In Proceedings of 3rd CREST/ISWC Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing, pp.46-53, October 31st 2004.

Wearable Camera Device for Supporting Object-Triggered Memory Augmentation

Takahiro Ueoka † Tatsuyuki Kawamura † Shigeyuki Baba ‡
Shinichi Yoshimura ‡ Yasuyuki Kono † Masatsugu Kidode †

{taka-ue,kawamura,kono,kidode}@is.naist.jp
† Graduate School of Information Science, Nara Institute of Science and Technology

Shigeyuki.Baba@jp.sony.com, shinichi@krc.sony.co.jp
‡ Sony-Kihara Research Center, Inc.

Abstract

This paper describes a camera device named ObjectCam2 that extracts color and active IR images of a nearby object from its background under both indoor and outdoor environments. ObjectCam2 is employed in a wearable object-triggered memory augmentation system as a wearable video camera device to capture a user's viewpoint video. ObjectCam2 is designed to cut out an object image for the purpose of identifying an object held by the user. In our experiments, the extraction accuracy of a target object image provided by ObjectCam2 is evaluated in both indoor and outdoor environments.

1 Introduction

We have developed a new camera device named *ObjectCam2*, which can extract an active infrared(IR) image and a color image of a nearby object. The device is suitable for identifying an object held by a user employing the device as a head-mounted camera. By using this device, a certain target object is easily extracted from its background image in real-time. The *ObjectCam2* is also designed to be light enough to be used as a wearable camera in everyday life.

We have conducted experiments to evaluate the performance of extracting the region around the object. Furthermore, these experiments were done under several light environments including an outdoor scene illuminated by the sunlight and an indoor scene. Experimental results in this paper show the effectiveness of *ObjectCam2*.

In our daily lives, we interact with objects in various ways; for example, we pick up objects, hold them in our hands; we manipulate objects, and then we put objects in certain places. We call events that entail these actions, "object-centered events." An object-centered event is inevitably driven by an object-triggered memory activity. Since we sometimes forget the memory activity, such as recalling where an object is placed, we have designed a new system to support the recollection of a memory related to an object-centered event. We call this system an "object-triggered memory augmentation system." Here, we employ the *ObjectCam2* in the augmentation system to extract an image of the target object held by the user to identify the object. This study is part of the research currently being conducted to accomplish an "augmented memory" system[1].

In this paper, sections 2 and 3 present design concepts of the "object-triggered memory augmentation system" and of *ObjectCam2*, respectively. The experiments, results, and discussions are presented in section 4.

2 Supporting Object-triggered Memory Augmentation

We focus on support for a user's object-triggered memory augmentation. Many objects trigger memories in the user's everyday life. The following presents examples of the memory recollections triggered by objects.

[A person (e.g. "Tom") can't recall...]

1. a nostalgic episode or a memorandum related to an object he handles
2. a person to whom he has lent an object he handles
3. a person from whom he has borrowed an object he handles
4. how he has used an object he handles last time
5. what objects he has to take with himself for a trip
6. what objects he needs to do a certain task
7. where he placed an object
8. the price of an object sold in another shop to compare it with another similar one sold in front of him

