Vol.2009-CVIM-167 No.6
2009/6/9

# Fast Shading and Shadowing and Handling Occlusions for Asuka-Kyo MR Contents

Tetsuya Kakuta ,[†1] Takeshi Oishi [†1]
and Katsushi Ikeuchi [†1]

We developed Mixed Reality (MR) contents that reconstructed the ancient capital of Asuka-Kyo and applied a fast shading and shadowing method that used shadowing planes. The proposed shadowing method can support both dynamic changes of illumination and movements of user's viewpoint. We also propose an effective foreground segmentation and shadow removal method to solve the occlusion problem in outdoor MR. The impression evaluation test of Asuka-Kyo MR Contents confirmed that shading and shadowing increased the audience's favorable responses to the evaluating factors of "Realistic," "Spectacular," and "Entertaining." And the subjective evaluation experiment showed that displaying these contents increased the audience's knowledge of both Asuka-Kyo and MR technologies.

## 1. Introduction

Virtual Reality (VR) technologies enable us to digitalize cultural heritage objects as 3D reconstructed models[5),33)]. Using internet and VR systems, we can release to the public the digital contents of these objects that are normally confined to a museum or located in their original sites. These digital contents can then be utilized by educational, entertainment, and tourism industries.

Superimposing reconstructed 3D models of cultural heritage objects on their original sites using Mixed Reality (MR) has been widely carried out in recent years[1),2)]. There are some restrictions on conventional VR contents. One limit is that we cannot appreciate them outside PC monitors and VR theaters. Another is that they lack a sense of reality compared with the original structures in their natural environments. On the other hand, MR allows us to make use of the landscape and atmosphere of the real environment and provide better experience for the users. Discovered ruins should be kept as close to the originals as possible for purposes of archaeological studies. MR can achieve effective representation of a reconstructed 3D model without having much effect on the original site and can thus strike a balance between preservation and presentation.

Many MR systems have been developed to virtually reconstruct 3D models of lost cultural heritage objects. Remains of ancient Greece have been reconstructed by the ARCHEOGUIDE project[34)]. The LIFEPLUS project proposes 3D reconstruction of ancient frescos and paintings of Pompeii[24)]. Some old wooden Japanese buildings, e.g., the Nara palace site[31)], are also represented by using MR technologies.

For the MR systems that aim to restore archaeological sites outside, it is important to achieve consistency of illumination. The shading of the virtual objects needs to match that of other objects in real environments. The virtual objects should also cast a correct shadow onto the real scene. Several methods have been proposed to solve this photometric problem in MR. Jacobs et al. provided a detailed survey of illumination techniques for MR[12)]. However, most methods are demonstrated on indoor scenes, and few of them are carried out at interactive update rates.

Another major problem of the MR system is the occlusion. Realistic image composition requires that the virtual objects be correctly occluded by foreground objects. Several research groups have worked on this occlusion problem. Kanbara et al. and Kim et al. have suggested the stereo vision-based method for estimating depth information of real surroundings[16),18)]. The method developed by Gordon et al. can correctly render interaction devices into the scene using foreground mask image extracted by the background subtraction method[10)]. Fischer et al. presented a real-time static occlusion handling method for medical intervention using preprocessed visual hull volume[9)]. A method for handling occlusion in non-real-time MR was proposed by Lepetit et al.[20)]. However, it is difficult to apply their method to an outdoor scene because of both the complicated geometry and changing illumination.

The evaluation of subjective psychological factors related to the illuminant consistency in MR is not emphasized in present research projects. Sato et al. reported an evaluation of shadow error between real and virtual occlusion ob-

†1 The University of Tokyo

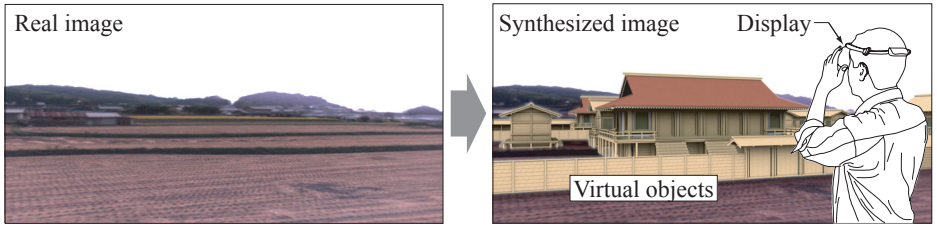



**Fig. 1** Current photo of Asuka-Kiyomiharanomiya and its reconstruction image.

jects[27]. Sugano el al. also pointed out the importance of shadow representation for increasing object presence[29]. Photometric inconsistency on human faces was evaluated by Takemura et al.[21]. Though they performed a quantitative evaluation test for illuminant consistency, there was no discussion on what subjective psychological factors contributed to the sense of reality and presence, or how we could increase the overall satisfaction level of users with the contents. Furthermore, these cases were executed in an indoor environment, and subjective evaluation for actual MR contents in outdoor scenes was not carried out.

In this paper, we propose a fast shadowing method for interactive MR application. We generate basis shadow images to express the shadows of virtual objects and set them onto other planate objects (hereafter called the shadowing planes) so that they correspond to both the arbitrary viewpoints and changing illumination of the real environment. In this method, we can make use of GPU (Graphics Processing Unit) for the synthetic process of basis shadow images. We achieved real-time computation of soft shadow on each shadowing plane by the use of GPU.

We also propose effective foreground segmentation and shadow removal method to solve the occlusion problem in outdoor MR. We have to estimate the depth of foreground objects and synthesize them in front or back of virtual objects. And also we have to remove shadow from the foreground area. We estimate depth of foreground objects from their position in the panoramic image obtained by a spherical vision camera. The advantage of using this kind of camera is that the illuminant information of the real scene is available from a single spherical image. On the other hand, conventional methods[26] need another camera to obtain a spherical image for illuminant estimation. We apply the proposed method to a video sequence and show the accuracy of segmentation in a synthesized image.

Finally we report an advanced evaluation test for illuminant consistency in MR. The motivation of this research is to identify psychological factors that influence evaluation of the contents and evaluate the subjective effectiveness of shadow representation using actual MR contents in an outdoor scene. We developed the MR contents of Japan's oldest capital, Asuka-Kyo which is said to have existed in the Asuka district around the sixth and seventh centuries, and applied the real-time illumination method to this content. The evaluation test was carried out as part of an open demonstration for the public on site, and the evaluators were selected from general tourists. Then we compared the learning effect before and after experiencing the MR contents and evaluated the improvement of their knowledge and interest in Asuka-Kyo and MR technologies. The effectiveness of shadow was evaluated by the semantic differential method and factor analysis using synthesized images with and without shade and shadow on the MR system.

The rest of the paper is organized as follows, In Section 2, we describe the fast shading and shadowing method. In Section 3, we introduce the foreground segmentation and shadow removal method for solving occlusion. In Section 4, we introduce the outline of Asuka-Kyo and the process of constructing its MR contents. The effectiveness of shading and shadowing for MR system and the educational effect of the Asuka-Kyo MR contents are discussed in Section 5. Finally, in Section 6, we present concluding remarks.

## 2. Fast Shading and Shadowing Method Using Shadowing Planes

We propose a fast shadowing method using shadowing planes. The major contributions of this method are as follows.

- Model-based shadowing that allows user to move the viewpoint.
- Hardware acceleration is available in the synthesis of basis images.

As shown on Fig.2, shadowing planes are arranged on the surface of the scene so that they can support the movement of viewpoint unlike the existing methods[23),25]. Additionally we can make use of GPU for the synthetic process of basis images.

The processing flowchart of the proposed method is described as below.

( 1 ) In the preprocessing stage, shadowing planes are generated from the convex

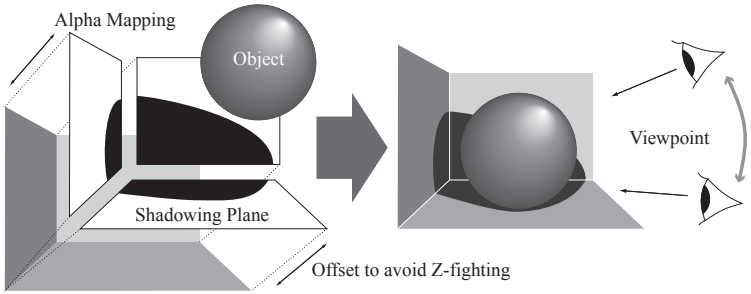



**Fig. 2** Mechanism of Shadowing Plane.

hull of each object in the scene.

