PAPER

Instant Casting Movie Theater: The Future Cast System

Akinobu MAEJIMA^{†a)}, Shuhei WEMLER^{††}, Tamotsu MACHIDA^{†††}, Masao TAKEBAYASHI^{††††}, *Nonmembers,*
and Shigeo MORISHIMA[†], *Member*

SUMMARY We have developed a visual entertainment system called "Future Cast" which enables anyone to easily participate in a pre-recorded or pre-created film as an instant CG movie star. This system provides audiences with the amazing opportunity to join the cast of a movie in real-time. The Future Cast System can automatically perform all the processes required to make this possible, from capturing participants' facial characteristics to rendering them into the movie. Our system can also be applied to any movie created using the same production process. We conducted our first experimental trial demonstration of the Future Cast System at the Mitsui-Toshiba pavilion at the 2005 World Exposition in Aichi Japan.

key words: *automatic face modeling, real-time character direction, rendering cluster system*

1. Introduction

Conventionally, most visual attraction systems in event halls or amusement parks present a non-interactive image to a visitor, such as a movie with a fixed scenario. Sometimes, these systems, such as IMAX or Motion ride systems, offer an immersive experience where the audience can experience realistic sensations. In addition, Virtual reality and Mixed reality systems can provide an interactive experience for the user by utilizing real-time CG, based on a user-controlled scenario. For many of us, though, this kind of experience does not go far enough in helping us to achieve our dream of actually appearing in a movie alongside famous movie stars. In this paper, we will detail a type of visual entertainment system that we have developed that enables visitors to appear as a 3DCG member of the cast of a pre-recorded film. We have named this the Future Cast System. Future Cast was first exhibited at the Mitsui-Toshiba pavilion at the 2005 World Exposition in Aichi Japan. The Future Cast System offers two key features. First, it provides each member of the audience with the amazing experience of participating in a movie in real time. Second, the system can automatically perform all of the processes required to make this happen: from capturing the shape of each audience member's face,

to generating a corresponding CG face in the movie.

To create a CG character which closely resembles the visitor's face with our system, a 3D range scanner first acquires the 3D geometry and texture of the visitor's face. Second, we extract facial feature points, such as eyes, eyebrows, nose, mouth, and the outline of the face from the frontal face image. At the same time, the system also estimates the visitor's gender and age, which are required for the system to allocate a suitable role for the participants in the story. Next, a generic facial mesh is adjusted to smoothly fit over the feature points of the visitor's face. We acquire the scanned depth of the visitor's face which corresponds to vertices on the deformed generic facial mesh, replace this with the scanned depth, and then attach the visitor's frontal face image to the mesh. The entire process, from the face scanning stage to the CG character generation stage, is performed automatically as soon as possible before the show starts. Then, scanned data, including the face model, age, gender information, and so on, are transferred to a file server. Next, attendants working in the pavilion judge whether each CG character's face is precisely modeled enough to appear in the film. This judgment depends on the degree of polygon collapse for each face model. If the face modeling process is incomplete, default face models from former viewers' faces are generated to complete the cast.

At the time of the show, we call up the visitor's character information, such as facial geometry, texture, and so on, from the file server and control the face using Cast/Environment Scenario Data which is defined for each individual cast member and scene of the movie. Then, the face, which is produced by facial expression synthesis and re-lighting, is rendered into the face region in the scene image. We can therefore generate animation in which the CG character speaks and expresses his or her feelings as if it were the visitor. This processing is performed by Image Generation PCs (IGPC) dedicated to image composition. The synthesized images from the IGPCs are transferred to an image Projection PC (PPC) and the movie is then played on the screen.

This paper is organized as follows: in Sect. 2, we indicate what type of performance we required from our system and related work. We then describe the hardware design of our system in Sect. 3. In Sect. 4, we explain how each visitor's 3D face model is generated, and describe the process of the real-time direction of a visitor's face. In Sect. 5, we explain the experimental results of the Future Cast System. In

Manuscript received April 16, 2007.

Manuscript revised November 20, 2007.

[†]The authors are with the Waseda University, Tokyo, 169-8555 Japan.

^{††}The author is with the Sliconstudio Corp., Tokyo, 150-0013 Japan.

^{†††}The author is with the Little Studio Inc., Tokyo, 150-0033 Japan.

^{††††}The author is with the DENTSU TEC Inc., Tokyo, 104-8477 Japan.

a) E-mail: akinobu@mlab.phys.waseda.ac.jp

DOI: 10.1093/ietisy/e91-d.4.1135

Sect. 6, we discuss our current and future work, and, finally, we note our conclusions. In this section, we describe our performance requirements for our system. Then we give an overview of related work in the areas of facial feature point-based face recognition, facial modeling and animation, and facial relighting techniques.

1.1 System Requirements

During the design phase of the Future Cast System, we identified our major performance requirements. In order to create a satisfactory version of the Future Cast System, it was essential that we meet these technical standards.

Easy and Instant Casting We needed to design a system that would automatically perform all processes, from capturing facial images to generating the final movie as quickly as possible, because it would be necessary for us to create CG characters for many people at the same time.

Personal Identification Here, the most important issue was that a visitor should be able to identify his/her own CG character with 90 ~ 100% accuracy, when viewing the movie in which his/her CG character was embedded.

Robust Rendering System During the movie show-times, the system has to keep working without experiencing any system failures. It was therefore important for us to construct a robust rendering system.

1.2 Related Work

Face geometry acquisition 3D range scanners such as the Cyber-ware 3030/RGB, Minolta Vivid 910, NEC DANAE-R and Sanyo PIERIMO [1] are already commercially available. These scanners are able to capture face geometry with a high degree of accuracy. However, it is difficult for the Cyber-ware 3030/RGB, Vivid 910, and NEC DANAE-R to capture the 3D geometry of areas which absorb lasers or sine-wave patterns such as the eyes, the inside of the mouth, and hair. In addition, these scanners are too expensive to use for our work. However, if a face is illuminated by rainbow stripes and the facial silhouette is complete, the PIERIMO can measure the facial geometry of these areas. In addition, this scanner is comparatively inexpensive. With our system, it was necessary to acquire the depth of pupils to estimate the center position of eyes in order to create eye model, so we developed a scanner based on [1].

Face recognition Wiskott et al developed elastic bunch graph matching using Gabor wavelets [2]. Kawade and Lee also developed techniques to improve this method [3], [4], and have been able to perform facial feature point extraction with a high degree of accuracy. This method can reliably extract facial features such as ethnicity, gender, age, facial hair, and accessories from frontal face images. Li et al have evaluated this technique using their facial database and, as

a result, 77% of automatically extracted feature points were within 5 pixels of the ground truth, and only 2% of the points were more than 10 pixels in error. Baback and Lian have developed age and gender estimation techniques [5], [6] that are able to estimate age and gender with more than 80% accuracy.

