A New View Management Method
for Wearable Augmented Reality Systems
- Emphasizing the User-viewed Object
and the Corresponding Annotation -

R. Tenmoku[1] and M. Kanbara[1] and N. Yokoya[1]

[1]Nara Institute of Science and Technology, Japan

Abstract

This paper describes a new view management method for annotation overlay using augmented reality(AR) systems. The proposed method emphasizes the user-viewed object and the corresponding annotation in order to present links between annotations and real objects clearly. This method includes two kinds of techniques for emphasizing the user-viewed object and the annotation. First, the proposed method highlights the object which is gazed at by the user using a 3D model without textures. Secondly, when the user-viewed object is occluded by other objects, the object is complemented by using an image made from a detailed 3D model with textures. This paper also describes experiments which show the feasibility of the proposed method by using a prototype wearable AR system.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation]: Artificial, augmented, and virtual realities

1. Introduction

Since computers have made remarkable progress in resent years, a wearable computer can be realized [SWP97, Man99, VSGP03]. At the same time, augmented reality(AR), which merges the real and virtual worlds, has received a great deal of attention as a new method for displaying information [FMS93, ABB+01, KY04]. Wearable AR systems which merge these two technologies make it possible to intuitively present location-based information to the user. To realize wearable AR systems [SKC+01, HFT+99, KKS01, TKY03, KTO04], measuring the user's position in a wide area is necessary. The research area of wearable AR systems has progressed as developing of positioning methods [WJH97, NF02, SYI+03, NKY05].

In recent years, annotation overlay systems with wearable computers have been studied in various respects. For example, Makita constructed a shared database of annotations in a networked server [MKY04]. By using the database, the renewal of annotations will be reflected efficiently for wearable computers of multiple users. Some other researchers proposed annotation presentation methods considering depths and occlusions of annotations [BH04, UMKT05]. Authoring tools [GF03] for wearable AR systems has also proposed in recent years.

On the other hand, view management methods for wearable AR systems have also proposed. Azuma proposed a view management method for AR scenes which rearranges multiple annotations without mutual overlap [AF03]. Bell proposed a method for VR and AR scenes, which can rearrange annotations using image regions where no object is projected on based on 3D models of scenes [BFH02]. Leykin proposed an automatic algorithm which determines positions of annotations by analyzing background textures [LT04]. However, most of previous works focused on how to arrange annotations on augmented scenes when there exists multiple annotations. The proposed method focuses on how to intuitively present links between annotations and their target objects.

In the proposed method, the user-viewed object and the corresponding annotation are presented with emphasis us-

Figure 1: *Outline of wearable annotation overlay system.*

ing 3D models of real scenes. In this paper, two kinds of techniques for emphasizing annotations are described. First, the proposed method highlights the object gazed at by the user using a 3D model without textures. Secondly, when the user-viewed object is occluded by other objects, the object is complemented by using an image made from a detailed 3D model with textures on the overlaid image. These annotation emphasis images support the user's intuitive understanding of links between annotations and their target objects.

This paper is structured as follows. Section 2 describes the proposed method of view management in detail. In Section 3, an experiment with a prototype system is described. Section 4 describes another experiment which evaluates the proposed method. Finally, Section 5 summarizes the present work and also gives future work.

2. Emphasizing Annotations Using 3D Models

In this section, the generation of emphasized annotation overlay images is described. The proposed method emphasizes user-viewed objects using 3D models of real scenes. First, Section 2.1 briefly describes a wearable annotation overlay systems to which the proposed method is assumed to be applied. The following sections then describe the proposed method in detail.

2.1. Wearable Annotation Overlay System

An outline of wearable annotation overlay systems which employ the proposed method is illustrated in Figure 1. The system measures the user's position and orientation by using sensors in real time. Moreover, the system merges annotations on a real scene image which is captured by a viewpoint camera based on the user's position and orientation. Generated images are presented to the user via a video see-through AR display device. Note that targets of annotations are real objects just like buildings.

2.2. Outline of Proposed Method

Figure 2 illustrates the dataflow of the conventional annotation overlay method and the proposed method. The proposed method includes the following two kinds of annotation emphasis techniques with wearable annotation overlay system mentioned in Section 2.1.

- **A) Highlighting user-viewed object and annotation**
  The user-viewed object and annotation are highlighted to be effectively presented to the user.
- **B) Complementing occlusions of user-viewed object**
  In order to intuitively represent a link between user-viewed object and annotation, occlusions of objects are complemented by using a 3D model with textures.

