View Management of Projected Labels on Nonplanar and Textured Surfaces

Daisuke Iwai, Tatsunori Yabiki, and Kosuke Sato, *Member*, *IEEE*

Abstract—This paper presents a new label layout technique for projection-based augmented reality (AR) that determines the placement of each label directly projected onto an associated physical object with a surface that is normally inappropriate for projection (i.e., nonplanar and textured). Central to our technique is a new legibility estimation method that evaluates how easily people can read projected characters from arbitrary viewpoints. The estimation method relies on the results of a psychophysical study that we conducted to investigate the legibility of projected characters on various types of surfaces that deform their shapes, decrease their contrasts, or cast shadows on them. Our technique computes a label layout by minimizing the energy function using a genetic algorithm (GA). The terms in the function quantitatively evaluate different aspects of the layout quality. Conventional label layout solvers evaluate anchor regions and leader lines. In addition to these evaluations, we design our energy function to deal with the following unique factors, which are inherent in projection-based AR applications: the estimated legibility value and the disconnection of the projected leader line. The results of our subjective experiment showed that the proposed technique could significantly improve the projected label layout.

Index Terms—Projection-based augmented reality, view management, label layout, projected character's legibility

1 Introduction

Labels are essential components in most augmented reality (AR) applications, because they help to explain the labeled physical objects. In optical and video see-through AR research fields, much research has been done for the past decade on superimposing digital labels, annotations, or legends on associated physical objects in a head-mounted display (HMD) [1]. Designing well-labeled augmentation requires skill and effort. Previous works focused on automatically determining a label layout by taking into account that the legibility of the characters should not be significantly degraded and that there should be a clear visual correspondence between the text and the anchor region it is labeling.

Compared to the see-through AR studies, much less research has been done so far in the field of projection-based AR research. In the projection-based approach, the following two unique issues must be taken into account, because they inherently arise from the fact that labels are superimposed directly onto the surfaces of physical objects. First, the legibility of the projected label is easily degraded because the characters are geometrically deformed, photometrically disturbed, and hidden by shadows when projected on nonplanar and textured surfaces. Second, a leader line, which connects a text and its anchor region, is disconnected when there is a geometrical discontinuity under it. Despite these difficulties in the view management of projected labels, the projection-based approach still holds several advantages over the approach used for see-through AR [2]. In particular, it offers multiple viewers wide field-of-view imagery at the same time. Consequently, projected labels on a physical object's surface are useful in many application fields, particularly those in which the object is statically placed and shared by multiple viewers who are moving freely around it. Projection objects of such application fields include dioramas or historically important objects in a museum, anatomical models in a hospital or medical school, and 3D terrain maps for civil engineering meetings. We regard these as our target application fields.

In this paper, we introduce a new label layout technique for a projection-based AR application that determines the placement of each label superimposed directly onto an associated physical object with a surface that is normally not suitable for projection. Central to our technique is a new legibility estimation method that evaluates how easily people can read a label that is projected on nonplanar and textured surface from arbitrary viewpoints. The estimation relies on the results of a psychophysical study in which we investigated the legibility of a projected label in various cases, considering the characters' deformation, contrast decrease, and whether or not a shadow was cast on them. In the same manner as in the case of conventional label layout solvers, we computed a label layout by minimizing an energy function. The terms in the function quantitatively evaluate different aspects of layout quality. In conventional solvers, the geometric relationships among labels, anchor regions, and leader lines are evaluated. For example, the terms penalize when one label overlaps another and when the leader line is lengthy. In addition to these evaluations, we designed our energy function to deal with the following factors, inherent in projection-based AR applications: estimated label legibility and disconnection of the projected leader line. Fig. 1 shows an overview of the proposed method.

• *The authors are with the Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan.*

Manuscript received 13 Aug. 2012; revised 19 Nov. 2012; accepted 5 Dec. 2012; published online 12 Dec. 2012.
Recommended for acceptance by A. Steed.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2012-08-0156.
Digital Object Identifier no. 10.1109/TVCG.2012.321.

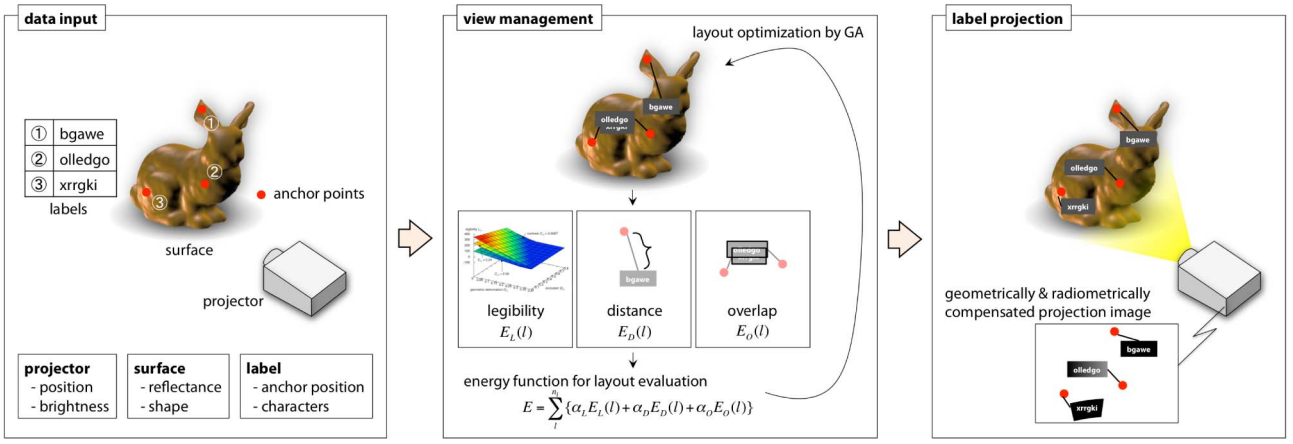


Fig. 1. Overview of the proposed method.

To summarize, this paper makes the following contributions:

- We describe a label layout method for projection-based AR applications to determine the placement of each label superimposed directly onto a nonplanar and textured surface.
- We investigate, through a psychophysical study, the legibility of the labels that are projected onto surfaces of various shapes and textures.
- We introduce new terms to the general energy function used to solve the label layout problems that are inherent in projection-based AR applications: i.e., the legibility degradation of projected labels and the disconnection of projected leader lines.

1.1 Related Works

Illustrations or graphs in scientific and technical textbooks and maps commonly use labels, legends, or annotations containing texts to help viewers understand images. However, finding the optimal solution of the general labeling problem has proven to be NP-hard [3]. Several approximation methods have been developed to reduce the computational complexity. One implements a real-time label layout [4], while another presents an approach for learning label layout styles by example to create high-quality layouts [5].