Facial animation Since the earliest work in facial modeling and animation [7], generating realistic faces has been the central goal. A technique to automatically generate realistic avatars has been developed using facial images [8]. The CG face model is generated by adjusting a generic face model to target facial feature points, which are then extracted from the image. Other researchers in this field have been able to construct facial geometry directly from images using morphable facial models [9]–[11].

Current major facial synthesis techniques include: image-based, photo-realistic speech animation synthesis using morphed visemes [12], a muscle-based face model has been developed [13]–[15], as has a blend shape technique where facial expressions are synthesized using linear combinations of several basic sets of pre-created faces for blending [11], [16]–[19]. Additionally, a facial expression cloning technique has also been developed [20]. Although each technique is able to synthesize highly realistic human facial expressions, in especially the muscle-based approach, it is difficult to create target expressions using the same muscle activations for different people, since each person's muscular layout is slightly different. In addition, an animator needs experience and knowledge of muscular characteristics in order to create target facial expressions. Moreover, if there are many characters in one scene of the movie, it is difficult to synthesize facial expressions by contracting facial muscle in real time.

The blend shape-based approach shares the same problem when applied to individuals. However, we can prepare a variety of basic shape patterns to represent the target facial expression in advance. In addition, animators can easily create target expressions, and little time is required for facial expression synthesis. Consequently, we decided to use the facial expression synthesis technique based on the blend shape-based approach with our system.

Face relighting To achieve photorealistic face rendering, we need to measure human skin's Bi-directional Reflectance Distribution Function (BRDF) and subsurface scattering of light by human skin, and construct a human skin reflectance model. [10] measured a uniform analytic BRDF from photographs. [21] developed a Bi-directional Sub-surface Scattering Reflectance Distribution Function (BSSRDF) with parameters based on biophysically-based spectral models of human skin. [22] published the method which generates a photorealistic re-lightable 3D facial model by mapping image-based skin reflectance features onto the scanned 3D geometry of a face. [23] combined an image-based model, an analytic surface BRDF, and an image-space approximation for subsurface scattering to create highly-realistic fa-

cial models for the movie industry. [24] developed a variant based on this method for real-time rendering on recent graphics hardware. [25] measured 3D facial geometry, skin reflectance characteristics and subsurface scattering using custom-built devices. They developed a new skin reflectance model where parameters can be estimated from measurements. The model breaks measured skin data down into a spatially-varying BRDF, a diffuse albedo map, and a diffuse subsurface scattering.

However, it is difficult for us to measure BRDF, skin subsurface scattering, and a diffuse albedo using the Light Stage or the Face Scanning Dome [22], [25], because of the prohibitive costs of these large scale measuring devices. Furthermore, if we measure reflectance characteristics of individual faces with these devices, huge data storage areas would be required. It is also difficult to measure and reproduce light from the subsurface scattering of human skin. The custom-built device proposed by [25] is able to capture skin subsurface scattering, but because the device is a contact type, we were unable to use this device due to hygiene factors.

To render faces, we use only the available face texture (diffuse texture) captured by a range scanner. Furthermore, in "Grand Odyssey" there are scenes where a maximum of five characters appear simultaneously. Thus, the rendering time for each character's face is restricted. An important role of our face renderer is to generate images that reduce the differences between visitors' faces and the pre-rendered story images. For skin rendering, we first use the texture of the scanned face together with uniformly specular lighting as a diffuse albedo. We then use a hand-generated specular map on which speculars on the forehead, cheeks, and tip of the nose become more non-uniformly pronounced than on other parts of the face. Additionally, the reflectance characteristics of the face, for example subsurface scattering and the appearance of laungo hair, are intentionally represented by varying light intensity on the outline of the face. We can thereby efficiently render an image similar in quality to that of the pre-rendered background CG image.

2. Hardware Design

2.1 System Structure

Figure 1 shows an overview of the Future Cast System. The system itself is divided into three sub-systems. The Face Modeling sub-system handles the process from capturing the visitor's face to automatically generating CG characters. The Rendering sub-system handles the rendering process where visitor's CG characters are embedded into the movie. The System Control sub-system, deals with the process of scan checks by supervisory attendants. Here, the quality of scans of visitors' faces are confirmed to be of sufficient quality before the movie begins. Additionally, the theater control staff cue the Rendering sub-system using a control command in order to start the movie.

The Face Modeling sub-system consists of five range

scanners which can measure the 3-D characteristics of objects, and twenty PCs for modeling 3-D shapes from images captured by the scanners (visitors' CG characters are generated by the same PCs after 3D shape modeling). The specifications of PCs for 3D shape modeling (called Modeling PCs) are equivalent to a Pentium 4, 3.0 GHz, with 1 GB of memory. The Modeling PCs are connected by 10/100 Base Ethernet. The Rendering sub-system consists of a Rendering Cluster System and a DLP Projector. The System Control sub-system consists of a PC for attendants to use to check the visitor's character model, a file server for storage of visitors' data, a time code generator for generating the movie's time code, and an audio matrix mixer for selecting the voice track.

2.2 Range Scanner

Using our system, the most important factor in creating a truly memorable experience for each member of the audience was to reproduce visitors' faces so accurately that each person could identify his/her own CG face in the movie. In addition, the 3D modeling process needed to be fully automatic with low computational requirements. This was because 240 CG face models were required to be generated in a pre-show time limit of 12 minutes, and there were to be no restrictions based on the visitor's age; so that elderly people and even babies could take part in the Future Cast experience.

For these reasons we needed to develop a range scanner which could instantly capture the characteristics of a human face placed into the scanner window and kept still for a few seconds. The range scanner consisted of seven 800×600 pixel CCD size digital cameras and two slide projectors which were able to project rainbow stripe patterns. Figure 2 shows the appearance of the range scanner. The seven digital cameras were placed in a semi-circular pattern around a window where each visitor would place his/her face, with three cameras positioned on each side and one directly facing the window. Two slide projectors, responsible for projecting rainbow stripe patterns onto the visitors' faces, were placed at the 10 o'clock and 2 o'clock positions as viewed from the window. The three cameras on both sides captured two images each as rainbow stripe patterns were projected onto targets under fluorescent lighting conditions. Figure 3 shows captured images used for 3-D shape modeling. The 3-D shapes were modeled using the Hybrid method (Active Stereo technique and Shape-from-Silhouette) [1]. Figure 4 shows the reconstructed 3D geometry of a face. As the depth of a captured face corresponds to a frontal face image, if we assign (u, v) coordinates to the frontal face, we can determine the depth of each corresponding pixel of the depth map. The time required to complete this 3-D shape modeling process on the modeling PC is approximately 2-3 minutes.

Fig. 1 System overview.

