EURASIP Journal on Applied Signal Processing 2004:11, 1663–1671

Digital Reminder: Real-World-Oriented Database System

Atsuo Yoshitaka
Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
Email: yoshi@isl.hiroshima-u.ac.jp

Yasuhiro Hori
Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
Email: yassan@isl.hiroshima-u.ac.jp

Hirokazu Seki
Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
Email: zhiro@isl.hiroshima-u.ac.jp

Received 30 June 2002; Revised 29 November 2003

Digital Reminder is a system which manages a person's view images based on gaze detection. A real-world-oriented database that consists of one's view is constructed by one's implicit behavior of watching objects. Digital Reminder provides a user with WYSIWYR- (what-you-see-is-what-you-retrieve-)based interaction for view image retrieval. WYSIWYR is a new framework in information retrieval. In retrieving image data in a database, we generally enter a query by specifying one or more image attributes such as color or shape with a keyboard and/or a pointing device. Unlike former image retrieval, WYSIWYR framework identifies the user's current view of watching as a query condition, and the user's behavior of watching an object is regarded as a trigger for the retrieval. Since the context of watching, that is, the type of visual information and gaze duration, is also recognized as a view frame attributes (VFA), a more appropriate result is presented to the user. The result of performance evaluation showed that the view images similar to one's current view of gazing were ranked higher in the result of retrieval, compared with the retrieval without VFAs.

Keywords and phrases: WYSIWYR framework, gaze detection, image database, context aware.

1. INTRODUCTION

Personal digital assistants (PDAs) are widely used in our daily life. One of the purposes of using them is to enforce and enhance our memory. They basically manage textual information such as meeting schedules. It may also be possible to store nontextual information, for example, digital photo image, illustration, or sound; however, accessing such data is often performed via textual description associated with it. Compared with textual information management, audio/visual information management in mobile environment is still an open issue. We think it lies in two areas; one is the method of data storage and the other is that of accessing audio/visual data.

Digital still/video cameras are widely used in our daily life not only for long-term archiving of special events but also for the enhancement or enforcement of one's memory. However, utilizing such devices for compensating one's memory is not an ideal solution, because this method provides neither autonomous information acquisition nor context-/content-aware data access beyond browsing. For example, suppose one is watching an object with interest but they do not take any picture of it because they thought it was not worth it at that time. If the necessity of a picture of the object is realized after the object was away, it is quite difficult to recover its information. One solution to avoid this problem is to record one's view continuously without manual start-stop operations video camera in order to cover all one's view. However, since the video stream is associated with no content-/context-aware information, all one can do is time-consuming browsing from the beginning to the end of the stream for accessing a scene one wishes to recall.

The results of content-based retrieval for image/video database [1, 2, 3, 4] are applicable for view retrieval in mobile environments as well; however, there are essential differences between it and ordinary video database retrieval. One is that the computation time for retrieval needs to be short enough for immediate access under mobile environments. This is because the visual data that relates to the user's current view may be the target of instant retrieval. Another reason, which is more important to be taken into account, is the fact that a query condition is explicitly specified with color, shape, and/or trajectory of object in ordinary image/video database retrieval. This type of querying may also be applicable for the retrieval of the user's view; however, it is independent from the current state of the user.

In this paper, we describe a system named *Digital Reminder*, which captures a user's view every time they watch objects. Since capturing one's view is performed based on gaze detection, a real-world-oriented database (RWODB) is created without one's explicit operations, and thus the problems mentioned above are evaded. In addition, we propose a WYSIWYR (what you see is what you retrieve) framework for view image retrieval. Instead of retrieving video data by submitting a query specifying conditions with input devices such as a keyboard and a mouse, retrieval is triggered simply by watching an object that is similar to what one wants to retrieve. That is, the user's behavior of watching the object is regarded as "Find past views which contain an object that is similar to what I'm currently watching." Context, that is, how the user concentrates on watching the object, as well as the features of the object, such as color, is captured from a head-mounted camera. Therefore, the user does not have to enter query conditions explicitly. Since the WYSIWYR is based on gaze detection, it fits to mobile computing environments, and the accessibility and the quality of retrieval are improved.

There are studies on storing a person's views into a mobile computer for the recall of visual information. WareCAM [5] and DyPRES [6] are examples of context-aware systems which provide a person with information associated with a situation. The information relates to the situation which the person faces; however, they do not capture the information the person once concentrated on.

Forget-me-not [7] is a system which acquires the context of a person's activity with a mobile computer. The context information is stored into a server computer via a network and is provided as the supplement of the person's memory. Since this is a large system working on the network, it is not a portable system.

