}_{\tilde{\epsilon}}$ and $\mathbf{c}_{\tilde{\epsilon}}$ are the weights computed using this pose.

This bilinear system can be solved using linear least squares. Principle component analysis may also be applied to better constrain the systems in shape and texture. In our experiments, using Levenberg-Marquardt [17] for texture gave slightly better results. We found our flow-based matching algorithm to be robust and reasonably fast to demonstrate light-field manifold reconstruction; however faster techniques for optimized model matching, direct parameter search, and dimensionality reduction may also be employed [14, 19, 9].

5. Results

We built a light field morphable model of the human head by capturing light fields of 50 subjects using a real-time light field camera array [23]. 48 views (6 x 8) of each individual were collected and the head was manually segmented from each light field (see Figure 2). The frontal view of every subject is displayed in Figure 3.

A multi-resolution Lukas-Kanade optical flow algorithm was used to compute the shape and texture of each prototype light field. The average face was computed using the bootstrapping algorithm described in Section 3. Similar to [21], convergence is tested by observing the average change in pixel gray value in each view of the reference light field. Ideally this change is zero at convergence, however, we found a limit of 0.01 to be satisfactory in our experiments. Using this limit the algorithm typically converged within 3 to 5 iterations. The 6x8 average light field is displayed in Figure 4(b)[1]. Its shape and texture of the average light field does not favor the characteristics of any one individual. Also, comparing it with Figure 4(a) demonstrates that it is a convincing light field of a human head.

Using the manifold constructed with all subjects we rendered a sequence that demonstrates the expressiveness of a light field morphable model. Images from this sequences are displayed in Figure 5. In the video, we traverse paths along the manifold between different prototype subjects, while varying the position of the light field virtual camera. The glasses of the second subject in this video demonstrates the view-dependent texture effects that may be achieved using our light field morphable model.

A set of randomly selected subjects were matched against a morphable model built using the remaining 49 subjects to test the models ability to generalize to a novel person. The 2D input to the light field morphable model and the corresponding matched light field is displayed in Figure 6 for each subject. A comparison to ground truth is also provided. Note our method built a 3D model from a single 2D image, in which 48 views of the novel subject were inferred. Comparing the matches of Figure 6 one finds that our coarse algorithm performs well in matching novel 2D images to the head manifold. Namely, the skin color, facial hair, and overall shape and expression of each novel subject are well approximated. Overall we found our method works well for computing automatic correspondence between human heads and the majority of examples in our prototype set were favorably matched.

6. Conclusion and Future Work

We introduced a novel active appearance modeling method based on an image-based rendering technique. Light field morphable models overcome many of the limitations presented by current 2D and 3D appearance models. They easily model complex scenes, non-lambertian surfaces, and view variation. We demonstrated the automatic construction of a model of the human head using 50 subjects. In future work we hope to construct a camera array with a wider field of view that utilizes a non-planar camera configuration. We expect our approach to scale directly to the construction of a dynamic light-field morphable model, since our capture apparatus works in real-time.

Acknowledgments

We acknowledge Tony Ezzat for the Lukas-Kanade optical flow implementation. We would also like to thank Jason Yang for his assistance in maintaining and utilizing the

[1]Minor color calibration artifacts are present in some views and are visible in the figures with some display devices. This did not negatively affect our algorithm.

light field camera array, Tony Ezzat for his helpful comments, and those who participated in our data collection.

References

[1] E. H. Adelson and J. Bergen, *Computation Models of Visual Processing*. Cambridge: MIT Press, 1991, ch. The Plenoptic Function and the Elemnts of Early Vision.

[2] A. P. B. Moghaddam, "Probabilistic visual learning for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[3] J. Barron, D. Fleet, S. Beauchemin, and T. Burkitt, "Performance of optical flow techniques," *CVPR*, vol. 92, pp. 236–242.

[4] D. Beymer and T. Poggio, "Face recognition from one example view, Tech. Rep. AIM-1536, , 1995.

[5] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Siggraph 1999, Computer Graphics Proceedings*, A. Rockwood, Ed. Los Angeles: Addison Wesley Longman, 1999, pp. 187–194.

[6] M. Brand, "Morphable 3d models from video," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, May 2001.

[7] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen, "Unstructured lumigraph rendering," in *SIGGRAPH 2001, Computer Graphics Proceedings*, E. Fiume, Ed. ACM Press / ACM SIGGRAPH, 2001, pp. 425–432.

[8] S. E. Chen and L. Williams, "View interpolation for image synthesis," *Computer Graphics*, vol. 27, no. Annual Conference Series, pp. 279–288, 1993.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *Lecture Notes in Computer Science*, vol. 1407, pp. 484–98, 1998.

[10] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol. Volume 20, pp. 657–664, August 2002.

[11] G. Edwards, C. Taylor, and T. Cootes, "Interpreting face images using active appearance models," in *3rd International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 300–305.

