Image-based Photorealistic 3-D Face Modeling

In Kyu Park, Hui Zhang**, Vladimir Vezhnevets***, and Heui-Keun Choh***

** School of Information and Communication Engineering, INHA University, Incheon, KOREA*
*** Multimedia LAB., Samsung Advanced Institute of Technology, Yongin, KOREA*
****Dept. of Computational Mathematics and Cybernetics, Moscow State University, Moscow, RUSSIA*
{Email: pik@ieee.org, hui_zhang@sait.samsung.co.kr, vvp@graphics.cs.msu.su, hkchoh@samsung.com}

Abstract

In this paper, we describe an automatic system for 3-D photorealistic face modeling from frontal and profile images taken by an uncalibrated handheld digital camera. The system employs a generic model adaptation framework. That is, after the fiducial features are detected from both images, a generic head model is deformed to match with the detected features. Realistic texture is created by combining the augmented facial textures from input images and synthesized texture, and mapped onto the deformed generic head model. Our shape deformation and texture generation algorithms have several advantages: (1) They generate photorealistic models even when the orthogonal assumption of input image pairs is not satisfied. (2) The generated ear has very accurate smooth shape and improves the quality of the whole face model dramatically. (3) A few supportive techniques on texture processing are employed to improve the visual quality.

1. Introduction

Automatic generation of realistic 3-D human head models has been a challenging task in computer vision and computer graphics for many years. Various applications like virtual reality, computer games, video conferencing, and 3-D animation could benefit from photorealistic human face models. Although there exist hardware devices like laser range scanner and structured-light range finder that can capture accurate 3-D shape of complex objects, they are very expensive and uneasy to use. Our goal is to make photorealistic human head models very easily for a common PC user. In order to achieve this, it is required to use affordable devices like digital camera for data acquisition, to maximize possible level of automation along with controls over the creation process, and to develop robust algorithms which generate plausible results even in case of imperfect input data.

Image-based face modeling has been explored by many researchers. Several works are devoted to creation of face models from two or three views [1, 2]. The views should be strictly orthogonal, which is hard to achieve when using common handheld camera without special set-up. A few systems [3] rely on user-specified feature points in several images, which is laborious procedure. Among other approaches, methods that utilize optical flow and stereo methods seem to be the farthest step towards the full automatic reconstruction process. However, due to the inherent problems of the stereo matching, the resulting models are often very noisy and exhibit unnatural deformations of the face surface. To avoid undesirable artifacts and to increase reconstruction robustness, a few researchers constrain the deformation to a limited set and then matched it to reconstructed 3-D points [4], to the source images [5], or use it in bundle adjustment for shape recovery [6, 7]. These approaches rely heavily on how accurate and extensive the set of possible model deformations is. Some researches tried to build a model from just one frontal facial image by using depth estimation from 2-D facial feature points [8], but the resulting model is yet far from photorealistic because of the lack of both shape and texture information.

The contribution of this paper is to develop a practical automatic system for generating a 3-D photorealistic facial model from frontal and profile images, without imposing strict picture taking conditions like illumination, view angle, and camera calibration. In this aspect it is similar to [9], but we can create a detailed 3-D model that is much more photorealistic. The proposed model deformation procedure works well even in case of imperfect imaging condition, in which the frontal and profile

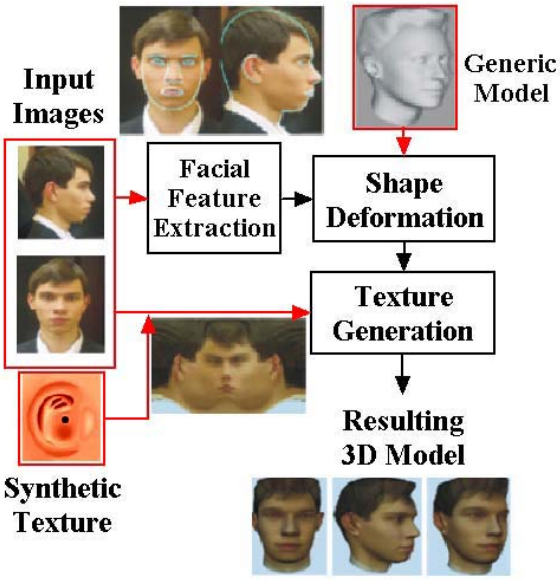


Figure 1. Overview of the proposed system.

images are not strictly orthogonal. The novelty of the proposed system also lies in the visually accurate modeling of ears and hair, which affects the appearance dramatically but has not received much attention in the community yet. The block diagram of the proposed system is shown in Figure 1.