Utilizing human biosignals for information acquisition is studied in [8]. StartleCam [8] captures a person's views by detecting the state of surprise from the transition of the person's electrical skin resistance. A method of capturing one's view by the trigger of the head movement such as rotation is discussed in [9]. In these two methods, a lot of unnecessary and redundant information is filtered. However, capturing visual information only when the person is surprised does not always capture noticed information, and view capturing based on head movement detection contains inevitable "noise." Retrieval of one's view triggered by one's current view is also studied in [10], which aims at the same objective as our study. However, the view object is associated with the movement of head, and the eye movement of the user is not captured. Therefore, neither the duration of watching objects nor the area of interest (AoI) of the person's view is taken into account. There is a study to capture a person's view with that person's reaction, which is based on brain wave analysis, in [11]. This method enables to capture the user's interest for visual information; however, it is hard to recognize the type of visual information. As far as we know, there is no system except Digital Reminder that captures one's view selectively by gaze detection. We think this is one of the promising approaches for detecting one's attention for visual objects.

The organization of this paper is as follows. Overview of Digital Reminder is described in Section 2. Section 3 introduces the fundamental idea of detecting a person's behavior by eye movement. Section 4 discusses WYSIWYR framework. The experimental results related to gaze detection and retrieval performance are described in Section 5. Concluding remarks are described in Section 6.

2. DIGITAL REMINDER

2.1. System organization

The system organization of Digital Reminder is illustrated in Figure 1. A head-mounted CCD camera is placed below an eye, directed toward the eye. Video is captured by 160×120 pixels, full color, 10 fps. A mirror is located above the eye so that the small CCD camera shoots both eye movement and the view of the user. Each of the video frames is separated into a view region and an eye image region. The eye image region is referred to for iris extraction, and a user's gaze is detected as described in Section 3.

As a default, Digital Reminder is operated in "store mode." When a gaze is detected, the user's view is turned over horizontally and stored into an RWODB as the visual information which the user noticed. Then, view frame attributes (VFAs), a gaze duration, and a time stamp are associated with the view frame. The view frame may also be associated with PDA data by the user's explicit operations. The prototype system is shown in Figure 2. Note that the mirror is a bit large since the physical configuration between camera (lens) and mirror is not optimized yet, and we will improve it in the future version of Digital Reminder.

When one wants to recall view frames by watching an object around oneself as a clue, one switches Digital Reminder from "store mode" to "retrieval mode." When one watches an object, view frames containing objects that are similar to the object one is watching are presented on an LCD display. If the user changes the viewpoint from one object to another, another retrieval operation is invoked so as to present the user with view frames containing objects that are similar to the current object in concentration.

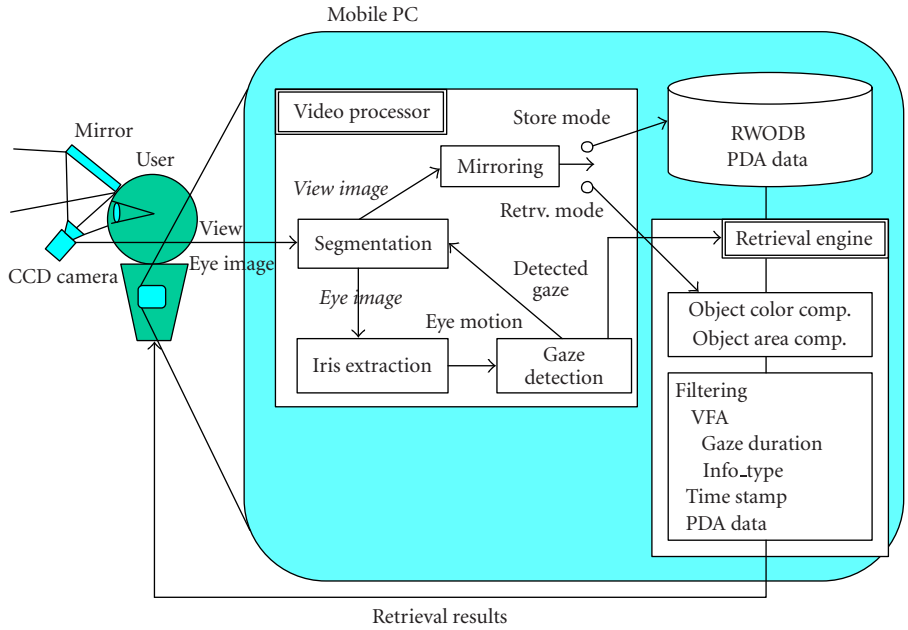


Figure 1: System organization.