[12] P. Fua and C. Miccio, "From regular images to animated heads: a least squares approach," in *Fifth European Conference on Computer Vision*, B. N. H. Bukhardt, Ed., Springer,Berlin, pp. 188–202.

[13] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *Computer Graphics*, vol. 30, no. Annual Conference Series, pp. 43–54, 1996.

[14] M. J. Jones and T. Poggio, "Multidimensional morphable models," in *ICCV*, 1998, pp. 683–688.

[15] M. Levoy and P. Hanrahan, "Light field rendering," *Computer Graphics*, vol. 30, no. Annual Conference Series, pp. 31–42, 1996.

[16] F. H. Pighin, R. Szeliski, and D. Salesin, "Resynthesizing facial animation through 3d model-based tracking," in *ICCV (1)*, 1999, pp. 143–150.

[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, second edition ed. New York, NY: Cambridge University Press, 2002.

[18] S. Romdhani, S. Gong, and A. Psarrou, “A multi-view nonlinear active shape model using kernel pca,” 1999.

[19] S. Sclaroff and J. Isidoro, “Active blobs,” in *Proc. International Conference on Computer Vision*, Mumbai,India, 1998.

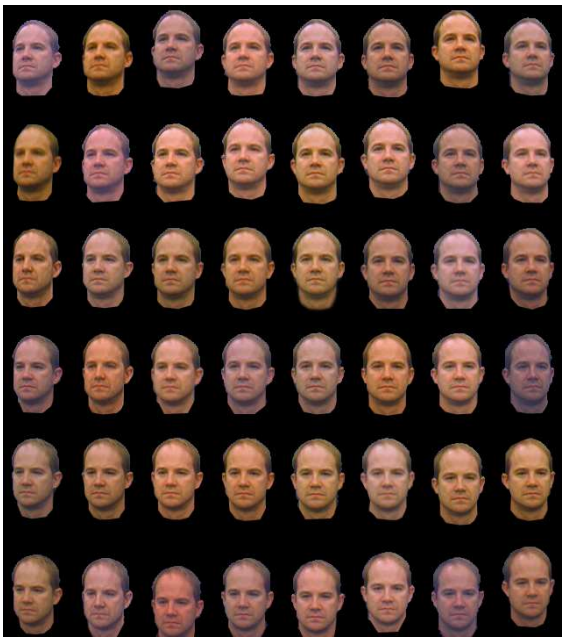
[20] S. M. Seitz and C. R. Dyer, “View morphing,” *Computer Graphics*, vol. 30, no. Annual Conference Series, pp. 21–30, 1996.

[21] T. Vetter, M. J. Jones, and T. Poggio, “A bootstrapping algorithm for learning linear models of object classes, Tech. Rep. AIM-1600, 1997.

[22] B. Wilburn, M. Smulski, H.-H. K. Lee, and M. Horowitz, “The light field video camera,” in *Proceedings of Media Processors 2002, SPIE Electronic Imaging*, 2002.

[23] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, “A real-time distributed light field camera,” in *Eurographics Workshop on Rendering*, 2002, pp. 1–10.

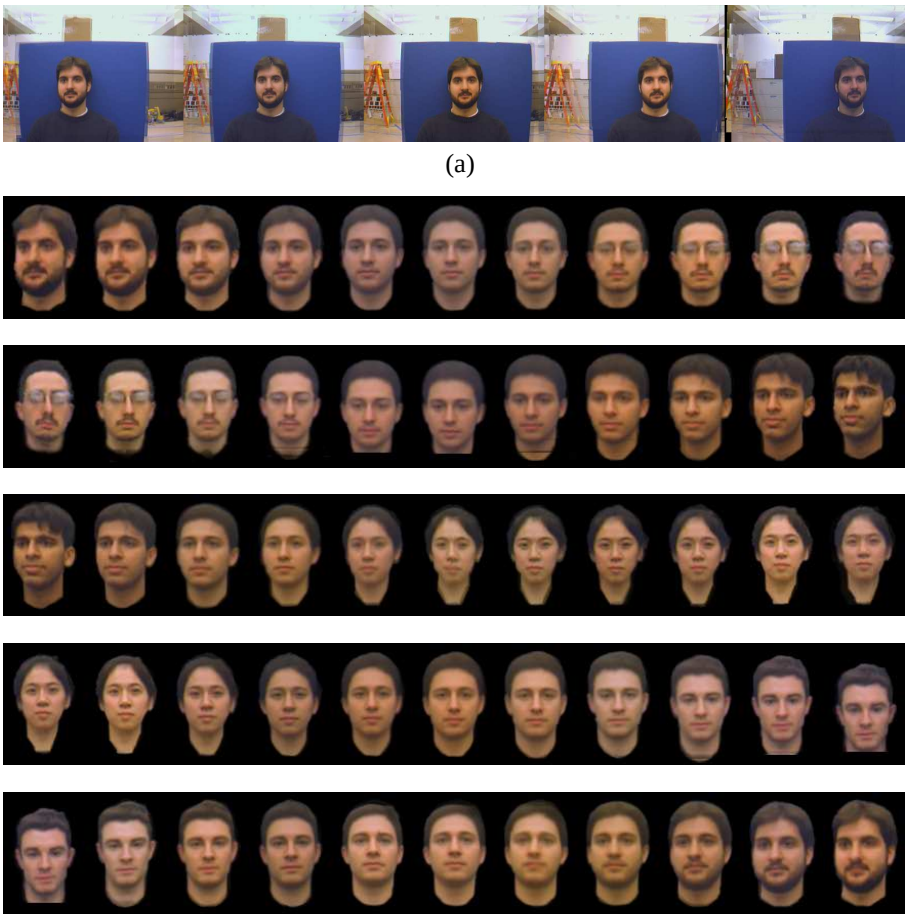
[24] Z. Zhang, L. Wang, B. Guo, and H.-Y. Shum, “Feature-based light field morphing,” in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. ACM Press, 2002, pp. 457–464.