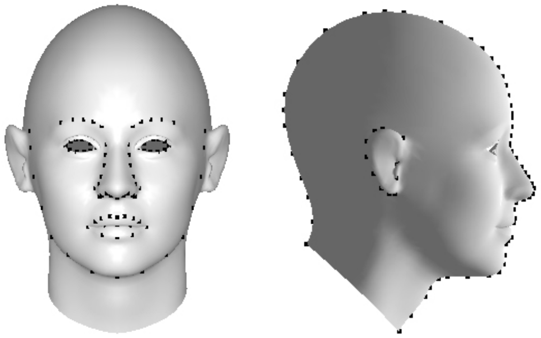
The paper is organized as follows. Section 2 and 3 are devoted to shape deformation and texture generation, respectively (We will present only the facial feature detection results in Section 2. Please refer to [11] for detailed discussion on algorithms). Section 4 describes the method of creating realistic appearance for ears and hair. Conclusion is given in Section 5.

2. Shape deformation

In the proposed system, new models are created by deforming the generic head model. The generic model is created by averaging a few laser-scanned faces and simplifying it with approximately 20,000 triangles. In the shape deformation stage, the generic model is deformed to fit the extracted facial features by employing the Radial-Basis Functions (RBF).

In order to cope with the conflicting facial features in the frontal and profile images, we develop two algorithms. One is the preprocessing of the detected profile facial features; the other is the new deformation algorithm.

2.1. RBF interpolation



(a) (b)

Figure 2. The key points marked in the generic model, which are detected from frontal (a) and profile (b) views, separately.

Suppose there are two corresponding data set $\{\overline{u}_i\} \subset \mathbf{R}^3$ and $\{\overline{u}_i'\} \subset \mathbf{R}^3$ representing N pairs of vertices, *i.e.* the key points, before and after the data interpolation, we can determine the following $f(\overline{p})$ as the deformation function of the whole 3-D space [10].

$$f(\overline{p}) = \overline{c}_0 + \begin{bmatrix} \overline{c}_1 & \overline{c}_2 & \overline{c}_3 \end{bmatrix} \overline{p} + \sum_{i=1}^{N} \overline{\lambda}_i \varphi_i (|\,\overline{p} - \overline{u}_i\,|) \quad (1)$$

where $\overline{p}$ is any 3-D vector and φ_i is the RBF for $\overline{u}_i$. In our implementation $\varphi_i(r) = e^{-r/K_i}$ where K_i is a predetermined coefficient that defines the influence range of $\overline{u}_i$. $\overline{c}_0, \overline{c}_1, \overline{c}_2, \overline{c}_3$ and $\overline{\lambda}_i$ are all 3-D coefficients, which are determined by solving the following constraints.

$$\begin{aligned} f(\overline{u}_i) &= \overline{u}_i'\Big|_{i=1}^{N} \\ \sum_{i=1}^{N} \overline{\lambda}_i &= 0 \\ \sum_{i=1}^{N} \overline{u}_{i,j} \overline{\lambda}_i &= 0\Big|_{j=x,y,z} \end{aligned} \quad (2)$$

Based on the RBF-based deformation framework and the 3-D correspondences of the given key points between the feature points in the generic model and the input image, the general model is deformed so as to generate the complete individual model by feeding all the vertices in the generic model into (1). The key points are defined as fiducial facial features, as shown in Figure 2.

2.2. 2-D/3-D key points generation

The key points represent perceptually important facial feature points, by which most of the geometric characteristics of individual human can be determined. In our approach, 160 fiducial points are sampled from

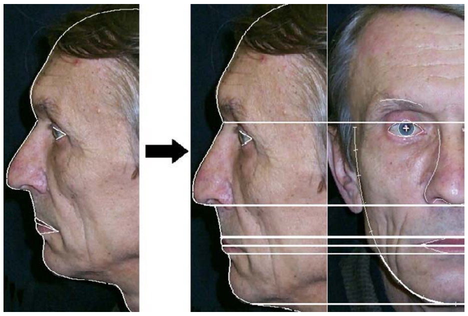


Figure 3. Local transformation on profile feature points.

the detected facial feature curves including lip, eye, eyebrow, chin and cheek contours. Note that the contour curves are extracted by using our previous algorithm [11]. The key points are sampled along each contour such that it has the same sampling rate in contour length as the key points on the corresponding contour of 3-D generic model. Note that the key points in 3-D generic model are predefined fixed points.

