To appear in IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes, 1999.

Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose and Lighting*

A. S. Georghiades Peter N. Belhumeur

Center for Computational Vision and Control
Yale University
New Haven, CT 06520-8267

David J. Kriegman

Beckman Institute
University of Illinois, Urbana-Champaign
Urbana, IL 61801

Abstract

We present an illumination-based method for synthesizing images of an object under novel viewing conditions. Our method requires as few as three images of the object taken under variable illumination, but from a fixed viewpoint. Unlike multi-view based image synthesis, our method does not require the determination of point or line correspondences. Furthermore, our method is able to synthesize not simply novel viewpoints, but novel illuminations conditions as well. We demonstrate the effectiveness of our approach by generating synthetic images of human faces.

1 Introduction

We present an illumination-based method for synthesizing images of an object under both novel pose and illumination conditions. This method uses as few as three images of the object taken under variable lighting but fixed pose to estimate the object's albedo and generate its geometric structure. Our approach does not require any knowledge about the light source directions in the modeling images, or the establishment of point or line correspondences.

In contrast, nearly all approaches to view synthesis or image-based rendering take a set of images gathered from multiple viewpoints and apply techniques akin to structure from motion [16, 28, 6], stereopsis [20, 21], image tranfer [3], image warping [17, 19, 24], or image morphing [7, 23]. Each of these methods requires the establishment of correspondence between image data (e.g. pixels) across the set. (Unlike other methods, the Lumigraph [11, 18] exhaustively samples the ray space and renders images of an object from novel viewpoints by taking $2-D$ slices of the $4-D$ light field at the appropriate directions.) Since dense correspondence is difficult to obtain, most methods extract sparse image features (e.g. corners, lines), and may use multi-view geometric constraints (e.g. the trifocal tensor [2, 1]) or scene-dependent geometric constraints [21, 8] to reduce the search process and constrain the estimates. By using a sequence of images taken at nearby viewpoints, incremental tracking can further simplify the process, particularly when features are sparse.

For these approaches to be effective, there must be sufficient texture or viewpoint-independent scene features (e.g. albedo discontinuities or surface normal discontinuities). From sparse correspondence, the epipolar geometry can be established and stereo techniques can be used to provide dense reconstruction. Underlying nearly all such stereo algorithms is a constant brightness assumption – that is, the intensity (irradiance) of corresponding pixels should be the same. In turn, constant brightness implies two seldom stated assumptions: (1) The scene is Lambertian, and (2) the lighting is static with respect to the scene – only the viewpoint is changing.

In the presented illumination-based approach, we also assume that the surface is Lambertian, although this assumption is very explicit. As a dual to the second point listed above, our method requires that the camera remains static with respect to the scene – only the lighting is changing. As a consequence, geometric correspondence is trivially established, and so the method can be applied to scenes where it is difficult to establish multi-viewpoint correspondence, namely scenes that are highly textured (i.e. where image features are not sparse) or scenes that completely lack texture (i.e. where there are insufficient image features).

At the core of our approach for generating novel viewpoints is a variant of photometric stereo [27, 29, 13, 12, 30] which simultaneously estimates geometry and albedo across the scene. However, the main limitation of classical photometric stereo is that the light source positions must be accurately known, and this necessitates a fixed lighting rig as might be possible in an industrial setting. Instead, the proposed method *does not* require knowledge of light source locations, and so illumination could be varied by simply waiving a light around the scene.

In fact, our method derives from work by Belhumeur and Kriegman in [5] where they showed that a small set of images with unknown light source directions can be used to generate a representation – the illumination cone – which models the complete set of images of an object (in fixed pose) under all variation in illumination. This method had as its pre-cursor the work of Shashua [25] who showed that in the absence of shadows the set of images of an object lies in a $3-D$ subspace in the image space. Generated images from the illumination cone representation accurately

*P. N. Belhumeur and A. S. Georghiades were supported by a Presidential Early Career Award, an NSF Career Award IRI-9703134, and ARO grant DAAH04-95-1-0494. D. J. Kriegman was supported by NSF under NYI, IRI-9257990 and by ARO grant DAAG55-98-1-0168.

depict shading and attached shadows under extreme lighting; in [10] the cone representation was extended to include cast shadows for objects with non-convex shapes. Cast shadows are global effects, as opposed to attached shadows, and their prediction requires the reconstruction of the object's surface.