Labeling is also regarded as a very important issue in AR research fields. Following the pioneering work on view management of labels in the context of AR [1], much research has been conducted since then on label layouts displayed in an HMD for optical and video see-through AR. Some researchers focused on geometric relationships among labels, anchor regions, and leader lines to avoid undesirable layouts, such as overlapping labels or intersections of leader lines [6], [7], [8], while others addressed the legibility degradation of a text that has been digitally overlaid on a textured background [9], [10], [11].

A relatively small number of researchers have worked on the view management of projected images. Siriborvornratanakul and Sugimoto proposed projecting images onto a surface that is suitable for projection by avoiding cluttered regions [12]. They focused on detecting the cluttered regions by analyzing the captured image of a scene without explicitly taking into account the 3D shape information of the scene. Uemura et al. focused on the view management of annotations for a wearable projection-based AR application in the context of assembly work support. They proposed projecting annotations on a surface next to a working place so that the user's hands do not block the projected light [13]. On the other hand, in this paper, we address the view management of projected labels by explicitly taking into account the shape and texture of the projection surface and evaluating the legibility of projected characters.

What makes view management of labels in projection-based AR application difficult is the legibility degradation of the projected label, which is particularly caused by the shape deformation, the decrease in the contrast of the label, and the cast shadows. There are various approaches that geometrically correct the deformation of a projected image on a nonplanar surface [2]. However, these compensation techniques correct the image for only one tracked viewer, while the other viewers still see a deformed one. Because we assume multiple viewers share projected labels at the same place, the techniques are not useful in our context. Researchers have also developed several radiometric compensation approaches that correct photometrically disturbed colors of a projection image reflected on a textured surface [14]. However, in general, the compensation is not always perfect due to environmental lighting and the limited dynamic range of the projector (i.e., a low maximum intensity and a black offset). The imperfection causes a decrease in the contrast of the projected label. The shadow of the projected label is also inevitable. Although the shadow can be compensated for in a multiprojection environment [15], the number of projectors is generally limited and consequently not always sufficient to remove the entire shadow. We focus on the view management of projected labels, regardless of the problems that cause the legibility degradation of the labels.

2 Legibility Estimation of Projected Labels on Nonplanar and Textured Surfaces

This section describes our computational model, which estimates the legibility of a label projected onto a nonplanar and textured surface. First, we explain what kind of parameters we take into account for the estimation. Then

Fig. 2. Geometrically corrected projection characters observed (a) from the optimal viewpoint and (b) from another viewpoint.

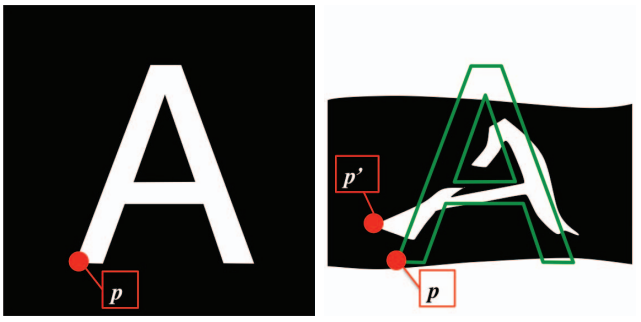


Fig. 3. Pixel correspondence of a character for computing its deformation: (a) a pixel p from the original character, (b) the corresponding pixel p' from the projected character on a non-planar surface (the green outline indicates the original character).

we describe the psychophysical study we conducted to investigate how these parameters affect legibility, and we explain the computational model we used to estimate it. It is important to note that we assume that the reflectance property of a projection surface is Lambertian and that we apply a billboard representation to our labels, made up of a uniform dark background and bright texts. A previous study proved that the billboard representation provides the best legibility among various representations [16].

2.1 Parameters for Estimating the Legibility of Projected Labels

We conducted an informal preliminary experiment in which we projected labels onto various types of nonplanar and textured surfaces and observed them from various viewpoints. We found that three relevant factors strongly correlate to the legibility of labels: the occlusion and the geometric deformation of projected characters, and the contrast of the label. Therefore, we considered that the legibility of a projected label l observed from a viewpoint v can be estimated through the function of these parameters as $L_{l,v}(O_{l,v}, D_{l,v}, C_l)$, where, $L_{l,v}$, $O_{l,v}$, $D_{l,v}$, and C_l represent the legibility of the label, the degree of the occlusion of label characters, the degree of deformation of the characters, and the label's contrast, respectively. In this paper, we compute these parameters for 2D images, because a label on a nonplanar and textured surface is projected onto the retinas of human eyes and observed as a 2D image. We assume that the depth variation of the surface in the label area is not large enough for a human to perceive its shape variation caused by binocular stereopsis. The following paragraphs explain how we model each parameter.

2.1.1 Occlusion

The complex shape of the projection surface causes the occlusion of projected characters. Even though multiple projectors can compensate for shadows, the number of projectors is not always sufficient to remove the occlusions completely. In this paper, we use one projector for displaying labels. Because we focus on occlusion areas of projected characters rather than the number of projectors, the results obtained with this setup can directly be applied to a multiprojection environment. We model the degree of occlusion of characters $O_{l,v}$ as the ratio of the visible (i.e., unoccluded) area of characters ${}^{v}S_{l,v}$ to the area of the original characters ${}^{o}S_{l,v}$. Therefore,

$$O_{l,v} = {}^{v}S_{l,v}/{}^{o}S_{l,v}. \tag{1}$$

2.1.2 Geometric Deformation

Characters in a label are geometrically deformed when projected onto nonplanar surfaces. Geometric correction techniques solve this problem by adjusting the shape of the original characters before projection so that the characters are not deformed while being projected on a nonplanar surface. However, the characters are geometrically correct only when viewed from a certain viewpoint, and still deformed when viewed from the other viewpoints (see Fig. 2). We model the degree of deformation of characters in label l, which is represented as $D_{l,v}$, as follows: As shown in Fig. 3, by comparing a deformed character with its original, we define $D_{l,v}$ as the averaged 2D euclidian distance between the 2D position of point p in the original character, where $\mathbf{x}(p) = (x_p, y_p)$, and that of the corresponding unoccluded point p' in the deformed character, where $\mathbf{x}(p') = (x_{p'}, y_{p'})$:

$$D_{l,v} = \frac{1}{{}^{v}S_{l,v}} \sum_{p} \|\mathbf{x}(p) - \mathbf{x}(p')\|. \tag{2}$$