2.3. Preprocessing of the profile key points

Before performing model deformation, the detected profile key points are preprocessed to cope with the incoherent frontal and profile facial features due to the non-orthogonal picture-taking condition. Two steps, a global transformation and a local one, are followed to fulfill this task.

During the global processing, all profile features key points are scaled and rotated to match with the frontal image. The scaling is necessary to compensate for different focal length and viewing distance. We rely on the vertical distance of two fiducial points, the nose bridge top (NBT) and the chin point (CP), to determine the scale coefficient. A rotation is also needed because the camera may be slightly rotated around the optical axis, or the pitch angle of the specific head is changed during picture-taking stage. However, the rotation angle is hard to determine because of the lack of available information for this purpose. In our system we assume the angle between the vertical axis and the line connecting NBT and CP to be constant for individuals and try to rotate the profile key points such that the angle between the line and the vertical axis is equal to the angle which is observed in the 3-D general model.

During the local transformation, the profile line is partitioned to different segments based on the key points including the nose, mouth and chin points. Each segment is scaled separately to fit with frontal information as shown in Figure 3.

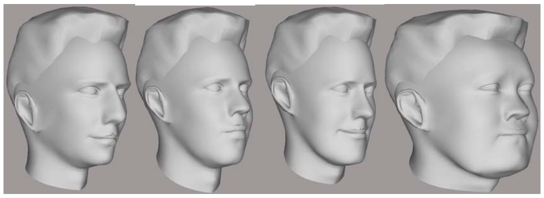


Figure 4. Examples of created head models.

2.4. Model deformation

In order to deform the generic model and generate photorealistic 3-D face model, a three-step deformation procedure is developed, which is described as follows.

First, we apply RBF interpolation with the frontal key points. Assume that the generic model is facing to **Z** direction, while the horizontal direction of the frontal face is aligned to **X** coordinate. The **X** and **Y** coordinates of displaced key points are set to their corresponding image positions under frontal view, and the **Z** coordinates remain the same as in the generic model. After RBF interpolation, the accurate match is obtained between the 3-D generic model and the facial features in the frontal image. We denote the deformation result of this step as **M1**.

Next, we utilize the profile key points to perform RBF interpolation. The **Y** and **Z** coordinates of displaced key points are set to their corresponding profile image positions, and the **X** coordinates remain the same as in the generic model. During this step we determine the **Z** values for all model vertexes. We denote the deformation result of this step as **M2**.

Finally all the key points are used to determine the resulting head shape. In this step the profile key points remain their positions in **M2**. As for the frontal key points, we set their **Z** coordinates according to **M2**, and set **X** and **Y** coordinates to **M1**. The final resulting shape matches with the detected 2-D facial features in both view angles. A few examples of created head shape are shown in Figure 4.

3. Texture generation

Being mapped onto the deformed 3-D wire-frame model, texture provides the major contribution to the visual appearance and perception of the model. In this section, we describe the texture generation method, including texture mapping to a public plane and blending frontal and profile texture.

3.1. Texture mapping to a public plane

In order to combine the texture from different view angles, we create a public UV plane containing texture coordinates of model vertices. It is a 2-D plane where the points are matched with vertex positions of the 3-D model. The plane has the normalized coordinate space, *i.e.* $\mathrm{M} = \{(u,v) | u,v \in [0,1]\}$. Such a public coordinate space makes it easy to combine the texture from real photo and synthesized texture.

Since human head is similar to a sphere-like shape, the spherical texture mapping is used to maximize the uniform distribution of the model vertices across the texture coordinate space. It is required to repair the directly generated space manually to solve the overlap problem (*i.e.* in ear areas) on the UV plane. However, since it is done once for the generic model in offline processing, no manual user interaction is required in the modeling stage.

To create a textured model, we map the color-adjusted photos onto the UV plane. Since the model key points have been fitted to the image feature points, the correspondences between model vertices and the image feature positions are already known as described in Section 2.2.