In generating the geometric structure, multi-viewpoint methods typically estimate depth directly from corresponding image points [20, 21]. It is well known that without sub-pixel correspondence, stereopsis provides a modest number of disparities over the effective operating range, and so smoothness or regularization constraints are used to interpolate and provide smooth surfaces. The presented illumination-based method estimates surface normals which are then be integrated to generate a surface. As a result, very subtle changes in depth are recovered as demonstrated in the synthetic images in Figures 4 and 5. Those images also show the effectiveness of our approach in generating realistic images of faces under novel pose and illumination conditions.

2 Illumination Modeling

In [5] Belhumeur and Kriegman have shown that, for a convex object with a Lambertian reflectance function, the set of all images under an arbitrary combination of point light sources forms a convex polyhedral cone in the image space $\mathbb{R}^n$ which can be constructed with as few as three images.

Let $\mathbf{x} \in \mathbb{R}^n$ denote an image with n pixels of a convex object with a Lambertian reflectance function illuminated by a single point source at infinity. Let $B \in \mathbb{R}^{n\times 3}$ be a matrix where each row in B is the product of the albedo with the inward pointing unit normal for a point on the surface projecting to a particular pixel in the image. A point light source at infinity can be represented by $\mathbf{s} \in \mathbb{R}^3$ signifying the product of the light source intensity with a unit vector in the direction of the light source. A convex Lambertian surface with normals and albedo given by B, illuminated by $\mathbf{s}$, produces an image $\mathbf{x}$ given by

$$\mathbf{x} = \max(B\mathbf{s}, \mathbf{0}), \tag{1}$$

where $\max(B\mathbf{s}, \mathbf{0})$ sets to zero all negative components of the vector $B\mathbf{s}$. The pixels set to zero correspond to the surface points lying in an *attached shadow*. Convexity of the object's shape is assumed at this point to avoid *cast shadows* (shadows that the object casts on itself). It should be noted that when no part of the surface is shadowed, $\mathbf{x}$ lies in the 3-D subspace $\mathcal{L}$ given by the span of the matrix B.

If an object is illuminated by k light sources at infinity, then the image is given by the superposition of the images which would have been produced by the individual light sources, i.e.,

$$\mathbf{x} = \sum_{i=1}^{k} \max(B\mathbf{s}_i, \mathbf{0}) \tag{2}$$

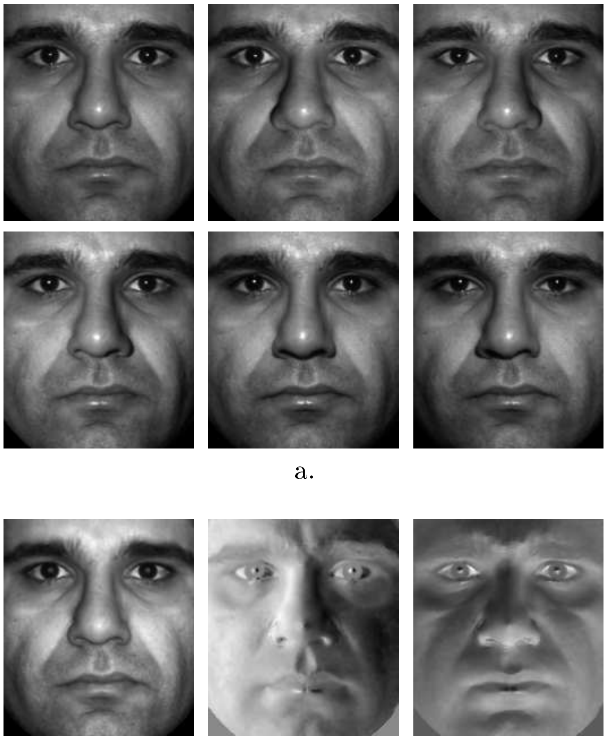
where $\mathbf{s}_i$ is a single light source. Due to the inherent superposition, it follows that the set of all possible images $\mathcal{C}$ of a convex Lambertian surface created by varying the direction and strength of an arbitrary number of point light sources at infinity is a convex cone. It is also evident from Equation 2 that this convex cone is completely described by matrix B.

