Expressive Expression Mapping with Ratio Images

Zicheng Liu   Ying Shan   Zhengyou Zhang

Microsoft Research*

Abstract

Facial expressions exhibit not only facial feature motions, but also subtle changes in illumination and appearance (e.g., facial creases and wrinkles). These details are important visual cues, but they are difficult to synthesize. Traditional expression mapping techniques consider feature motions while the details in illumination changes are ignored. In this paper, we present a novel technique for facial expression mapping. We capture the illumination change of one person's expression in what we call an *expression ratio image* (ERI). Together with geometric warping, we map an ERI to any other person's face image to generate more expressive facial expressions.

Keywords: Facial animation, Morphing, Animation

1 Introduction

Facial expressions exhibit not only facial feature motions, but also subtle changes in illumination and appearance (e.g., facial creases and wrinkles). These details are important visual cues, but they are difficult to synthesize.

One class of methods to generate facial expressions with details is the morph-based approaches and their extensions [2, 14, 16, 3]. The main limitation is that this approach can only generate expressions in-between the given expressions through interpolation. If we only have someone's neutral face, we would not be able to generate this person's facial expressions using morph-based methods.

Another popular class of techniques, known as expression mapping (performance-driven animation) [4, 10, 20, 13], does not have such limitation. It can be used to animate 2D drawings and images, as well as textured or non-textured 3D face models. The method is very simple. Given an image of a person's neutral face and another image of the same person's face with an expression. The positions of the face features (eyes, eye brows, mouths, etc.) on both images are located either manually or through some automatic method. The difference vector is then added to a new face's feature positions to generate the new expression for that face through geometry-controlled image warping [21, 2, 10]. One problem with such geometric-warping-based approach is that it only captures the face feature's geometry changes, completely ignoring illumination changes. The resulting expressions do not have the expression details such as wrinkles. These details are actually very important visual cues, and without them, the expressions are less expressive and convincing.

Figure 1: *Expression mapping with expression details. Left: neutral face. Middle: result from geometric warping. Right: result from our method.*

As an example, Figure 1 shows a comparison of an expression with and without the expression details. The left image is the original neutral face. The one in the middle is the expression generated using the traditional expression mapping method. The image on the right is the expression generated using the method to be described in this paper. The feature locations on the right image are exactly the same as those on the middle image, but because there are expression details, the right image looks much more convincing than the middle one.

In this paper, we present a novel technique to capture the illumination change of one person's expression and map it to any different person to generate more expressive facial expressions. The critical observation is that the illumination change resulting from the change of surface normal can be extracted in a skin-color independent manner by using what we call an *expression ratio image* (ERI). This ERI can then be applied to any different person to generate correct illumination changes resulted from the geometric deformation of that person's face.

The remainder of this paper is organized as follows. We describe the related work in next section. We then introduce the notion of the expression ratio image (ERI) in Section 3. In Section 4, we describe techniques to filter ERI to remove the noises caused by pixel mis-alignment. Some experimental results are shown in Section 5. We conclude the paper with a discussion of the limitations of our approach and a proposal for future research directions.

*zliu@microsoft.com,yshan@ieee.org,zhang@microsoft.com

ACM SIGGRAPH 2001, 12-17 August 2001, Los Angeles, CA, USA

2 Related work

Besides the work mentioned in the introduction, there has been a lot of other work on facial expression synthesis. For an excellent overview, see [13]. The physically-based approach is an alternative to expression mapping. Badler and Platt [1] used a mass-and-spring model to simulate the skin and muscles. Extensions and improvements to this technique have been reported in [19, 18, 9].

Marschner et al. [11] used the color ratio between the rendered image pairs under the old and new lighting conditions to modify photographs taken under the old lighting condition to generate photographs under the new lighting condition. In a similar spirit, Debevec [5] used the color difference (instead of ratio) between the synthesized image pairs with and without adding a synthetic object to modify the original photograph.