(2) Basis shadow images and diffuse component images are rendered with basis lights that are indicative of approximated lighting environment.

(3) Acquiring the omnidirectional image of the real scene and compute the light scale parameters.

(4) Setting the intensity of light source in accordance with light scale parameters.

(5) Computing the linear combination of basis shadow images and diffuse component images with light scale parameters.

(6) Mapping generated soft shadow images onto shadowing planes.

(7) Rendering the scene and synthesizing it with real image.

We explain the feature of the proposed method and implementation process of shadowing plane in the following sections.

### 2.1 Proposed Method

The shadowing planes can represent simulated shadows of virtual objects over both the real image and objects themselves. And also, they can express the shadows on virtual objects casted by real objects using basis images which store shadows generated from objects corresponding to the real scene.

#### 2.1.1 Generation of shadowing planes

Shadowing planes are generated from the geometry of the scene. To approximate the surface of object in the scene, we compute convex hulls of each object. In case of complicated object, we need to divide the object into some clusters previously. Fig.3 shows the generating process of shadowing planes from complex

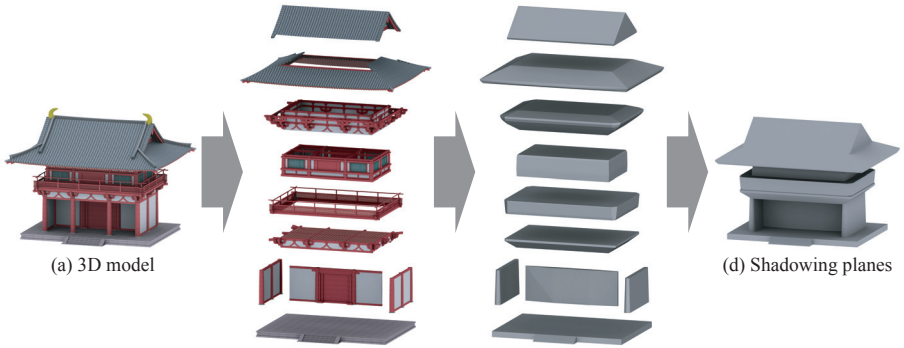



**Fig. 3** Generation of shadowing planes from a complicated model using convex hulls.

architectural model. Each shadowing plane covers virtual objects roughly, and is offset a little in the direction of a user's viewpoint to avoid Z-fighting. Therefore they are put between the objects and user's viewpoint.

#### 2.1.2 Shadow by Alpha Blending

Let us consider that we set $m$ shadowing planes $P_j^{shadow}\{j = 1, 2, \ldots, m\}$ on the scene, then we can generate $n$ basis shadow images $I_{j,i}^{basis}$ and diffuse component images $d_{j,i}^{basis}$ for each shadowing plane. The light scale parameters also described as $S_i$. When we suppose the irradiance of a point on the object surface as $E$ and that after computing shadow by shadowing plane $P_j^{shadow}$ as $E'$, $E'$ can be described as below[26].

$$E' = E\frac{\sum I_j^{basis}}{\sum d_j^{basis}} \tag{1}$$

$\sum I_j^{basis}$ indicates the synthesized image from basis shadow images and $\sum d_j^{basis}$ indicates that from diffuse component images. We define the shadow image $I_j^{shadow}$ as below.

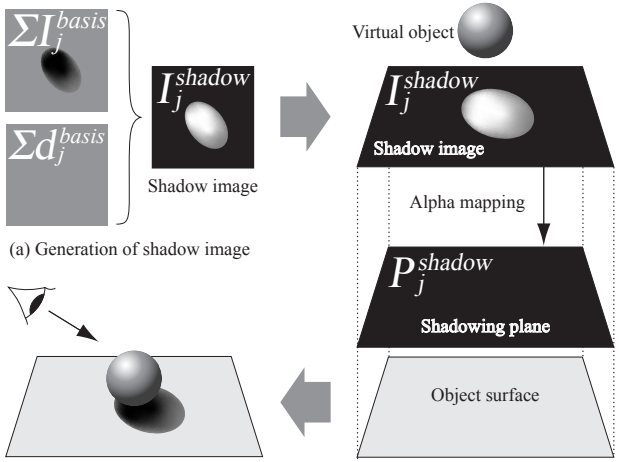



**Fig. 4** Generation of shadow images and mapping onto shadowing planes.

$$I_j^{shadow} = 1 - \frac{\sum I_j^{basis}}{\sum d_j^{basis}} \tag{2}$$

In the $I_j^{shadow}$, a pixel of value 1 means complete shadow area and value 0 means shadowless area (Fig.4(a)). As shown in Fig.4(b), we express shadow by mapping shadow images on shadowing planes as an alpha texture. The irradiance $E'$ can be expressed as

$$E' = E(1 - I_j^{shadow}) + P_j^{shadow} I_j^{shadow} \tag{3}$$

The reflectance of shadowing planes is zero then $P_j^{shadow} I_j^{shadow} = 0$.

$$E' = E(1 - I_j^{shadow}) \tag{4}$$

The equation (1) can be derived from the equation (2) and (4) . The shadowing planes attenuate the irradiance of the surface and express shadow effect from other objects.

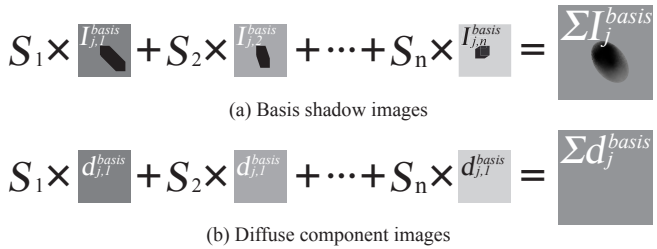



**Fig. 5** Linear combination of basis shadow images and diffuse component images.

#### 2.1.3 Synthesis of Basis Shadow Images and Diffuse Component Images on GPU

There is a linear relation between the radiance of lights and the reflectance of object surface. Then we can obtain the soft shadow image $\sum I_j^{basis}$ from the linear sum of light scale parameters $S_i$ and basis shadow images $I_{j,i}^{basis}$ as shown in Fig.5. In the same way, we can obtain the linear sum $\sum d_j^{basis}$ from diffuse component images $d_{j,i}^{basis}$. The key point here is we can make use of the GPU for the computations noted above.

Fig.6 shows the synthesis process of basis shadow images on the fragment shader. The fragment shader is used for computing color and other attributes of each pixel. In this stage, we arrange basis shadow images like tiles and transport them to the texture memory. Then we compute the linear combination of tiled basis shadow images and light scale parameter using the fragment shader with a number of texture units. Finally, we pick up pixel values from the synthesized shadow image and map them on the basis shadow images according to the corresponding texture coordinate.

### 2.2 Implementation

In this section, we explain how to generate the basis shadow images and estimate the illumination of the scene.

#### 2.2.1 Approximation of the Illumination

We can assume that the illumination of the scene is a hemispheric surface light source as illustrated in Fig.7(a). In this model, we can compute the illuminance $E$ of the minimal surface $A$ with whole surface light source as

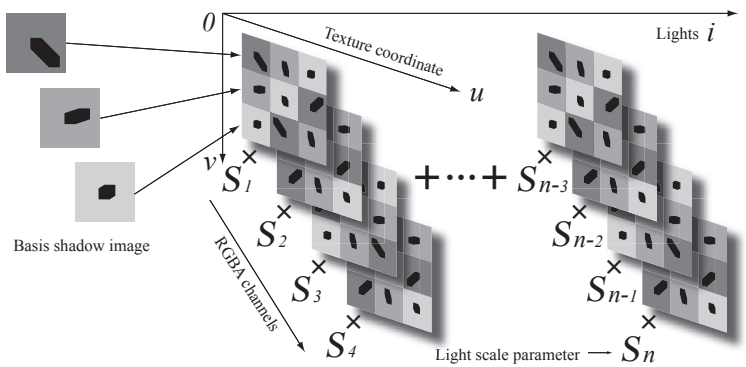



**Fig. 6** Synthesis of basis shadow images using fragment shader.

$$E = \int_{-\pi}^{\pi} \int_{0}^{\frac{\pi}{2}} L_0(\theta_i, \phi_i) \cos\theta_i \sin\phi_i d\theta_i d\phi_i \tag{5}$$

where $L_0(\theta_i, \phi_i)$ is the luminance per unit solid angle from the direction of $(\theta_i, \phi_i)$, and $\cos\theta_i$ is the parameter which means the attenuation relating to the direction of incidence.