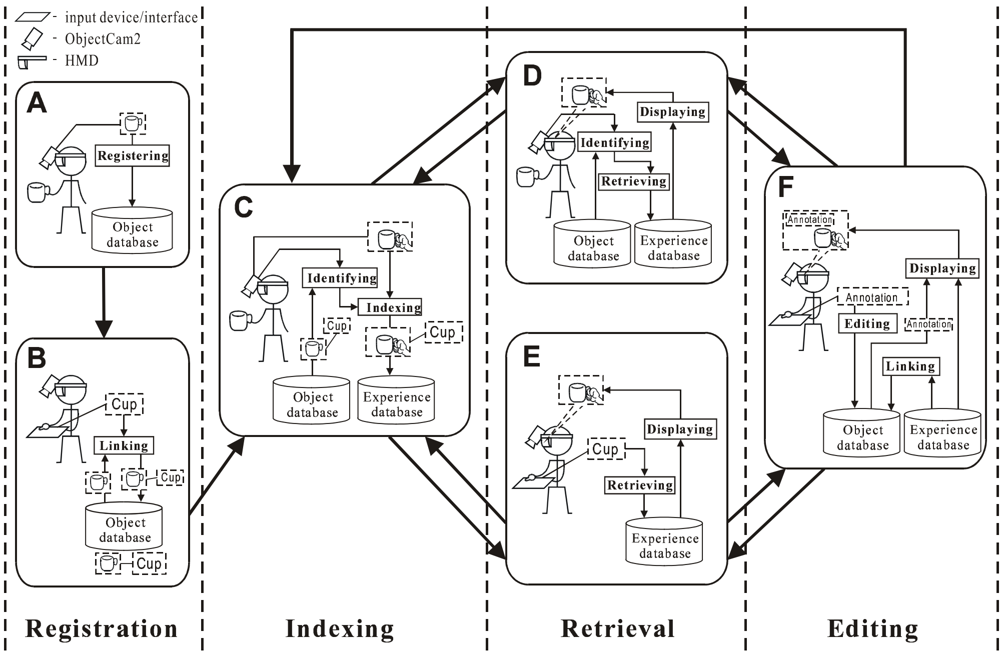

Figure 1: Phases of an object-triggered memory augmentation system
(Note that processes of annotation display in states D and E are the same as in F although they are omitted.)

We have proposed an "object-triggered memory augmentation system" to support each example. The system preliminarily records a video of Tom's (the user's) viewpoint when he handles each related object. Using the object as a retrieval key, the system retrieves the video to show him. Such an object-triggered memory augmentation system must provide both a video capturing function and an object identification function.

In this study, we propose an active IR color-camera named *ObjectCam2* as a wearable camera device to detect and identify a trigger object handled by a user. Detecting and identifying the object are the most important functions in all phases for the object-triggered memory augmentation system. *ObjectCam2* assists both functions.

Figure 1 illustrates an example of a design for the object-triggered memory augmentation system. There are four phases in the figure: registering, indexing, retrieving, and editing. The example system is composed of a PC, a Head-Mounted Display (HMD), a wearable camera that can capture a video from a user's viewpoint, and an input device/interface, e.g., a wearable keyboard device [3] or a *Wearable Virtual Tablet* interface [4]. In the example, the system records a video as experience data by using the wearable camera. The wearable camera is also employed to detect an event such as when a trigger object is handled, and to identify the object. The system can get an identification (ID) number after the object identification process.

Registration Phase

- **State A:** The user registers a trigger object as preparation for recollection of his/her experiences.
- **State B:** The user enters the data, e.g. an object name, an owner name, and attributes like a category class by using the input device/interface.

Indexing Phase

- **State C:** The system continuously records a video by using the wearable camera. When the user handles a certain registered object, the system detects the event and identifies the object by using the input device/interface. The system then indexes the ID number of the object to the experience data.

Retrieval Phase

- **State D:** The user handles a registered object to recall his/her experience. The system identifies the object. The system then retrieves an interval of a video associated with the ID number of the object. Finally, the system displays the required video to the HMD. This type of retrieval supports only in the case that he/she handles the registered object, such as the example 1, 2, 3 and 4.

- **State E:** The user enters a query, e.g. the name of a registered object, by using the input device/interface. The system then retrieves a video associated with the query. Finally, the system displays the video to the HMD. The user can use this type of retrieval even if he doesn't have the registered object at hand as the case denoted in the example 5, 6, 7 and 8.

Editing Phase

- **State F:** The user edits the attributes of the object retrieved in the retrieval phase and/or inputs annotation data by using the input device/interface. After the operation(s), the system renews the attribute and/or links the annotation data with the retrieved video.

A recent study describes an object-triggered memory augmentation system using a wearable camera and the RFID tag reader in order to detect and identify the object [5]. The system using only *ObjectCam2* for detecting and identifying the object has the following advantage:

- Simple system construction needless wearing RFID tag reader.
- Needless to attach/keep attaching RFID tags to objects.