Fig. 2 The face modeling system: Seven digital cameras (gray arrows) and two slide projectors (white arrows) are arranged in a semi-circle.

Fig. 3 The images are obtained with six cameras arranged on both sides of the 3D scanner. Top (left to right) are images with a rainbow stripe pattern on the face. Bottom (left to right) are images taken under fluorescent lighting conditions. 3D geometry is reconstructed from these images.

Fig. 4 Modeling result from the face modeling system. Left: a personal frontal face image. Middle left: Personal 3D geometry viewed from the front. Middle Right: Personal 3D geometry viewed from the left side. Right: Personal 3D geometry viewed from the right.

2.3 Rendering Cluster System

The most important requirement for the rendering system was to efficiently divide the computational workload between computers to ensure reliability. We developed a real time rendering system that would run as consistently as possible throughout the movie. We called this the Rendering Cluster System.

Figure 5 shows the Render Cluster System. This system consists of eight PCs used for frame Image Generation (Image Generator PC; IGPC) and a PC for sending the image to one Projector per screen (Projection PC; PPC). The IGPC had a 3.2 GHz CPU, 512 MB memory, NVIDIA Geforce FX5950, and RAID to improve the transfer speed of the hard disks. The specifications of the PPC were equivalent to Xeon 3.0 GHz, 4 or 2 GB memory, with NVIDIA Quadro FX3000G. IGPCs and PPC were connected to each other by gigabit networks. In addition we used JumboFrame

Fig. 5 Composition of rendering cluster system.

to improve the transfer rate on each ethernet port. At first, the PPC receives control commands and time codes from the control system. The PPC also receive visitor's CG character information from the file server (including face geometry, texture, and parameters for facial expressions) then the PPC sends this information to each IGPC. Each IGPC then renders the story image containing the visitor's CG character. The PPC receives the image rendered by each IGPC, and then sends frame images to the projector in synchronization with the time code.

In addition, we used a time division method for generating movie frames in the cluster system. The rendering processes of the movie is divided into 30 frames per second. The eight IGPCs render the movie's frames, and transmit them to the PPC. When the number of IGPCs is assumed to be N , the i -th IGPC creates the image of an $i + N \times m$ frame where m is an arbitrary integer. Since the number of IGPCs may change during operation as system errors can sometimes occur, the PPC can manage the IGPCs to ensure that images are generated continuously without any blank frames. If five out of the eight Image Generator PCs are running, the Rendering Cluster System can run a film in real time at 30 fps without skipping a single frame. If there is a problem with an IGPC, it is automatically re-booted and then restored to the rendering line.

3. Applications

3.1 CG Character Modeling

We developed the automatic CG character modeling module by customizing and combining previously developed computer vision-graphics techniques. Our CG character modeling module is capable of creating a CG character in approximately 5 seconds on the Modeling PC. This can be divided into four processes: facial feature point extraction, gender and age estimation, face fitting and face normalization. These processes automatically generate a personal CG character from a frontal face image and 3-D shape geometry.

3.1.1 Facial Feature Point Extraction

To generate a visitor's CG face using captured face texture and 3D facial geometry, it is first necessary to estimate the location of each person's features. From the captured face

Fig. 6 The result of feature points extraction (a) 89 feature points. (b) Interpolated 128 feature points.

texture, we extract facial feature points including the outline of the eyes, eyebrows, nose, lips, and face using the Graph Matching technique with Gabor Wavelet features [3], [4]. Figure 6 (a) shows the 89 facial feature points which are defined and placed to create a Facial Bunch Graph. We then register Gabor Wavelet features extracted from any facial images (including gender, ethnic characteristics, facial hair, etc) into each node of the Facial Bunch Graph. To enable our feature point detector to correctly extract facial feature points, we collected a total of 1,500 facial images as training data.

In the face fitting phase, which is explained in the following chapter, a generic face model is deformed using a scattered data interpolation method which utilizes extracted feature points as targets. This method enables us to deform the model smoothly. However deformed vertices on the model do not always correspond to outlines of face parts other than the target points. It is important that the vertex of the model conforms closely to the outline of eyes and the lip-contact line on the face image because these locations tend to move a lot. When a model that has been deformed somewhat roughly performs an action, such as speaking or blinking, it becomes difficult to obtain a natural appearance. We therefore need to adjust the model with a high degree of accuracy for the feature points on the outline of the eyes, lips, and especially the lip-contact line.

Consequently, we interpolated the areas between the 89 feature points using a cubic spline curve and then 128 re-sampled points on an approximated curve to achieve correspondence between the feature points. This process was performed for eyes, eyebrows, and the outline of the lip. Figure 6 (b) shows the result of our facial feature point extraction process.

3.1.2 Gender and Age Estimation

To estimate gender and age in order to automatically decide a visitor's role in the movie, gender and age estimation are performed by non-linear Support Vector Machine classifiers which classify Gabor Wavelet features based on Retina Sampling [5], [6]. The age information from the classifier's output is categorized into five groups: under 10, 10-20, 20-

40, 40-60, and over 60. Our facial gallery for age and gender recognition consists of 400 images for the under 10 category, 400 images for the 10-20 category, 600 images for the 20-40 category, 900 images for the 40-60 category and 800 images for the over 60 category (numbers of images apply equally to both males and females respectively). In addition, gender information is used to select audio tracks pre-recorded by actors/actresses.

3.1.3 Face Fitting

We fit the deformable generic face mesh with 1264 triangle polygons which are small enough in number to facilitate real-time rendering. In turn these 128 facial feature points are fitted to a common topology for facial expression synthesis. Facial feature points do not have a one-to-one correspondence to the vertices of the generic face mesh, so it is necessary to deform the model using scattered data interpolation techniques such as Radial Basis Functions [26]–[30].

$$R(\mathbf{x}_j) = \sum_{i=1}^N w_i \phi(\mathbf{x}_j - \mathbf{x}_i) \quad (0 \leq j \leq N) \quad (1)$$

$$\phi(t) = \begin{cases} 1 - \sqrt{1 - (1 - \frac{t}{\sigma})^2} & (0 \leq t \leq \sigma) \\ 0 & (t > \sigma) \end{cases} \quad (2)$$

The generic face mesh smoothly adjusts to the 128 facial feature points using Radial Basis Functions with a Quarter-Circle basis $\sigma = 600.0$ (Fig. 7 (b)). There are several processes involved in generating the visitor's 3DCG face for their role in the movie. First, we extract the 3D geometry (x, y, z) from the depth image by specifying (u, v) coordinates which are decided by facial feature points (Fig. 7 (a)). We then transform the vertices of the generic face mesh according to the characteristics of the extracted 3D geometry (Fig. 7 (c)). Then finally, we overlay the frontal face image (Fig. 7 (d)) onto the now individualized face mesh (Fig. 7 (e)).