In this research, we apply a geodesic dome to this hemisphere, and approximate the surface light source by the assembly of basis lights $L_i$ located on the every face of the dome. Suppose that the frequency of geodesic dome is high sufficiently and the radiance of each basis light is even, the illuminance $E$ of the minimal surface $A$ can be expressed by the sum of radiance of $L_i$.

### 2.2.2 Generation of Basis Shadow Images

We render the scene in advance and record the shadows of virtual objects projected onto shadowing planes as basis shadow images. Virtual cameras, which look towards each shadowing plane perpendicularly, capture the shadows of virtual objects cast on the shadowing planes. As shown in Fig.7(b), we obtain shadow of virtual objects on the shadowing plane $P_j^{shadow}$ using basis lights $L_i$ on the geodesic dome as basis shadow images $I_{j,i}^{basis}$. And also we render the scene without virtual objects and obtain diffuse component images $d_{j,i}^{basis}$.

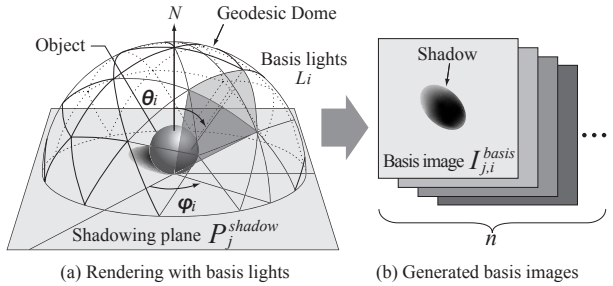



**Fig. 7** Generation of basis images.

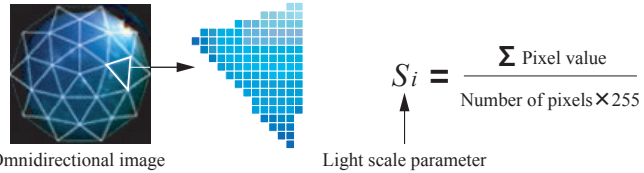



**Fig. 8** Acquiring the luminance.

### 2.2.3 Acquiring the luminance of the Scene

We obtain the information in the luminance of the scene with an omnidirectional image taken by a video camera with a fisheye lens as shown in Fig.8. Then we project the geodesic dome noted above onto the omnidirectional image, and compute the sum total value of internal pixels per each triangular region. At this point, we bring in the light scale parameter $S_i(i = 1, 2, .., n)$ to represent the intensity of each light source.

For the shading of virtual objects, we set six virtual directional lights in the scene. The intensity of every light is determined by the parameter $S_i$. With that we can express correct shadings of virtual objects responding to the real scene.

### 2.3 Experimental Result

We tested our proposed method in both an indoor scene and an outdoor scene.

Our system is mainly based on Canon's MR Platform system[32], which includes a video see-through head mounted display and the Polhemus's Fastrak, six degree-of-freedom (DOF) electromagnetic tracking sensor. We also used a Windows PC (2.40GHz Core 2 Duo E6600 CPU, 1024MB RAM, nVIDIA GeForce7950GT

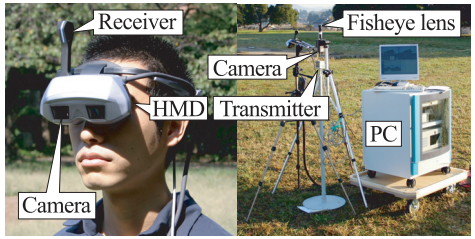
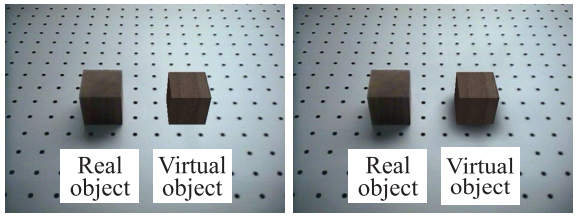



**Fig. 9** Appearance of our MR-system.




(a) before shadowing (b) after shadowing

**Fig. 10** Results for an indoor scene

GPU). The appearance of our equipment is shown in Fig.9.

Fig.10 and 11 show the comparison of synthesized images. We can see that both the shading and shadows of virtual objects are reasonably matched to the illumination in the real scene.

In the experiment of an outdoor scene (Fig. 11), we used 1520 basis images generated from 40 directional lights and 38 shadowing planes. The size of each basis image is $128 \times 128$ pixels and the resolution of the synthesized image is $640 \times 480$ pixels. We implemented the computation of the linear combination of basis images on the fragment shader to making use of GPU acceleration. The virtual objects consist of 58500 polygons and we successfully achieved about 18fps frame rate.

## 3. Detection of Moving Objects and Cast Shadows Using a Spherical Vision Camera

We propose effective foreground segmentation and shadow removal method to solve the occlusion problem in outdoor MR. We have to estimate the depth of foreground objects and synthesize them in front or back of virtual objects. And also Fig.12 shows the need for removing shadow from the foreground area. We estimate depth of foreground objects from their position in the panoramic image obtained by a spherical vision camera.

(a) before shadowing (b) after shadowing

**Fig. 11** Results for an outdoor scene

The advantage of using this kind of camera is that the illuminant information of the real scene is available from a single spherical image. Conventional methods[15),26)] need another camera to obtain a spherical image for illuminant estimation. We apply the proposed method to a video sequence and show the accuracy of segmentation in a synthesized image.

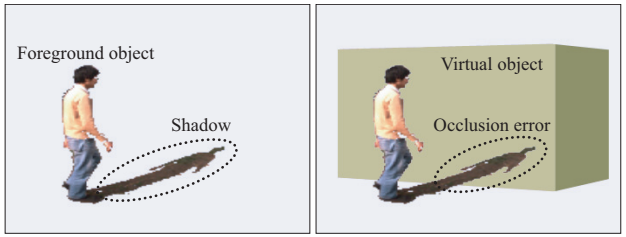
The processing flow of the proposed method is described below. First, we extract foreground areas from an input image. Second, the shadow of foreground objects is removed. Third, we estimate the depth of each foreground object and compare these to the depth of virtual objects. Last, the virtual objects are superimposed on an input image in consideration of occlusion.

### 3.1 Foreground Extraction

In this section, we describe the foreground extraction method based on a probabilistic model. Our method is extended from the method proposed by Shiota et al.[28)], which is combined Kolmogorov's method[19)] and Sun's method[30)]. It is assumed that the background image is known. In addition to the color and contrast information, the method uses temporal probability as a prior knowledge for improving accuracy.

#### 3.1.1 Probabilistic Model

For the foreground/background segmentation, we should minimize the energy




(a)Foreground segmentation (b)Synthesis with virtual object

**Fig. 12** Problem of shadow region.

term $E^t$ at time $t$ according to[28].

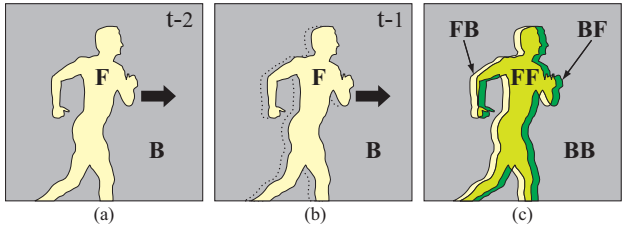
$$E^t = \alpha V^T(X^t, X^{t-1}, X^{t-2}) + \beta V^S(X^t, I^t) + U^C(X^t, I^t) \quad (6)$$

where $I^t = (I_1^t, I_2^t, \cdots, I_N^t)$ is the image extracted from video sequences at time $t$ and $X = \{x_1, \cdots, x_N\}$ is the label in the input image. $V^T$ is the temporal prior term and $V^S$ is the spatial prior term, and they are defined by the prior probability. $U^C$ is the color likelihood term and can also be defined by the color likelihood observed from an input image. $\alpha, \beta$ are the weight variables between energies and are given experimentally. We describe the overview of these terms in the remainder of this section.