Two kinds of annotation emphasis images are shown in Figures 2 (a) and (b). In this paper, these images are referred to as an "object highlight image" and an "occlusion complement image", respectively. A main flowchart of the proposed method is illustrated in Figure 3(a). Figures 3(b) and (c) show sub-flowcharts to generate object highlight images and occlusion complement images, respectively. These sub-flowcharts correspond to the "STEP4 Generation and presentation of annotation emphasis image" in Figure 3(a).

To generate object highlight images and occlusion complement images, 3D models of real scene are necessary. To recognize the user-viewed object, a 3D model of real scene which is painted different colors on every object is used. This model is called a "structure model" below. The structure model can be made from CAD data of buildings easily. Moreover, in order to generate occlusion complement images, another 3D model which has surface textures called a "texture model" is needed. By using the texture model of the real scene, it can be realized to complement occluded regions of the user-viewed object so as to present the whole object to the user. Two kinds of generated images support the user's intuitive understanding of links between annotations and their target objects.
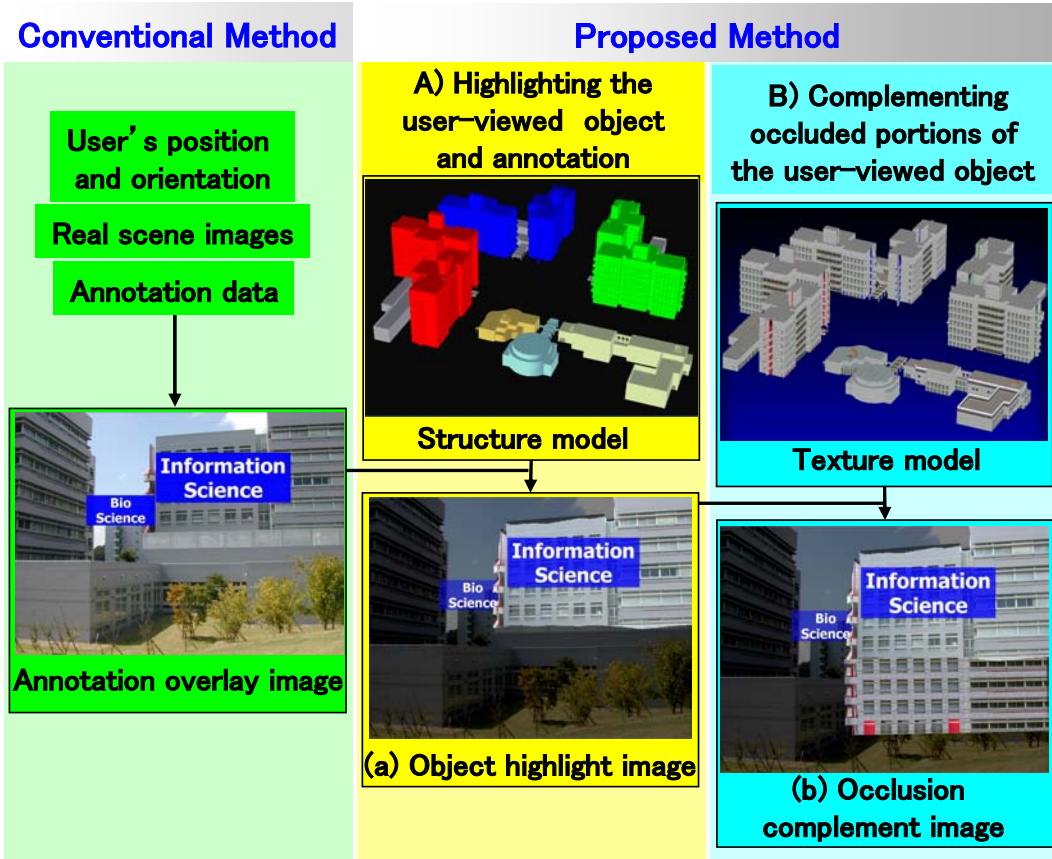
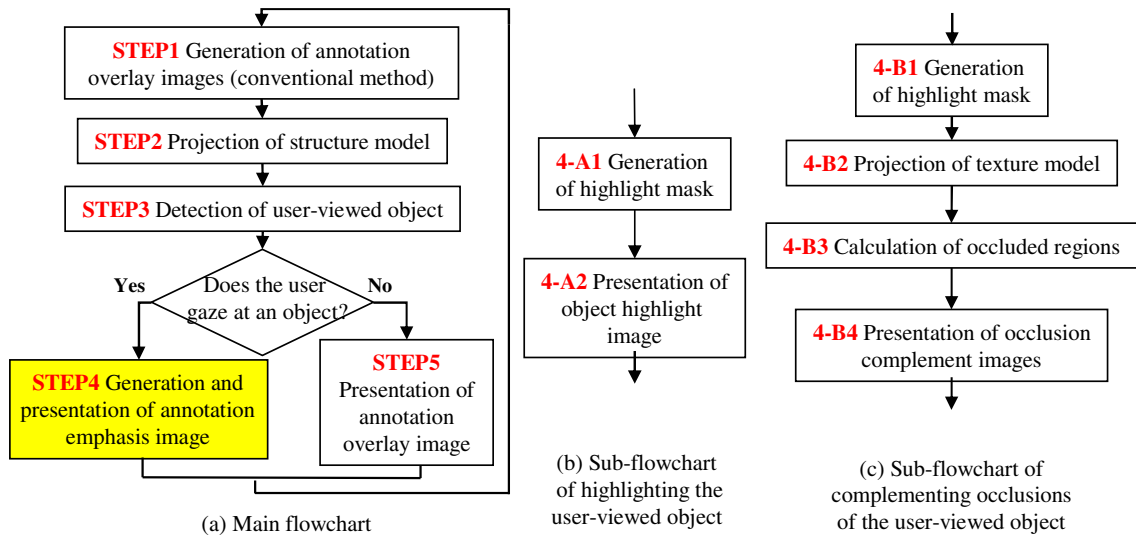**Figure 2:** *Dataflow of conventional and proposed methods.*