3.1.4 Normalizing Faces

The last step in the CG character modeling process is to adjust the size, position and orientation of the visitor's face mesh to the generic one. The facial posture and size of the cast member's face in the movie are defined according to the generic face mesh. Furthermore, the facial posture varies according to the timing of the face capturing. Therefore, face normalization is essential. The process is as follows:

Position First, we calculate the center position between the eyes of both the visitor's face and that of the generic face model, and then overlay the faces with the visitor's on top. The center position is determined by the barycentric coordinates of all vertices taken from the outline of the eyes of the visitor's face mesh, because our feature point detector can reliably extract these positions.

Fig. 7 Visitor's CG character generation process (a) Visitor's 3D geometry (depth image). (b) Deforming a generic face model to fit facial feature points using radial basis functions. (c) The deformed mesh with visitor's depth. (d) Visitor's face image for texture mapping. (e) Finally, the visitor's face model is created. In addition, eye and inner-mouth models are arranged at this time.

Rotation We estimate the regression plane from vertices on the personal face mesh. The normal vector of this regression plane can be assumed to be current direction of a face. We then estimate the rotational matrix that makes the angle between the normal vector and the z-directional vector $(0, 0, 1)$ be 0 by [31]. However, vertices around the facial contour can not always be extracted precisely using our scanner, which may involve errors in depth. We only use pre-labeled vertices except facial outline during regression plane estimation, because individual faces have the same mesh topology. For the refinement of this normalization, unlike Hirose's approach [31], we iteratively perform this process until the difference of angles converges to 0 using pre-labeled vertices. Additionally, because the proposed system requires real-time processing, we also have to set the iteration threshold.

Scale Finally, we adjust the size of the individual face model. The individual face model is expanded or reduced in size in order to ensure that the ratio between the eye's inner edges equals that of the generic face model. This scaling ratio is also used to ensure that the facial width remains proportionate. After this scaling correction, the width ratio w_x and height ratio w_y between the individual face model and the generic face one are calculated. Especially when $(w_x + w_y)/2.0$ exceeds 1.0, the visitor's face can become too large compared with the generic face. In such situations, we perform a weighted reduction by using $2.0/(w_x + w_y)$, which prevents a particular face model from becoming too large.

Using this algorithm, the individual face model is normalized appropriately.

3.2 Controlling Real-Time CG Character

The process of controlling CG characters includes synthesizing facial expressions and relighting faces based on the light environment in each scene in the movie. The facial expression synthesis and the facial relighting should be per-

Fig. 8 Example of key shapes.

formed as quickly as possible, because it is necessary to render up to 5 CG characters at the same time in the movie.

3.2.1 Facial Expression Synthesis

Our facial expression synthesis module had to satisfy the following three requirements:

1. Synthesize natural facial expressions without polygon collapse.
2. Represent each visitor's facial expressions using a set of universal key shapes.
3. Operate in real time.

First, we create 34 universal key-shapes ($K = k_1, \dots, k_{34}$) based on "generic face mesh" (Fig. 8). Our module then calculates displacement vectors ($D = d_1, \dots, d_{34}$) between the neutral mesh (generic face mesh f) and the specific expression's mesh (one of 34 key-shapes k_i) such as mouth opening. Note that these displacement vectors can not represent individual facial expressions completely, so synthesized facial expressions are often stereotyped expressions.

Unlike conventional blend shape implemented on generic 3DCG software, our facial expression module automatically adjusts magnitudes for each "generic displacement vector" using weights w decided by the ratio of the face width and height between the individual face mesh and the generic one. This customization process is performed

after the face modeling process. The customized displacement vectors D are then transferred from the modeling PC to the IGPC through the file server before the movie starts.

$$b = f + \sum_{i=1}^{34} \beta_i^t w d_i \quad (3)$$

During the movie, the CG character's facial expressions are synthesized by blending customized displacement vectors $w d_i$ according to the pre-defined blending weight of each cast member's displacement vectors β_i^t on IGPC by Eq. (3). In this paper, β_i^t is referred to as "Cast Scenario Data", and is synchronized with each movie frame. Compared with other methods, this approach is effective in preventing unexpected polygon collapse and can run in real-time.

3.2.2 Relighting Face

Our goal in relighting the face is to generate a matching image between the visitor's face and the pre-rendered CG movie. In order to satisfy the above requirement, we first created a specular map which approximates the specular reflection of human skin. Second, we approximated the reflectional characteristics of the face, such as subsurface scattering and the appearance of laungo hair, by varying the intensity of light appearing on outlines of the face.

As background for our project, we utilized a specially created movie called "Grand Odyssey", in which all aspects of the movie, except for the character's faces, were generated in advance using 3DCG technology. This type of movie was more useful for our purposes than a live action movie because we were able to limit the scanning and rendering requirements to suit our specific focus. Also, it would have been difficult for us to measure each visitor's BRDF and BSSRDF, due to scanning time and financial constraints. Additionally, as up to five characters could appear in the movie at the same time, we needed to be mindful of rendering time constraints. Consequently, we chose to use only the reflectional characteristics of visitors' skin that could be acquired from an image of the front of the face. However, these images still contain a combination of ambient, diffuse and specular reflection. We therefore devised a lighting method to ensure a uniform specular reflection for the whole of the face, by installing an alignment of florescent lights fitted with white diffusion filters.

To represent skin gloss, we developed a hand-generated specular map on which speculars on the forehead, cheeks, and tip of the nose become more non-uniformly pronounced than other parts of the face. The intensity and sharpness of the speculars were generated by white gaussian noise. The map can be applied to all visitors' faces. Additionally, we approximated subsurface scattering and laungo hair lighting effects by varying the light intensity on the outlines of the characters' faces. Our system is able to render images between the visitor's 3D face model and the pre-rendered movie image without any disparity in lighting.

Fig. 9 An example of layer composite.

Environmental light parameters such as the number of lights, their location, types of light and the color of light for each scene in the movie, were defined when the movie was first created. These parameters are referred to as Environment Scenario Data. The faces of CG characters are illuminated in real time by the IGPC using this scenario data.

3.3 Automatic Casting and Voice Selection

In our system, it is necessary to automatically allocate movie roles to visitors who would like to join the cast of the movie. Roles are allocated at random, based on the cast list once the visitor's age and gender have been estimated (according to the process described in Sect. 4.1.2). Thus the cast of the movie varies each time it is shown, according to the composition of the visitors' group (number of male and female participants, etc). We pre-recorded a male and a female voice track for each character in the movie using actors. Voices are automatically assigned to characters according to the gender estimation results prior to the start of the movie.