**Temporal prior term**: $V^T(X^t, X^{t-1}, X^{t-2})$ is computed under the assumption that the prior knowledge of the labeling is represented by the second-order Markov chain model for temporal domain. The labeling result at a pixel is moving to temporal direction when the foreground object moves in an image as shown in Fig.13. This term imposes a tendency to temporal continuity of segmentation labels.

**Spatial prior term**: $V^S$ indicates the tendency that the same label is assigned to the neighboring pixels in an image. Generally, the labels are spatially continuous on the foreground area but they are different on the segmentation boundaries. Therefore, we define the energy term $V^S$ due to the contrast between neighboring pixels. Figure 14 shows the diagram of pixel-pairs. When the labels of pixel-pairs are different, the contrast of this pixel-pair becomes larger.

**Color likelihood term**: $U^C(X^t, I^t)$ is the likelihood that the color is $I_n$ when




(a) (b) (c)

**Fig. 13** Temporal transition.

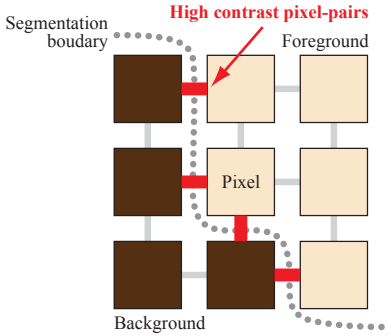



**Fig. 14** Contrast of neighboring pixel-pairs.

the label of $n \in \mathcal{V}$ is $x_n$. The likelihood is computed based on the color distribution of foreground and background. These color distributions are represented by Gaussian Mixture Models[3),19)] and are estimated by using the Expectation Maximization (EM) algorithm. The color distribution of the background region is learned from the background image $I^B$. A combined model of the color distribution over an input image and the distribution of each pixel are used to improve the local accuracy and robustness.

### 3.1.2 Energy Minimization

We compute the labels $(\hat{X}^1, \cdots, \hat{X}^t)$ that minimize the energy $E(X^1, \cdots, X^t, I^1, \cdots, I^t)$ shown in equation (6). However, it is impossible to compute all $(\hat{X}^1, \cdots, \hat{X}^t)$ at one time for the reason of the limit of computation time. We compute current label $\hat{X}^t$ using old labels $\hat{X}^1, \cdots, \hat{X}^{t-1}$ that are already

estimated.

$$\hat{X}^t = \arg\min E(X^t, \hat{X}^{t-1}, \hat{X}^{t-2}, I^1) \tag{7}$$

Optimum label $\hat{X}^t$ can be estimated by graph cut[4]. First, non-directed graph $\mathcal{G} = (\mathcal{V}', \mathcal{E})$ consisting of two nodes is generated from the input image. Here each edge $\mathcal{E}$ has nonnegative weight $w_e$ that corresponds to the energy function. Then we divide the graph $\mathcal{G}$ in order to separate two nodes using edge class $\mathcal{C} \subset \mathcal{E}$. Therefore the graph $\mathcal{G}(\mathcal{C}) = (\mathcal{V}', \mathcal{E} - \mathcal{C})$ consists of two graph where nodes are separated. The cost of graph cut $|\mathcal{C}|$ is represented as the sum of edge weights belonging to class $\mathcal{C}$. Then we compute the graph cut that minimizes the cost $|\mathcal{C}|$ and obtains the segmentation result $\hat{X}^t$ with this minimum cut.

#### 3.1.3 Shadow Removal

As shown in Fig.12, the shadow in the foreground will cause an occlusion problem when it is overlaid on virtual objects. Therefore, we remove those shadows from a foreground by introducing an illumination-invariant value, which can be obtained from pixel values[17].

Let us define the chromaticity as

$$i_r = \frac{I_R}{I_B} \quad , \quad i_g = \frac{I_G}{I_B}. \tag{8}$$

where $I_R, I_G, I_B$ are the image intensity. Then we assume

$$i_r = s_r e_r \tag{9}$$
$$i_g = s_g e_g \tag{10}$$

where $s_c$ and $e_c$ correspond to the chromaticities of the surface reflectance and the illumination.

Here, we introduce an assumption that blackbody radiation can predict the daylight illumination colors[8),14]. If we assume that the camera sensitivity function is sufficiently narrow and apply Wien's approximation to Planck's formula, the illumination chromaticity $(e_r, e_g)$ can be formulated in a simple manner[7),22]:

$$e_r = w e_g^A \tag{11}$$

where $A = \left(\frac{1}{\lambda_R} - \frac{1}{\lambda_B}\right) / \left(\frac{1}{\lambda_G} - \frac{1}{\lambda_B}\right)$, $w = \frac{\lambda_G^{5A} / \lambda_B^{5A}}{\lambda_R^5 / \lambda_B^5}$, and both are constant numbers characterizing the camera. $\lambda_c$ (where index $c = \{R, G, B\}$) is the center wavelength of the camera sensitivity, which can be obtained using a monochromator and a spectrometer[35].

If we substitute equation (11) into equations (9) and (10), we obtain

$$i_r / (i_g)^A = s_r / (s_g)^A \equiv F. \tag{12}$$

The equation means that if the surface color is identical, the value of $i_r/(i_g)^A$ will always be the same, regardless of the illumination color, the pixels of the ground have the same value whether in shadow or sunlight. The value $s_r/(s_g)^A$ is referred to as $F$[22].

We can remove shadowed pixels by using this $F$-value. If each pixel satisfies the conditions below, the pixel should be shadowed.

(1) The difference between $F$-values of the foreground and background is less than the threshold $t_F$.
(2) The brightness of the pixel of the foreground is less than that of the background.

### 3.2 Depth Estimation and Synthesis

To express correct occlusion, we then estimate the depth of foreground objects. We proposed a handy depth estimation method using a camera whose external parameters are known. Fig.15(a) shows the geometrical setup of the camera. We assume the ground is totally flat and the optical axis of the camera is parallel to the ground. When there are moving objects on the ground, referring to the input image, their depth can be computed from the bottom line of each foreground region as shown in Fig.15(b). If it is supposed that the scene is projected equidistantly by the lens and there is no distortion, we can compute the depth of moving objects $D$ as

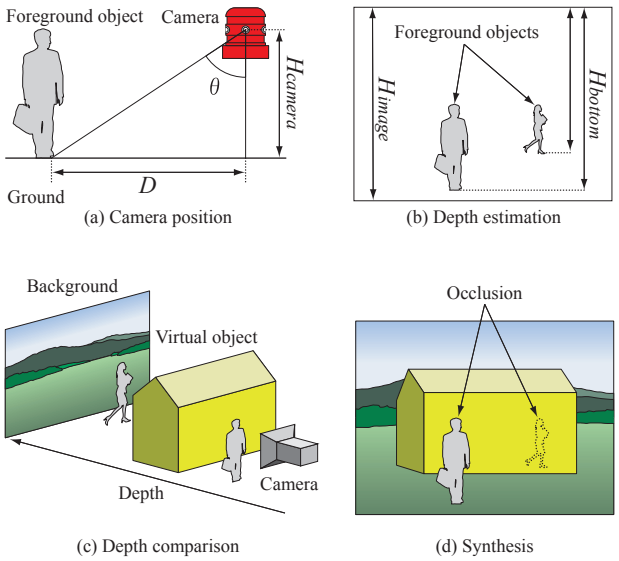



**Fig. 15** Depth estimation of foreground objects.

$$D = H_{camera} tan\{\frac{\pi(H_{image} - H_{bottom})}{H_{image}}\} \quad (13)$$

where $H_{camera}$ is the height of the camera, $H_{image}$ is the height of the input image and $H_{bottom}$ is the height of the bottom line of the foreground region.