Given a face under two different lighting conditions and another face under the first lighting condition, Riklin-Raviv and Shashua [15] used color ratio (called quotient image) to generate an image of the second face under the second lighting condition. Stoschek [17] combined this technique with image morphing to generate the re-rendering of a face under continuous changes of poses or lighting directions.

The work reported in this paper handles geometrical deformations under constant lighting rather than changing lighting condition with fixed geometry as in the previous works. As far as we know, this is the first technique that is capable of mapping one person's facial expression details to a different person's face.

3 Expression ratio image

For any point $\mathbf{p}$ on a surface Π, let $\mathbf{n}$ denote its normal. Assume there are m point light sources. Let $\mathbf{l}_i, 1 \le i \le m$, denote the light direction from $\mathbf{p}$ to the ith light source, and I_i its intensity. Suppose the surface is diffuse, and let ρ be its reflectance coefficient at $\mathbf{p}$. Under the Lambertian model, the intensity at $\mathbf{p}$ is

$$I = \rho \sum_{i=1}^{m} I_i \mathbf{n} \cdot \mathbf{l}_i \tag{1}$$

After the surface is deformed, the intensity at $\mathbf{p}$ becomes

$$I' = \rho \sum_{i=1}^{m} I_i \mathbf{n}' \cdot \mathbf{l}'_i \tag{2}$$

where $\mathbf{n}'$ is the normal at $\mathbf{p}$ after deformation, and $\mathbf{l}'_i$ is the light direction after deformation.

From Equations (1) and (2), we have

$$\frac{I'}{I} = \frac{\sum_{i=1}^{m} I_i \mathbf{n}' \cdot \mathbf{l}'_i}{\sum_{i=1}^{m} I_i \mathbf{n} \cdot \mathbf{l}_i} \tag{3}$$

We denote

$$\Re \equiv \frac{\sum_{i=1}^{m} I_i \mathbf{n}' \cdot \mathbf{l}'_i}{\sum_{i=1}^{m} I_i \mathbf{n} \cdot \mathbf{l}_i}. \tag{4}$$

$\Re$, which is a real function defined over Π, is called the *expression ratio image* (ERI) of Π. From (3), we have

$$I' = \Re I \tag{5}$$

for each point on the surface Π.

Notice that $\Re$ is independent of the reflectance coefficients. Equation (5) holds for any reflectance function of the surface Π. So for any unknown reflectance function, if we know the illumination before deformation, then we can obtain its illumination after deformation by simply multiplying the expression ratio image $\Re$ with the illumination before deformation.

Let us now consider how to map one person's expression to another. Given two people's faces A and B, assume for every point on A, there is a corresponding point on B which has the same meaning (eye corners, mouth corners, nose tip, etc). By applying Equation (3) to A and B at each point, we have

$$\frac{I'_a}{I_a} = \frac{\sum_{i=1}^{m} I_i \mathbf{n}'_a \cdot \mathbf{l}'_{ia}}{\sum_{i=1}^{m} I_i \mathbf{n}_a \cdot \mathbf{l}_{ia}} \tag{6}$$

and

$$\frac{I'_b}{I_b} = \frac{\sum_{i=1}^{m} I_i \mathbf{n}'_b \cdot \mathbf{l}'_{ib}}{\sum_{i=1}^{m} I_i \mathbf{n}_b \cdot \mathbf{l}_{ib}} \tag{7}$$

Since human faces have approximately the same geometrical shape, if they are in the same pose, their surface normals at the corresponding positions are roughly the same, that is, $\mathbf{n}_a \approx \mathbf{n}_b$ and $\mathbf{n}'_a \approx \mathbf{n}'_b$, and the lighting direction vectors are also roughly the same, that is, $\mathbf{l}_{ia} \approx \mathbf{l}_{ib}$ and $\mathbf{l}'_{ia} \approx \mathbf{l}'_{ib}$. Under this assumption, we have

$$\frac{I'_a}{I_a} \approx \frac{I'_b}{I_b} \tag{8}$$

Of course, if A and B have exactly the same shape, the above equation is exact, not approximate. The approximation error increases with the shape difference between two faces and with the pose difference when the images are taken.