3.4 Real-Time Movie Rendering

As it is only necessary to render the visitor's face into the movie, we were able to reduce required computational processing time. Each of the movie's image frames can be divided into four layers: the face layer, face mask layer, (faceless) background layer, and effect layer. The face mask layer, background layer and effect layer are pre-rendered and stored on the IGPC and PPC's HDDs. During real-time rendering, the face layer is rendered first. It is then cut out using the face mask layer. Finally, the composite face layer, background layer, and effect layer are merged. In Fig. 9 above, we show an example of the pre-completed image. This image is composed of the face layer and the background layer. The effect layer is added to complete the image frame. This process is controlled by a pixel shader created with High-Level Shader Language on a GPU. The use of the GPU enables the layer composition process to be performed quickly; reducing computational requirements.

4. Results

We performed two types of experiments with our Future

Cast System. The purpose of the first was to verify the robustness of the system by testing its long-term operation at the Mitsui-Toshiba pavilion at the 2005 World Exposition in Aichi, Japan. The second experiment was to determine whether the system could scan and embed visitors' likenesses into the movie accurately enough for the visitors to be able to identify themselves on screen.

4.1 Large Scale Theater

We implemented trials of the Future Cast System in two locations. The first was at the 2005 World Exposition in Aichi, Japan, from March to September 2005, and the second was at the National Museum of Emerging Science and Innovation in Tokyo, from November 25th to 27th. At the World Expo we utilized a large theater system using combinations of smaller 20-person theaters. These small theaters consisted of a screen, the Face Modeling, Rendering, and System Control sub-systems. We used the 3DCG movie "Grand Odyssey" as the backdrop to our Future Cast system. We limited the number of characters appearing in this 15-minute movie to twenty, so as to provide a reasonable amount of screen time for each visitor. Twelve 20-person theaters were constructed in the pavilion, thus allowing 240 people to experience the movie at the same time when divided into smaller groups. Figure 10 shows screen shots from one of the 20-people theaters. The face of each cast member is large enough for the visitors to identify.

In the latter half of the movie, the side walls of four

Fig. 10 Frames from the movie shown at the 20-person theater. Each character appears in a close-up on a 200 inch screen.

of the 20-person theaters open to create a larger 80-person theatre which is four times wider than the original. The four screens have now become one large 800-inch wide screen. The visitors are now presented with a more impressive image. The projection boundaries of each of the four screens merge together seamlessly. Figure 11 shows a shot of this large screen. At the end of the movie, the screen wall in front of the audience disappears gradually below the floor to reveal two other 80-person audiences who, in combination with the first group, are now arranged in a triangular seating formation. During this exposition, a total of 1,630,000 visitors experienced the Future Cast system.

The operational procedure of this system at Mitsui-Toshiba pavilion is as follows:

1. Explain the facial scanning process to visitors (4-5 minutes).
2. Scan each visitor's 3D face geometry and texture using 5 range scanners (7-8 minutes). Visitor's CG face is simultaneously created by our CG character modeling module.
3. Attendant confirms visitor's 2D scanned image and automatically-generated face model using attendant PC (2-3 minutes).
4. Provide the prologue video of "Grand Odyssey" to visitors (5 minutes).
5. The movie "Grand Odyssey" is shown (15 minutes).

While "Grand Odyssey" is running, the next group of visitors is being helped through steps 1-4 of the process.

To ensure the visitors' experience of the movie was suitably impressive, it was important for us to confirm that the quality of the initial scan of the visitors' face was satisfactory. To this end we placed theater staff in each Face Scanning Booth to provide visitors with specific guidance to ensure a clear scan. The theater attendants then judged, according to their own subjective criteria, whether they considered the face model to be natural enough, or if they thought the scan should be repeated only once.

The result of this checking procedure was a strong indication of how successful the CG face generation stage was (including face scanning, facial feature point detection, and face fitting). We maintained records of all attendant checks at all theaters from September 1st to 25th (the final day of the Expo). From these results, we calculated our success

Fig. 11 A scene from the real-time movie "Grand Odyssey". The Future Cast system can provide audiences with an amazing experience. Every member of the audience can participate in the movie as a "movie star".

rate in generating CG characters automatically. The success rate was defined by dividing the number of successfully generated faces N by the total number of audience members M shown in Eq. (4). Figure 12 shows the modeling success rate over 25 days from September 1st to September 25th along with the minimum, maximum and average success rate for each day.

$$\text{ModelingSuccessRate} = \frac{N}{M} \times 100[\%] \quad (4)$$

Our results indicated that the Future Cast System successfully generated CG characters with a maximum probability of 99.4%, a minimum of 80.8%, and average of 93.5%. In addition, our rendering system ran without any system failures during the Expo.

4.2 Small Scale Theater

As a result of policy restrictions at the Mitsui-Toshiba pavilion, it was difficult to carry out a questionnaire survey to evaluate our system. We were therefore unable to verify the point of most interest to us: whether the audience could identify themselves in the movie or not. We were, however, able to use a questionnaire when we exhibited Future Cast at the National Museum of Emerging Science and Innovation in Tokyo from November 25th to 27th. To evaluate the effectiveness of our work and find points to improve, we tried to gather as much data from the audience members as possible. At this time, 548 visitors saw the movie and provided feedback based on our one-page questionnaire (the contents of which appear below). Responders selected from prepared answers for each question.

- Age
- Gender
- Have you experienced Future Cast before?
- What feature did you focus on to identify yourself during the movie?[†]
- What feature did you focus on to identify others during the movie?[†]
- To what degree did you feel you were an actor/actress in the movie?^{††}

Fig. 12 The modelling success rate - September 1st - 25th, 2005. The result tends to improve gradually as the final day of the Expo approaches.

- How clearly could you recognize yourself?^{††}
- How natural did you feel your facial expressions were?^{††}
- Regarding your character's body shape and action, would you prefer the character to resemble yourself more?^{††}
- Regarding your character's skin color and texture, would you prefer the character to resemble yourself more?^{††}
- How much discomfort did you feel regarding your character's voice?^{††}
- Comments

Tables 1~8 show the results by age, specific mean score, and overall mean score for each question. Here, N describes the number of answers available for each question. Responders who felt uncomfortable with the voice of their character selected one reason from the following to account for their discomfort: "type of voice", "speech style" or "other". Of the 548 people who answered the questionnaire, 212 were satisfied with their character's voices. 336 indicated that they were not satisfied with the voice of their character. Of these, 71% selected "voice type", 16% "speech style" and 13% "other" to account for this feeling. In the next section of this paper we will move on to discuss the next-generation of the Future Cast System.

Table 1 What feature did you focus on to identify yourself during the movie?

Ranking	Male (N = 241)		Female (N = 211)	
	Parts	Score	Parts	Score
1	Face	2.72	Face	2.79
2	Expression	2.26	Expression	2.20
3	Voice	2.11	Hairstyle	2.11
4	Hairstyle	1.98	Voice	2.05
5	Action	1.96	Body Shape	1.98
6	Body Shape	1.93	Action	1.91

Table 2 What feature did you focus on to identify others during the movie?