Fig.15 (c) illustrates the comparison of depth of foreground objects and virtual objects. Then, as shown in Fig.15 (d), we synthesize them by considering their occlusion. At the stage of synthesis, the background image is rendered to the color buffer first. Then a mask image is generated from foreground regions which are in front of virtual objects. The mask image is rendered to the stencil buffer in order to avoid rendering virtual objects to this area. Finally, virtual objects are rendered to the color buffer and we can express collect occlusion. The shading and shadows of virtual objects are computed from the spherical image obtained from a camera using the method proposed by Section 2.


**Fig. 16** Process of segmentation and rendering. (a) Input image, (b) foreground extraction, (c) shadow removal, (d) synthesized image.

### 3.3 Experimental Result

Our MR-system is based on a spherical vision camera Ladybug2 developed by Point Grey Research Inc. Ladybug2's head unit consists of six CCD cameras, and it enables the system to collect video from more than 75% of the full sphere. The main advantages of using this spherical vision camera are that we can obtain illuminant information at one time.

We set this camera on a tripod in an outdoor scene horizontally using a level gauge. The height of the camera from the ground is 1.5 meters, which is almost as high as a human's eye. In the experiment, a video sequence of an outdoor scene has been obtained with the camera Ladybug2 with a resolution of $1024 \times 512$ pixels. We superimpose virtual objects and apply the proposed method to this video sequence offline. The spec of the PC is, OS: Windows XP (SP2), CPU: Core2Duo E6850 3.0GHz, RAM: 4GB, GPU: nVIDIA GeForce8800GTS 640MB.

Fig.16 shows each process of segmentation and rendering. We set the parameters of GMM as $K^b = 15, K^B = 2, K^f = 5$. As shown in Fig.16 (b), we extract foreground area from the input image first (Fig.16 (a)). Then we remove shadow from the foreground area using $F$-value(Fig.16 (c)). Second, the depth of the foreground is estimated from the segmented image and rendered to the stencil buffer. Finally, we render virtual objects and generate the synthesized image. Fig.16 (d) shows the accurate occlusion of real and virtual objects. The shadow of foreground objects is correctly removed and the shading of virtual objects is match to the illumination of the scene.

The proposed method can process 897.4 msec/frame on the MR-system noted above. Especially the spatial prior term for segmentation seems to takes much computation time. We are trying to make the process faster and handle the

occlusion of MR in real time.

## 4. Development of Asuka-Kyo MR Contents

In this section, we present a brief overview of Asuka-Kyo and the process of creating MR contents. The fast shading and shadowing method that we apply to the contents is explained in the next section.

### 4.1 Reconstruction of CG model of Asuka-Kyo

Asuka-Kyo indicates the generic name of the capital where the palaces of various emperors were assumed to be in the Asuka area from the late sixth to the late seventh centuries. Though many tourists visit the historic sites of Asuka village, most of them feel already lost and there is no way other than to rely on our imagination to visualize the spectacle of Asuka at that time. Since Asuka village is a part of the application region of the "Law Concerning Special Measures for Preservation of Historic Natural Features in Ancient Cities" and there is strict development control, it is difficult to build a new museum or replica for tourists. Moreover, excavation still continues and there is a possibility that the restoration plan will change with a newly-discovered archaeological fact. Against such a background, MR technology is very effective in reconstructing the ruins because it can overlay virtual buildings on the real site without harmful effect.

The restoration area in this project extends about 1350 meters in the north-south direction and about 600 meters in the east- west direction. Our MR contents contain the Asuka-Kiyomiharanomiya palace, the Kawaradera temple, the Asukadera temple, the Ishigami site, and the Asuka Enchi site.

It is known that there are multilayered remains in the palace area. We also decided to restore the Den-Asuka-Itabukinomiya III-B term (672-694) because it is in the upper layer and the position of the buildings is already known. The III-B term called Asuka-Kiyomiharanomiya differs from the III-A term called Nochino-Asuka-Okamotonomiya in the existence of the additional Ebinoko great hall in the southeast region. The Asuka-Kiyomiharanomiya palace expands 720 meters in the north-south direction and 100-450 meters in the east-west direction. It is said that there were government offices and gardens around the palace.

To create the CG model of Asuka-Kyo, we referred to the reconstruction drawings provided from the Nara National Research Institute for Cultural Properties and the Archaeological Institute of Kashihara. We modeled CG buildings of Asuka-Kyo with 3dsMax by Autodesk Inc. manually. The painted surface, tiled roofs of temples and the plain wood material of palace buildings are expressed by texture images. Fig.17 shows the reconstruction model of the Asuka-Kiyomiharanomiya palace, and Fig.18 shows that of the Ebinoko great hall. The total number of polygons of these models is about 686,000. The requirements for the Asuka-Kyo MR content are

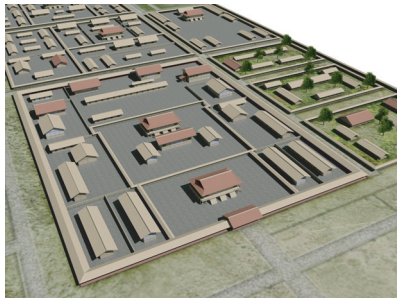


**Fig. 17** CG model of Asuka-Kiyomiharanomiya palace.

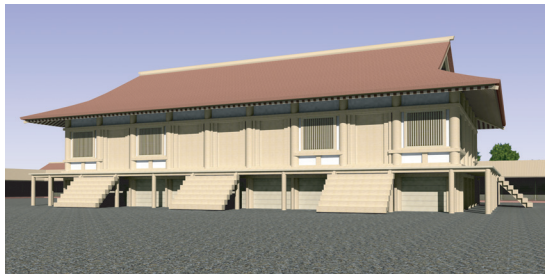


**Fig. 18** CG model of Ebinoko great hall.

- Rendering large-scale complex models in real time.
- Realistic rendering for a close look by users.
- Realistic illumination corresponding to the change of illumination outside.

There is a general trade-off between quality and efficiency in the rendering process. Conventional methods[11),13)] can express only hard shadows in fixed illumination. We applied the fast shadowing method to the proposed contents in order to fulfill these three requirements. We found it is easy to apply the shadowing plane method to the models of Asuka-Kyo because they are basically static and have a simple form without complex wooden parts. Also they consist of only diffuse materials without specular components, so it is possible to generate correct shadows from basis shadow images.

## 5. Evaluation

In this section, we report the advanced evaluation test about the illuminant consistency in MR. The motivation of this research is to identify the psycholog-

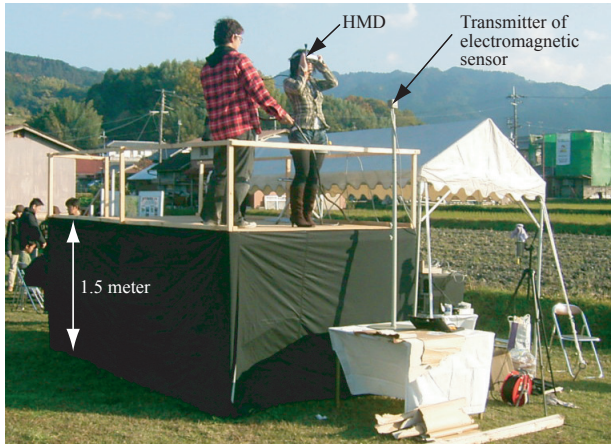



**Fig. 19** Appearance of evaluation experiment.

ical evaluation factor for the contents and make out the subjective effectiveness of shadow representation using actual MR contents in outdoor scene. The effectiveness of shadow is evaluated by the semantic differential method and factor analysis using synthesized images with and without shade and shadow on the MR system. And also we compared the learning effect before and after experiencing the MR contents and evaluated the improvement of their knowledge and interest in Asuka-Kyo and MR technologies.

### 5.1 Experimental Setup

#### 5.1.1 MR System

Our MR system is mainly based on Canon's MR Platform system as with Section 1. This MR system enables us to acquire the image of the real site, align virtual objects to the image, and synthesize them. The size of the image displayed on HMD is 640×480 pixels. The evaluation tests to be described are performed on a standard computer system with an Intel Core 2 Duo E6850 CPU, nVIDIA GeForce 8800GTS GPU and 4GB RAM. For obtaining the illumination, we used Victor's CCD video camera KY-57 equipped with a fisheye lens (Fit Corporation's FI-19).