Figure 3: *Flowchart of the proposed method.*

(a) Annotations

(b) Real scene image

(c) Structure model projection image

(d) Highlight mask

(e) Texture model projection image

(f) Occluded regions

(g) Object highlight image
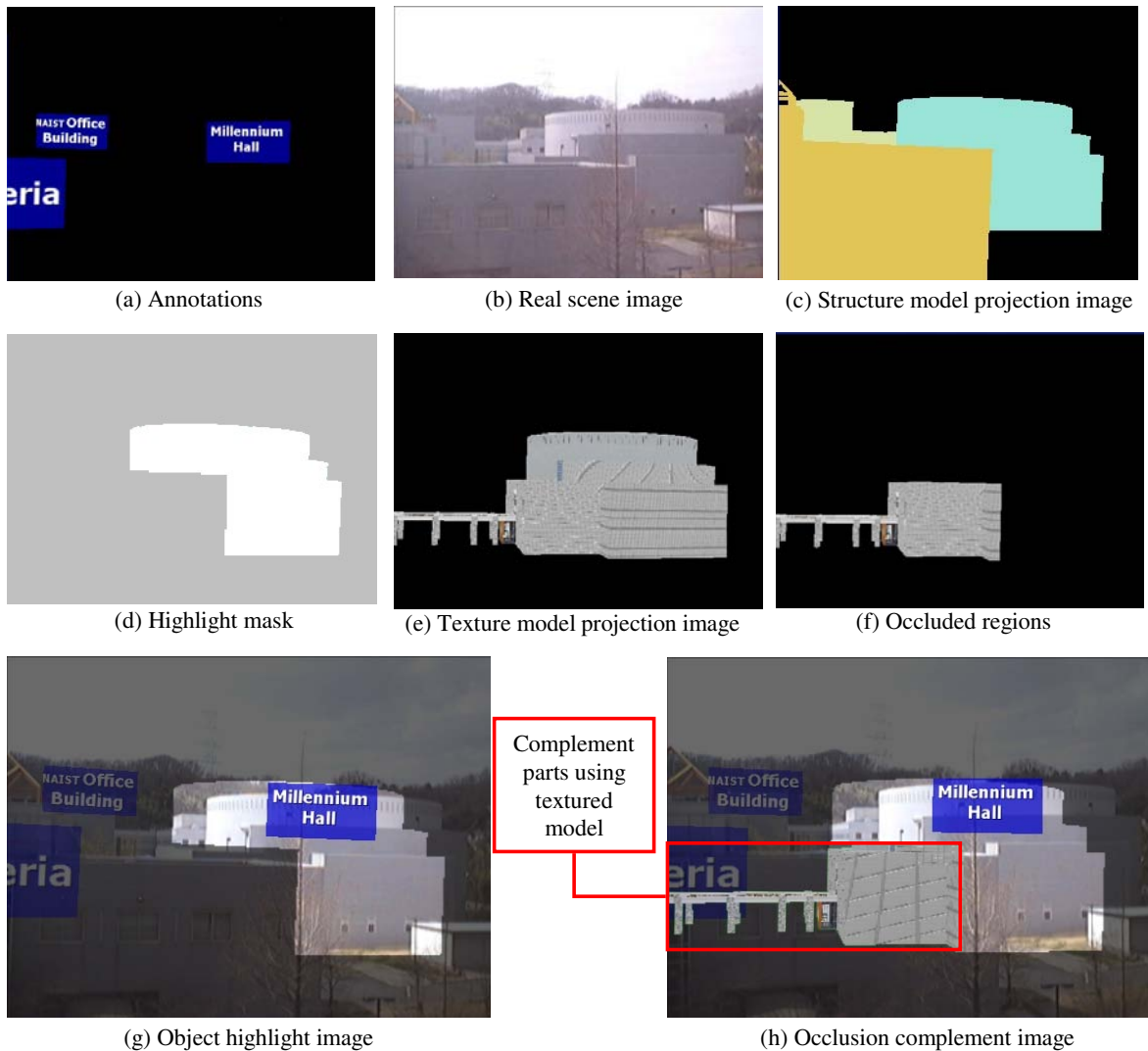
(h) Occlusion complement image