3 *ObjectCam2* Design Concepts

This section shows the features of *ObjectCam2* customized for identifying a trigger object under everyday conditions. The *ObjectCam2* has the following features:

- small and lightweight
- dividing a nearby object image from its background by 30Hz
- employing color and active IR images
- controlling blinking IR LEDs
- controlling the exposure time to respond to the problem of dropping out the highlighted surface of the nearby object

We certify the specifications of the *ObjectCam2* and describe a vision-based method for extracting the object image robustly under the everyday illumination conditions affected by sunlight.

3.1 *ObjectCam2*

Figure 2: A wearable camera *ObjectCam2*

Table 1 indicates the specifications of the *ObjectCam2*. The table also indicates the specifications of *ObjectCam*[6], a prototype of *ObjectCam2*. Each camera is a kind of active IR camera with an IR LED array in its front. Each camera is used as a head-mounted camera mounted on the front of a helmet. Figure 2 indicates a scene using the *ObjectCam2*. The *ObjectCam2* is more miniaturized and made to be more lightweight than the *ObjectCam* so as to be employed as a wearable camera device. As opposed to *ObjectCam*'s continuous lighting of the IR LED array, *ObjectCam2* controls blinking in the IR LED array to capture a color image with reflected active IR light and to capture a color image without the lights. The CMOS unit of the *ObjectCam2* has no IR eliminating filter so that *ObjectCam2* can capture an image that has luminance within the color and the IR light range. Using the images, *ObjectCam2* can divide a nearby object image from its background by 30Hz.

3.2 Integrating functions for identifying a trigger object

The *ObjectCam2* realizes the two functions shown below to identify a trigger object. 1) The function for employing multiple features of the object including color and active IR images [7]. 2) The function of blinking IR LEDs for extracting a region of the nearby target image, including the image of the trigger object itself [8]. By integrating these two functions, *ObjectCam2* has an advantage among the other active IR cameras in extracting and recognizing trigger objects in the everyday environments of its user.

Table 1: Specification of *ObjectCam2* and *ObjectCam*

	ObjectCam2	*ObjectCam*
Weight (g)	390 (head mounted device only)	1,810 (670 (head mounted device), 1,140 (waist mounted device))
Size of camera head (mm) (width×height×depth)	95 × 70 × 90	80 × 170 × 120
Photo acceptance unit	1 CMOS (color and IR) (exposure time controllable)	2 CCD (1 color and 1 IR)
Output image	color and IR (extracted region of nearby object)	color and IR (full view only)
Shutter frequency (Hz)	30	30
number of IR LEDs	32 (blinking controllable)	98

3.3 Extracting region of nearby target image

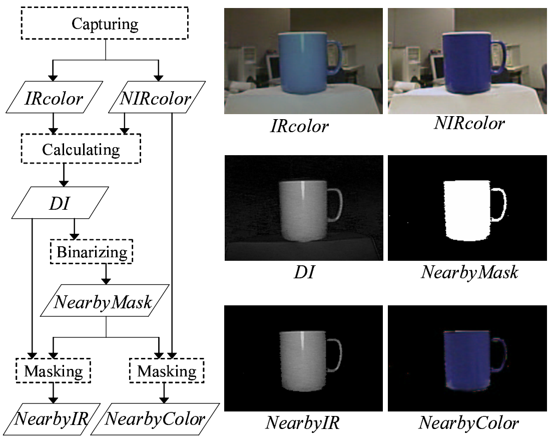

Figure 3: Process diagram for extracting a nearby target object

Figure 3 indicates a diagram of the integrated process for extracting the nearby target image. The *IRcolor* is a color image with a reflected active IR light and the *NIRcolor* is a color image without the light. The *ObjectCam2* captures the *IRcolor* and the *NIRcolor* by turns. In the equation below, *ObjectCam2* subtracts the IR luminance caused by environment IR illumination in both *IRcolor* and *NIRcolor*:

$$DI = Y_{IRcolor} - Y_{NIRcolor}. \qquad (1)$$

DI is an active IR image indicating the IR luminance only illuminated by active IR light. $Y_{IRcolor}$ and $Y_{NIRcolor}$ denote images only indicating luminance(Y) in YUV color feature of *IRcolor* and *NIRcolor*. A mask image *NearbyMask* is simply created by binarizing *DI*. The *ObjectCam2* finally creates an active IR image *NearbyIR* and a color image *NearbyColor* from *DI* and *NIRcolor* by simply masking with *NearbyMask*. Extracted region of nearby target images are illustrated in *NearbyIR* and *NearbyColor*.