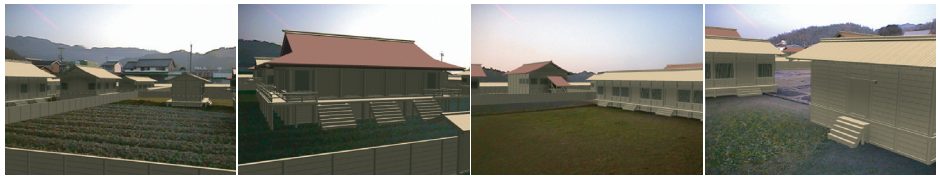


**Fig. 20** Synthesized image of Asuka-Kiyomiharanomiya palace on MR system.

#### 5.1.2 Displaying Procedure

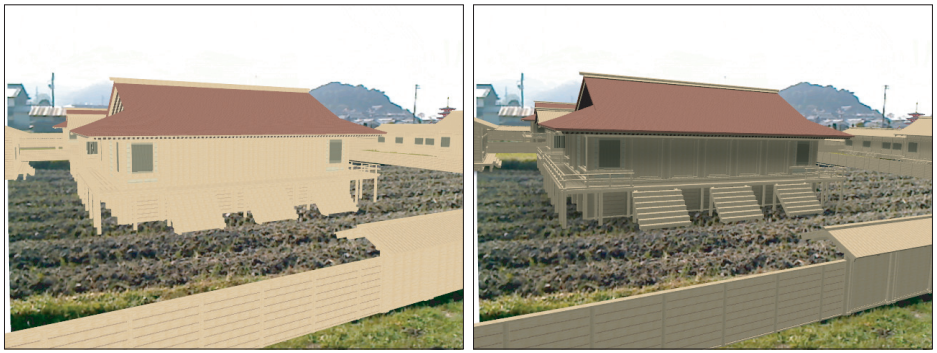
We carried out subjective evaluation tests on the historical site of Asuka village. Since the reconstructed CG model of Asuka-Kiyomiharanomiya contains wooden fences 2 meters high in the inner bailey, we put up a scaffolding 1.5 meters high so the users can overlook the palace as shown in Fig.19. We installed the MR system on a northeast direction of the main temple because there are no real buildings and the users can take a look around the site. The camera for illumination acquisition was placed 5 meters away from the MR system to avoid capturing users and equipment. Fig.20 shows the synthesized images on HMD. Natural soft shadows and real-time image synthesis were achieved by application of fast shadowing method.

### 5.2 Evaluation Experiment for User's Impression with and without Shading and Shadowing

First we conducted a subjective evaluation test to measure the user's impression of the MR contents and clear up the potential psychological factors on evaluating the synthesized image. Then we studied whether significant differences came about by adding correct shade and shadow.

#### 5.2.1 Experimental Design

Fifty-six observers (28 male, 28 female, average age 51.88, range 12-81). We asked them to report the impression of the synthesized image on the MR system. The semantic differential method is used to analyze the result, and we begin with 24 pairs of adjectives that are considered to adequately describe the impression of a synthesized image based on[6]. The examinees were asked to rate them using a scale from one to six. The observers were divided into two groups and we showed two types of synthesized images on the MR system with and without shading and shadowing to each group (Fig.21). Note that we did not teach this difference



(a) Synthesized image (b) Synthesize image with shading and shadowing

**Fig. 21** Synthesized images for evaluation experiment for user's impression with and without shading and shadowing.

to each group. The experience was performed under a clear sky within an hour to minimize the change of illumination condition. The experience time of each group was two minutes.

### 5.2.2 Result and Discussion

Fig.22 shows the average score of each adjective pair with and without shade and shadow. The left side of the figure contains negative adjectives and right side contains positive adjectives. When we compared each average score, the group of "with shading and shadowing" clearly moved to the right (positive) side. To express shade and shadow of virtual objects might improve the evaluation of the contents. In particular, there is significant difference over 1.0 in the adjective pairs, "invisible - visible" and "incomplete - complete." This could be due to the fact that the shade and shadow emphasize the appearance of solidity and the models seem to be authentic. By contrast, there is little difference in "light - heavy" and "changeable - stable." It would appear that the shade and shadow have an insignificant effect on the model's projection of stability.

The factor analysis reveals the potential common factors related to the contents evaluation as shown in Table 1. In the calculation process, we extract the initial factors by the principal factor method and adopt a promax rotation. The adjective pairs in Table 1 are sorted by their absolute value of factor loadings. The result shows that the three extracted factors explain almost all the adjective

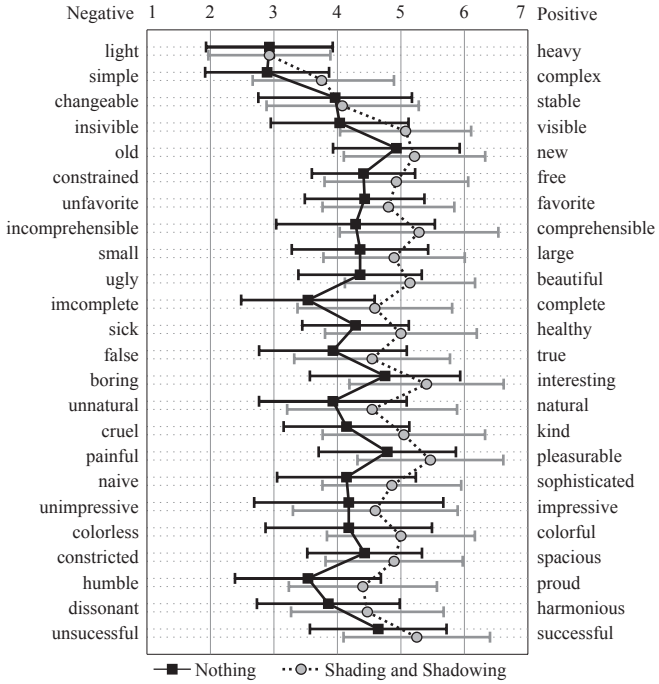



**Fig. 22** Average score of adjective pairs.

pairs.

In the first factor, the factor loading of "false - true" pair is especially high (0.947) and it contains "dissonant - harmonious" and "unnatural - natural" pairs. Therefore we assume this factor is related to the harmony and consistency of the image and it can be said to represent the "Realistic" aspect. In the second factor, the factor loadings of "colorless - colorful" and "unimpressive - impressive" are high. We assume this means the appeal and power that is "Spectacular." In the third factor, the factor loadings of "incomprehensible - comprehensible", "constrained - free" and "unfavorite - favorite" are high, so we consider this means an acceptance and appeal of the contents, in other words "Entertaining."

The factor scores of three factors described above for each group are plotted in Fig.23. The ellipses in this figure mean the 95% confidence region given a

| Adjective pairs | Factor loading | | |
|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 |
| Factor 1 : "Realistic" | | | |
| false - true | **0.947** | 0.075 | −0.171 |
| incomplete - complete | **0.677** | −0.060 | 0.316 |
| dissonant - harmonious | **0.628** | 0.052 | 0.095 |
| humble - proud | **0.619** | 0.079 | 0.287 |
| unnatural - natural | **0.546** | 0.134 | 0.287 |
| sick - healthy | **0.535** | 0.204 | 0.206 |
| Factor 2 : "Spectaculer" | | | |
| colorless - colorful | 0.057 | **0.840** | 0.034 |
| unimpressive - impressive | 0.187 | **0.815** | −0.064 |
| unsuccessful - successful | −0.108 | **0.706** | 0.228 |
| naive - sophisticated | 0.489 | **0.551** | −0.090 |
| constricted - spacious | 0.190 | **0.517** | 0.041 |
| Factor 3 : "Entertaining" | | | |
| incomprehensible - comprehensible | −0.096 | 0.175 | **0.868** |
| constrained - free | 0.140 | −0.093 | **0.629** |
| unfavorite - favorite | 0.065 | 0.236 | **0.545** |