Figure 4: *Elements of generated images.*

In the proposed method, the system presents annotation emphasis images only when the user gazes at a certain object as shown in Figure 3(a). When the system detects that the user does not gaze at any objects, conventional annotation overlay images are displayed to the user. Each step of the main flowchart excepting for STEP4 is described in detail in the following:

[STEP1,5 Generation and presentation of annotation overlay images]
The system overlays annotations (Figure 4(a)) on real scene images (Figure 4(b)) based on the user's position and orientation and annotation data (annotation images and positions) to present them to the user [TKY03].

[STEP2 Projection of structure model]
The system generates the structure model projection images (see Figure 4(c) for an example). The structure model projection images are generated by projecting the structure model to a virtual camera which aligns the same position and orientation as the user's viewpoint.

[STEP3 Detection of user-viewed object]
The system determines the user-viewed object using the structure model projection image shown in Figure 4(c). In the present method, the system recognizes that the user is gazing at an object when it is in a central area of a generated user's view image. To put it concretely, the user-viewed object is determined based on the color of the central area of the structure model projection image.

2.3. Highlighting User-viewed Object and Annotation

Highlighting user-viewed objects consists of the following two steps.

[4-A1 Generation of highlight mask]

The system generates a highlight mask which is shown in Figure 4(d). In the proposed method, a semitransparent black mask is used as a highlight mask. In this mask, only image regions on which a user-viewed object is projected are entirely transparent to present the object with an emphasis.

[4-A2 Presentation of object highlight image]

The proposed method realizes to generate an object highlight image as shown in Figure 4(g) by overlaying the highlight mask (Figure 4(d)) on the annotation overlay image.

2.4. Complementing Occlusions of User-viewed Object

This section describes each step for complementing occlusions of the user-viewed object. Three steps (4-B2, 4-B3, and 4-B4 in Figure 3(a)) of this process are described below. The step 4-B1 is the same operation as the step 4-A1 mentioned in the previous section.

[4-B2 Projection of texture model]

The system generates texture model projection images shown in Figure 4(e). Textured model projection images are generated by projecting the texture model to a virtual camera.

[4-B3 Calculation of occluded regions]

By subtracting a user-viewed object area of the texture model projection image from a projection image of user-viewed object model with texture, occlusion area of the user-viewed object (Figure4(f)) can be obtained.

[4-B4 Presentation of occlusion complement images]

An occlusion complement image (Figure 4(h)) is generated by overlaying an occluded region image (Figure 4(f)) on the highlight mask (Figure 4(d)) and the annotation overlay image.