3.2. Texture blending

Before texture blending, the frontal and profile images are preprocessed to compensate for different color scale, by employing the correction algorithm [12]. Then, the frontal and profile textures are blended together to generate the final texture on the UV plane, which will be mapped onto the created 3-D model during rendering.

For each point $p \in \{(x,y) | x,y \in [0,1]\}$ in the UV plane, we obtain its color of the blended texture by the interpolation as follows

$$C(p) = k_f(p)C_f(p) + k_s(p)C_s(p) \qquad (3)$$

where C_f and C_s are the colors of point p in the frontal and profile textures, respectively; k_f and k_s are normalized weights for different textures satisfying $k_f(p) + k_s(p) = 1$ for every p.

The weights of texture blending are generated with a multi-resolution spline algorithm [13]. Based on Gaussian pyramid decomposition, this image blending algorithm can achieve smooth transition between images without blurring or degrading finer image details. It also benefits the utilization of blending boundary having arbitrary shape.

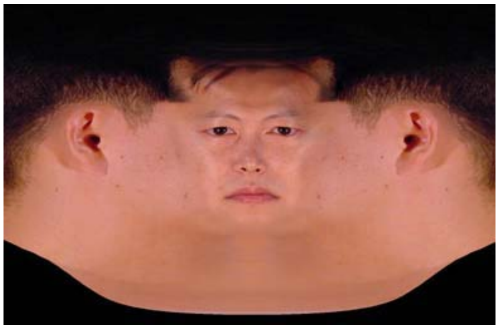


Figure 5. Blended texture.

In our approach, a couple of additional techniques are employed to improve the texture quality. One is the utilization of synthesized texture to improve texture around ear area, which will be described in next section. The other is to use artificial pixels for neck area, where strong shadow and self-occlusion are often observed. The artificial pixel colors are randomly generated according to the individual skin color statistics. Gaussian convolution is also applied to such area to remove possible noise.

A typical result of blended texture is shown in Figure 5, while some textured individual head models are shown in Figure 9.

4. Ear and hair modeling

Ears and hair affect the visual quality of resulting models greatly, so we pay special attention to generating accurate shape and texture for them.

The ear shape can be determined automatically during the model deformation stage. However, this usually gives unsatisfactory results since the ear vertexes are too far away from the head center, and sometimes RBF interpolation generates strange shapes for extrapolation. In fact, the ear shape is so complex that it is very difficult to be modeled accurately from image information. In our system we use a fixed ear shape from the generic model, scale it to fit the image ear size, and combine it smoothly with the created head models. Such a simple strategy avoids the ugly ear shape.

A key issue of this strategy is the smooth shape combination between the scaled generic ear and the deformed head model. This is also solved with RBF data interpolation. In detail, we manually specify two boundaries on the ear patch, as shown in Figure 6. The outside boundary is the blending boundary between the deformed head model and the ear patch, while the inside boundary separates the ear patch into a facial part and a stitching-out one. Then for each deformed

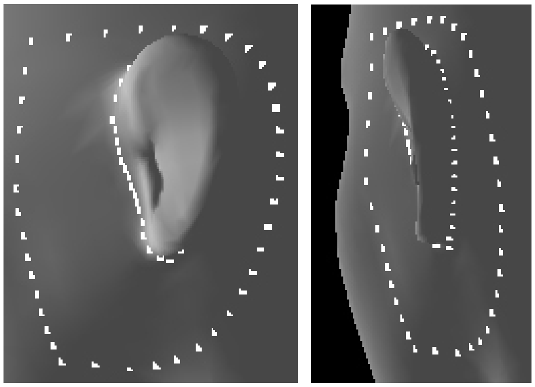


Figure 6. Outside and inside boundaries for ear shape combination (observed from different view angles).

model, we first perform a global scale transform on the ear patch to make the lengths of the blending boundaries on two models match with each other. The ear patch is also translated to make the center of blending boundary vertexes coincide to that of the head model. Afterwards a RBF interpolation is performed for ear vertexes between two boundaries. We displace the blending boundary vertexes to the corresponding positions on the head model, and remain inside boundary vertexes at their original positions. Finally, the stitching out part can be scaled to match with the image ear size. This fully automatic algorithm creates smooth blending results in our experiment.