Let $\mathcal{A}$ and $\mathcal{A}'$ denote the images of A's neutral face and expression face, respectively. Let $\mathcal{B}$ denote the image of person B's neutral face, and $\mathcal{B}'$ the unknown image of his/her face with the same expression as $\mathcal{A}'$. Furthermore, if we assume these images have been aligned, then by (8), we have

$$\frac{\mathcal{B}'(u,v)}{\mathcal{B}(u,v)} = \frac{\mathcal{A}'(u,v)}{\mathcal{A}(u,v)} \tag{9}$$

where (u,v) are the coordinates of a pixel in the images. Therefore, we have

$$\mathcal{B}'(u,v) = \mathcal{B}(u,v)\frac{\mathcal{A}'(u,v)}{\mathcal{A}(u,v)} \tag{10}$$

More realistically, the images are usually taken in different poses with possibly different cameras, and so are usually not aligned. In order to apply the above equations, we have to align them first. In summary, given images $\tilde{\mathcal{A}}$, $\tilde{\mathcal{A}}'$, $\tilde{\mathcal{B}}$ which have not been aligned, we have the following algorithm for facial expression mapping which captures illumination changes.

Step 1. Find the face features of $\tilde{\mathcal{A}}$, $\tilde{\mathcal{A}}'$ and $\tilde{\mathcal{B}}$ (either manually or using some automatic method)

Step 2. Compute the difference vector between the feature positions of $\tilde{\mathcal{A}}'$ and $\tilde{\mathcal{A}}$. Move the features of $\tilde{\mathcal{B}}$ along the difference vector, and warp the image accordingly. Let $\mathcal{B}_g$ be the warped image. This is the traditional expression mapping based on geometric warping.

Step 3. Align $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{A}}'$ with $\mathcal{B}_g$ through image warping, and denote the warped images by $\mathcal{A}$ and $\mathcal{A}'$.

Step 4. Compute ratio image $\Re(u,v) = \frac{\mathcal{A}'(u,v)}{\mathcal{A}(u,v)}$.
Step 5. Set $\mathcal{B}' = \Re\mathcal{B}_g$ for every pixel.

This algorithm requires three warping operations for each input image $\mathcal{B}$. When applying the same expression to many people, we can save computation by pre-computing the ratio image with respect to $\mathcal{A}$ or $\mathcal{A}'$. During expression mapping for a given image $\mathcal{B}$, we first warp that ratio image to $\mathcal{B}_g$ and then multiply the warped ratio image with $\mathcal{B}_g$. In this way, we only perform warping twice, instead of three times, for every input image.

3.1 Colored lights

In the above discussion, we only considered the monochromatic lights. For colored lights, we apply exactly the same equations to each R, G and B component of the color images, and compute one ERI for each component. During expression mapping, each ERI is independently applied to the corresponding color component of the input image.

If all the light sources have the same color but only differ in magnitude, it is not difficult to find that the three ERIs are equal to each other. To save computation, we only need to compute one ERI in this case.

3.2 Comments on different lighting conditions

If the lighting conditions for $\mathcal{A}$ and $\mathcal{B}$ are different, Equation (10) does not necessarily hold. If, however, there is only an intensity scaling while the light directions are the same, the equation is still correct. This probably explains why our method works reasonably well for some of the images taken under different lighting environment.

In other cases, we found that performing color histogram matching [8] before expression mapping is helpful in reducing some of the artifacts due to different lighting conditions. In addition, we found that the result is noisy if we directly apply the three color ERIs. A better solution that we use is to first convert RGB images into the YUV space [7], compute the ratio image only for the Y component, map it to the Y component of the input image, and finally convert the resulting image back into the RGB space.

An even better solution would be to use sophisticated relighting techniques such as those reported in [6, 12, 15].