This suggests a way to construct the illumination model for an individual: gather three or more images of the face without shadowing under varying but unknown illumination but under fixed pose and use them to estimate the three-dimensional illumination subspace $\mathcal{L}$. This can be done by normalizing the images to be of unit length and then estimating the best three-dimensional orthogonal basis B^* using a least-squares minimization technique such as singular value decomposition (SVD). Note that the basis B^* differs from B by an unknown linear transformation, i.e., $B = B^*A$ where $A \in GL(3)$ [9, 12, 22]; for any light source, $\mathbf{x} = B\mathbf{s} = (B^*A)(A^{-1}s)$. Nevertheless, both B^* and B define the same illumination cone and represent valid illumination models.

Unfortunately, using SVD in the above procedure leads to an inaccurate estimate of B^*. For even a convex object whose Gaussian image covers the Gauss sphere, there is only one light source direction (the viewing direction) for which no point on the surface is in shadow. For any other light source direction, shadows will be present. If the object is non-convex, such as a face, then shadowing in the modeling images is likely to be more pronounced. When SVD is used to find B^* from images with shadows, these systematic errors bias its estimate significantly. Therefore, an alternative way is needed to find B^* that takes into account the fact that some data values should not be used in the estimation.

We have implemented a variation of [26] (see also [28, 15]) that finds a basis B^* for the 3-D linear subspace $\mathcal{L}$ from image data with missing elements. To begin, define the data matrix for c images of an individual to be $X = [\mathbf{x}_1 \ldots \mathbf{x}_c]$. If there were no shadowing, X would be rank 3 (assuming no image noise), and we could use SVD to factorize X into $X = B^*S^*$ where S^* is a $3\times c$ matrix the columns of which are the light source directions scaled by the light intensities $\mathbf{s}_i$ for all c images.

Since the images have shadows (both cast or attached), the following method is used: without doing any row or column permutations sift out all the full rows (with no invalid data) of matrix X to form a full sub-matrix $\tilde{X}$. Perform SVD on $\tilde{X}$ and get an initial estimate of S^*. Fix S^* and estimate each of the rows of B^* independently using least squares. Then, fix B^* and update each of the light source direction $\mathbf{s}_i$ independently. Repeat these last two steps until estimates converge. In our experiments, the algorithm is very well behaved, converging to the global minimum within 10-15 iterations. Though it is possible to



a.

b.

Figure 1: a) Six of the original single light source images used to estimate B^*. Note that the light source direction in each image varies only slightly about the viewing direction. b) The basis images of B^*.

converge to a local minimum, we never observed this either in simulation or in practice.

Figure 1 demonstrates the process for constructing the illumination model. Figure 1.a shows six of the original single light source images of a face used in the estimation of B^*. Note that the light source direction in each image varies only slightly about the viewing axis. Figure 1.b shows the basis images of the estimated matrix B^*. These basis images encode not only the albedo (reflectance) of the face but also its surface normal field. They can be used to construct images of the face under arbitrary and quite extreme illumination conditions. But the image formation model in Equation. 1 does not account for cast shadows of non-convex objects such as faces. In order to determine which parts of the image are in cast shadows for a given light source direction, we need to reconstruct the surface of the face (see next section) and then use ray-tracing techniques.

3 Surface Reconstruction

In this section we demonstrate how we can generate an object's surface from B^* after enforcing the integrability constraint on the surface normal field. It has been shown [4, 31] that from multiple images where the light source directions are unknown, one can only recover a Lambertian surface up to a three-parameter family given by the generalized bas-relief (GBR) transformation. This family scales the relief (flattens or extrudes) and introduces an additive plane. It has also been shown that the family of GBR transformations is the only one that preserves integrability.

3.1 Enforcing Integrability

The vector field B^* estimated in Section 2 may not be integrable, so prior to reconstructing the surface up to GBR, the integrability of B^* must be enforced. Since no method has been developed to enforce integrability during the estimation of B^*, we enforce it afterwards. That is, given B^* estimate a matrix $A \in GL(3)$ such that B^*A corresponds to an integrable normal field; the development follows [31].