2.1.3 Contrast

The texture or the spatial variance of the reflectance property of a projection surface lowers the contrast of a projected label. To achieve a billboard representation, which comprises a uniform dark background and bright characters, we apply a radiometric compensation technique, i.e., the method proposed by Bimber et al. [17]. The radiance of the reflection R_p of an unoccluded point q in the label l can be computed by the following equation:

$$R_q = I_q F_q M_q + E_q M_q, \tag{3}$$

where I_q, F_q, M_q, and E_q represent the input intensity value for the projector, the form factor that linearly scales the input intensity to the irradiance from the projector, the reflectance of the pigment, and the irradiance of environmental lighting and the black level of the projector, respectively. After F_q, M_q, and E_q are calibrated, we can compute the input value I_q to display the desired reflection R_q. Based on this model, even when we input zero to a projector (i.e., $I_q = 0$), both the projector's black offset and the environmental lighting elevate the black

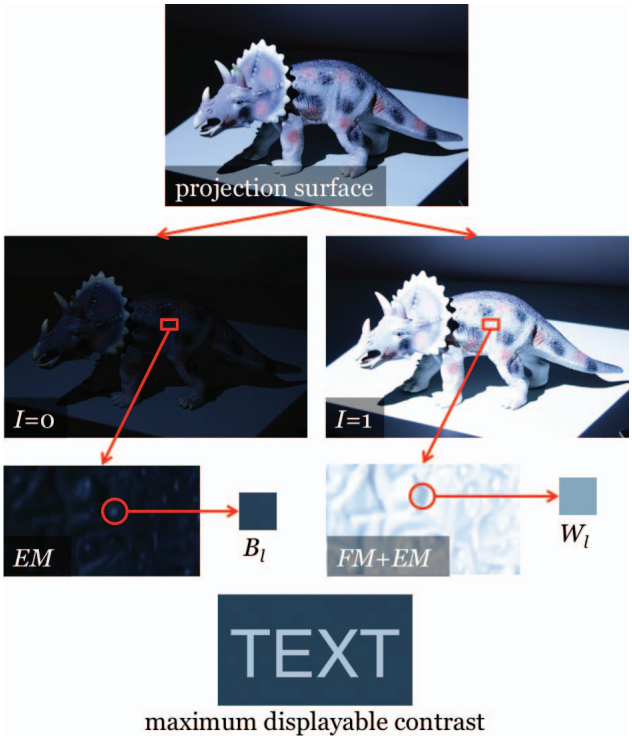


Fig. 4. Contrast of a projected label.

level of reflection. We define the maximum radiance of black level reflections over the area of a label l as the maximum black level B_l:

$$B_l = \max_q (E_q M_q). \tag{4}$$

In addition, the radiance of the reflected white projection light (i.e., $I_q = 1$) becomes small when the light is reflected by a dark pigment. We define the minimum radiance of reflected white light over the area of a label l as the minimum white level W_l:

$$W_l = \min_q (F_q M_q + E_q M_q). \tag{5}$$

As described above, we apply a billboard representation to our label, which comprises a uniform dark background and bright characters. To achieve the uniform radiance of the background and of the characters over a label region, we apply the maximum black level B_l as the radiance of the background and the minimum white level W_l as the radiance of the characters (see Fig. 4). Therefore, the contrast of a projected label l can be computed as follows:

$$C_l = B_l / W_l. \tag{6}$$

As described above, we assume that the reflectance property of a projection surface is Lambertian, where the intensity and color of the reflected light depend on the incident angle of incoming light but not on the viewing angle. Therefore, the contrast of a projected label is view-independent.

2.2 Psychophysical Study for Computational Model of Legibility Estimation

We conducted a psychophysical study to decide the function $L_{l,v}$ for estimating the legibility of projected labels.

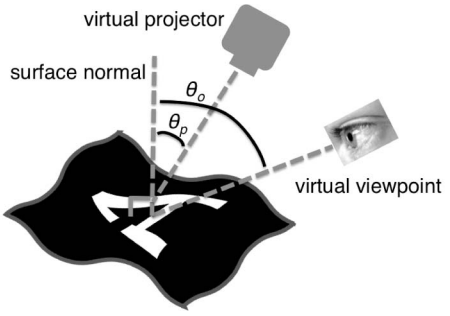


Fig. 5. Virtual environment for CG simulation.

2.2.1 Method

In the experiment, we rendered the CG simulation of a projected label and added Gaussian noise to it. Subjects were asked to observe the rendered image on a 2D monitor and to read the characters in the label. The legibility of a simulated label can be controlled by adjusting the amount of noise. Therefore, if a character in a simulated label can be correctly read even when a strong Gaussian noise has been added, we regarded its legibility as high. On the other hand, if a character with a small amount of added noise cannot be correctly read, we regarded its legibility as low.

Note that we do not use Gaussian noise to simulate textured backgrounds. We assume that the texture of the surface can be canceled by a radiometric compensation technique [17]. The Gaussian noise is used just for decreasing the legibility of a displayed character by changing its spatial intensity pattern. When the difference between the intensity pattern of the noise-added character and that of the original is too large, people cannot correctly read the character. Because the intensity pattern of a deformed, low contrast, and/or occluded character is already changed, people cannot correctly read it, even with a small amount of noise. Therefore, if people can read a character with a large amount of noise, we regard its legibility as high, and vice versa.

Fig. 5 shows the virtual environment used in our CG simulation. A virtual projector projected a label onto a nonplanar surface s and rendered the projected scene from a virtual viewpoint. We defined the projector's view angle as θ_p and the virtual observer's view angle as θ_o as shown in Fig. 5. We applied a geometric correction technique [18] by which the projected character is undistorted when viewed from the direction of the surface normal vector n (i.e., $\theta_o = 0$). Because the surface is nonplanar, projected characters are deformed when viewed from other viewpoints (i.e., $\theta_o \neq 0$) and some parts of the projected label are occluded. We rendered the label as an 8-bit grayscale image. To represent different contrast levels, we fixed the brightness of the character as white (i.e., $W_l = 255$) and changed the gray level of background B_l.

The Gaussian noise added to the rendered label is computed as follows:

$$G(\kappa) = \kappa \sqrt{-2 \log u_1} \cos 2\pi u_2, \tag{7}$$

where $0 < u_1,\ u_2 \leq 1$ are uniform pseudorandom numbers. The amount of noise can be controlled by changing κ. When pixel values are above or below the range of the 8-bit image (0-255), they are clipped.

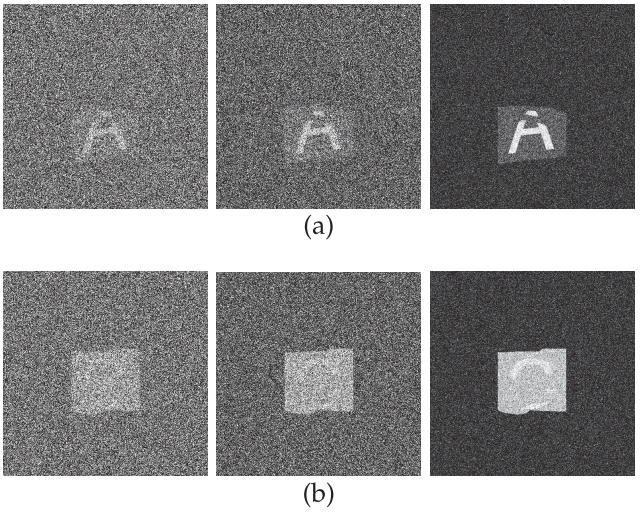


Fig. 6. Stimuli (from left to right, $\kappa = 400, 240, 80$): (a) $C_l = 0.01$ and (b) $C_l = 0.56$.