3.4 Controlling Exposure Time

We have solved the problem of the highlighted area dropping out from nearby object image extracted by the *ObjectCam2*. The problem is caused by luminance saturation in *IRcolor*. When the exposure time of *IRcolor* and *NIRcolor* are the same without being controlled, the luminance of *IRcolor* is larger than that of *NIRcolor* due to the active IR illumination. The luminance of the highlighted part of an object, such as the white-colored surface, tends to be saturated in *IRcolor*. The luminance of the part in *DI* is calculated to a smaller value than the proper value.

To avoid the saturation of *IRcolor*, the *ObjectCam2* controls exposure time at the time of capturing *IRcolor*. When $\delta_{IRcolor}$ and $\delta_{NIRcolor}$ indicate the exposure time of *IRcolor* and *NIRcolor*, equation (1) should be fixed as below:

$$DI = \frac{\delta_{NIRcolor}}{\delta_{IRcolor}} Y_{IRcolor} - Y_{NIRcolor}. \qquad (2)$$

The illumination condition of $Y_{IRcolor}$ and $Y_{NIRcolor}$ has been coordinated with a common environment illumination condition to calculate proper *DI*.

Figure 4 illustrates the result of controlling the exposure time. Without controlling the exposure time, only the luminance of the white-colored right-hand surface of the cube is calculated as smaller in *DI*, and the surface has been lost in *NearbyColor*. The *ObjectCam2* controls $\delta_{IRcolor}$ to a half of $\delta_{NIRcolor}$, and avoids dropping the part from the nearby object image with proper *DI* created from unsaturated *IRcolor* and *NIRcolor*.

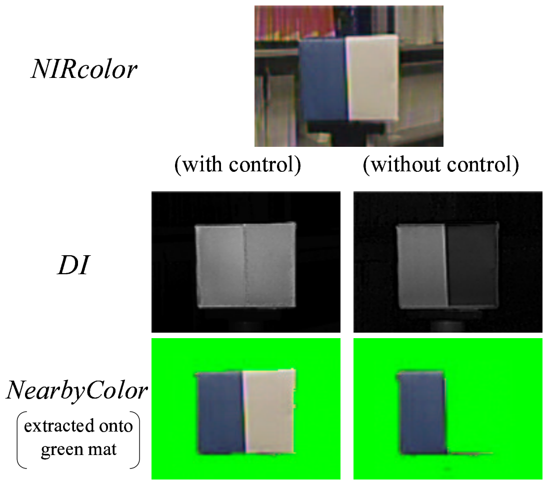

Figure 4: Result of controlling exposure time

4 Experiments and Results

We have evaluated the extraction accuracy of a nearby object image by *ObjectCam2* in experiments showing that *ObjectCam2* is suited for use in an everyday environment. The extraction accuracy of a nearby object image is indicated in the $NearbyMask$. By comparing the quality of the $NearbyMask$ created by *ObjectCam2* with that of the $NearbyMask$ created by *ObjectCam*, we have demonstrated the functional effects of the *ObjectCam*.

4.1 Experiments

Experiments were carried out at two locations, indoors at night without sunlight illumination and outdoors at 2:30 – 3:30 p.m. with indirect sunlight illumination. We used 41 objects (illustrated in figure 5), consisting of 21 standard objects (0 – 20) and 20 utility objects(21 – 40). Standard objects consist of spheres(0 – 6), corns(7 – 13) and cubes(14 – 20). Objects in each shape group are given the colors cyan, magenta, yellow, white, and three gradient steps (30%, 60%, and 100%) of black. Utility objects were arbitrary samples with follwing policies. 1) A set of utility objects was made to include wide variety of features (surface material, color pattern, and shape) so as to avoid a specific result. 2) Maximal diagonal length was constrained from 5 cm to 25 cm so as to be captured fully in the frame of cameras and not to be captured in too small a size.