Consider a continuous surface defined as the graph of $z(x,y)$, and let $\mathbf{b}$ be the corresponding normal field scaled by an albedo field. The integrability constraint for a surface is $z_{xy} = z_{yx}$ where subscripts denote partial derivatives. In turn, $\mathbf{b}$ must satisfy:

$$\left(\frac{b_1}{b_3}\right)_y = \left(\frac{b_2}{b_3}\right)_x$$

To estimate A such that $\mathbf{b}^T(x,y) = \mathbf{b}^{*^T}(x,y)A$, we expand this out. Letting the columns of A be denoted by A_1, A_2, A_3 yields

$$\begin{aligned}(\mathbf{b}^{*^T}A_3)(\mathbf{b}_x^{*^T}A_2) - (\mathbf{b}^{*^T}A_2)(\mathbf{b}_x^{*^T}A_3) = \\ (\mathbf{b}^{*^T}A_3)(\mathbf{b}_y^{*^T}A_1) - (\mathbf{b}^{*^T}A_1)(\mathbf{b}_y^{*^T}A_3)\end{aligned}$$

which can be expressed as

$$\mathbf{b}^{*T}S_1\mathbf{b}_x^* = \mathbf{b}^{*T}S_2\mathbf{b}_y^* \tag{3}$$

where $S_1 = A_3A_2^T - A_2A_3^T$ and $S_2 = A_3A_1^T - A_1A_3^T$.

S_1 and S_2 are skew-symmetric matrices and have three degrees of freedom. Equation 3 is linear in the six elements of S_1 and S_2. From the estimate of B^* discrete approximations of the partial derivatives ($\mathbf{b}_x^*$ and $\mathbf{b}_y^*$) are computed, and then SVD is used to solve for the six elements of S_1 and S_2. In [31], it was shown that the elements of S_1 and S_2 are cofactors of A, and a simple method for computing A from the cofactors was presented. This procedure only determines six degrees of freedom of A. The other three correspond to the generalized bas relief (GBR) transformation [4] and can be chosen arbitrarily since GBR preserves integrability. The surface corresponding to B^*A differs from the true surface by GBR, i.e., $z^*(x,y) = \lambda z(x,y) + \mu x + \nu y$ for arbitrary λ, μ, ν with $\lambda \neq 0$.

3.2 Generating a GBR surface

After enforcing integrability, we can now reconstruct the corresponding surface $\hat{z}(x,y)$. Note that $\hat{z}(x,y)$ is not a Euclidean reconstruction of the face, but a representative element of the orbit under a GBR transformation.

To find $\hat{z}(x,y)$, we use the variational approach presented in [14]. A surface $\hat{z}(x,y)$ is fit to the given components of the gradient p and q by minimizing the functional

$$\int\int_{\Omega} (\hat{z}_x - p)^2 + (\hat{z}_y - q)^2 \, dx \, dy.$$

the Euler equation of which reduces to $\nabla^2 z = p_x + q_y$. By enforcing the right natural boundary conditions and employing an iterative scheme that uses a discrete approximation of the Laplacian, we can reconstruct the surface $\hat{z}(x,y)$ [14].

As stated before, we can only recover the surface of the object up to three parameter family given by the GBR transformation. To use this surface for synthesizing images of the face from novel viewpoints, we need to somehow resolve this ambiguity. Recall that a GBR transformation scales the relief (flattens or extrudes) and introduces an additive plane. Since we are dealing with human faces which constitute a well known class of objects, we can choose an appropriate set of GBR parameters that transforms the relief of a face into some canonical class shape. (In the case when the class of objects is not well defined, the problem of resolving the GBR ambiguity becomes more subtle.) Note that this operation (which is also a GBR transformation) does not completely resolve the ambiguity of the relief. It nevertheless comes very close to that effect.

Figure 2 shows the reconstructed surface of the face (of Figure 1) after resolving the GBR ambiguity. The first basis image of B^* shown in Figure 1.b has been texture-mapped on the surface. Even though we can never hope to recover the exact Euclidean structure of the face (i.e. resolve the ambiguity completely), we can still generate synthetic images of a face under variable pose where the shape distortions due to the GBR ambiguity cannot be discerned. Moreover, since shadows are preserved under GBR transformations [4], images synthesized under an arbitrary light source from a surface whose normal field has been GBR transformed will have correct shadowing.