4 Filtering

Since image alignment is based on image warping controlled by a coarse set of feature points, misalignment between $\mathcal{A}$ and $\mathcal{A}'$ is unavoidable, resulting in a noisy expression ratio image. So we need to filter the ERI somehow to clean up the noise while not smoothing out the wrinkles. The idea is to use an adaptive smoothing filter with little smoothing in expressional areas and strong smoothing in the remaining areas.

Since $\mathcal{A}$ and $\mathcal{A}'$ have been roughly aligned, their intensities in the non-expressional areas should be very close, i.e., the correlation is high, while their intensities in the expressional areas are very different. So for each pixel, we compute a normalized cross correlation c between $\mathcal{A}$ and $\mathcal{A}'$, and use $1-c$ as its weight.

After the weight map is computed, we run an adaptive Gaussian filter on the ERI. For pixels with a large weight, we use a small window Gaussian filter so that we do not smooth out the expressional details. For pixels with a small weight, we use a large window to smooth out the noise in the ERI.

We could descretize the weights into many levels and assign a different window size for each level. But in practice we found that it is enough to use just two levels.

5 Results

In this section, we show some test results. For each image, we manually put mark points on the face features. Figure 2 shows an example of the mark points. Currently only the points are used as features for image warping while the line segments in the figure are only for display purpose. We use the texture mapping hardware to warp an image from one set of markers to another by simply applying Delauney triangulation to the mark points. This method is fast but the resulting image quality is not as good as with other more advanced image warping techniques [2, 10].

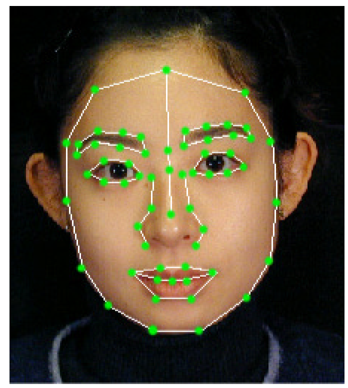


Figure 2: *The mark points.*

For the first example, we map the thinking expression of the middle image in Figure 3 to a different person's neutral face which is the left image of Figure 4. The middle image of Figure 4 is the result from the traditional geometrical warping. The right image is the result of our method. We can see that the wrinkles due to the skin deformation between the eyebrows are nicely captured by our method. As a result, the generated expression is more expressive and convincing than the middle one obtained with geometric warping.

Figure 3: *An expression used to map to other people's faces. The image on the right is its expression ratio image. The ratios of the RGB components are converted to colors for display purpose.*

Figure 1 in Section 1 shows an example of mapping the expression displayed in Figure 5(a). The left image is her

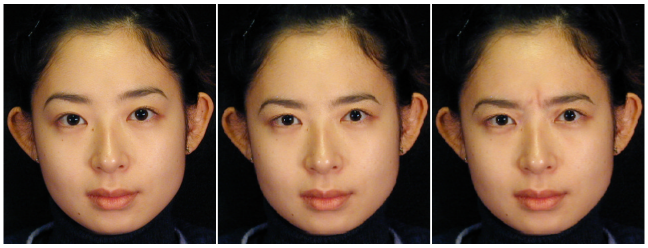


Figure 4: *Mapping a thinking expression. Left: neutral face. Middle: result from geometric warping. Right: result from ERI.*

neutral face. Our result (the right image of Figure 1) contains the visual effects of skin deformations around the nose and mouth region. It is clearly a more convincing smile than the middle image which is generated by geometric warping.