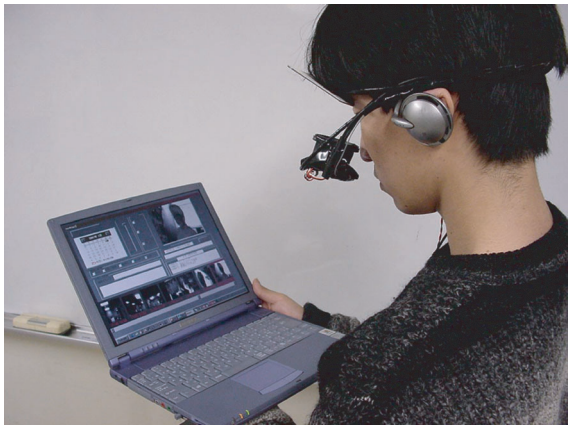


Figure 2: An overview of digital reminder.

2.2. Real-world-oriented database

A RWODB is created in Digital Reminder based on an object-oriented data model. The object-oriented data model is suitable for managing multimedia information including image, video, as well as alphanumeric data with the facility of polymorphism.

The RWODB is a database that stores occurrences in the real world into the database. Every occurrence is represented by various types of data such as video or alphanumeric data. The purpose of RWODB is to store the history of the real-world events and enable oneself to access it in order to enhance one's memory.

It is also possible to create an RWODB simply by capturing all of the user's view as a long stream of continuous video data. However, inevitable redundancy of the data is not negligible and less context is referred to in information retrieval. Therefore, it is preferable to introduce a mechanism to acquire the state of gazing implicitly for eliminating the redundancy of information and gathering all the gazed information, part of which may be missed under explicit, manual operations by the user.

The database schema in Digital Reminder is illustrated in Figure 3. Ovals represent classes and gray circles correspond to objects. Instance variables are represented by links between objects, which are identified by instance variable names attached to the links. As shown in the figure, a time stamp and personal schedule information are associated with a gazing frame. An VFA is composed of the type of visual information, *info_type*, such as an article, a figure/picture/landscape, and a *gaze_duration* which corresponds to the degree of importance/complexity of the visual information.

2.3. User interface

The user interface of Digital Reminder in retrieval mode is shown in Figure 4. Figure 4a shows an interface for retrieving view frames by specifying conditions for VFA so as to find nondocumentary objects. The VFA consists of *information type* and *gaze duration*, both of which are acquired based on gaze detection. The upper-right video frame shows a user's current view. The video frames located at the bottom of the window correspond to the result of retrieval, which is obtained by specifying conditions on VFA. They are arranged in temporal order. PDA data management for the user's schedule or text-based retrieval is operated at the spaces in the middle (the right one under the current view frame is for entering PDA data, and the left one is for full text retrieval). Note that only the objects whose colors are close to that of the object in current gaze are presented as the result of retrieval.

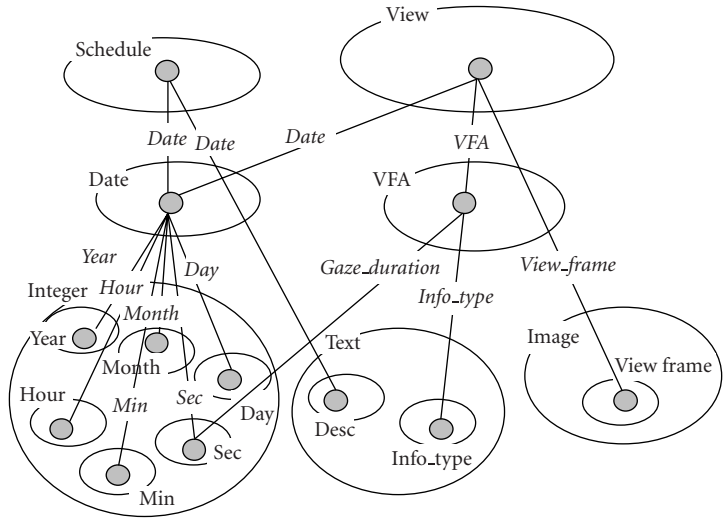
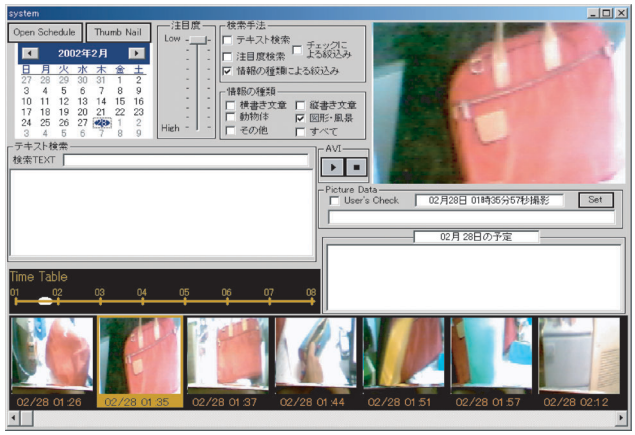
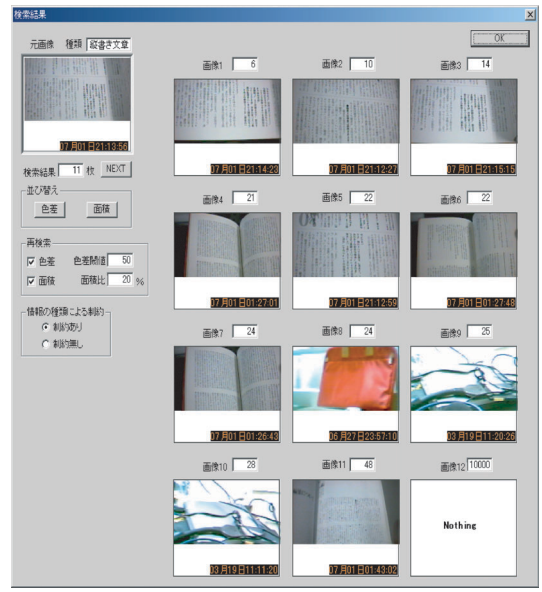