4 Image Synthesis

We first demonstrate the ability of our method to generate images of an object under novel illumination conditions but under fixed pose. Figure 3 shows sample single light source images of a face which have been corrected to account for cast shadows. We employed a ray-tracing technique that uses the reconstructed surface of the face to determine the cast shadow regions in the images. Observe that despite the near absence of shadows in the images of Figure 1.a, the sample images have strong attached and cast shadows.

Figure 4 displays a set of synthesized images of the the face viewed under variable pose but with fixed lighting. The images were created by rigidly rotating the reconstructed surface shown in Figure 2 first about the horizontal and then about the vertical axis. Along the rows from left to right, the azimuth varies (in 10 degree intervals) from 30 degrees to the right of the face to 10 degrees to the left. Down the columns, the elevation varies (again in 10 degree intervals) from 20 degrees above the horizon to 30 degrees below. For example, in bottom image of the second column from the left the surface has an azimuth of 20 degrees to the right and an elevation of 30 degrees below the horizon. The single light source illuminating the surface is following the face around as it changes pose. This implies that a patch on the surface has the same intensiry in all poses. It is interesting to see that the images look quite realistic with maybe the exception of the three right images in the bottom row which appear to be a little flattened. This is not due to any errors during the geometric or photometric modeling but probably due to our visual priors; we are not used to looking at a face from above.

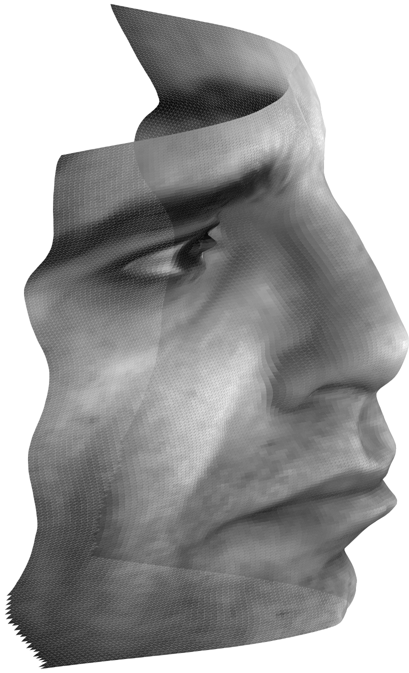


Figure 2: The reconstructed surface.

In Figure 5 we combine both variations in viewing conditions to synthesize images of the face under novel pose and illumination conditions. We used the same poses as in Figure 4 but now the light from the single point source is fixed to come along the gaze direction of the face in the top-right image. Therefore, as the face moves around and its gaze direction changes with respect to the light source direction, the shading of the surface changes and both attached and cast shadows are formed, as one would expect. The synthesized images seem to agree with our visual intuition.

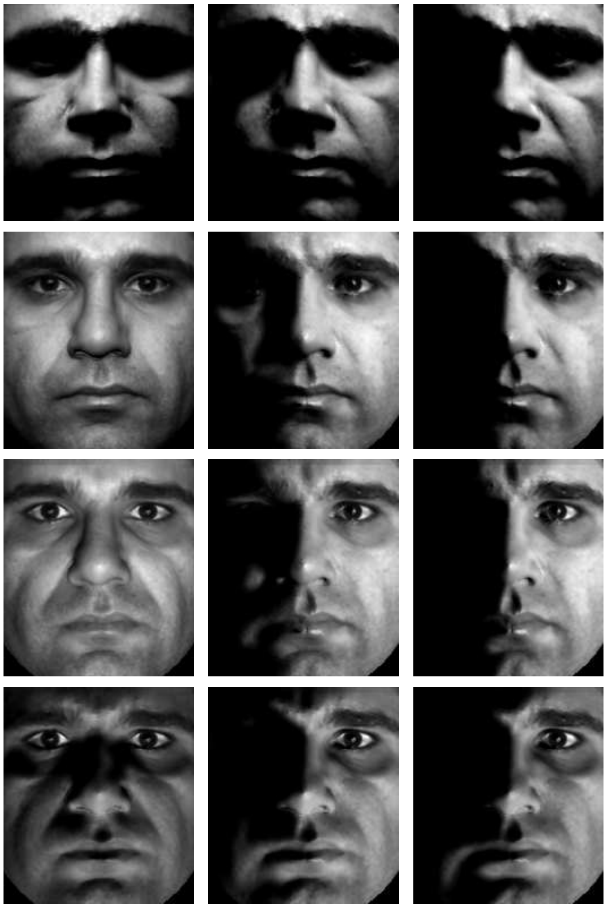


Figure 3: Sample images of the face under novel illumination conditions but fixed pose.