(a)

(b)

Figure 4: Full 6x8 light fields of (a) a prototype subject, (b) the average head.



(a)

(b)

Figure 5: Images from a rendered video sequence. In this video we traverse paths along the manifold between different prototype subjects, while varying the position of the light field virtual camera. The images from the sequence have been cropped and raster scanned from left to right above. Note the view-dependent texture effects apparent on the subject's glasses. (a) Demonstration of camera movement. (b) Traversal through light field manifold while changing the position of the virtual camera. Please see video attachment for a complete demonstration.

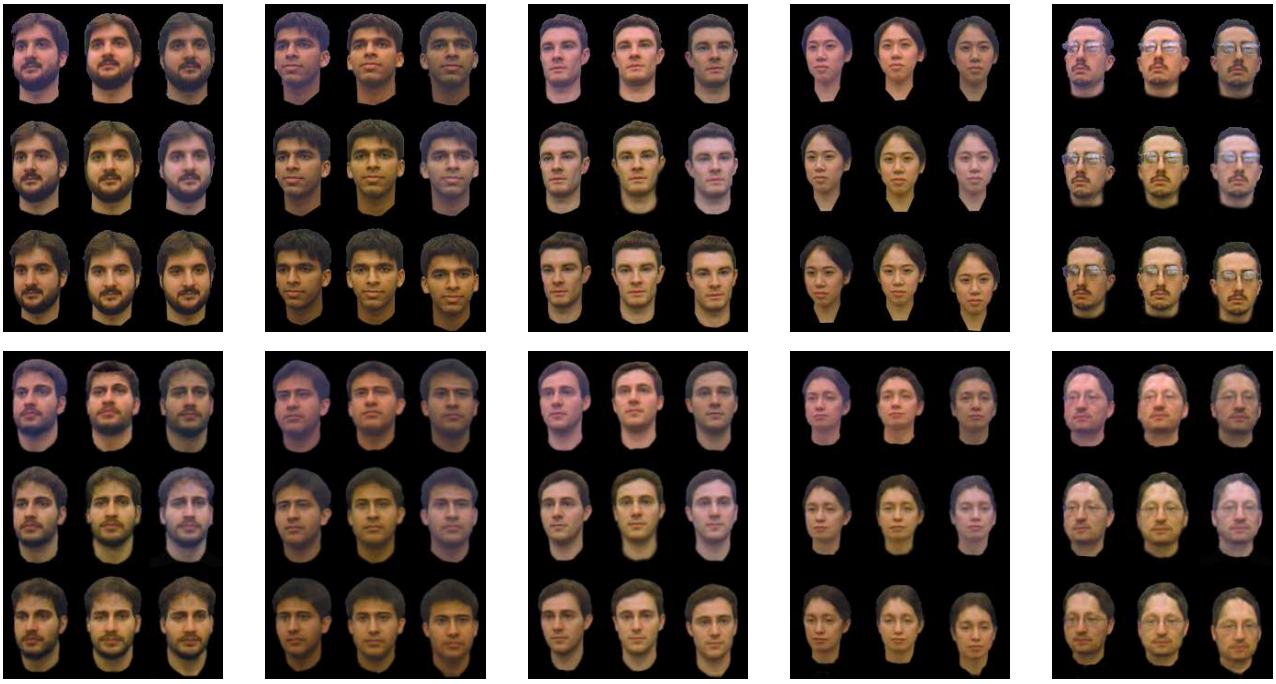


Figure 6: Top row shows portions of the light fields of five individuals. For each individual a 2D image alone (the central view) was used to reconstruct a light field using a morphable model constructed with that individual left out. The bottom row shows the reconstructed light fields.