Ranking	Male (N = 243)		Female (N = 213)	
	Parts	Score	Parts	Score
1	Face	2.75	Face	2.88
2	Expression	2.25	Expression	2.35
3	Body Shape	2.23	Voice	2.28
4	Voice	2.14	Body Shape	2.25
5	Hairstyle	2.11	Hairstyle	2.18
6	Action	1.84	Action	2.05

Table 3 How clearly could you recognize yourself?

	Male	Female	Mean
MOS	3.97 (N = 267)	3.20 (N = 237)	3.58 (N = 504)

[†]Responders indicated whether they "always pay attention", "sometimes pay attention", or "do not pay attention" to the following factors used to identify themselves and others: "body shape", "voice", "action", "facial expression", "hairstyle", and "face".

^{††}Responders answered based on a five point scale ranging from positive to negative.

Table 4 How natural did you feel your facial expressions were?

	Male	Female	Mean
MOS	3.17 (N = 259)	2.71 (N = 230)	2.94 (N = 489)

Table 5 Regarding your character's body shape and action, would you prefer the character to resemble yourself more?

	Male	Female	Mean
MOS	3.60 (N = 258)	3.30 (N = 232)	3.45 (N = 490)

Table 6 Regarding your character's skin color and texture, would you prefer the character to resemble yourself more?

	Male	Female	Mean
MOS	3.72 (N = 257)	3.39 (N = 231)	3.55 (N = 488)

Table 7 How much discomfort did you feel regarding your character's voice?

	Male	Female	Mean
MOS	3.18 (N = 255)	3.26 (N = 232)	3.22 (N = 487)

Table 8 To what degree did you feel you were an actor/actress in the movie?

	Male	Female	Mean
MOS	3.60 (N = 264)	3.11 (N = 234)	3.36 (N = 498)

5. Discussion and Future Work

We have developed a new type of visual entertainment the Future Cast System which over 1,630,000 people, regardless of age or gender, have experienced to date. Future Cast enables anyone to take part in a movie as a member of the cast, and, consequently, it is important to ascertain whether visitors are able to identify themselves as CG characters in the movie. The system's level of performance is determined by the accuracy of the face modeling process and the proportion of audience members who are able to identify themselves in the movie.

First, we discuss the reliability of face modeling based on the results of our large-scale theater experiment. Second, we consider the degree that audience members were able to identify their characters from a questionnaire survey. Third, as future work, we discuss ways to improve our system in order to more increase the personal identification ratio that audiences recognize themselves as CG characters. We additionally discuss how much audiences get absorbed in the "Grand Odyssey" provided by the Future Cast System. Finally, we discuss the effectiveness of our system from the viewpoint of theater operations.

5.1 Easy and Instant Casting

Figure 12 shows that our automatic face modeling technique satisfies our performance requirements. Consequently, we were able to demonstrate the effectiveness of our system, which enables people to participate as members of a movie cast easily and instantly. Especially on the final day of the

Expo, the Future Cast system was able to create suitable faces 99.4% of the time. This was largely because most of the audience members had already experienced the Future Cast scanning system, and the pavilion staff had gained a lot of experience in ensuring that the 3D capture of visitor's faces was successful. If a person moved at the moment the image was captured (e.g. by blinking, opening the mouth, or rotating the head), the face scan would fail due to a misalignment of corresponding points on the scanned image. To model a CG face exactly, it is necessary to measure the 3D geometry of the face precisely. Therefore, it is important that we ensure that all attendant staffs are fully trained on how to instruct visitors to avoid moving during the scan, and that they make sure visitors place their head in the optimal position in the scan window. We can also improve the performance of our face generation method by considering non-human factors, such as expanding available training data for the facial feature point detector to include a greater range of ages and more ethnic diversity.

5.2 Personal Identification

In the current Future Cast system, visitors can use the character's face, facial expressions, voice, hairstyle, body shape, and actions to identify themselves. Table 3 shows the results of whether audiences considered that they were able to identify themselves in the movie or not; with figures of 3.97 (male), 3.20 (female), and 3.58 (average). In addition, if we consider a score of 4 or more out of 5 to be a successful result, 69.66% of male respondents and 59.92% of female respondents were able to identify themselves clearly. These identification rates are lower than the face model generation success rates, and therefore need to be improved. We will now discuss the current version of the Future Cast system and some improvements we need to make in the future.

Face According to Tables 1 and 2, we found that the face is the part most used by visitors in identifying themselves and others, irrespective of gender. Since the face is a person's most distinctive physical characteristic (except in the case of identical twins), this would appear to be natural. It was therefore effective for us to focus on this aspect of appearance when attempting to render the likenesses of individuals into a movie during this project.

Our system makes it possible to model the face model precisely with 93.5% accuracy, thus satisfying our requirements for the Future Cast system. However, accurately representing human skin remains an important issue. Currently, it is difficult to completely and accurately reproduce the skin texture of an individual. As a result of rendering time constraints, we decided to apply a general lighting technique to the face. However, if we were able to reproduce the texture of personal skin more completely, it might enable audience members to recognize their rendered images more easily in the movie.

It is necessary, however, for us to consider how important it is to visitors that they see their skin rendered ac-

Fig. 13 Regarding your character's skin color and texture, would you prefer the character to resemble yourself more?

curately. Figure 13 shows that the older the visitor is, the less importance tends to be placed on accurate skin texture rendering. Responses from female respondents seem to indicate that they do not place as much interest in seeing their skin rendered as accurately as male visitors do. This trend becomes clearer as the age of respondents increases. Bearing these results in mind, and remembering that the Future Cast experience is designed for the purpose of entertainment, it is probably advisable to construct a skin shader that renders a more attractive, youthful skin texture than a completely accurate version of the visitors' own skin.

Facial Expression Facial expressions also play an important role in personal identification, because they tend to differ between individuals, even when expressing the same emotions. As the results in Tables 1, 2 show facial expressions were actually placed second in importance by respondents in our survey. In the current version of the Future Cast System, facial expressions are created using combinations of stereotypical key shapes. In Table 4, we can see that the naturalness score of facial expressions by visitors was low at 3.17 for males and 2.71 for females, with a mean score of 2.94. We would like to synthesize facial expressions with more personal characteristics by creating personalized key shapes with wrinkles. This would be achieved using a muscle-based approach. In doing so, we would enable visitors to recognize and identify with their characters more easily.