Figure 3: RWODB schema in Digital Reminder.



(a) Retrieving nondocumentary object by gaze pattern.



(b) WYSIWYR-based retrieval for vertical writing.

Figure 4: User interface for view image retrieval.

Figure 4b shows an example of WYSIWYR-based retrieval. In WYSIWYR-based retrieval, retrieval is invoked only by watching an object that is similar to what the user wants to retrieve. In this example, reading vertical writings on a paper is interpreted as "Find view frames that contain vertical writing," and retrieval is invoked without manipulating a keyboard or a mouse. As a consequence, view frames containing vertical writing are presented in the window. An upper and leftmost view frame corresponds to one's current view, and the rest of the view frames are the result of retrieval, where the upper-left one is evaluated as the highest match. If a user points one of the retrieved view frames, this user is able to access other information related to the view frames.

3. VIEW ACQUISITION IN DIGITAL REMINDER

3.1. Characteristic of eye movement

By psychological experiments, human eyes are known to have typical movement when one watches an object. When one watches a still object, the alternation of 300 milliseconds. of fixation and saccade completed within 30 milliseconds. is observed [12]. Such an alternation of fixation and saccade is not observed when a person is not gazing at any object. In case of watching a moving object, smooth pursuit occurs without saccades. Based on the characteristic of the eye movement stated above, it is possible to know the state of a person in the sense of whether they gaze at an object or not.

Moreover, the directions of saccades take a typical pattern depending on the type of information the person is watching. In the case where a person is reading horizontal writing, periodic, horizontally-long saccades directed from right to left are observed. Similarly, periodic, vertically-long saccades are observed in the case of reading vertical writing such as Japanese literature. When one is watching graphics, landscapes, or visual objects except writing, saccades of random directions are observed.

As stated above, the state of a person in the sense of whether they gaze at an object can be detected by eye movement. Therefore, it is possible to construct a visual database of the person's view, which consists of only the views that person watched. Moreover, gaze-based view detection is advantageous in classifying the type of visual contents, since the way the person is watching the information can be classified by the pattern of eye movement and the duration of a series of saccades.

3.2. Gaze detection

Eye movement is captured by a small video camera mounted below an eye. Video is captured by 160×120 pixels, 10 fps. First, each video frame is binarized to extract an iris. Since the binarized frame contains edges around eyelids, they are eliminated by excluding vertically narrow areas. The coordinates of the center of a pupil are then obtained as the center of the extracted iris. Note that this simple process is enough for extracting the occurrence and the direction of saccade.

The x and y coordinates of a pupil at time t (frame, $t > 0$) are denoted as $x(t)$ and $y(t)$, respectively. Let $d(t)$ denote the distance of saccades between t and $t-1$:

$$d(t) = \sqrt{(x(t) - x(t-1))^2 + (y(t) - y(t-1))^2}, \quad t = 2, 3, \ldots. \tag{1}$$

The state of gazing at objects is detected when the following condition is satisfied, for $t > 4$:

$$d(t-3) + d(t-2) + d(t-1) < h_f, \qquad d(t) > h_s. \tag{2}$$

In the above description, h_f and h_s are thresholds for determination of fixation and saccade, respectively. Under the video capturing condition stated above, h_f and h_s are set to 2 (pixels). These thresholds are determined by experiment, which results in the highest precision under the above-mentioned hardware configuration. When the above condition is satisfied, the eyes alternate fixation and saccade, which means the person gazes at objects. Therefore, it is possible to store only the visual information that the person gazed by storing the person's view if condition (2) is satisfied.