For each label l that is rendered under the conditions of θ_p, θ_o, and s, the model parameters $O_{l,v}$, $D_{l,v}$, and C_l are computed. A subject observes the rendered labels with different noise levels one by one, from the largest amount of κ to the smallest one. For each noise level, we asked the subject to read the character in the displayed label. We recorded the κ values when the subject could read the character correctly, and we regarded the maximum one κ_{max} as the legibility of label l. Therefore,

$$L_{l,v}(O_{l,v}, D_{l,v}, C_l) = \kappa_{max}. \quad (8)$$

2.2.2 Results

Simulated characters were displayed on a 2D monitor, and each subject observed each of the characters at a distance of 0.6 m from the monitor. The width of a rendered character corresponded to a view angle of two degrees. We extracted 32 different small surface patches from Stanford Happy Buddha as virtual projection surfaces. Five characters, A, B, C, D, and E were projected onto these surfaces from a virtual projector at two different angles ($\theta_p = 1/8\pi, 3/8\pi$ rad). The projected results were rendered by capturing images from three different view angles ($\theta_o = 1/8\pi, 1/4\pi, 3/8\pi$ rad) with a virtual camera. We prepared three different contrast levels by increasing the gray level of the background color while keeping the color of the character as white. We measured the actual contrast between these backgrounds and the white character by displaying these colors on the monitor and measuring their luminance using a luminance meter. As a result, the contrast values, C_l, were 0.01, 0.24, and 0.56. In total, 2,880 (= 32 surfaces $\times$5 characters $\times$2 projector angles $\times$3 view angles $\times$3 contrast values) different rendered labels were prepared for the experiment. For each rendered label, we added six different noise levels for κ of 80, 160, 240, 320, 400, and 480 (see Fig. 6).

Ten subjects were selected from our local university (nine males and one female, aged 21 to 24). All subjects had normal or corrected to normal vision. Participation in the experiments was voluntary and unpaid. Each subject

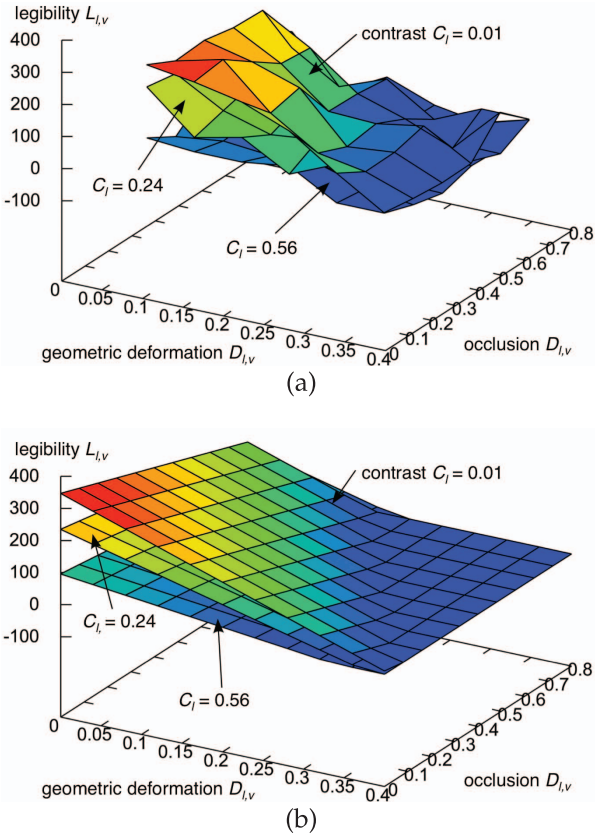


Fig. 7. Averaged $L_{l,v}$ (different surfaces represent different contrast levels): (a) raw data, (b) plane fitting results.

observed 100 different rendered labels out of the total 2,880. Consequently 1,000 trials were conducted in total (10 subjects $\times$100 labels). In each trial, a rendered label was observed with six different noise levels, one by one, from the largest amount of κ to the smallest. The character in the label was not changed when the κ value was decreased. For each noise level, we asked the subject to identify the character in the label. For each rendered label, we recorded the maximum κ value when the subject could identify the character correctly (i.e., κ_{max}). Fig. 7a shows the experimental results of $L_{l,v}$ (= κ_{max}) averaged among subjects.

Due to their relatively linear behavior, the results in each contrast level were fitted to planes using multidimensional linear regression. To apply the general plane equation [19],

$$L_{l,v}(O_{l,v}, D_{l,v}, C_l) = a_1(C_l)O_{l,v} + a_2(C_l)D_{l,v} + a_3(C_l), \quad (9)$$

for parameterizing the fitted functions, we needed to find continuous functions that approximated the discrete plane parameters (a_1, a_2, and a_3) over all contrast levels C_l. The one-dimensional curve fitting resulted in the following fitting functions:

$$a_1(C_l) = -273.1 \exp(-7.3C_l) - 100.7, \quad (10)$$

$$a_2(C_l) = 11.2C_l - 9.1, \quad (11)$$

$$a_3(C_l) = 390.3C_l^2 - 753.7C_l + 395.4. \quad (12)$$

The parameters a_1 and a_2 correspond to the gradients of the planes in directions $O_{l,v}$ and $D_{l,v}$, and a_3 represents the shift