3. Experiment of Emphasizing User-viewed Object with a Prototype System

We have constructed a prototype system using the proposed method to carry out an experiment of presenting emphasized annotations to the user in our campus. Figure 5 shows the hardware configuration of the prototype system and specifications of each device used in this experiment. In the prototype system, the user equips an IrDA receiver which can identify the position IDs from IrDA markers on the ceiling. IrDA sensors can specify the user's position within 1 meter [TKY03]. Figure 6 illustrates the experimental environment. In this environment, three annotations were overlaid on their target buildings: "NAIST Office Building" (which will be called "Office" hereafter), "Millennium Hall" ("Hall"), and
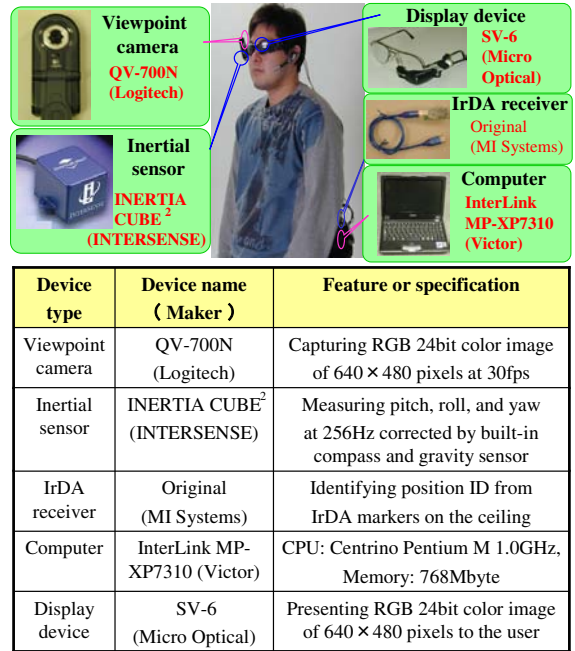

Device type	Device name (Maker)	Feature or specification
Viewpoint camera	QV-700N (Logitech)	Capturing RGB 24bit color image of 640×480 pixels at 30fps
Inertial sensor	INERTIA CUBE^2 (INTERSENSE)	Measuring pitch, roll, and yaw at 256Hz corrected by built-in compass and gravity sensor
IrDA receiver	Original (MI Systems)	Identifying position ID from IrDA markers on the ceiling
Computer	InterLink MP-XP7310 (Victor)	CPU: Centrino Pentium M 1.0GHz, Memory: 768Mbyte
Display device	SV-6 (Micro Optical)	Presenting RGB 24bit color image of 640×480 pixels to the user

Figure 5: *Hardware configuration of prototype system and specifications of devices.*
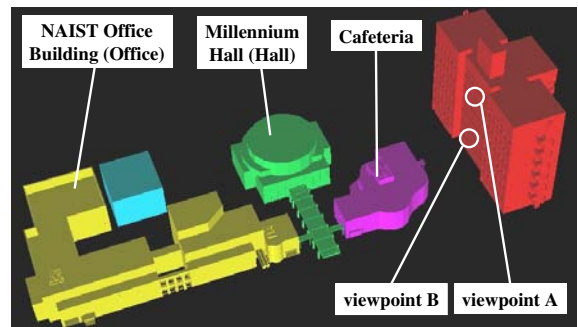

Figure 6: *Experimental environment.*

"Cafeteria". This experiment had two trials. In the first trial, object highlight images were presented to the user based on a structure model of this experimental environment. The second trial presented occlusion complement images to the user based on a texture model of this environment.