5 Discussion

Appearance variation of an object caused by small changes in illumination under fixed pose can provide enough information for estimating (under the assumption of Lambertian reflectance function) the object's surface normal field scaled by its albedo. In the presented method, as few as three images with no knowledge about the light source directions can be used in the estimation. The estimated surface normal field can then be integrated to reconstruct the object's surface. Unlike multi-view based image synthesis, our approach does not require the determination of point or line correspondences to do the reconstruction. Given that we are dealing with a well known class of objects, we can acceptably resolve the GBR ambiguity of the reconstructed surface. Then, the surface together with the surface normal field scaled by the albedo are sufficient for synthesizing images of the object under novel pose and lighting. We have demonstrated the effectiveness of this approach by generating synthetic images of human faces.

References

[1] S. Avidan, T. Evgeniou, A. Shashua, and T. Poggio. Image-based view synthesis by combining trilinear tensors and learning techniques. In *ACM Symposium on Virtual Reality Software and Technology*, 1997.

[2] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 1034–1040, 1997.

[3] E. Barett, M. Brill, N. Haag, and P. Payton. Invariant linear methods in photogrammetry and model matching. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 277–292. MIT Press, 1992.

[4] P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 1040–1046, 1997.

[5] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions? In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 270–277, 1996.

[6] R. Carceroni and K. Kutulakos. Shape and motion of 3-d curves from multi-view image scenes. In *Image Understanding Workshop*, pages 171–176, 1998.

[7] S. Chen and L. Williams. View interpolation for image synthesis. In *Computer Graphics (SIGGRAPH)*, pages 279–288, 1993.

[8] G. Chou and S. Teller. Multi-image correspondence using geometric and structural constraints. In *Image Understanding Workshop*, pages 869–874, 1997.

[9] R. Epstein, A. Yuille, and P. N. Belhumeur. Learning and recognizing objects using illumination subspaces. In *Proc. of the Int. Workshop on Object Representation for Computer Vision*, 1996.

[10] A. Georghiades, D. Kriegman, and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1998.

[11] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The Lumigraph. In *Computer Graphics (SIGGRAPH)*, pages 43–54, 1996.

[12] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *JOSA-A*, 11(11):3079–3089, Nov. 1994.

[13] B. Horn. *Computer Vision.* MIT Press, Cambridge, Mass., 1986.

[14] B. Horn and M. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 35:174–208, 1992.

[15] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and characterizing intensity images. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1997.

[16] J. Koenderink and A. Van Doorn. Affine structure from motion. *JOSA-A*, 8(2):377–385, 1991.

[17] S. Laveau and O. Faugeras. 3-D scene representation as a collection of images and fundamental matrices. Technical Report 2205, INRIA-Sophia Antipolis, February 1994.

[18] M. Levoy and P. Hanrahan. Light field rendering. In *Computer Graphics (SIGGRAPH)*, pages 31–42, 1996.

[19] W. R. Mark, L. McMillan, and G. Bishop. Post-rendering 3d warping. In *Computer Graphics (SIGGRAPH)*, pages 39–46, 1997.

[20] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. Computer Vision*, 3:293–312, 1989.

[21] D. P.E., C. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Computer Graphics (SIGGRAPH)*, pages 11–20, 1996.

[22] R. Rosenholtz and J. Koenderink. Affine structure and photometry. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 790–795, 1996.

[23] S. Seitz and C. Dyer. View morphing. In *Computer Graphics (SIGGRAPH)*, pages 21–30, 1996.

[24] J. Shade, S. Gortler, L. wei He, and R. Szeliski. Layered depth maps. In *Computer Graphics (SIGGRAPH)*, pages 251–258, 1998.

[25] A. Shashua. *Geometry and Photometry in 3D Visual Recognition*. PhD thesis, MIT, 1992.