**Table 1** Factor loadings for every adjective pairs.

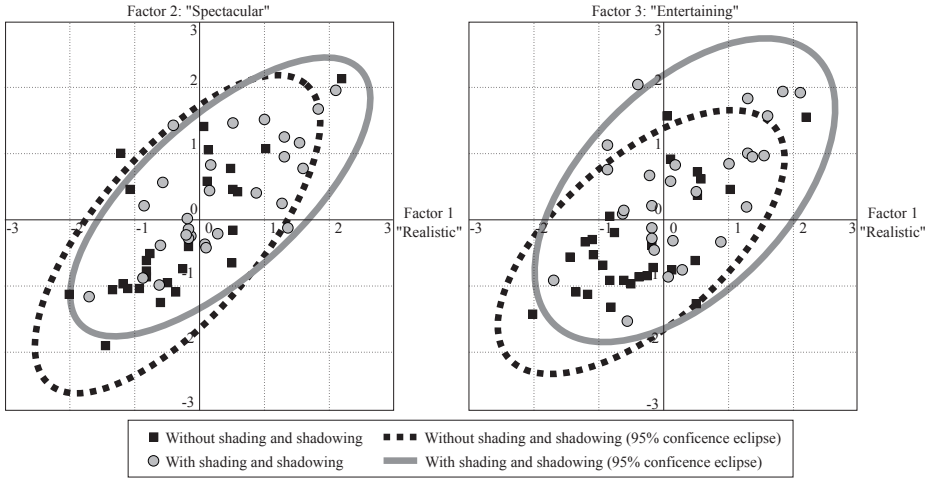



**Fig. 23** Factor score of examinees with and without shading and shadowing.

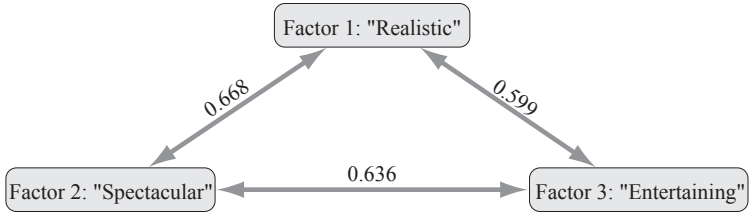



**Fig. 24** Correlation coefficient between the pair of factors.

2-dimension normal distribution of factor scores. When we compare the distribution of each group, "with shading and shadowing" group makes a high score than the other. Particularly, there is great difference between two groups in the axis of the "Entertaining" factor as shown in Fig.23 (b).

Fig.24 shows the coefficient of correlation of the three factors. There are fairly high correlations among these factors, so we can consider the correct shading and shadowing in MR are important to improve not only the reality and consistency of the synthesized image but also the sharpness and punch of models, as well as the entertainment value and interactivity of the contents.

### 5.3 Evaluation Experiment for Education Effect

In the next place, we performed an experiment to evaluate the education effect of the proposed contents by comparing the learning effectiveness of subjects before and after experiencing the contents. We wanted to investigate if and how much effect the contents have for users in respects of interests and concerns to the history of Asuka Period.

#### 5.3.1 Experimental Design

Nineteen observers (12 male, 7 female, average age 49.63, range 28-77) who were not engaged in the last experiment volunteered to participate in the experiment. Evaluation items are determined from the learning goals of Geography, History, and Information subjects in the ministry's curriculum guidelines for high schools. Specifically we presume four end points of 1) Interests in the remains, 2) Historical knowledge, 3) Conservation awareness of cultural heritage objects, 4) Understanding of MR technologies, which are expected to increase by expe-

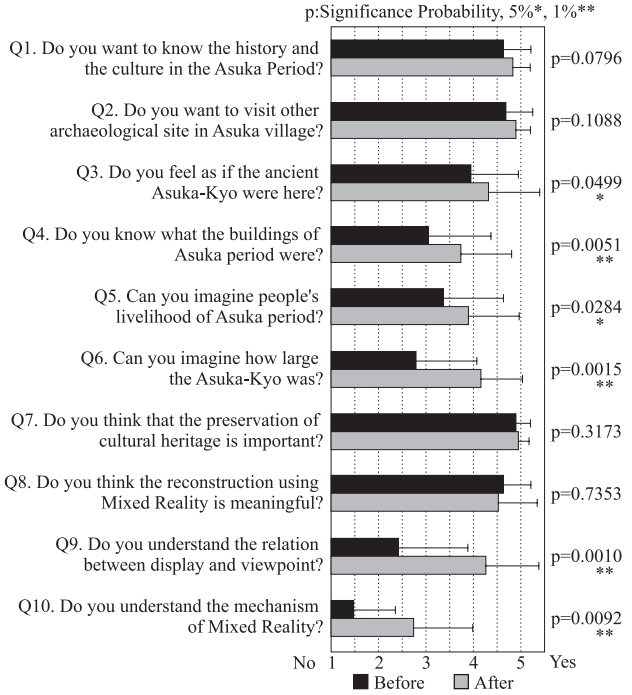



**Fig. 25** Average score of educational effects before and after experiencing the contents.

riencing the MR contents. Ten questions were prepared from these end points. These questions are randomly arranged on the survey forms and observers were asked to rate them using a scale from one to five before and after experiencing the contents. The experience time of each observer was two minutes.

### 5.3.2 Result and Discussion

As a result, we confirmed the score had been increased with almost all question terms. Fig.25 shows the average score of every question before and after going through the contents. The p-values in Fig.25 indicate the significance probability calculated by the Wilcoxon signed-rank test and they mean the probability of obtaining score by chance (two-sided test).

The symbol of * and ** indicates a significance level less than 5% ($p<0.05$) and 1% ($p<0.01$) respectively. We found a statistically significant improvement of less than 1% in Q 4, 6, 9, 10 and less than 5% in Q 3, 5.

The result shows the proposed contents have an education effect on learning goal 2) Historical knowledge, and 4) Understanding of MR technologies, previously-assumed. Q3 and Q6 show the effect that lets users recognize the scale of ancient Asuka-Kyo and its spatial corresponding relationship with current remains and landscape. Especially the score of Q6 is remarkably increased and we might guess it is because the observers could physically experience the speciousness of Asuka-Kyo through the MR system. The scores of Q9 and Q10 show that the contents contributed to users' understanding of the MR technologies and devices. The average scores of these questions are 2.42 and 1.47 respectively. They are relatively low compared to other questions but dramatically increased after experiencing the MR system. Getting hands-on experience should be considered to increase the understanding of MR technologies.

On the other hand, the scores of Q1, Q2, Q7 and Q8 show no significant education effect. We might assume this is because the interest and concerns of examinees are high enough from the beginning so that the average scores of these questions are positive and the contents have a very small effect on the observer.

## 6. Conclusion

This paper proposed fast shading and shadowing method for MR. We approximate the illumination in the scene and generate the basis images using the shadowing planes. Soft shadow images, correspond to the illumination of the real scene, are synthesized from basis shadow images. Then we map these shadow images onto shadowing planes as alpha texture and express soft shadows of virtual objects in real-time. However the proposed method is applicable only to static scene, it is effective for a specific application (e.g. MR-based restoration of cultural heritages in outdoor scene). Our method can achieve the consistency of illumination and improve the quality of synthesized image in MR-systems.

We have also presented a novel occlusion method for MR. The proposed approach uses a probabilistic model-based segmentation and shadow removal with $F$-value. The depths of foreground objects are estimated from their position in the panoramic image obtained from a spherical vision camera. Making use of the stencil buffer, we can express the occlusion of foreground objects and virtual

objects correctly. The spherical vision camera makes it possible to obtain the illuminant information from a single spherical image. Therefore we can estimate the distribution of illumination at the same time and express correct shading and shadow of virtual objects. We have shown the effectiveness of the approach in an outdoor scene with complex illumination. The result clearly showed the advantage of the occlusion method. Our approach is of value in increasing the reality of synthesized images in MR-systems.

Finally, we developed and evaluated MR contents of the Japanese ancient capital Asuka-Kyo. We successfully improved the reality of the synthesized image in this content by applying the fast shading and shadowing method. The result of the first experiment regarding users' impressions reveals three common factors in evaluating MR contents, "Realistic,""Spectacular," and "Entertaining." We confirmed that the shading and shadowing of virtual objects improved these evaluating factors and the illuminant consistency is essential for the subjective evaluation. Correct shading and shadowing are also important to increase the entertainment value of the contents in addition to the reality and consistency. Our second subjective evaluation experiment shows the education effect of the contents. MR applications that reconstruct lost remains on historical sites are shown to increase the interests and concerns of the subjects. Improving the educational effect of the Asuka-Kyo MR contents is our task for the future. Especially the end points of "Historical knowledge" and "Conservation awareness for cultural heritage" are minimally affected by the MR contents as described in Section 5. Multiple information service such as text and audio commentary might improve these aspects.