As for the ear texture, we assign some synthesized texture (as shown in Figure 1) to the occluded ear areas. First it is color adjusted to match with the individual skin color. Then spline based texture blending algorithm is employed again to add this patch into the profile texture image. The improved ear texture is shown in Figure 7. The ear boundary is detected by matching a curve template to the profile image. The complete ear detection consists of three steps: (1) profile image normalization to compensate for different scale and orientation, (2) ear initialization to match the template with the image ear and to translate it to an initial position, and (3) ear refinement to deform the template to match the accurate ear boundary.

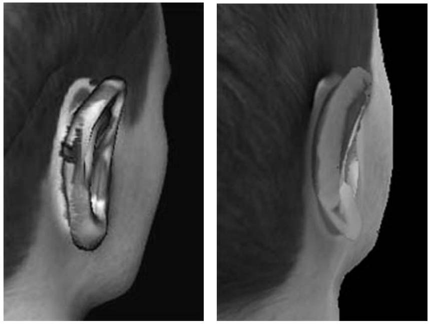


Figure 7. Improved ear with synthesized texture.

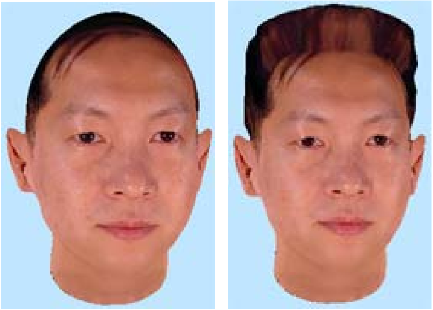


Figure 8. Created model with and without hair shape.

Currently we use a very simple method to generate individual hair. In the generic model we create some polygons to represent hair shape, and deform it during the RBF interpolation stage. The hair texture is obtained automatically since we assign the texture coordinates on the public UV plane for hair vertexes. This simple method really improves the appearance, as shown in Figure 8.

5. Conclusion

A face modeling system that produces a polygonal textured head model from frontal and profile images is described. Since it is designed to be used by a common user with cheap equipments, it is highly automated and robust. A deliberately designed shape deformation algorithm is proposed to achieve this goal. A variety of the generated 3-D face models are shown in Figure 9.

6. Acknowledgement

We would like to thank Eugene Lisitsin, Anna Degtiareva, Denis Krasavin, Stanislav Soldatov, Alla Andreeva in Moscow State University, and Artem Andreev in Samsung Advanced Institute of Technology, for their contribution to the development and the implementation of the face modeling system.

Also, we would like to thank Dr. Yang-Seock Seo, Dr. Seok Yoon Jung, and Mr. Mahnjin Han in Samsung Advanced Institute of Technology, for their consistent support to the project.

7. References

[1] B. Dariush, S. Kang, and K. Waters, "Spatiotemporal analysis of face profiles: Detection, segmentation, and registration," *Proc. of Third International Conference on Automatic Face and Gesture Recognition*, pp. 248-253, Nara, Japan, April 1998.