[26] H. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 17(9):854–867, September 1995.

[27] W. Silver. *Determining Shape and Reflectance Using Multiple Images*. PhD thesis, MIT, Cambridge, MA, 1980.

[28] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. J. Computer Vision*, 9(2):137–154, 1992.

[29] R. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 17:117–140, 1981.

[30] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 207–218, 1998.

[31] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 158–164, 1997.

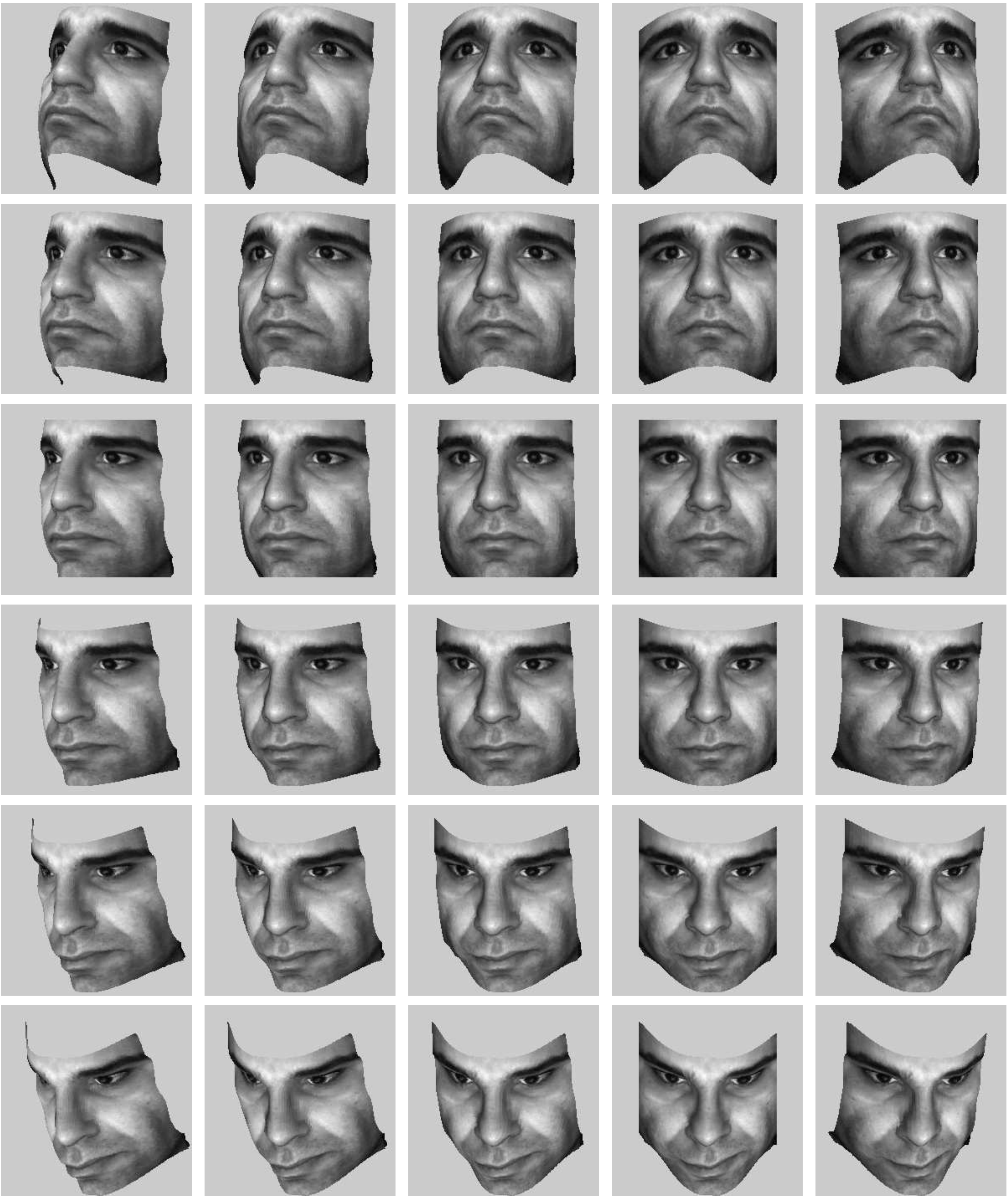


Figure 4: Synthesized images under variable pose but with fixed lighting; the single light source is following the face.