3.3. State classification by the pattern of eye movement

The pattern of eye movement is evaluated to classify the type of information the person gazes. The classification is based on frequencies, C_x and C_y, that satisfy the condition $x(t) - x(t-1) > n_x$, $y(t) - y(t-1) > n_y$ for a period, respectively. Parameters n_x and n_y are calibration factors which depend on a CCD camera.

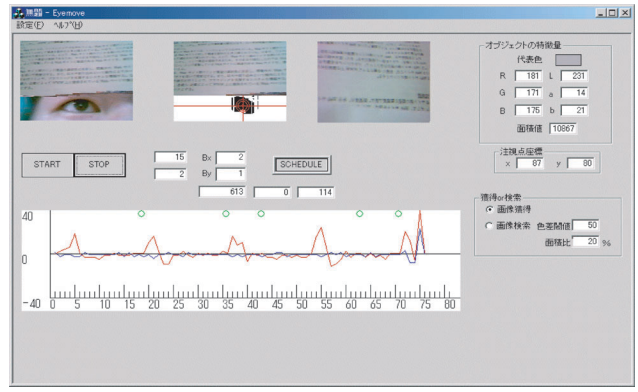


Figure 5: Eye movement in reading horizontal writing.

3.3.1. Gazing at formatted sentences

In case of reading horizontal writing such as articles in a newspaper or a book in English, relatively longer saccades are periodically observed, which move from right to left. Meanwhile, long-distance vertical saccades are hardly observed. In case of reading vertical writing such as a book of Japanese literature, saccades of long distance directed from bottom to top of a page are periodically observed. Therefore, we regard a person to be in the state of reading a document/writing when either of the following conditions is satisfied:

$$C_x = 0, \qquad C_y > b_y, \tag{3}$$

$$C_x > b_x, \qquad C_y = 0. \tag{4}$$

In the above expression, b_x and b_y are thresholds. Both of them take 1, which is determined by experiment.

Figure 5 is an example of detecting the state of reading horizontal writing. The upper-right video frame shows the current object (i.e., book) which a user is reading, and the upper-middle video frame shows the user's view and iris. The lower-middle graphs show the user's eye movement. One that takes conspicuous periodic local maxima corresponds to horizontal movement with respect to x-axis, and the other one corresponds to vertical movement along y-axis. In the figure, we can observe periodic long saccades along x-axis, though there are no conspicuous long saccades along y-axis. That is, this movement satisfies condition (4), and it is detected that he is reading a horizontal writing.

3.3.2. Gazing at figures or landscapes

When a person gazes at figures, pictures, or landscapes, saccades that follow the edges of them are observed. Since this is the movement of random direction, there are few periodic vertical or horizontal saccades. We classify the state of gaze into "gazing at figures/landscapes" if the following condition is satisfied:

$$C_x > b_x, \qquad C_y > b_y. \tag{5}$$

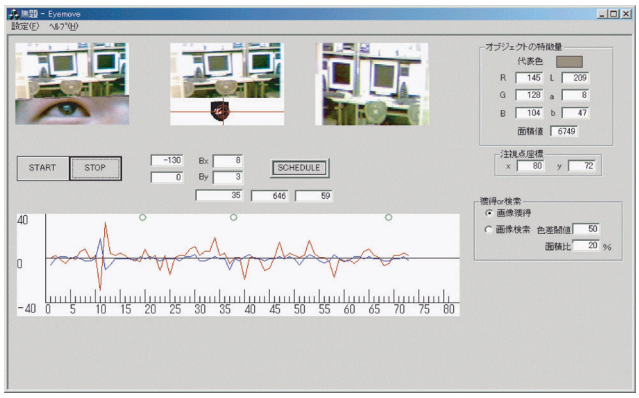


Figure 6: Eye movement in watching objects except writing.

Another example of gaze detection is shown in Figure 6. In this case, the object of interest is not writing but a computer display. Therefore, saccades of random direction are observed instead of periodic horizontal or vertical saccades. Note that circles on top of the graphs denote view frames stored into the database.

3.3.3. Gazing at a moving object

Smooth pursuit does not arise in gazing at static objects; however, it arises in the case of gazing at a moving object. Therefore, we extract smooth pursuit in order to detect the gaze for a moving object. That is, the state of a person is detected as gazing at moving objects if the following condition is satisfied for more than 7 frames (i.e., 0.7 seconds). The threshold, b_c, is set to 2, which is determined by experiment:

$$0 < d(t) < b_c. \tag{6}$$

3.4. View frame attributes

In this section, we describe view frame acquisition and its attributes. When a video frame satisfies condition (2), it is temporarily held in a video buffer. Hereafter, we call a video frame which satisfies condition (2) a *gazing frame*. If the following gazing frame is not detected within the interval v_{idle}, it is regarded as a *noise* or a state of *little attention* to an object. Therefore, the gazing frame is discarded from the video buffer.