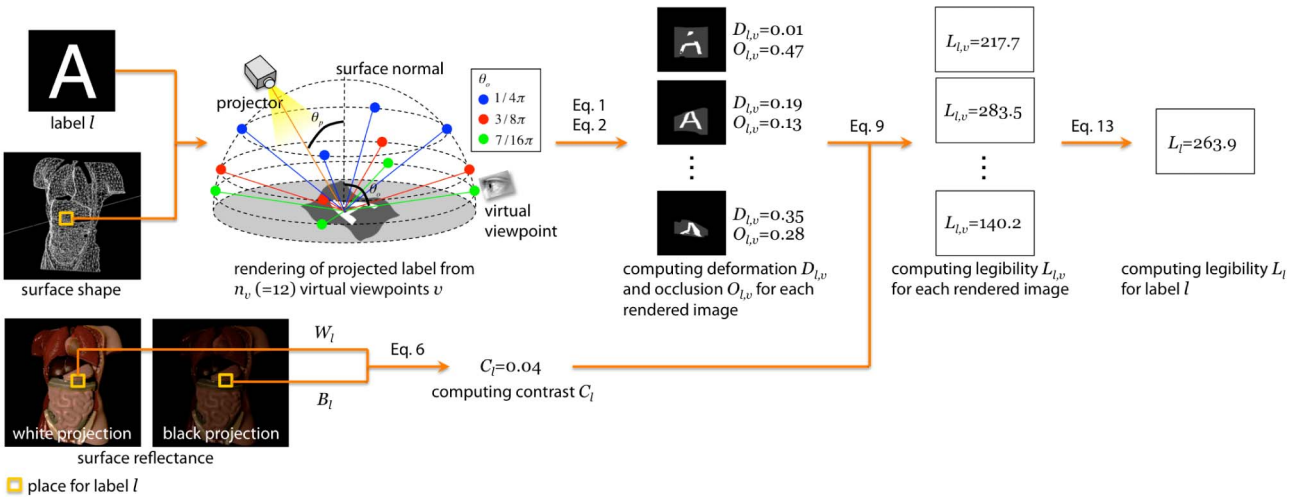


Fig. 8. Process flow of legibility estimation.

of the legibility $L_{l,v}$. The fitted planes are shown in Fig. 7b, where the values below zero are clipped to zero.

2.3 Legibility Estimation

Using (9), we estimate the legibility of a label l observed from a viewpoint v. However, because we do not assume that the observer's viewpoint is known, an estimated legibility for a fixed viewpoint v is not useful. Furthermore, we assume that multiple viewers simultaneously observe the same object from different viewpoints in our target applications, as described in Section 1. Therefore, we regard it as good when the legibility is not only high from a single viewpoint but also generally high from multiple viewpoints. We average the estimated multiple legibility values of a label l as observed from different viewpoints v_i as follows:

$$L_l = \frac{1}{n_v}\sum_{i}^{n_v} L_{l,v_i}, \tag{13}$$

where n_v represents the number of viewpoints.

In this paper, we sample 12 viewpoints (i.e., $n_v = 12$), and average the estimated legibility of a projected label from these viewpoints. The viewpoints are the combinations of four longitude ($\phi_o = 0,\ \pi/2,\ \pi,\ 3\pi/2$) and three latitude ($\theta_o = \pi/4,\ 3\pi/8,\ 7\pi/16$) angles. We apply such uneven latitude sampling because the deformation of projected characters is significant when observed near the viewing angle of $\pi/2$. In addition, characters are not significantly deformed when observed near the normal of a projection surface.

The entire process of legibility estimation is depicted in Fig. 8.

2.4 Validation of Proposed Computational Model

We validated the proposed computational model of legibility estimation using a subjective experiment. The aim of this experiment was to evaluate the legibility of projected labels estimated by our proposed computational model by comparing them with those ranked by subjects. When ranking legibility, subjects observed labels projected onto real nonplanar and textured surfaces.

The surfaces used in the experiment were those of an anatomical model and a 3D terrain map. Fourteen labels were projected onto different locations on the surface. The locations were fixed throughout the experiment. Each label contained a single character, which was randomly selected from 10 different characters: A, B, C, D, E, F, G, H, J, or K. Different characters were selected for each subject. Projected labels were geometrically corrected and radiometrically compensated for by the techniques explained in Section 2.1, based on the shapes and reflectance properties of the surfaces, which were measured in advance. Consequently, the legibility of the projected characters was estimated with the proposed computational model (13).

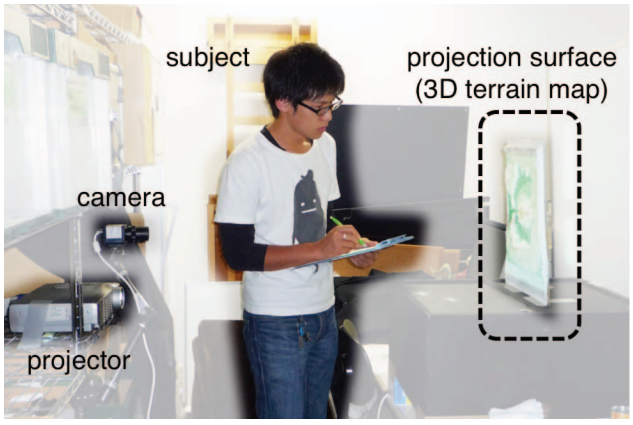
Ten subjects were selected from our local university (eight males and two females, aged 22 to 24). All subjects had normal or corrected to normal vision. Participation in the experiment was voluntary and unpaid. They were allowed to move their bodies and heads to observe projected labels from arbitrary viewpoints. Fig. 9 shows the overview of the experiment. Fig. 10 shows the relationships between the rankings for projected characters averaged among subjects and the estimated legibility for each label location. A linear regression showed that there were associations between the estimated legibility and subjective rankings ($R^2 = 0.59,\ p < 0.01$ for the anatomical model and $R^2 = 0.65,\ p < 0.01$ for the 3D terrain map). Therefore, we considered that the proposed computational model can correctly estimate the legibility of a projected label.

3 View Management of Projected Labels

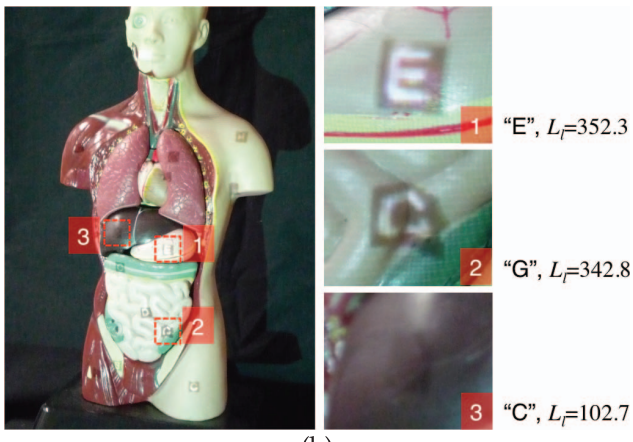
We computed a label layout by minimizing an energy function based on genetic algorithm (GA).

3.1 Energy Function

In a conventional label layout solution, the geometric relationships among labels, anchor regions, and leader lines are evaluated. For example, the terms penalize overlapping of labels and lengthy leader lines. In addition to these evaluations, we design our energy function to deal with the following unique factors that are inherent in projection-based AR applications: the legibility of a pro-



(a)



(b)

Fig. 9. Overview of the experiment: (a) experimental setup and (b) projected labels on anatomical model and their estimated legibility.

jected character and the disconnection of a projected leader line. When there are n_l labels to be projected, we formulate the energy E as follows:

$$E = \sum_{l}^{n_l} (\alpha_L E_L(l) + \alpha_D E_D(l) + \alpha_O E_O(l)), \tag{14}$$

where $E_L(l)$, $E_D(l)$, and $E_O(l)$ represent the energy terms for a label l, i.e., the legibility of the characters in the label, the distance of the leader line, and the label's overlap, respectively, while α_L, α_D, and α_O represent weights. The weights allow a user to balance the importance of each term.