**Acknowledgments** This research was, in part, supported by the Ministry of Education, Culture, Sports, Science and Technology, under the program, "Development of High Fidelity Digitization Software for Large-Scale and Intangible Cultural Assets." We also grateful to Asuka Histrical National Government Park, Asuka Village, Nara National Research Institute for Cultural Properties, and Kashihara Archaeological Institute, Nara prefecture.

## References

1) Azuma, R.: A Survey of Augmented Reality, *Presence: Teleoperators and Virtual Environments*, Vol.6, No.4, pp.355–385 (1997).
2) Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B.: Recent Advances in Augmented Reality, *IEEE Computer Graphics and Applications*, Vol.21, No.6, pp.34–47 (2001).
3) Blake, A., Rother, C., M.Brown, P.Perez and P.Torr: Interactive Image Segmentation using an adaptive GMMRF model, *In Proceedings of European Conference on Computer Vision* (2004).
4) Boykov, Y., Veksler, O. and Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.Vol. 23(11), pp.1222 – 1239 (2001).
5) Callieri, M., Debevec, P., Pair, J. and Scopigno, R.: A realtime immersive application with realistic lighting: the Parthenon, *Computer & Graphics*, Vol.30, No.3, pp.368–376 (2006).
6) C.E.Osgood, Suci, G.J. and Tannenbaum, P.: *The Measurement of Meaning*, University of Illinois (1967).
7) Finlayson, G.D. and Hordley, S.D.: Color constancy at a pixel, *Journal of Optical Society of America A.*, Vol.Vol. 18, No.No. 2, pp.pp. 253–264 (2001).
8) Finlayson, G.D. and Schaefer, G.: Solving for color constancy using a constrained dichromatic reflection model, *International Journal of Computer Vision*, Vol. 42, No.3, pp.127–144 (2001).
9) Fischer, J., Bartz, D. and Straβer, W.: Occlusion Handling for Medical Augmented Reality using a Volumetric Phantom Model, *Proc. Symp. on Virtual Reality Software and Technology (VRST'04)*, pp.174–177 (2004).
10) Gordon, G., Billinghurst, M., Bell, M., Woodfill, J., Kowalik, B. and Erendi, A.: The Use of Dense Stereo Data in Augmented Reality, *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR02)*, pp.14–23 (2002).
11) Haller, M., Drab, S. and Hartmann, W.: A real-time shadow approach for an augmented reality application using shadow volumes, *Proc. Symp. on ACM Virtual Reality Software and Technology(VRST'03)*, pp.56–65 (2003).
12) Jacobs, K. and Loscos, C.: Classification of Illumination Methods for Mixed Reality, *Proc. Eurographics State of the Art Report(STAR)*, pp.95–118 (2004).
13) Jacobs, K., Nahmias, J.D., Angus, C., Reche, A., Loscos, C. and Steed, A.: Automatic Generation of Consistent Shadows for Augmented Reality, *Proc. Graphics Interface 2005*, pp.113–120 (2005).
14) Judd, D.B., MacAdam, D.L. and Wyszecky, G.: Spectral distribution of typical daylight as a function of correlated color temperature, *Journal of the Optical Society of America*, Vol.54, No.8, pp.1031–1040 (1964).
15) Kakuta, T., Oishi, T. and Ikeuchi, K.: Real-time Soft Shadows in Mixed Reality

using Shadowing Planes, *Proc. IAPR Conference on Machine Vision Application (MVA2007)*, pp.195–198 (2007).

16) Kanbara, M., Fujii, H., Takemura, H. and Yokoya, N.: A stereo vision-based mixed reality system with natural feature point tracking, *Proc. Int. Symp. on Mixed Reality (ISMR01)*, pp.56–63 (2001).

17) Kawakami, R., Tan, R.T. and Ikeuchi, K.: Consistent Surface Color for Texturing Large Objects in Outdoor Scene, *Proc. Int. Conf. on Computer Vision (ICCV'05)*, pp.1200–1207 (2005).

18) Kim, H., Yang, S. and Sohn, K.: 3D reconstruction of stereo images for interaction between real and virtual worlds, *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality (ISMAR03)*, pp.169–176 (2003).

19) Kolmogorov, V., Criminisi, A., Blake, A., Cross, G. and Rother, C.: Bi-Layer Segmentation of Binocular Stereo Video, *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. Vol.2, pp.407 – 414 (2005).

20) Lepetit, V. and Berger, M.O.: A Semi-Automatic Method for Resolving Occlusion in Augmented Reality, *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR00)*, pp.174–177 (2000).

21) M.Takemura, I.Kitahara, Y.O.: Photometric Inconsistency on a Mixed-Reality Face, *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR06)* (2006).

22) Marchant, J.A. and Onyango, C.M.: Shadow-invariant classification for scenes illuminated by daylight, *Journal of Optical Society of America A.*, Vol.Vol. 17, No.No. 11, pp.pp. 1952–1961 (2000).

23) Nimeroff, J.S., Simoncelli, E. and Dorsey, J.: Efficient re-rendering of naturally illuminated environments, *Proc. Eurographics Workshop on Rendering (EGWR94)*, pp.359–373 (1994).

24) Papagiannakis, G., Schertenleib, S., O'Kennedy, B., Arevalo-Poizat, M., Magnenat-Thalmann, N., Stoddart, A. and Thalmann, D.: Mixing virtual and real scenes in the site of ancient Pompeii, *Computer Animation and Virtual Worlds*, Vol.16, No.1, pp.11–24 (2005).

25) Sato, I., Hayashida, M., Kai, F., Sato, Y. and Ikeuchi, K.: Fast Image Synthesis of Virtual Objects in a Real Scene with Natural Shading, *The Institute of Electronics, Information and Communication Engineers, D-II (in Japanese)*, Vol.J84-D-II, No.8, pp.1234–1242 (2001).

26) Sato, I., Sato, Y. and Ikeuchi, K.: Acquiring a radiance distribution to superimpose virtual objects onto a real scene, *IEEE Trans. on Visualization and Computer Graphics* (1999).

27) Sato, I., Sato, Y. and Ikeuchi, K.: Illumination from shadows, *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2003).

28) Shiota, K.: *Probabilistic model-based foreground extraction and application to 3D shape reconstruction*, Master thesis, The University of Tokyo (in Japanese) (2007).

29) Sugano, N., Kato, H. and Tachibana, K.: The Effects of Shadow Representation of Virtual Objects in Augmented Reality, *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR03)*, pp.76–83 (2003).

30) Sun, J., Zhang, W., Tang, X. and Shum, H.-Y.: Background Cut, *ECCV*, Vol.Vol.2, pp.628 – 641 (2006).

31) Tenmoku, R., Nakazato, Y., Anabuki, A., Kanbara, M. and Yokoya, N.: Nara palace site navigator: Device-independent human navigation using a networked shared database, *Proc. Int. Conf. on Virtual Systems and Multimedia (VSMM2004)*, pp.1234–1242 (2004).

32) Uchiyama, S., Takemoto, K., Satoh, K., Yamamoto, H. and Tamura, H.: MR Platform: A Basic Body on Which Mixed Reality Applications Are Built, *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR02)*, pp.246–253 (2002).

33) Utsugi, K., Moriya, T. and Takeda, H.: Making of a Virtual World of Heijokyo from Historical Knowledge, *Proc. Int. Conf. on Virtual Systems and Multimedia (VSMM2001)* (2001).

34) Vlahakis, V., Ioannidis, N., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., Gleue, T., Daehne, P. and Almeida, L.: Archeoguide: An Augmented Reality Guide for Archaeological Sites, *IEEE Computer Graphics and Applications*, Vol.22, No.5, pp.52–60 (2002).

35) Vora, P.L., Farrell, J.E., Tietz, J.D. and Brainard, D.H.: Digital color cameras - 2 - Spectral response, *HP Technical Report* (1997).