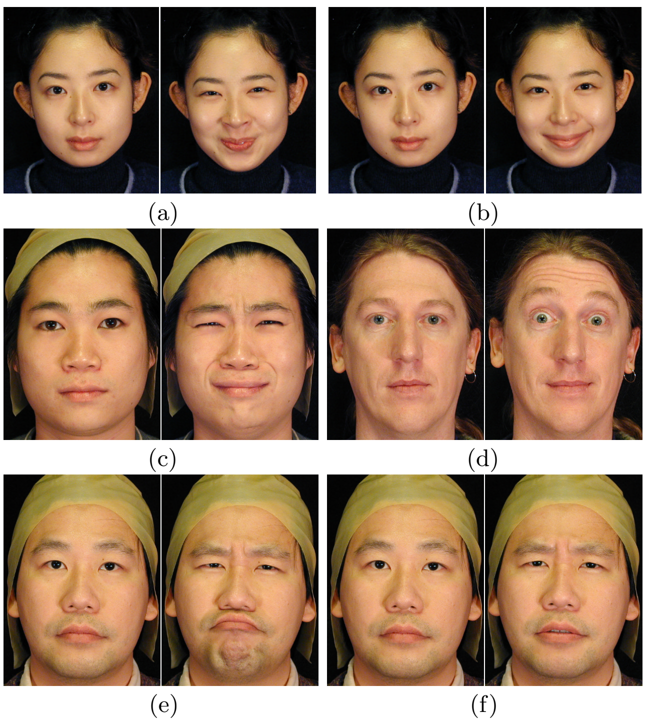


Figure 5: *Expressions used to map to other people's faces.*

Figure 6 shows the result of mapping the sad expression in Figure 5(c). The right image in Figure 6 is the result generated by using our method, while the result from geometric warping is shown in the middle. The right image clearly shows a sad/bitter expression, but we can hardly see any sadness from the middle image.

Figure 7 shows the result of mapping a raising-eyebrow expression in Figure 5(d). We can see that the wrinkles on the forehead are mapped nicely to the female's face.

Figure 8 shows the result of mapping the frown expression in Figure 5(e) to an already smiling face (the left image in Figure 8). Because the frown expression is in a separate face region from the existing smile expression, the mapping works quite well and the resulting expression is basically the sum of the two different expressions.

Figure 9 shows an example of mapping expressions under different lighting conditions. The thinking expression

Figure 6: *Mapping of a sad expression. Left: neutral face. Middle: result from geometric warping. Right: result from ERI.*

Figure 7: *Mapping of a raising-eyebrow expression. Left: neutral face. Middle: result from geometric warping. Right: result from ERI.*

in Figure 5(f) is mapped to the neutral face in Figure 9. These two images were taken in different lighting environment. Again, the image on the right is the result using our method. We can see that the wrinkles between and above the two eyebrows are mapped quite well to the target face. The resulting expression clearly exhibits the visual effects of eyebrow crunching.

In Figure 10, 11, and 12, we show the results of mapping the smile expression in Figure 5(b) to different faces. Figure 10 shows this smile expression being mapped to a male's face. The left image is the neutral face. The middle image is generated using geometric warping and we can see that the mouth-stretching does not look natural. The image on the right is generated using our method. The illumination changes on his two cheek bones and the wrinkles around his

Figure 8: *Mapping a frown expression to a smile expression. Because the two expressions are in separate face regions, the mapping is almost equivalent to the sum of the two expressions. Left: the existing expression. Middle: result from geometric warping. Right: result from ERI.*

Figure 9: *Expression mappings with different lighting conditions. Left: neutral face. Middle: result from geometric warping. Right: result of ERI.*

mouth create the visual effects of skin-bulging. It exhibits a more natural and convincing smile.

Figure 10: *Mapping a smile. Left: neutral face. Middle: result from geometric warping. Right: result from ERI.*

Figure 11 shows the result of mapping the same smile expression in Figure 5(b) to Mona Lisa. The left image in Figure 11 is the image generated by Seize and Dyer [16] using their view morphing technique (we scanned the picture from their paper). The image on the right is the result generated using our method. The wrinkles around her two mouth corners make her smile look more natural and convincing than the one in the middle which is generated using geometric warping.