When one or more gazing frames are detected within v_{idle}, the first gazing frame is regarded as the view frame which is to be stored into a database. The view frame is associated with a *view frame attribute*, which is composed of *gaze duration* and *visual information type* (denoted as *info_type* in Figure 3). The number of gazing frames is counted as far as gazing frames are detected within v_{idle}, and it is regarded as the gaze duration that corresponds to the *importance/complexity* of the visual information for the user. This is based on an observation that the longer the user watches an object, the more importance or complexity it implies.

In case of watching still objects, the visual information type is also evaluated based on the criteria in Section 3.3. That is, its value takes one of "vertical writing," "horizontal writing," and "graphics/landscapes." For example, suppose the user watches the leaflet of a product which consists of horizontal writing and photos. In this case, a gazing frame is extracted since condition (5) is satisfied, and the visual information type of the view frame is recognized as "graphics/landscapes."

The evaluation of the importance of gazing frames of moving objects follows another procedure: video frames are stored in the video buffer when the coordinates of iris vary in three consecutive frames. The video frames are continuously stored as far as the coordinates vary. When the number of the stored frames exceeds 6, they are stored into database as a sequence of gazed visual information.

4. WYSIWYR FRAMEWORK

4.1. Overview

WYSIWYR framework is one of the interaction methods in information retrieval, especially for image retrieval. Ordinary retrieval is invoked by specifying a query by input devices such as a keyboard and a mouse. Retrieval based on WYSIWYR framework is triggered by spontaneous eye movement of gazing at objects, which does not require explicit operations such as pointing/entering a query condition by a mouse/keyboard. This way of triggering is possible since a user's state, that is, whether the user gazing at an object or not, is detected by eye movement. Since retrieval operation for visual database does not require hand-operated input devices, we think WYSIWYR framework fits well to mobile computing environments.

We categorize the retrieval based on WYSIWYR as a class of *Query-by-Example* (QBE) in the sense that a view frame acquired by eye movement analysis corresponds to the example (i.e., condition) of a query. However, WYSIWYR is different from ordinary QBE [13, 14] in the following senses.

(a) The example (i.e., view frame) is automatically acquired by gaze detection.
(b) The context of the example, that is, the attributes of visual information, is extracted from the user's spontaneous, implicit behavior.

When a gazing frame is detected as described in Section 3.2, the AoI is first extracted. During the user's gaze, a sequence of coordinates of the viewpoints is held to obtain the AoI. The AoI is one of the regions segmented by luminance, where vertices correspond to the coordinates when an eye is in fixation.

The size and the dominant color of the AoI are then calculated in order to compare them with those of view frames stored in the database. Note that those features of AoI for each view frame in the database are stored when the view frame is obtained.

In addition to the size and the dominant color of the AoI, view frames in the database are filtered by one or the combination of the following attributes:

(i) VFA (visual information type and gaze duration);
(ii) time stamp (date and time);
(iii) PDA data.

4.2. Feature extraction

When a view frame is obtained in retrieval, regions are extracted by merging the pixels whose difference of luminance is less than a threshold. Then one of the regions where most of the coordinates of viewpoints are located is regarded as the AoI.

The dominant color and the size of AoI are then calculated. The size of AoI is the number of pixels of the AoI. The dominant color of the AoI is represented by *Lab* color representation, which is determined as the mode of color histogram of the AoI.

4.3. Visual object retrieval

When a view frame is obtained in retrieval, the similarity of size and dominant color of an AoI is first evaluated. The similarity of dominant color is evaluated as a Euclidean distance between the dominant color of a view object in the database and that of detected view frame as follows:

$$Dc = \sqrt{(L_d - L_v)^2 + (a_d - a_v)^2 + (b_d - b_v)^2}. \quad (7)$$

In the above formula, Lab values of the AoI in the database are denoted as L_d, a_d, and b_d, and those of the current view frame are denoted as L_v, a_v, and b_v, respectively.

The similarity of the AoI is evaluated as the ratio of the size of AoI of a view frame in the database to that of the current view frame. That is, the ratio Ra is represented as

$$Ra = \frac{S_d}{S_v}, \quad (8)$$

where S_d denotes the size of AoI of a view frame in the database and S_v denotes that of the current view frame acquired as a query condition.

The candidates of the result of retrieval are a set of view frames which satisfy the condition

$$Dc < h_{Dc}, \qquad |1 - Ra| < h_{Ra}. \quad (9)$$

Thresholds for Dc and Ra are denoted as h_{Dc} and h_{Ra}, respectively, which are specified by a user.

In order to take the *context* of retrieval into account, the candidates are further filtered by the attributes associated with the video frames. This is described in the next section.