Each object was set on an object table at a distance of 35cm from the installation location of each camera. *ObjectCam2* and *ObjectCam* captured the object to create a $NearbyMask$ of the object image. The $NearbyMask$ of *ObjectCam2*, covering only a region of the nearby object image, was used in the masking phase in figure 3. The $NearbyMask$ of *ObjectCam* was created by simply binarizing the IR image *ObjectCam* itself captures [6]. Each threshold for binarizing was adjusted to an appropriate value in each environment based on heuristics.

To evaluate the quality of the $NearbyMask$ in comparing *ObjectCam2* and *ObjectCam*, we have calculated a noise ratio and the missing ratio of each $NearbyMask$. The area of noise S_n and that of the missing S_m indicate the quality of the $NearbyMask$. These values are calculated using a $CorrectMask$, which is a template for evaluating the $NearbyMask$. We manually create the $CorrectMask$ for each $NearbyMask$ from the source image of the object. S_n is defined as a partial area value of the $NearbyMask$ covered by the background area of the $CorrectMask$. S_m is defined as the partial area value of the $CorrectMask$ excluding the area covered by the $NearbyMask$. When S_c is a common area value between $CorrectMask$ and $NearbyMask$, the noise ratio P_n and the missing ratio P_m of the $NearbyMask$ are denoted as below:

$$P_n = \frac{S_n}{S_c}. \quad (3)$$

$$P_m = \frac{S_m}{S_c + S_m}. \quad (4)$$

4.2 Results

Table 2 denotes P_n and P_m of the $NearbyMask$ created by each camera. In the outdoor experiment, we found that *ObjectCam2* has an advantage in the case of 39 objects, excluding object (29) and (38). In these cases, all P_n with *ObjectCam2* are smaller than those with *ObjectCam*. This fact indicates that *ObjectCam2* has made a better quality of the $NearbyMask$ for extracting these objects than that made by *ObjectCam*. In the case of object(29) and (38), each P_m with *ObjectCam2* is a large value over 99%.

In the indoor experiment, both *ObjectCam2* and *ObjectCam* have scored a lower P_n than 3%, excluding the case of the black 100% colored objects. "No value" included in Table 2 represents the result of division by zero in equation (3). The value indicates there is no extracted area in the correct region of the nearby object.

Figure 6 illustrates the extracted images of object(36). Throughout these experiments, we found that *ObjectCam2* has eliminated the background image more robustly than *ObjectCam*.

4.3 Discussion

The noise ratio of the $NearbyMask$ must be decreased to achieve a robust identification of a trigger object using *ObjectCam2*. From equation (3), the