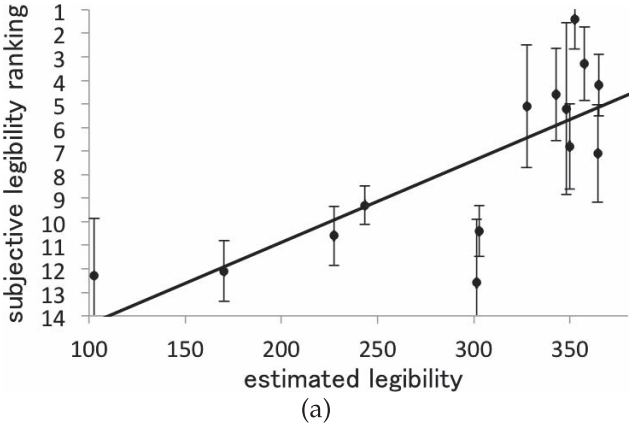
3.1.1 Legibility

The legibility of projected characters in a label is estimated by (13). The energy term for the legibility is represented as the reciprocal of the legibility.

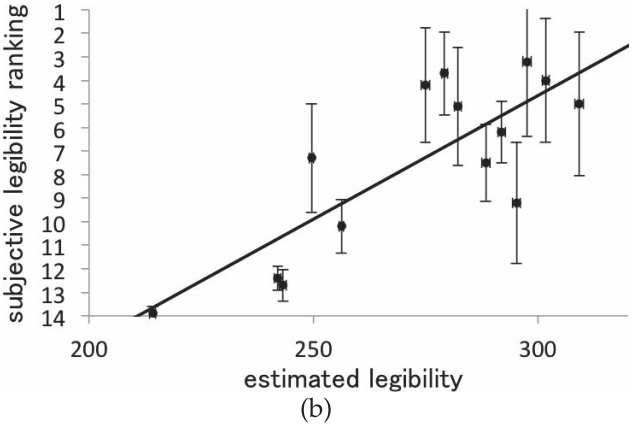
$$E_L(l) = 1/L_l. \tag{15}$$

3.1.2 Distance

In general, the distance from an anchor point to the corresponding label is one of the most important parameters when deciding an optimal label layout. In addition, when considering the layout of projected labels, we need to take into account the disconnection of projected leader lines caused by the nonplanarity of the projection surface.



(a)



(b)

Fig. 10. Relationships of subjective legibility rankings and the estimated legibility with a fitted line through a linear regression: (a) anatomical model and (b) 3D terrain map.

To solve this issue, we propose taking into consideration the length on the surface, instead of a simple euclidian distance. Because we measure the shape of the projection surface for the geometric correction, we already have mesh data (vertices and edges) of the surface. Thus, we can apply Dijkstra's Algorithm to search for the shortest path along mesh edges from the vertex of an anchor to that of the corresponding label, and regard the length of the path as the distance d_l. If there is no path from the anchor to the label, we give a large constant value c_D to the energy term

$$E_D(l) = \begin{cases} d_l, & (\text{if connected}), \\ c_D, & (\text{otherwise}), \end{cases} \tag{16}$$

3.1.3 Overlap

When a label overlaps another label, we add a penalty to the energy function. This is represented as follows:

$$E_O(l) = \begin{cases} 1, & (\text{if overlapped}), \\ 0, & (\text{otherwise}). \end{cases} \tag{17}$$

3.2 Optimization

We search for the optimum label layout which minimizes the energy E of (14). As mentioned in Section 1, the projection object is statically placed and shared by multiple viewers who are moving freely around it in our target applications. In addition, projection-based AR is suitable for indoor applications because the projection light is too dark

to show images under the sunlight. Therefore, we assume that our projection object is placed indoors. The indoor environment light is generally regarded as static. In this condition, the optimum label layout changes only when a new label is added or an existing one is removed. In most of our target application fields (e.g., museums, medical schools, and civil engineering meetings), it is reasonable to say that users request to change labels infrequently, for example once a minute at the most frequent, and the optimal label layouts are computed in offline. That is, labels for all the anchor regions are projected simultaneously. Therefore, we design our optimization framework by taking into account that the proposed technique is going to be used in such situations.

We choose GA as our optimization method, in particular the minimal generation gap model [20], because it gives us a globally optimum solution. Although GA normally does not work in real time, we do not need a real-time solution for the reasons described above. We consider the layout of N_{label} labels. The index of each vertex of the projection surface where label l is projected is represented as v_l. We regard v_l as a gene in our GA. An individual solution comprises these genes ($I = (v_1, v_2, , \ldots, v_{N_{label}})$). In the initialization, we generate a population that comprises N_{seed} individual solutions by randomly selecting the placements of labels. Then the following process is repeated N_{loop} times.

Two parent solutions are randomly selected from the existing population. Crossover and mutation are applied to the selected parents N_{family} times to generate $2N_{family}$ child solutions. We choose two solutions that fit the best from the parent and child solutions (i.e., $2 + 2N_{family}$ solutions), where the fitting function is identical to our energy function. A new population is created with the chosen ones by replacing the parent solutions.

After N_{loop} repetitions, we choose the best solution as the optimal layout of the projected labels.

4 Evaluations

A user study was conducted to evaluate our proposed label layout technique. We compared a label layout generated by minimizing the proposed energy function (14) with non-zero α_L (proposed condition), to that generated with $\alpha_L = 0$ (conventional condition). In other words, only conventional constraints (the distance from anchors to labels and overlap of labels) were considered in the conventional condition.

We asked a subject to read the characters of the projected labels under both conditions as quickly and as correctly as possible, and we measured the task completion time, the percentage of correct answers, and the subjective preference of the layout. We used a 3D terrain map as the projection surface on which we chose 20 anchor points. A label consisting of three different characters was assigned to each anchor point. The characters were randomly selected from the Roman alphabet. "Q" and "V" were excluded because we thought they might be confused with "O" and "U," respectively.

Each task was started when the subject pressed the start button to commence the evaluation program. Then, labels were projected onto the surface and were connected to their anchors by leader lines. Four randomly selected anchors were red, and the others were blue. The subject was required to read out loud the labels connected to the red anchors and press the stop button to terminate the program when finished reading. The task completion time was measured as the duration time between each key press action. After each task, the subject rated the preference of the layout, according to a 7-point Likert scale from -3 (least preferred) to 3 (most preferred). We did not give any feedback to the subjects about the accuracy of their answers, and the optimal label layout varied among different tasks because the estimated legibility varied according to the difference of the projected characters for each task.