Character's Voice Results from our survey indicate that the character's voice is considered to be the third or fourth most important characteristic used in identifying people (Tables 1, 2). A person's voice recorded by microphone sounds different than the same voice directly perceived "in the head". It is therefore difficult for people to recognize their own voice immediately. The voice of a character might therefore help to identify other people, but it may be of little help in recognizing oneself. In our system, we use the pre-recorded voice of either an actor or actress as a substitute for the voice of the visitor. Therefore, responders feel some sense of discomfort if they perceive the voice of the character to be different from their own, or the people they

know. Table 7 shows that responders tended to feel some discomfort with the voice of their characters.

To make it easier for participants to identify themselves and others, it is necessary for personal speech characteristics to be synthesized from short vocal samples provided by the audience. In order to make this possible, it is necessary to obtain a wide range of voice samples which must be classified semi-automatically according to their characteristics.

Using these voices in movies is difficult because automatically synthesized speech tends to "drop" and sound unnatural to a greater degree than raw recorded speech. Therefore, in our current system we chose to use pre-recorded voices rather than those of the visitors. This decision was effective when we consider time constraints and voice quality. In order to generate a character whose voice is more similar to the participant's own, in the future, many audio tracks will be recorded in advance and the most similar voice will be selected and used.

Hairstyle, Body Shape, and Action Compared with body shape and physical actions, hairstyle seems to attract more attention for the purpose of identifying oneself. This is because hairstyle is observed on a daily basis far more often than either body shape or behavior. To help visitors recognize themselves easily, we need to create a hair model similar to the visitor's own personal hairstyle. However, in the current version of the Future Cast System, we have decided to focus more attention on the face than on hairstyles, etc, because the face is much more personal. We therefore use all common models for the hairstyle, physique and actions of a character. In addition, because hair motion is difficult to render and animate in real-time, all characters in the movie wear helmets. These have consequently been modeled as background images. As future work, we would like to model these parts directly to create a more realistic character model.

5.3 Immersive Feeling

Our system allows participants to co-star in movies. When evaluating this kind of entertainment system, it is important to get a sense of how the participants feel about appearing in the movie. The questionnaire results which relate to this question are shown in Table 8. The results were 3.60 for male respondents and 3.11 for female respondents, with a mean score of 3.36. The scores indicate that the present system does not enable participants to completely relive the experience of being a cast member in a movie. For example, audience members must watch themselves objectively and can not look around in the movie and see what the characters see. To provide audiences with a more impressive experience, we would like to develop a visual system in which the participants can view the world of the movie from any of the viewpoints of the movie cast.

5.4 Operation of Future Cast System

During the Expo, over 8,000 visitors per day visited the Mitsui-Toshiba pavilion, experiencing the show in groups of 240 at 20-minute intervals. As a result of time limitations, it was clearly impossible to manually create a face model, displacement vectors for facial expressions, and to allocate a role in the movie for each individual visitor. However, our Future Cast System allows these processes to be performed automatically. We are therefore able to conclude that our system is effective in generating a large number of face models and displacement vectors, and in assigning roles to the cast. Moreover, when the face modeling module fails to generate visitor's face model, default face models from former viewer's faces are replaced to complete the cast. It is therefore necessary to reduce the time for facial geometry modeling by improving our facial range scanner for re-scanning the visitor's face failed at first scanning.

Additionally, because it was necessary for us to adhere to the personal data protection policy regarding the personal data of visitors, each visitors' facial texture and geometry had to be completely removed from the file server after the show. However, for the purpose of future research, we need to collect this type of information, with the visitor's consent, because of its value in the fields of computer graphics and vision.

While the show was running at the Expo, our rendering system was required to continuously provide the movie for many visitors waiting for the next show. Our rendering system operated throughout the Expo except during a blackout, and was able to run the movie without dropping any frames when five or more of the eight IGPCs were operating. Our results clearly demonstrate the high levels of performance, reliability and scalability of our rendering system.

6. Conclusion

In this paper, we have described a type of visual entertainment system, the Future Cast System. The Future Cast System has two key features. First, the Future Cast system provides audiences with an amazing experience: the ability to take part in a movie as a 3DCG member of the cast co-starring alongside famous movie stars, singers and other celebrities in pre-created CG movies. Additionally, they are able to sit down and watch a movie in which they themselves are the stars. Second, the Future Cast system can automatically perform all of the processes required to make this possible; from capturing visitor's facial images and recreating them as 3DCG characters, to actually running the movie. Our face modeling process, the core technology of our system, can automatically generate visitors' CG faces with an accuracy of 93.5%.

In addition, our rendering system is robust enough to show movies without dropping any frames. The current version of the Future Cast System provides audiences with a movie in which only their faces are embedded. As future

tasks, we will consider how to improve the audience's sense of personal identification while watching the movie. The key to this possibly lies in generating CG characters which share the visitors' physical characteristics, such as hairstyle, hair motion, body shape, body motion, skin characteristics, and even accessories such as glasses. Finally, it is our hope that the Future Cast System we have developed will enable audiences to enjoy an experience that surpasses conventional movie entertainment.

Acknowledgments

This study was supported by the Special Coordination Funds for Promoting Science and Technology of Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors thank Tatsuo Yotsukura and Shoichiro Iwasawa for the helpful suggestions on feasibility studies. Also, we would like to thank Ken Anjyo, Aiko Nakatani, Richard Sproston, and the anonymous reviewers for their helpful feedback and comments. Finally, special thanks to all staff of "Grand Odyssey" and Mitsui-Toshiba pavilion.

References

- [1] K. Fujimura, Y. Matsumoto, and E. Tetsuchi, "Multi-camera 3d modeling system to digitize human head and body," *Proc. SPIE*, pp.40–47, 2001.
- [2] L. Wiskott, J. Fellous, N. Kruger, and C. der Malsburg, "Face recognition and gender determination," *Proc. International Workshop Automatic Face and Gesture Recognition*, pp.92–97, 1999.
- [3] M. Kawade, "Vision-based face understanding technologies and applications," *Proc. 2002 International Symposium on Micromechanics and Human Science*, pp.27–32, 2002.
- [4] L. Zhang, H. Ai, S. Tsukiji, and S. Lao, "A fast and robust automatic face alignment system," *Tenth IEEE International Conference on Computer vision 2005 Demo program*, 2005.
- [5] B. Moghaddam and M. Yang, "Gender classification with support vector machines," *Proc. fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp.306–311, 2000.
- [6] H. Lian, B. Lu, E. Takikawa, and S. Hosoi, "Gender recognition using a min-max modular support vector machine," *International Conference on Natural Computation 2005(2)*, pp.438–441, 2005.
- [7] F.I. Parke, "Computer generated animation of faces," *ACM'72: Proc. ACM annual conference*, 1972.
- [8] M. Lyons, A. Plante, S. Jehan, S. Inoue, and S. Akamatsu, "Avatar creation using automatic face processing," *Proc. sixth ACM international conference on Multimedia*, pp.427–434, 1998.
- [9] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D.H. Salesin, "Synthesizing realistic facial expressions from photographs," *SIGGRAPH '98: Proc. 25th annual conference on Computer graphics and interactive techniques*, 1998.
- [10] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," *International Conference on Computer Graphics and Interactive Techniques, Proc. 26th annual conference on Computer graphics and interactive techniques*, pp.187–194, 1999.
- [11] V. Blanz, C. Basso, T. Poggio, and T. Vetter, "Reanimating faces in images and video," *Comput. Graph. Forum*, 2003.
- [12] T. Ezzat, G. Geiger, and T. Poggio, "Trainable videorealistic speech animation," *SIGGRAPH '02: Proc. 29th annual conference on Computer graphics and interactive techniques*, pp.388–398, 2002.
- [13] Y. Lee, D. Terzopoulos, and K. Waters, "Realistic modeling for facial animation," *Proc. SIGGRAPH '95*, pp.55–62, 1995.