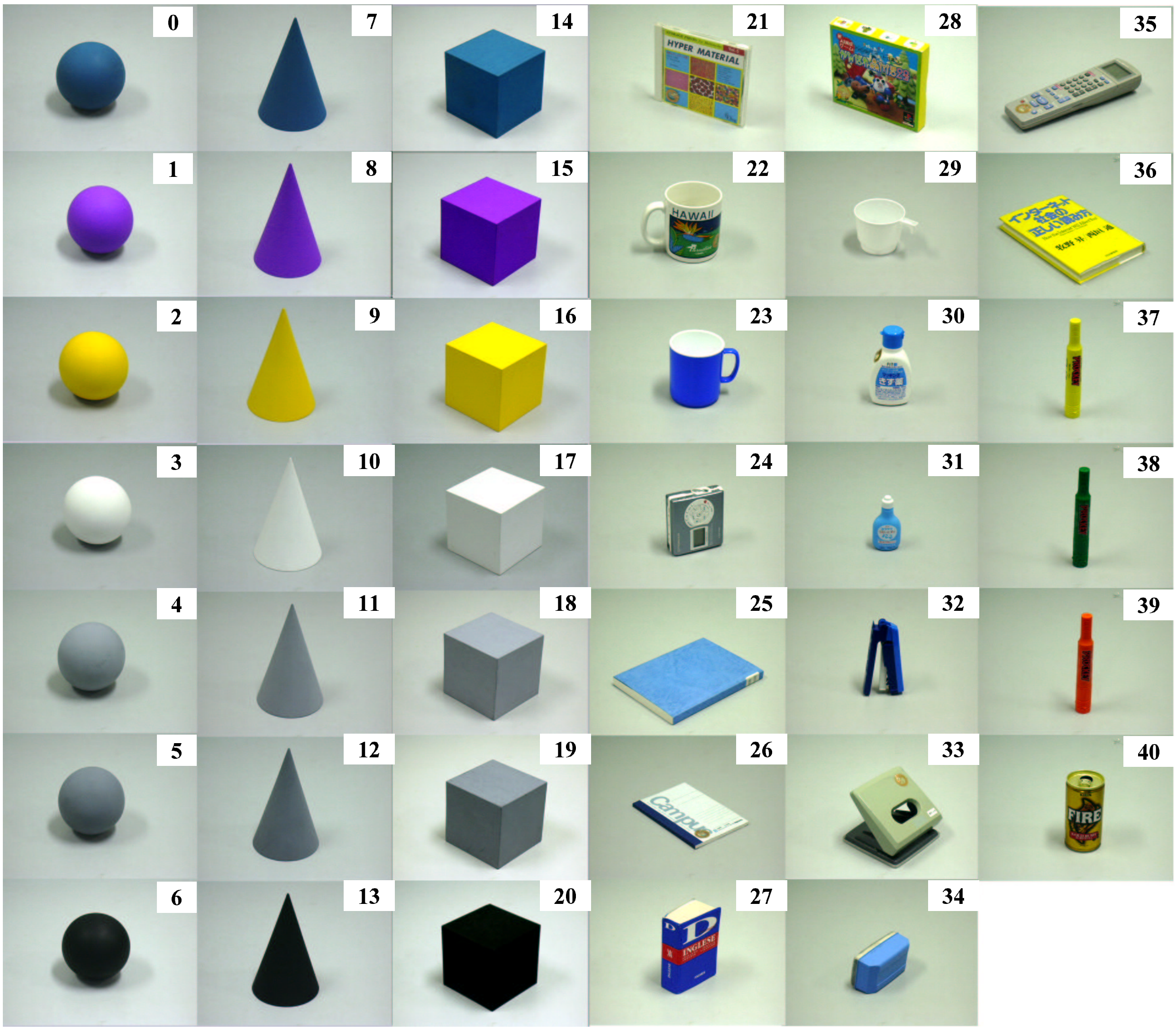

Figure 5: Target objects in experiments

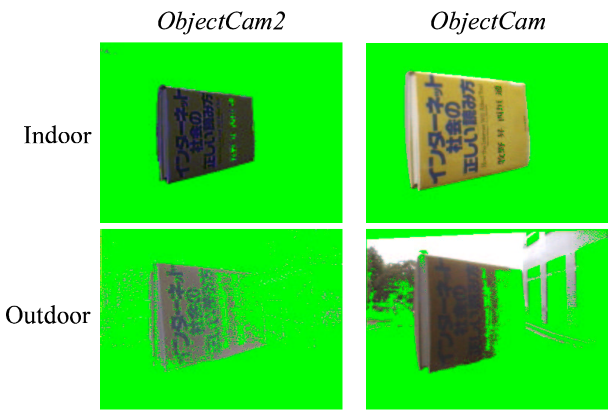

Figure 6: Experimental result
($NearbyColor$ of object(36) extracted onto green mat)