Figures 7 and 8 show output images at the viewpoint A (the sixth floor of the right building in Figure 6) and the viewpoint B (the third floor of the same building), respectively. In both figures, (a) shows a generated image when the user does not gaze at any object. Figures 7(b), (c), (e), 8(b), and (c) represent object highlight images. These images show that the proposed method can highlight almost correct parts of the user-viewed object so as to support the
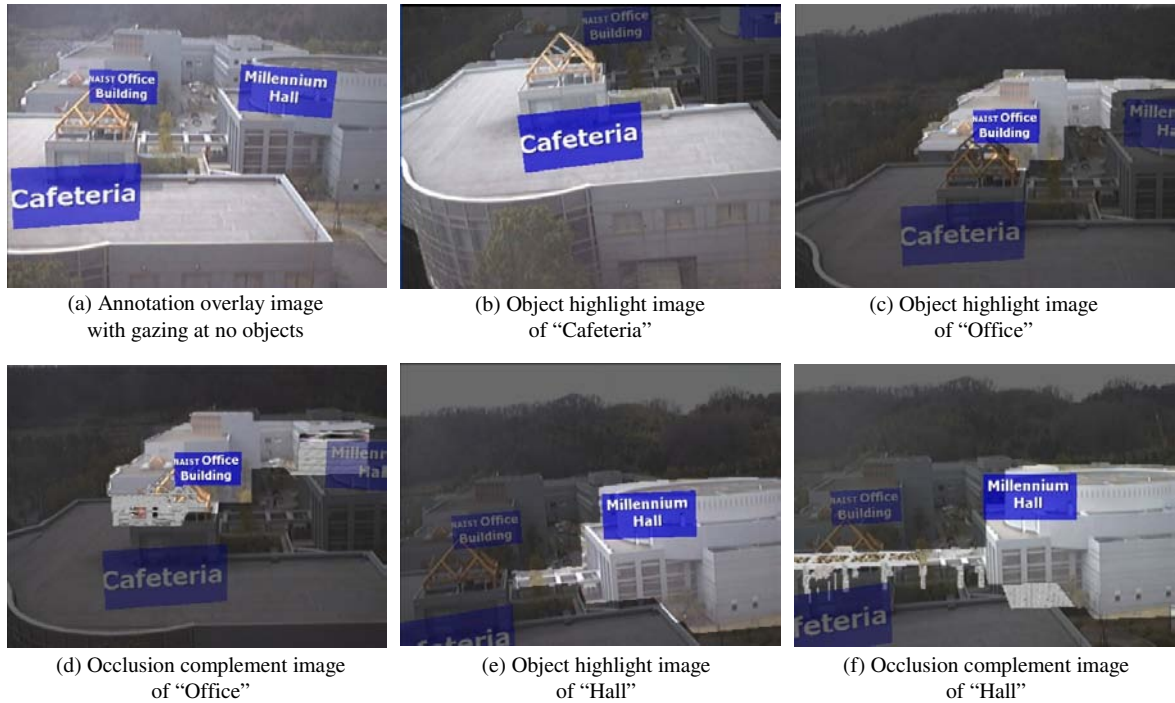
(a) Annotation overlay image
with gazing at no objects

(b) Object highlight image
of "Cafeteria"

(c) Object highlight image
of "Office"

(d) Occlusion complement image
of "Office"

(e) Object highlight image
of "Hall"

(f) Occlusion complement image
of "Hall"

Figure 7: *Generated images at the viewpoint A.*

(a) Annotation overlay image
with gazing at no object

(b) Object highlight image
of "Cafeteria"

(c) Object highlight image
of "Office"

(d) Occlusion complement
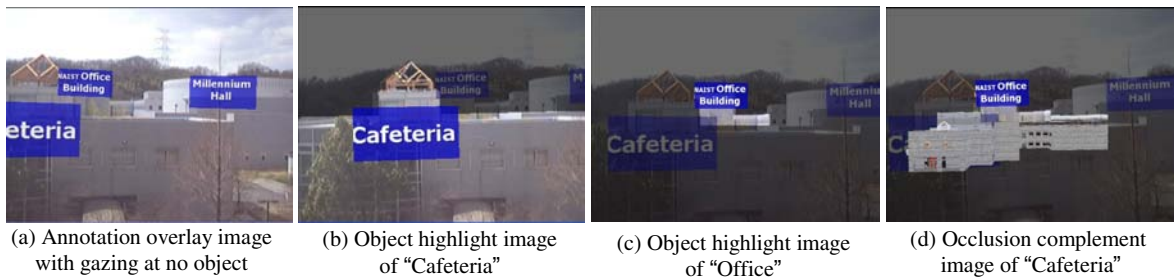image of "Cafeteria"

Figure 8: *Generated images at the viewpoint B.*

user. Occlusion complement images are shown in Figures 7(d), (f), and 8(d). Especially, Figure 8(d) shows that the proposed method realizes intuitive annotation on a mostly occluded object. Through this experiment, we have found that the proposed method can present the correspondence between real and virtual worlds to the user clearly. Through this experiment, object highlight images were generated at 20fps and occlusion complement images were generated at 15fps, approximately.

4. Evaluating Experiments

In the same experimental environments, we have also carried out another experiment which evaluates how correctly the user-viewed object regions are highlighted by the proposed method. By comparing estimated regions by the pro-

posed method with manually selected regions of the simultaneous input image, this experiment evaluated the rates of correctly highlighting the user-viewed object. In this experiment, both of grand truth images and estimated images were 320 width and 240 height sized images. Figure 9 show examples of an input image, a grand truth image, and an estimated image. We have five times comparisons in 6 cases(3 objects x 2 viewpoints), respectively. Table 1 shows average values of pixel counts of (a), (b), and product sets of (a) and (b) in these 6 cases. Table 1 also shows average accuracy rates in each case. Accuracy rates are given by the following formula.