- [14] B. Choe, H. Lee, and H. Ko, "Performance-driven muscle-based facial animation," *Proc. Computer Animation*, pp.67–79, 2001.
- [15] E. Sifakis, I. Neverov, and R. Fedkiw, "Automatic determination of facial muscle activations from sparse motion capture marker data," pp.417–425, 2005.
- [16] D. DeCarlo and D. Metaxas, "The integration of optical flow and deformable models with applications to human face shape and motion estimation," *CVPR '96: Proc. 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp.231–238, 1996.
- [17] P. Joshi, W.C. Tien, M. Desbrun, and F. Pighin, "Learning controls for blend shape based realistic facial animation," *2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp.187–192, 2003.
- [18] J. Xiang Chai, J. Xiao, and J. Hodgins, "Vision-based control of 3d facial animation," *SCA '03: Proc. 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pp.193–206, 2003.
- [19] D. Vlasic, M. Brand, H. Pfister, and J. Popovic, "Face transfer with multilinear models," pp.426–433, 2005.
- [20] J. yong Noh and U. Neumann, "Expression cloning," *SIGGRAPH '01: Proc. 28th annual conference on Computer graphics and interactive techniques*, pp.277–288, 2001.
- [21] A. Krishnaswamy and G.V.G. Baranoski, "A biophysically-based spectral model of light interaction with human skin," *Comput. Graph. Forum*, pp.331–340, 2004.
- [22] P. Debevec, T. Hawkins, C. Tchou, H.P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," *SIGGRAPH '00: Proc. 27th annual conference on Computer graphics and interactive techniques*, pp.145–156, 2000.
- [23] G. Borshukov and J.P. Lewis, "Realistic human face rendering for "the matrix reloaded"," *ACM SIGGRAPH 2003 Technical Sketch*, 2003.
- [24] P.V. Sander, D. Gosselin, and J.L. Mitchell, "Real-time skin rendering on graphics hardware," *SIGGRAPH '04: ACM SIGGRAPH 2004 Sketches*, p.148, 2004.
- [25] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H.W. Jensen, and M. Gross, "Analysis of human faces using a measurement-based skin reflectance model," *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pp.1013–1024, 2006.
- [26] J.C. Carr, R.K. Beatson, J.B. Cherrie, T.J. Mitchell, W.R. Fright, B.C. McCallum, and T.R. Evans, "Reconstruction and representation of 3d objects with radial basis functions," *SIGGRAPH '01: Proc. 28th annual conference on Computer graphics and interactive techniques*, pp.67–76, 2001.
- [27] N. Arad and D. Reisfeld, "Image warping using few anchor points and radial functions," *Comput. Graph. forum*, pp.35–46, 1995.
- [28] S. Lee, G. Wolerg, and S.Y. Shin, "Scattered data interpolation with multilevel b-spline," *IEEE Trans. Vis. Comput. Graphics*, pp.228–244, 1997.
- [29] J.P. Lewis, M. Cordner, and N. Fong, "Pose space deformation: A unified approach to shape interpolation skelton-driven deformation," *International Conference on Computer Graphics and Interactive Techniques*, *Proc. 27th annual conference on Computer graphics and interactive techniques*, pp.165–172, 2000.
- [30] J.Y. Noh, D. Fidaleo, and U. Neumann, "Animated deformations with radial basis functions," *Virtual Reality Software and Technology*, *Proc. ACM symposium on Virtual reality software and technology*, pp.166–174, 2000.
- [31] M. Hirose, *Application for Robotic Vision of Range Image Based on Real-time 3D Measurement System*, Ph.D. thesis, Chukyo University, 2002.

Akinobu Maejima received the B.S., M.S., degrees, all in Electrical Engineering and Electronics from Seikei University, Tokyo, Japan, in 2002, 2004, respectively. Currently, he is a research associate of Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan. His current research interests include facial modeling using computer vision and graphics. He is a member of ACM and IEEE.

Shuhei Wemler received the B.S. and M.S. in physics from Tokyo University of Science in 1997 and 1999. Currently, he is an engineer of Silicon Studio Corp Ltd. He is working on development of Realtime 3D computer graphics softwares. His contribution to the Future Cast System is development of the robust rendering system. He is a member of ACM SIGGRAPH.

Tamotsu Machida received the B.S. and M.S. degrees, all in Precision Engineering from the University of Tokyo, Tokyo, Japan, in 1981 and 1983. From 1983 to 1989, he was a senior technical director of LINKS corp. From 1990 to 1996, he was a freelance computer graphics director. In 1996, he founded an independent Computer Graphics production ViNo azul, inc. Currently, he is a producer and a manager of Little Studios, Inc. He is a member of ACM SIGGRAPH. He received the Best paper award in 1985, from the Japan Society for Precision Engineering.

Masao Takebayashi is a producer of DENTSUTEC INC. He is an expert on direction for visual and audio systems specialized for exhibition spaces. He is a technical producer of the Future Cast System.

Shigeo Morishima received the B.S., M.S. and Ph.D. degrees, all in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1982, 1984, and 1987, respectively. From 1987 to 2001, he was an associate professor and from 2001 to 2004, a professor of Seikei University, Tokyo. Currently, he is a professor of Faculty of Science and Engineering, Waseda University. His research interests include Computer Graphics, Computer Vision, Multimodal Signal Processing and Human Computer Interaction. He is a member of the IEEE, ACM SIGGRAPH and the Institute of Electronics, Information and Communication Engineers Japan (IEICE-J). He is a trustee of Japanese Academy of Facial Studies. He received the IEICE-J achievement award in May, 1992 and the Interaction 2001 best paper award from the Information Processing Society of Japan in Feb. 2001. Also he is a visiting researcher of ATR Spoken Language Communication Laboratory from 2001. He is a leader of Digital Animation Labs.