Table 2: Quality of $NearbyMask$

Noise Ratio (%)				
	Outdoor		Indoor	
Object ID	*ObjectCam2*	*ObjectCam*	*ObjectCam2*	*ObjectCam*
0	**84.67**	375.20	0.21	**0.01**
7	**79.39**	284.20	1.44	**1.21**
14	**202.28**	231.75	0.24	**0.19**
1	**33.76**	319.02	0.77	**0.00**
8	**21.00**	277.63	**0.55**	2.28
15	**18.47**	141.81	**0.13**	0.42
2	**29.23**	309.73	0.28	**0.02**
9	**20.68**	275.17	**2.09**	3.58
16	**9.87**	134.48	0.67	**0.32**
3	**36.43**	299.84	0.52	**0.24**
10	**27.47**	275.81	**2.80**	4.90
17	**9.26**	122.64	0.69	**0.68**
4	**221.33**	1204.92	0.09	**0.00**
11	**684.40**	5572.96	0.72	**0.60**
18	**2695.00**	12539.58	0.17	**0.08**
5	**811.49**	3110.13	0.78	**0.00**
12	**4941.67**	22915.15	0.44	**0.10**
19	**5063.64**	10965.84	0.11	**0.01**
6	**3271.05**	8598.56	1300.00	**0.00**
13	**15500.00**	17282.03	(no value)	**0.00**
20	**10670.00**	55330.30	(no value)	**0.00**
21	**21.75**	125.47	0.50	**0.16**
22	**37.98**	300.05	0.61	**0.59**
23	**59.94**	323.05	**0.57**	0.92
24	**90.32**	1021.43	0.81	**0.00**
25	**81.09**	1051.60	**0.27**	1.16
26	**13.96**	295.16	1.07	**0.64**
27	**106.64**	659.68	**0.63**	1.32
28	**32.92**	327.48	0.79	**0.58**
29	7250.00	**461.75**	2.43	**0.78**
30	**44.04**	341.32	**1.35**	1.70
31	**97.43**	855.81	**2.94**	3.14
32	**155.50**	848.28	**3.18**	5.63
33	**168.88**	620.49	**1.32**	2.53
34	**340.44**	1235.17	**0.29**	1.26
35	**2145.45**	15306.00	0.26	**0.00**
36	**8.71**	106.35	1.61	**0.97**
37	**151.88**	1530.72	0.74	**0.25**
38	6436.84	**5636.38**	5.17	**0.37**
39	**151.95**	1412.64	**0.21**	0.37
40	**199.44**	2071.15	2.31	**0.00**

Missing Ratio (%)				
	Outdoor		Indoor	
Object ID	*ObjectCam2*	*ObjectCam*	*ObjectCam2*	*ObjectCam*
0	72.04	18.59	4.49	4.03
7	73.95	3.35	1.42	1.32
14	96.19	44.64	1.63	0.92
1	33.95	8.08	5.11	4.52
8	16.38	0.49	3.03	2.09
15	59.30	16.30	1.15	0.92
2	24.69	5.89	2.99	2.27
9	8.69	0.49	0.73	0.66
16	19.62	8.20	1.56	1.33
3	30.44	1.39	5.51	2.20
10	29.91	1.09	0.02	0.46
17	5.92	1.84	0.75	0.40
4	91.46	75.84	16.10	7.74
11	96.90	95.19	9.59	5.08
18	99.68	99.09	2.84	2.61
5	97.40	90.93	34.51	15.65
12	99.66	98.89	20.02	8.33
19	99.82	99.02	45.08	12.59
6	99.43	96.73	99.96	98.29
13	99.89	98.56	100.00	99.74
20	99.92	99.79	100.00	99.87
21	55.95	18.60	9.69	6.20
22	60.69	36.67	11.25	3.88
23	69.85	28.49	11.84	3.81
24	75.51	73.99	61.10	40.41
25	92.24	93.82	7.31	1.16
26	34.75	54.23	11.13	3.96
27	85.96	75.12	1.88	0.74
28	76.14	73.54	4.86	1.50
29	99.53	2.50	7.50	4.17
30	28.97	3.66	2.57	1.06
31	24.03	4.10	3.40	2.09
32	75.83	44.29	9.75	3.73
33	91.49	61.28	68.45	46.74
34	91.20	64.53	17.61	5.74
35	99.51	98.79	46.35	14.55
36	38.86	26.25	2.21	1.67
37	65.39	37.92	15.67	11.20
38	99.19	80.67	70.70	13.73
39	67.32	29.58	16.65	9.24
40	89.39	83.89	70.24	54.42

noise ratio should be decreased by increasing S_c. Increasing S_c means decreasing the missing area at the same time.

We found a problem such that the dark colored surfaces of object images were dropped out in the outdoor environment. The problem has been indicated in experimental results of gradient black colored standard objects and cyan colored standard objects. Results in utility object (38) are also caused by similar problem. One of the ways to respond to the problem is to enhance the luminance of active IR LEDs to capture a reflected active IR image even if the surface reflects only a little IR light. To avoid extracting a background region illuminated by strong active IR light, the luminance of each pixel in the object image has to be corrected dynamically into a proper value by considering the color-specific reflectivity of IR light based on the color information of $NIRcolor$.