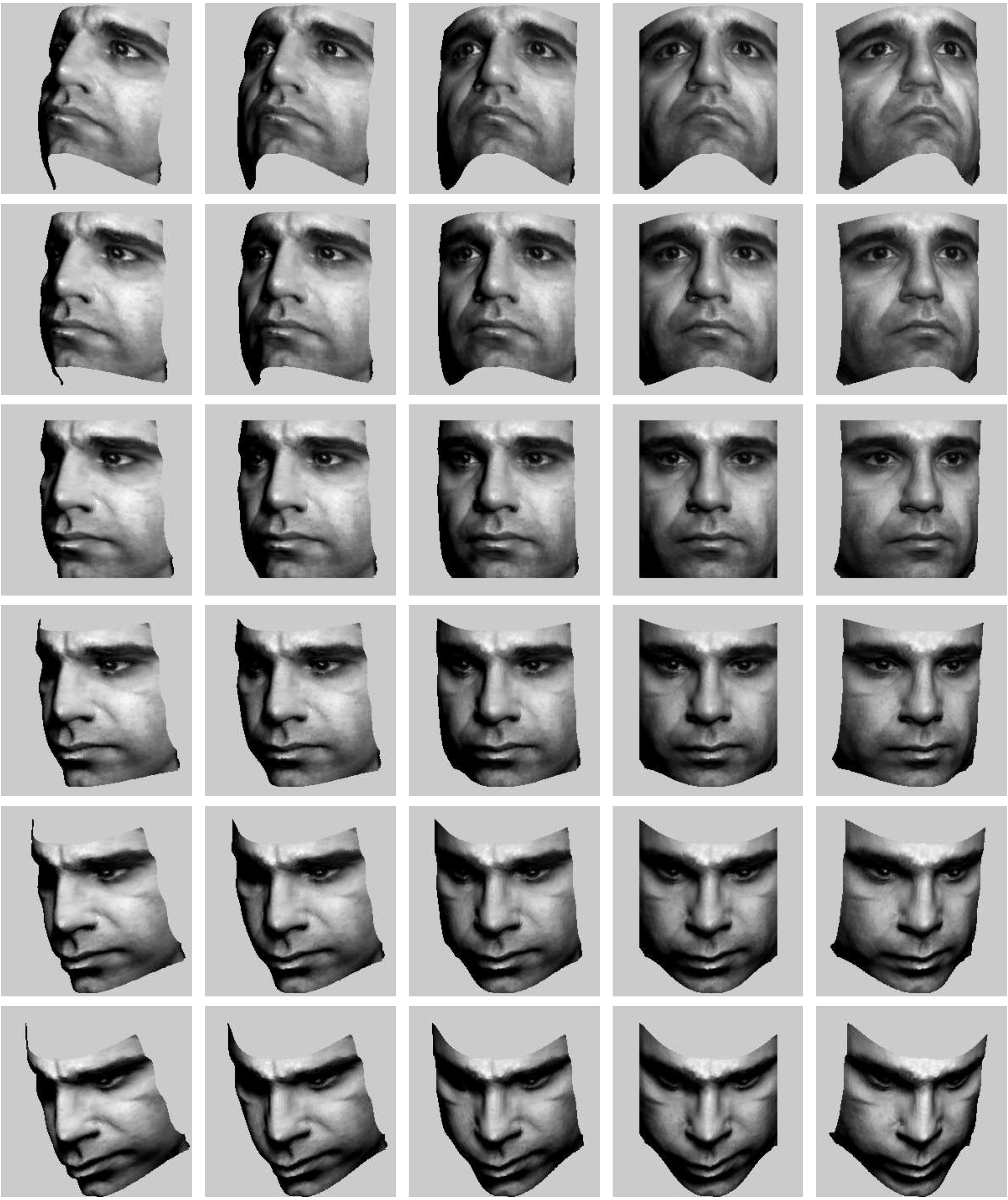


Figure 5: Synthesized images under *both* variable pose and lighting. As the face moves around the single light source stays fixed resulting to image variability due to changes in pose and illumination conditions.