4.4. Filtering by associated attributes

Gaze duration and time stamp are also evaluated so as to arrange the result in the order of importance or recentness. In some cases, the gaze duration implies the importance or the complexity of the information in the view frame; the rearrangement of the candidates by the gaze duration reflects the context of gazing.

Visual information type, which is a part of VFA, is also referred to so as to percolate only the same type of information to the result that is presented to a user. By evaluating the visual information type of the object as well as AoI and its color, it is possible to take the context of access to the database into account. That is, we can exclude visual objects whose color attributes are similar but whose contexts are different. This implicit percolation is achieved by WYSIWYR framework but not by ordinary QBE approaches.

A time stamp associated with the view object is referred to for rearranging the retrieval result in the ascending or descending temporal order.

It is also possible to retrieve view objects by specifying visual information type, gaze duration, and/or time stamp as a prime condition(s). In this case, the evaluation of dominant color and region size is set to be lower priority.

Table 1: Result of view classification.

	Recall	Precision
Sentences	67% (640/951)	79% (640/811)
Figures/landscapes	85% (397/468)	83% (397/478)
Moving object	74% (122/164)	76% (122/160)
Others	56% (270/480)	79% (270/342)
No attention	94% (1055/1125)	76% (1055/1397)

5. EXPERIMENTAL RESULTS

5.1. Gaze detection

The experimental result of gaze detection is shown in Table 1. The number of frames extracted/sampled is indicated in parentheses. We implemented Digital Reminder on a Pentium II 266 MHz mobile PC, and video data is captured via PCMCIA interface in NTSC format. One of the reasons why the precision and recall still need to be improved is due to video format and resolution of the CCD camera.

In the table, "others" corresponds to the state where frequent saccades are observed but the pattern of movement corresponds to none of "sentences," "figures," and "moving object." The state of "idle" denotes the state where a user is watching no objects with attention. Though the average precision and the recall of gaze detection are 79% and 70%, respectively, most of the user's views without gaze are effectively discarded since the recall for the state of no attention to visual object is about 94%. Since Digital Reminder stores view frames only when the user was in concentration on visual objects, it eliminates redundancy in storing a user's view. In addition, implicit gaze detection enables to capture most of the gazed information without explicit operations for storing the user's view into the database.

5.2. Performance improvement by view frame attribute

The experimental result for evaluating the effectiveness of VFAs in WYSIWYR is shown in Figure 7. In this experiment, objects which were already gazed and stored into the database are gazed again as query conditions for view object retrieval. The result of retrieval which evaluates the effectiveness of VFAs as well as the color and the area of object is denoted as "with VFA," and "without VFA" denotes the result of retrieval

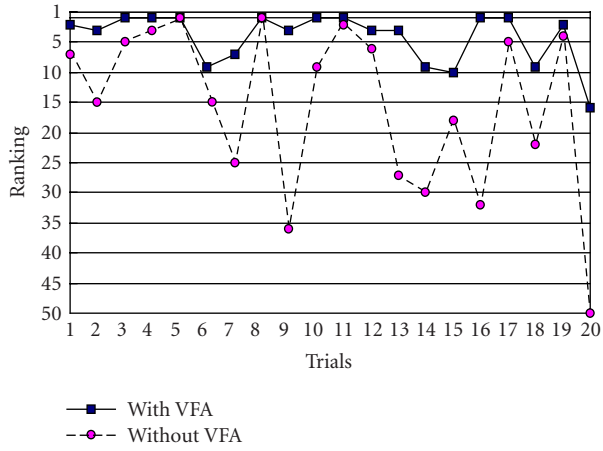


Figure 7: Performance of retrieval.

only by object color and area. Ranking denotes the position where the view frame containing the object that is the same as the one in the gazing frame is placed in the result.

In most of the cases, the view frame containing the same object as the user gazes is ranked at higher position by evaluating VFA in retrieval. Therefore, we conclude that the quality of retrieval is improved by the proposed method.

6. CONCLUSION

In this paper, we described the system named Digital Reminder, which captures view frames automatically and creates a real-world-oriented database. Since capturing view frames is performed based on gaze detection, database creation does not require one's explicit operation for storing one's view. Eight hours of continuous use of Digital Reminder accumulated approximately 8.3 MB of visual data, whereas continuous storage of view by 10 fps, 1.5 Mbps MPEG1 requires approximately 1.8 GB. This result shows that the proposed method effectively reduces the amount of storage by discarding views without one's attention. In addition, our approach is advantageous in the sense that the database is created without explicit operations and that even the visual object whose importance is realized later is also captured as far as the user gazes at the object. During the process of visual information acquisition, Digital Reminder recognizes the type of visual information by spatial pattern of saccades and the importance or the complexity of information based on gaze duration, which are associated with each of the view frames as a view frame attribute.