We found another problem with object (29) that only retains a half-transparent white surface. The object image of object (29) dropped out almost completely in the outdoor experiment of *ObjectCam2*. Strong sunlight was transmitted from the background of object (29), which caused a partial saturation condition of $NIRcolor$ in the region of the object. If $NIRcolor$ has been saturated partially, the luminance of the saturated area in DI will tend to be turned down. This "partial saturation condition" problem is common to the partial reflection of environmental light on a part of the object surface, and is also common to the sky region captured in an object image. To solve the partial saturation problem completely, not only $\delta_{IRcolor}$, but also $\delta_{NIRcolor}$ in equation (2), has to be constrained so that $NIRcolor$ does not have a saturated area.

A dilemma arose, however, in solving the darkcolored object extraction problem and the partial saturation problem simultaneously. The dark colored object extraction problem needs bright $NIRcolor$ to utilize the color information of $NIRcolor$. On the other hand, the constraint of $\delta_{NIRcolor}$ shows that $NIRcolor$ contains less color features than under the normal condition. To improve the robustness of *ObjectCam2* in extracting a nearby object image under sunlight illumination, we must solve the dilemma.

5 Concluding Remarks

We have developed an *ObjectCam2* camera device, which is worn by a user, to extract a nearby object image from its background image. An object-triggered memory augmentation system can simplify its system construction by using *ObjectCam2* for both video capturing and object identifying. Integrating the color/IR image-capturing function and the blinking function of IR LEDs, the *ObjectCam2* has the advantage of identifying an object held by a user. The saturation problem in extracting an object with a highlighted surface has been cleared by the exposure time controlling function of *ObjectCam2*.

We are planning to improve *ObjectCam2* so it can be employed by an actual wearable memory recollection system. A difference image using continuously captured images, such as the DI of the *ObjectCam2*, in principle contains time difference edges. We hope to correct the edges. At the same time, we are continuously investigating the functional effect of *ObjectCam2* in several environmental conditions, including under sunlight illumination.

Acknowledgement

This research is supported by Core Research for Evolutional Science and Technology (CREST) Program "Advanced Media Technology for Everyday Living" of Japan Science and Technology Agency (JST)

References

[1] Rhodes, R.J., "The Wearable Remembrance Agent: a System for Augmented Memory," *Proc. ISWC1997*, pp.123–128, 1997.

[2] Lamming, M. and Flynn, M., "Forget-me-not: Intimate Computing in Support of Human Memory," *Proc. FRIENDS21: International Symposium on Next Generation Human Interface*, pp.125–128, 1994.

[3] Fukumoto, M. and Tonomura, Y., "Body coupled FingerRing: wireless wearable keyboard," *Proc. CHI'97*, pp.147–154, 1997.

[4] Ukita, N. and Kidode, M., "Wearable Virutual Tablet: Fingertip Drawing on a Portable Plane-Object Using an Active-Infrared Camera," *Proc. International Conference on Intelligent User Interface 2004(IUI2004)*, pp. 169–176, 2004.

[5] Kawamura, T., Hukuhara, T., Takeda, H., Kono, Y. and Kidode, M., "Ubiquitous Memories : Wearable Interface for Computational Augmentation of Human Memory based on Real World Objects," *Proc. ICCS 2003*, pp. 273–278, 2003.

[6] Ueoka, T., Kawamura, T., Kono, Y. and Kidode, M., "I'm Here! : a Wearable Object Remembrance Support System," *Proc. MobileHCI 2003*, pp. 422–427, 2003.

[7] Mihara, I., Harajima, T. and Numazaki, S., "Pop.Eye : A Popout Video Camera System for Personal Use" (In Japaese), *Proc. WISS 2002*, pp. 73–79, 2002.

[8] Lee, C.M., Schroder, K.E. and Seibel, E.J., "Efficient image recognition of walking hazards using IR illumination in wearable low vision aids," *Proc. ISWC 2002*, pp. 127–128, 2002.