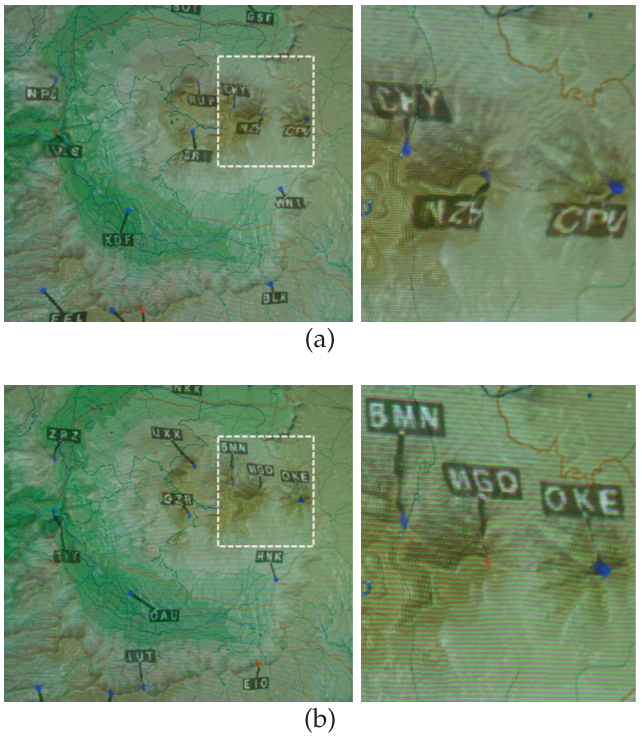


Fig. 11. Projection examples of (a) conventional condition and (b) proposed condition.

Ten subjects were selected from our local university (nine males and 1 female, aged 21 to 24). All subjects had normal or corrected to normal vision. Participation in the experiment was voluntary and unpaid. Note that three subjects participated in all the experiments described in this paper. We allowed subjects to freely move their bodies and heads to read the labels, while fixing the initial position of their faces at $\theta_o = 1/4\pi$ through the experiment. Each subject completed 50 (25 for each condition) tasks followed by four (two for each condition) exercise tasks. The order of the conditions of the tasks was chosen randomly. The parameters of the energy function and the GA were $\alpha_L = 0.005$ (only for the proposed condition), $\alpha_D = 0.1$, $\alpha_O = 0.2$, $N_{seed} = 500$, $N_{family} = 500$, and $N_{loop} = 1{,}000$. It took 7.3 s to compute each layout with a PC (Intel Core i7 3.2 GHz). Fig. 11 shows the projection examples of both conditions.

Fig. 12a shows the averaged task completion time for each condition. Subjects could read aloud the projected characters within 8.53 s in the proposed condition and 9.41 s in the conventional condition. A paired T test between them

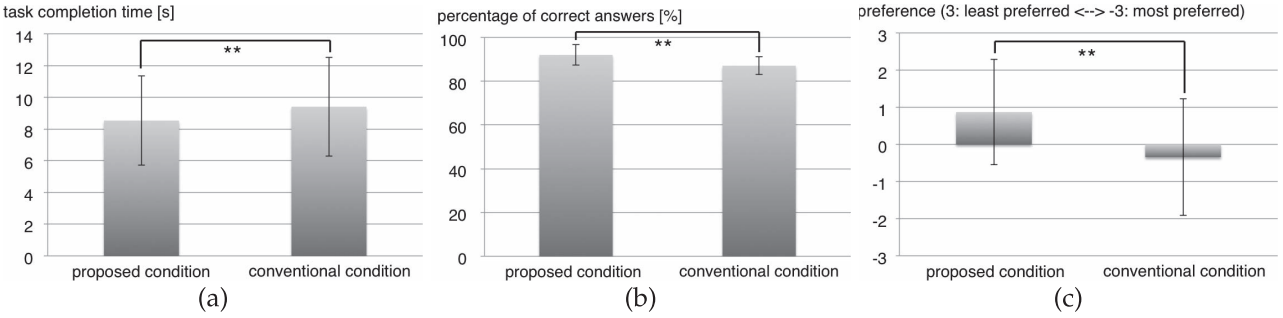


Fig. 12. Experimental results: (a) averaged task completion time, (b) averaged percentage of correct answers, and (c) averaged preferences (the bar is the standard deviation, ** : $p < 0.01$).

($t_{498} = 3.29,\ p < 0.01$) showed that the proposed condition could significantly shorten the time compared to the conventional condition.

Fig. 12b shows the averaged percentage of correct answers for each condition. Subjects could correctly read the projected characters during 92.0 percent of the tasks in the proposed condition and during 87.1 percent of the tasks in the conventional condition. A paired T test between them ($t_9 = 4.77,\ p < 0.01$) showed that the proposed condition could significantly improve the percentage of correct answers compared to the conventional condition.

Fig. 12c shows the averaged subjective preference for each condition. On average, the subjects reported a positive preference ($+0.87$ from median) for the proposed condition and a negative preference (-0.34 from median) for the conventional condition. An unpaired T test between them ($t_{498} = 9.06,\ p < 0.01$) showed that the proposed condition could provide a layout that was significantly more preferred than the conventional condition.

According to the experimental results, the label layout is surely improved by taking into account the legibility of characters in labels projected onto a nonplanar and textured surface when computing the optimal layout. In principle, the length of the reader lines is longer in the proposed condition because α_L is relatively larger in the energy function of the proposed condition than in that of the conventional condition. The actual averaged lengths in the experiment were 17.7 and 12.7 mm in the proposed and conventional conditions, respectively. Therefore, even though the leader lines are longer, we could reduce the cognitive load of the subjects in understanding labels by introducing the legibility of projected characters into the energy function.

As shown in Fig. 12a, variations are large in task completion time. The legibility of projected characters differs significantly among the 20 labels. Even when the layout is determined by the proposed method, some labels are easy to read while others are difficult to read. Because the proposed method optimizes the layout by taking into account not only the legibility but also the distance between the leader line and the label's overlap, the legibility of some projected labels is still low. This variation of legibility causes the large variation in task times. As shown in Fig. 12b, there is a small difference in the averaged percentages of correct answers. We think that the task was not difficult enough to show a great improvement when using the proposed method. On the other hand, through the experiment, we also noticed that people have the ability to correctly read considerably deformed characters if they spend a lot of time. To sum up, our proposed method could shorten the reading time of projected labels, but it did not improve the correctness.