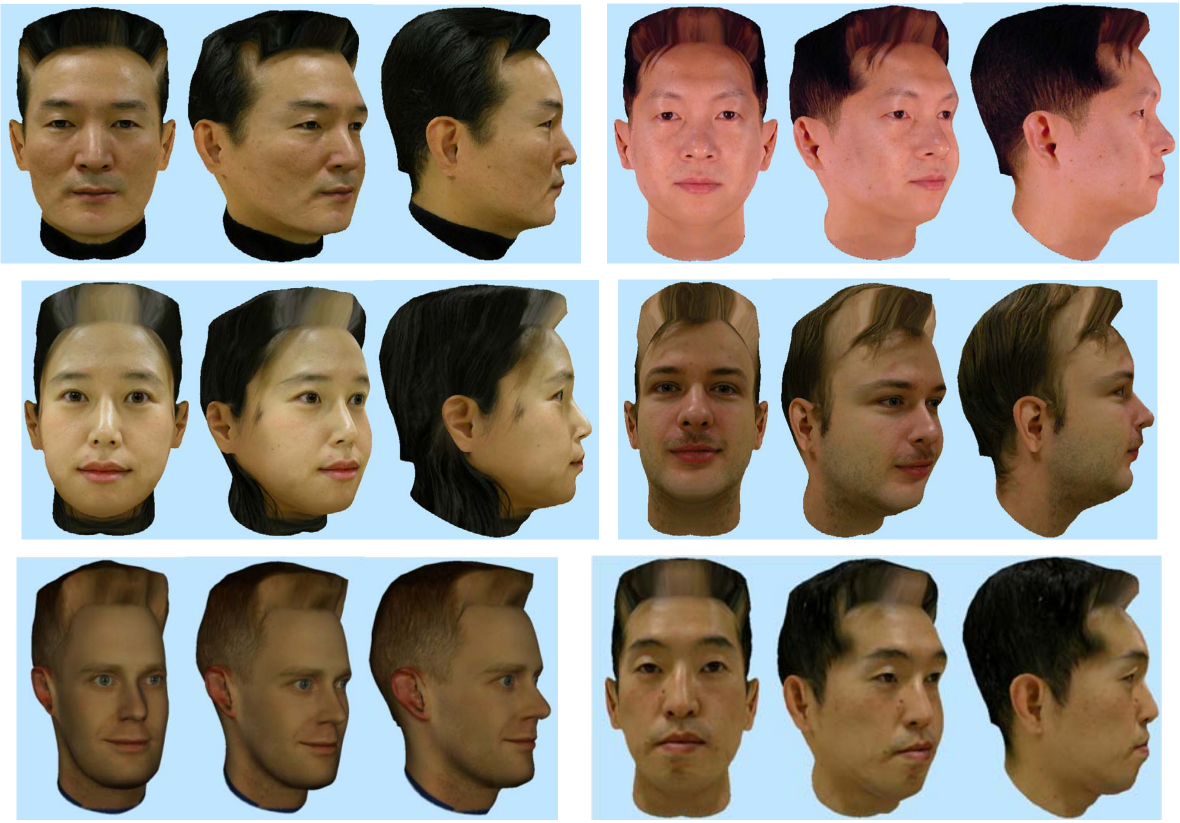


Figure 9. Face models generated by the proposed system.

[2] K. Lee, K. Wong, S. Or, and Y. Fung, "3-D face modeling from perspective-views and contour-based generic-model," *Real-Time Imaging*, vol. 7, no. 2, pp. 173-182, April 2001.

[3] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. Salesin, "Synthesizing realistic facial expressions from photographs," *Proc. of SIGGRAPH '98*, pp. 75-84, Orlando, USA, July 1998.

[4] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen, "Rapid modeling of animated faces from video," *Journal of Visualization and Compute Animation*, vol. 12, no. 4, pp. 227-240, September 2001.

[5] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3-D faces," *Proc. of SIGGRAPH '99*, pp. 187-194, Los Angeles, USA, August 1999.

[6] Y. Shan, Z. Liu, and Z. Zhang, "Model-based bundle adjustment with application to face modeling," *Proc. of IEEE International Conference on Computer Vision*, vol. II, pp. 644-651, Vancouver, Canada, July 2001.

[7] P. Fua, "Regularized bundle-adjustment to model heads from image sequences without calibration data," *International Journal of Computer Vision*, vol. 38, no. 2, pp. 153-171, July 2000.

[8] C. Kuo, R. Huang, and T. Lin, "3-D facial model estimation from single front-view facial image," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 183-192, March 2002.

[9] N. Sarris, N. Grammalidis, and M. Strintzis, "Building three dimensional head models," *Graphical Models*, vol. 63, no. 5, pp. 333-368, September 2001.

[10] J. Carr, R. Beaton, J. Cherrie, T. Mitchell, W. Fright, B. McCallum, and T.R. Evans. "Reconstruction and representation of 3-D objects with radial basis functions," *Proc. of SIGGRAPH 2001*, pp. 67-76, Los Angeles, USA, August 2001.

[11] V. Vezhnevets, S. Soldatov, A. Degtiareva, and I. Park, "Automatic extraction of frontal facial features for 3-D face modeling," *Proc of Sixth Asian Conference on Computer Vision*, pp. 1020-1025, Jeju, Korea, January 2004.

[12] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics & Applications*, vol. 21, no. 5, pp. 34-41, September/October 2001.

[13] Peter J. Burt and Edward H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics*, vol. 2, no. 4, pp. 217-236, October 1983.