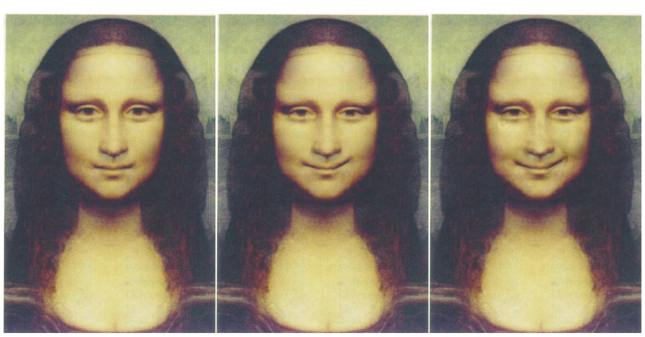


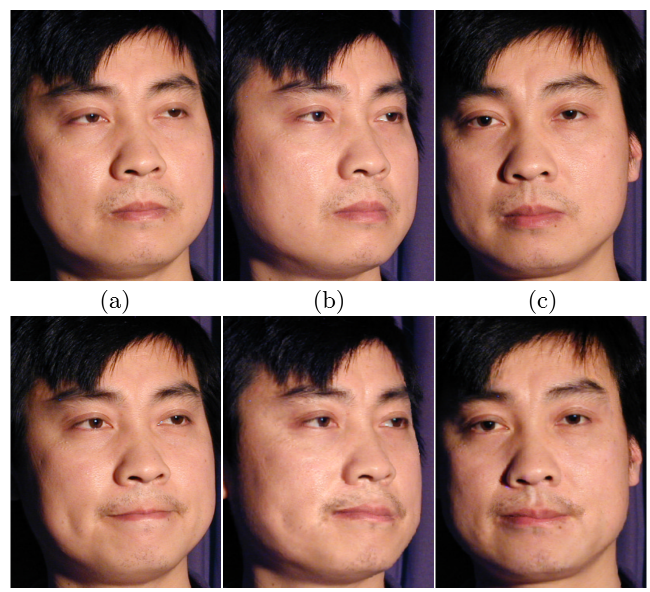
Figure 11: *Mapping a smile to Mona Lisa's face. Left: "neutral" face. Middle: result from geometric warping. Right: result from ERI.*

Figure 12 shows the results of mapping the smile expression of Figure 5(b) to two statues. The images of both statues are downloaded from the web. The wrinkle around her left mouth corner and illumination changes on the left cheek are mapped nicely to both statues. The more subtle wrinkle around her right mouth corner is mapped to (b) as well. However, it does not get mapped to (a) because of the shadow on this statue's right face.

(a) (b)

Figure 12: *Mapping expressions to statues. (a)Left: original statue. (a)Right: result from ERI. (b)Left: another statue. (b)Right: result from ERI.*

When the poses of the two faces are different, the mapping may fail. Figure 13 shows such an example. (a), (b) and (c) are the same neutral faces with different poses. (e) and (f) are the results of mapping expression (d) to (b) and (c), respectively. Notice that the original expression in (d) has a dimple on his right face. Because the pose of (b) is different from (a), the dimple in (e) is not as clear as in the original expression. The difference between the poses of (c) and (a) is even larger and the dimple does not get mapped at all.



(d) (e) (f)

Figure 13: *Expression mapping may fail when the poses are too far apart. (a), (b), and (c): neutral faces with different poses. (e): result of mapping (d) to (b). (f): result of mapping (d) to (c).*

6 Conclusions

We have shown that expression ratio image is an effective technique to enhance facial expression mapping with illumination changes. An expression ratio image can capture subtle but visually important details of facial expressions. The resulting facial expressions are significantly more expressive and convincing than the traditional expression mapping based on geometric warping.

The proposed approach can be extended to facial expression mapping for 3D textured face models. In this case, we only need to apply ERIs to the texture image, and we can obtain more expressive facial expressions for 3D facial animation.

This technique also applies to shaded 3D face models. One could map expressions between synthetic face models, as well as map between real face expressions and synthetic face models.