5 Conclusion

We explored a model for computing the layout of projection-based labels on nonplanar and textured surfaces. The model took into account the inherent problems of projection-based technologies, specifically, the legibility degradation of projected characters and the disconnection of a leader line connecting an anchor to its corresponding label. For the first issue, we proposed a technique for estimating the legibility of projected characters by considering their deformation, contrast loss, and the occlusion caused by nonplanarity and spatial variance of the reflectance of the projection surface. A subjective experiment showed that the technique performed well when estimating the legibility of projected characters. For the second issue, we proposed to calculate the path of the leader line from an anchor to the label on the projection surface using Dijkstra's Algorithm instead of simple euclidian distance. Through subjective experiments, we ascertained the fact that the proposed model could improve label layout, because the subjects could understand the labels better and preferred this layout over other layouts. Taking into account the processing time of 7.3 s, we believe that our GA-based optimization is useful in our target applications.

In the proposed technique, the shapes and reflectance of the projection surfaces need to be measured prior to the legibility computation. However, we do not regard this as a limitation of our approach. Since we assume that the projection objects in our target applications are static, the measurement is not required online.

Our method suffers from the disadvantages of conventional projection-based AR applications. Due to the limited dynamic range and the brightness of the projector, the contrast of the projected images defined by (6) might become too low for users to correctly read texts projected onto dark surface pigments. The usage of color information may improve legibility even when the brightness of the projected label is not sufficient. This being our first attempt to assess the legibility of projected characters, we focused only on the luminance because the luminance is more

dominant than chrominance in the human visual system. We believe that chrominance information is also useful for the legibility of projected characters. Extending the proposed computational model of legibility estimation to the chrominance domain could be one of our future projects.

We empirically determined the weights of (14) by computing label layouts with different weight values and checking the results. For various target objects with different shapes and textures, users need to search for the weights through the same procedure. On the other hand, we believe that this is useful for the users because they can optimize the weights for their individual purposes. For example, when a user would like to place the labels closer to the anchors, he/she can do so simply by setting α_D to be larger than to the other weights. Finding a more efficient way of determining the weights is also our future work.

REFERENCES

[1] B. Bell, S. Feiner, and T. Höllerer, "View Management for Virtual and Augmented Reality," *Proc. ACM Symp. User Interface Software and Technology,* pp. 101-110, 2001.

[2] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds.* A.K. Peters Ltd., 2005.

[3] J. Christensen, J. Marks, and S. Shieber, "An Empirical Study of Algorithms for Point-Feature Label Placement," *ACM Trans. Graphics,* vol. 14, pp. 203-232, 1995.

[4] K. Ali, K. Hartmann, and T. Strothotte, "Label Layout for Interactive 3D Illustrations," *The J. WSCG,* vol. 13, pp. 1-8, 2005.

[5] I. Vollick, D. Vogel, M. Agrawala, and A. Hertzmann, "Specifying Label Layout Style by Example," *Proc. ACM Symp. User Interface Software and Technology,* pp. 221-230, 2007.

[6] R. Azuma and C. Furmanski, "Evaluating Label Placement for Augmented Reality View Management," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality,* pp. 66-75, 2003.

[7] K. Makita, M. Kanbara, and N. Yokoya, "View Management of Annotations for Wearable Augmented Reality," *Proc. IEEE Int'l Conf. Multimedia and Expo,* pp. 982-985, 2009.

[8] F. Zhang and H. Sun, "Dynamic Labeling Management in Virtual and Augmented Environments," *Proc. Int'l Conf. Computer Aided Design and Computer Graphics,* pp. 397-402, 2005.

[9] J.L. Gabbard, J.E. Swan, and D. Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality," *Presence: Teleoperators and Virtual Environments,* vol. 15, pp. 16-32, 2006.

[10] J.L. Gabbard, J.E. Swan, D. Hix, S.-J. Kim, and G. Fitch, "Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications," *Proc. IEEE Virtual Reality Conf.,* pp. 35-42, 2007.

[11] A. Leykin and M. Tuceryan, "Automatic Determination of Text Readability Over Textured Backgrounds for Augmented Reality Systems," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality,* pp. 224-230, 2004.

[12] T. Siriborvornratanakul and M. Sugimoto, "Clutter-Aware Adaptive Projection Inside a Dynamic Environment," *Proc. ACM Symp. Virtual Reality Software and Technology,* pp. 241-242, 2008.

[13] K. Uemura, K. Tajimi, N. Sakata, and S. Nishida, "Annotation View Management for Wearable Projection," *Proc. Int'l Conf. Artificial Reality and Telexistence,* pp. 202-206, 2010.

[14] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer, "The Visual Computing of Projector-Camera Systems," *Computer Graphics Forum,* vol. 27, no. 8, pp. 2219-2254, 2008.

[15] M. Nagase, D. Iwai, and K. Sato, "Dynamic Defocus and Occlusion Compensation of Projected Imagery by Model-Based Optimal Projector Selection in Multi-Projection Environment," *Virtual Reality,* vol. 15, pp. 119-132, 2011.

[16] J. Jankowski, K. Samp, I. Irzynska, M. Jozwowicz, and S. Decker, "Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability," *Proc. Int'l Conf. Human Factors in Computing Systems,* pp. 1321-1330, 2010.

[17] O. Bimber, A. Emmerling, and T. Klemmer, "Embedded Entertainment with Smart Projectors," *IEEE Computer,* vol. 38, no. 1, pp. 56-63, Jan. 2005.

[18] G. Welch, H. Fuchs, R. Raskar, H. Towles, and M.S. Brown, "Projected Imagery in Your 'Office of the Future'," *IEEE Computer Graphics and Applications,* vol. 20, no. 4, pp. 62-67, July 2000.

[19] A. Grundhöfer, M. Seeger, F. Hantsch, and O. Bimber, "Dynamic Adaptation of Projected Imperceptible Codes," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality,* pp. 161-168, 2003.

[20] H. Satoh, M. Yamamura, and S. Kobayashi, "Minimal Generation Gap Model for GAS Considering both Exploration and Exploitation," *Proc. Int'l Conf. Soft Computing,* pp. 494-497, 1996.

Daisuke Iwai received the BS, MS, and PhD degrees from Osaka University, Japan, in 2003, 2005, and 2007, respectively. He was a visiting scientist at Bauhaus-University Weimar, Germany, from 2007 to 2008. He is currently an associate professor at the Graduate School of Engineering Science, Osaka University. His research interests include human-computer interaction and projection-based mixed reality.

Tatsunori Yabiki received the BS and MS degrees from Osaka University, Japan, in 2010 and 2012, respectively. He is currently a software development engineer at Cybozu Corporation. His research interests include projection-based mixed reality.

Kosuke Sato received the BS, MS, and PhD degrees from Osaka University, Japan, in 1983, 1985, and 1988, respectively. He was a visiting scientist at the Robotics Institute, Carnegie Mellon University, from 1988 to 1990. He is currently a vice dean of the Graduate School of Engineering Science, Osaka University. His research interests include image sensing, 3D image processing, digital archiving, and virtual reality. He is a member of the IEEE, the IEEE Computer Society, and the ACM.