WYSIWYR framework enables a user to retrieve image objects not by the explicit operation for retrieval but by the implicit action of gazing at an object that is similar to what the user wants to retrieve. The WYSIWYR framework is in the class of QBE (Query-by-Example) approach; however, it differs from other QBE studies in the sense of implicit query composition and visual object type evaluation.

These remarkable features are due to the method of acquiring the object of interest by eye movement. Experimental result showed that query evaluation with VFAs improves the quality of QBE, and users' review showed that about 10% of acquired view frames were helpful to recall or enforce their memory.

ACKNOWLEDGMENT

We would like to acknowledge Mr. Masashi Yoshida and Mr. Takashi Kawamura for their implementation work of the prototype system, and Prof. Masahito Hirakawa for his discussion.

REFERENCES

[1] A. Yoshitaka, S. Kishida, M. Hirakawa, and T. Ichikawa, "Knowledge-assisted content based retrieval for multimedia databases," *IEEE Multimedia*, vol. 1, no. 4, pp. 12–21, 1994.

[2] A. Yoshitaka, Y. I. Hosoda, M. Yoshimitsu, M. Hirakawa, and T. Ichikawa, "VIOLONE: Video retrieval by motion example," *Journal of Visual Languages & Computing*, vol. 7, no. 4, pp. 423–443, 1996.

[3] M. J. Egenhofer, "Query processing in spatial-query-by-sketch," *Journal of Visual Languages & Computing*, vol. 8, no. 4, pp. 403–424, 1997.

[4] P. Pala and S. Santini, "Image retrieval by shape and texture," *Pattern Recognition*, vol. 32, no. 3, pp. 517–527, 1999.

[5] T. Starner, S. Mann, B. Rhodes, et al., "Wearable computing and augmented reality," Tech. Rep. TR-355, MIT Media Lab Vision and Modeling Group, Cambridge, Mass, USA, 1995.

[6] T. Jebara, B. Schiele, N. Oliver, and A. Pentland, "DyPERS: Dynamic Personal Enhanced Reality System," Tech. Rep. 463, MIT Media Lab Perceptual Computing Section, Cambridge, Mass, USA, 1998.

[7] M. Lamming and M. Flynn, ""Forget-me-not:" intimate computing in support of human memory," in *Proc. FRIEND21 '94 International Symposium on Next Generation Human Interface*, pp. 125–128, Tokyo, Japan, February 1994.

[8] J. Healey and R. Picard, "StartleCam: A Cybernetic Wearable Camera," Tech. Rep. 468, MIT Media Lab Perceptual Computing Section, Cambridge, Mass, USA, 1998.

[9] Y. Nakamura, J. Ohde, and Y. Ohta, "Structuring personal experiences—analyzing views from a head-mounted camera," in *Proc. IEEE International Conference on Multimedia and Expo (ICME '00)*, vol. 2, pp. 1137–1140, New York, NY, USA, July–August 2000.

[10] T. Kawamura, Y. Kono, and M. Kidode, "A novel video retrieval method to support a user's recollection of past events aiming for wearable information playing," in *Proc. IEEE Second Pacific-Rim Conference on Multimedia (PCM '01)*, pp. 24–31, Beijing, China, October 2001.

[11] K. Aizawa, K. Ishijima, and M. Shiina, "Summarizing wearable video," in *Proc. IEEE International Conference on Image Processing*, vol. 3, pp. 398–401, Thessaloniki, Greece, October 2001.

[12] M. Ikeda, *What the Eyes are Watching*, Heibonsha Limited Publishers, Tokyo, Japan, 1988.

[13] M. M. Zloof, "QBE/OBE: A language for office and business automation," *IEEE Computer*, vol. 14, no. 5, pp. 13–22, 1981.

[14] M. Flickner, H. Sawhney, W. Niblack, et al., "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.

Atsuo Yoshitaka graduated from Hiroshima University in March 1989, and received his M.S. and Doctor of Engineering in March 1991 and March 1997, respectively. He is currently serving as a Research Associate in the Information Systems Laboratory at Hiroshima University in Japan. His research interests include content-based retrieval for multimedia databases, visual user interfaces for database retrieval, and real-world-oriented interfaces. He is a Member of the IEEE, the IEEE Computer Society, and the Information Processing Society of Japan.

Yasuhiro Hori received the B.S. degree in the Faculty of Engineering at Hiroshima University, Hiroshima, Japan, in 2002. He is currently studying at the Graduate School of Engineering of Hiroshima University.

Hirokazu Seki received his B.E. and M.E. degrees from Hiroshima University in 2001 and 2003, respectively. His research interests include multimedia databases and real-world-oriented interfaces. He is currently with Matsushita Electric Industrial Co., Ltd., and is designing and developing car navigation systems.