$$[Accuracy\ rate] = \frac{[Pixel\ counts\ of\ product\ set\ of\ (a)\ and\ (b)]}{[Pixel\ counts\ of\ (a)]}$$
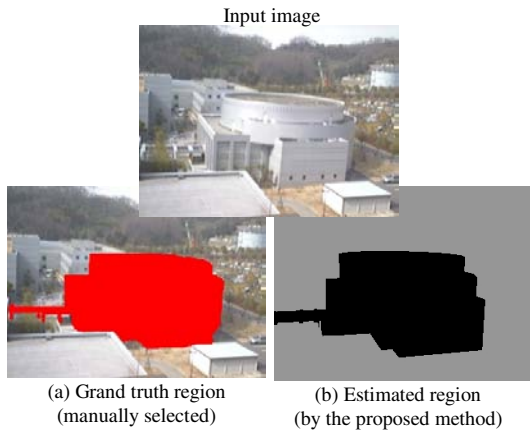
We can conclude that the proposed method is available

Input image

(a) Grand truth region
(manually selected)

(b) Estimated region
(by the proposed method)

Figure 9: *Image examples used in evaluation experiment.*

Table 1: *Evaluation results.*

Case	Pixel count of (a)	Pixel count of (b)	Pixel count of (a) $\wedge$ (b)	Accuracy rate [%]
Cafeteria (viewpoint A)	41432	41170	39565	95.5
Office (viewpoint A)	8297	8310	7108	85.6
Hall (viewpoint A)	19928	22304	18082	90.7
Cafeteria (viewpoint B)	42227	48913	40935	97.0
Office (viewpoint B)	2172	1684	1407	64.7
Hall (viewpoint B)	12751	14401	12015	94.2

(a) : the user-viewed object region manually selected
(b) : the user-viewed object region estimated by the proposed method

for user's intuitive understanding of relevance of annotations and real objects because Table 1 shows accuracy rates are over 60% in any cases. Accuracy rates increase as the user-viewed object regions become large. We found following two kinds of situations as estimation errors.

(1) Difference between shape of estimated region and that of grand truth region
(2) Misalignment of these regions on image

These estimation errors occur due to the differences between the 3D model and the real scene and errors of the user's position and orientation. Differences between the 3D model and the real scene cause (1). The user's positioning errors which were within 1 meter in this experiment lead to (1) mainly, however they can be slightly factors of (2). (2) is also caused by errors of the user's orientation. The inertial sensor which was used in this experiment has some errors on angle of yaw in consequence of the distorted terrestrial magnetism while it can reset accumulative errors [TKY03]. Because of errors on yaw direction, some gaps between real objects and CG components in generated images could be found as shown in Figures 4(g) and (h). The accuracy rates can be improved

by the improvement of measuring user's position and orientation or using more precise 3D models.

5. Summary and Future Work

This paper has proposed an annotation emphasizing technique using 3D models of real scenes as a new information presentation method for wearable annotation overlay systems. In the proposed method, in order to represent links between annotations and real objects effectively, the system generates two kinds of images: object highlight images and occlusion complement images. The former highlights the user-viewed object and corresponding annotation. The latter complements occluded regions of the user-viewed object using the textured 3D model. This paper has also described an experiment using a prototype system. In the future, we will consider to combine conventional view management methods [AF03, BFH02], information filtering methods [JLS+00], or depth representation methods [LSG+03] to the proposed method.

References

[ABB+01]   R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Juiler and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, Vol. 21, No. 6, pp. 34–47, 2001.

[AF03]   R. Azuma and C. Furmanski, "Evaluating label placement for augmented reality view management," *Proc. 2nd Int. Symp. on Mixed and Augmented Reality*, pp. 66–75, 2003.

[BFH02]   B. Bell, S. Feiner and T. Höllere, "Information at a glance," *IEEE Computer Graphics and Applications*, Vol. 22, No. 4, pp. 6–9, 2002.

[BH04]   R. Bane and T. Höllere, "Interactive tools for virtual x-ray vision in mobile augmented reality," *Proc. 3rd Int. Symp. on Mixed and Augmented Reality*, pp. 231–239, 2004.

[FMS93]   S. Feiner, B. MacIntyre and D. Seligmann, "Knowledge-based augmented reality," *Communications of the ACM*, Vol. 36, No. 7, pp. 52–62, 1993.

[GF03]   S. Güven and S. Feiner, "Authoring 3d hypermedia for wearable augmented and virtual reality," *Proc. 7th Int. Symp. on Wearable Computer*, pp. 118–126, 2003.