7 Limitations and future directions

One limitation of this method is in dealing with different lighting conditions. Even though we had some success of applying ERI to expressions under different lighting environment with the help of histogram matching, a more general solution is to use advanced relighting techniques such as [6, 12, 15].

Currently the image marks are very sparse. It is desirable to add line and curve features for better facial feature correspondences. We are planning to implement better image warping techniques such as those reported in [2, 10]. Better image warping should reduce the artifacts of the triangulation-based warping method that we currently use. We also hope that better image warping technique together with line and curve features will improve pixel correspondences. High-quality pixel correspondences could reduce the need of ratio image filtering, thus allowing more expression details to be captured in an ERI.

Acknowledgements

We would like to thank Ko Nishino for helpful discussions and many helps with image acquisition. We would like to thank Conal Elliott for carefully reading our manuscripts and providing valuable comments. We would like to thank Alex Colburn, Steve Harris, Chuck Jacobs, Brian Meyer, Sing Bing Kang, Sashi Raghupathy, and Emiko Unno for their help with image acquisition. We would like to thank Michael Cohen for his support.

References

[1] N. Badler and S. Platt. Animating facial expressions. In *Computer Graphics*, pages 245–252. Siggraph, August 1981.

[2] T. Beier and S. Neely. Feature-based image metamorphosis. In *Computer Graphics*, pages 35–42. Siggraph, July 1992.

[3] C. Bregler, M. Covell, and M. Slaney. Video rewrite: Driving visual speech with audio. In *Computer Graphics*, pages 353–360. Siggraph, August 1997.

[4] S. E. Brennan. *Caricature Generator*. M.S. Visual Studies, Dept of Architecture, Massachusetts Institute of Technology, Cambridge, MA., Sept. 1982.

[5] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Computer Graphics, Annual Conference Series*, pages 189–198. Siggraph, July 1998.

[6] P. E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Computer Graphics, Annual Conference Series*, pages 145–156. Siggraph, July 2000.

[7] J. Foley, A. van Dam, S. Feiner, and J. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, 1992.

[8] R. Gonzalez and R. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, 1992.

[9] Y. Lee, D. Terzopoulos, and K. Waters. Realistic modeling for facial animation. In *Computer Graphics*, pages 55–62. Siggraph, August 1995.

[10] P. Litwinowicz and L. Williams. Animating images with drawings. In *Computer Graphics*, pages 235–242. Siggraph, August 1990.

[11] S. R. Marschner and D. P. Greenberg. Inverse lighting for photography. In *IST/SID Fifth Colort Imaging Conference*, November 1997.

[12] S. R. Marschner, B. Guenter, and S. Raghupathy. Modeling and rendering for realistic facial animation. In *Rendering Techniques*, pages 231–242. Springer Wien New York, 2000.

[13] F. I. Parke and K. Waters. *Computer Facial Animation*. AKPeters, Wellesley, Massachusetts, 1996.

[14] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin. Synthesizing realistic facial expressions from photographs. In *Computer Graphics, Annual Conference Series*, pages 75–84. Siggraph, July 1998.

[15] T. Riklin-Raviv and A. Shashua. The quotient image: Class based re-rendering and recongnition with varying illuminations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 566–571, June 1999.

[16] S. M. Seize and C. R. Dyer. View morphing. In *Computer Graphics*, pages 21–30. Siggraph, August 1996.

[17] A. Stoschek. Image-based re-rendering of faces for continuous pose and illumination directions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 582–587, 2000.

[18] D. Terzopoulos and K. Waters. Physically-based facial modeling and animation. *Journal of Visualization and Computer Animation*, 1(4):73–80, March 1990.

[19] K. Waters. A muscle model for animating three-dimensional facial expression. *Computer Graphics*, 22(4):17–24, 1987.

[20] L. Williams. Performace-driven facial animation. In *Computer Graphics*, pages 235–242. Siggraph, August 1990.

[21] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, 1990.