[HFT$^+$99]   T. Höllerer, S. Feiner, T. Terauchi, G. Rashid and D. Hallaway, "Exploring mars: Developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers and Graphics*, Vol. 23, No. 6, pp. 779–785, 1999.

[JLS$^+$00]   S. Julier, M. Lanzagorta, S. Sestito, Y. Baillot, T. Höllerer, S. Feiner and L. Rosenblum, "Information filtering for mobile augmented reality," *Proc. Int. Symp. on Augmented Reality (ISAR00)*, pp. 3–11, 2000.

[KKS01]   M. Kourogi, T. Kurata and K. Sakaue, "A panorama-based method of personal positioning and orientation and its real-time applications for wearable computers," *Proc. Int. Symp. on Wearable Computers*, pp. 107–114, 2001.

[KTO04]   Y. Kameda, T. Takemasa and Y. Ohta, "Outdoor see-through vision utilizing surveillance cameras," *Proc. 3rd Int. Symp. on Mixed and Augmented Reality*, pp. 151–160, 2004.

[KY04]   M. Kanbara and N. Yokoya, "Real-time estimation of light source environment for photorealistic augmented reality," *Proc. 17th IAPR Int. Conf. on Pattern Recognition (ICPR2004)*, No. II, pp. 911–914, 2004.

[LSG$^+$03]   M. Livingston, J. Swan II, J. Gabbard, T. Höllerer, D. Hix, S. Julier, Y. Baillot and D. Brown, "Resolving multiple occluded layers in augmented reality," *Proc. 2nd Int. Symp. on Mixed and Augmented Reality*, pp. 56–65, 2003.

[LT04]   A. Leykin and M. Tuceryan, "Automatic determination of text readability over textured backgrounds for augmented reality systems," *Proc. 3rd Int. Symp. on Mixed and Augmented Reality*, pp. 224–230, 2004.

[Man99]   S. Mann, "Wearable computing: A first step toward personal imaging," *IEEE Computer*, Vol. 30, No. 2, pp. 25–32, 1999.

[MKY04]   K. Makita, M. Kanbara and N. Yokoya, "Shared annotation database for networked wearable augmented reality system," *Proc. 5th Pacific Rim Conf. on Multimedia*, Vol. 3, pp. 499–507, 2004.

[NF02]   L. Naimark and E. Foxin, "Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker," *Proc. of 1st Int. Symp. on Mixed and Augmented Reality*, pp. 27–36, 2002.

[NKY05]   Y. Nakazato, M. Kanbara and N. Yokoya, "Wearable augmented reality system using invisible visual markers and an ir camera," *Proc. 9th Int. Sympo. on Wearable Computers*, pp. 198–199, 2005.

[SKC$^+$01]   D. Stricker, J. Karigiannis, I. T. Christou, T. Gleue and N. Ioannidis, "Augmented reality for visitors of cultural heritage sites," *Proc. Int. Conf. on Cultural and Scientific Aspects of Experimental Media Spaces (CAST 01)*, pp. 89–93, 2001.

[SWP97]   T. Starner, J. Weaver and A. Pentland, "A wearable computer based american sign language recognizer," *Proc. Int. Symp. on Wearable Computers*, pp. 199–202, 1997.

[SYI$^+$03]   M. Sakata, Y. Yasumuro, M. Imura, Y. Manabe and K. Chihara, "Altair: Automatic location tracking system using active ir-tag," *Proc. Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems(MFI2003)*, pp. 299–304, 2003.

[TKY03]   R. Tenmoku, M. Kanbara and N. Yokoya, "A wearable augmented reality system using positioning infrastructures and a pedometer," *Proc. 7th Int. Symp. on Wearable Computers (ISWC2003)*, pp. 110–117, 2003.

[UMKT05]   K. Uratani, T. Machida, K. Kiyokawa and H. Takemura, "A study of depth visualization techniques for virtual annotations in augmented reality," *Proc. IEEE Virtual Reality (IEEE VR)*, pp. 295–296, 2005.

[VSGP03]   R. D. Vaul, M. Sung, J. Gips and A. S. Pentland, "Mithril 2003: Applications and architecture," *Proc. Int. Symp. on Wearable Computers 2003*, pp. 4–11, 2003.

[WJH97]   A. Ward, A. Jones and A. Hopper, "A new location technique for the active office," *IEEE Personal Communications*, Vol. 4, No. 5, pp. 42–47, 1997.