Page-Flipping Detection and Information Presentation for Implicit Interaction with a Book

Kaori Fujinami and Nobuhiro Inagawa
Department of Computer, Information and Communication Sciences,
Tokyo University of Agriculture and Technology
fujinami@cc.tuat.ac.jp

Abstract

In this article, we propose an augmented book cover and a bookmark as means for implicit interaction with a book. Implicit interaction is a style of human-computer interaction, in which a user's explicit command is not required to control a system. The augmented items sense their state of use by accelerometers and detect page-flipping events. Accumulation of the events is utilized by an application, a Virtual Illustration System, to estimate the current page number. Finally, the number identifies corresponding multimedia contents. Thus, a user reads a book in an ordinal way, and at the same time, her context is implicitly captured and given to an application as an input.

The performance of page-flipping detection and the usability of the devices are evaluated. In addition, an evaluation of the interruptive nature of the presentation is carried out using real contents. The performance of the virtual illustration system itself is also tested.

Keywords: *Page-Flipping Detection, Implicit Interaction, Book, Bookmark*

1. Introduction

The advancement of technologies such as wireless network connectivity, high performance downsized computation and devices with low power consumption, has allowed us to access huge amounts of information anytime, anywhere. However, if a system does not provide information in an appropriate way, e.g., taking into consideration appropriate timing and ease of understanding, then users may feel inconvenienced, and this inconvenience leads to confusion as to how best to utilize these technologies. This means that a user is required to make an effort to acquire useful information.

In order to fully take advantage of the availability of information, it is important to be able to filter out irrelevant sources in particular situations; this is a major research topic in ubiquitous/pervasive computing environments, referred to as *context-awareness* [1]. Contextual information about a user is obtained through various types of sensor. In particular, we have been interested in sensor-augmented daily objects (artefacts) as a mean of natural and lightweight context extraction. We primarily utilize an object to achieve a specific task; for example, we utilize an entrance door to exit or enter a room, and a room light could be turned on if the outside knob is touched first when opened. Sensing the state of use of an object, i.e. "opened with outside knob first", allows a system to infer a user's current or upcoming task, e.g., "entering into a room".

Having an idea of a user's task, a system can take an appropriate action that supports the task itself or related tasks without asking a user what to do. The input to a system is a user's normal activity, and thus it is very natural to a user. Moreover, sensing is self-contained and thus *lightweight*; no external infrastructure is required in terms of extracting a user's context. In contrast, other technologies that also enable the same service, e.g. location tracking and video-based activity recognition [2, 3], require dedicated sensing infrastructures in the environment. Such a style of human-computer interaction that does not require a user's explicit command is called *implicit interaction* [4] or *background interaction* [5]. In the implicit interaction paradigm, not only obtaining a user's context, but also providing a service in an unobtrusive manner needs to be considered. This is because of the implicit nature of the sensing method; a user is utilizing a particular object to achieve a task, not to provide contextual information. So the usual manner of execution of a primary task should be respected, while context-enhanced value is provided. We have been investigating a service model that is realized by context-sensitive objects [6, 7].

In this article, a paper-based book is augmented with sensors to literally provide "context-aware" multimedia information. A sensor augmented book cover and a bookmark have been investigated to track the current page. The page (number) currently being read is given to the system as a context, and related information is provided. This is considered as an extension of a traditional printed illustration. An illustration in a book is utilized to elucidate textual expression with visual presentation, which can help us understand things that are difficult to explain with text. In addition to such usefulness, a sense of immersion or an atmosphere is offered by an illustration. A reader could be given a rich experience with such a *digitally enhanced* illustration; for example, a person who has no idea about the Gion Festival of Kyoto could more richly imagine the festival, if pictures and festival music were presented and played according to the corresponding description. Although an electronic book could also present a reader with such information [8, 9], there is still a vast amount of paper-based books for which it would be worth developing such an augmented reading experience. Additionally, for many users, the tangible character of a paper-based book is still superior to that of an electronic book, even if electronic paper and flexible display technologies have been investigated to fill in the gaps [10, 11]. Thus, we focus on the augmentation of a traditional paper-based book. The contributions of this article are three-fold:

1. We present an augmented book cover and bookmark as means of implicitly interacting with a paper-based book. The devices detect page-flipping in a lightweight manner. The accumulated number of page-flipping events can approximate the current reading position.
2. We present an application prototype that presents multimedia information based on the reader's current reading position obtained from the two devices.
3. Finally, we present two types of information presentation for the application, and conduct user studies using real contents to assess the effectiveness and the burden of the presentation, as well as the concept of the application itself.

The rest of the paper is organized as follows. Related work is examined in the next section. In section 3, an augmented book device called a Sentient Book is proposed and evaluated regarding the page-flipping detection performance. Then, an application, a Virtual Illustration System, is proposed in section 4. Finally, in section 5, we conclude the article with a discussion.

2. Related work

To improve human-computer interaction, many attempts that aim at taking advantage of electronic and physical documents have been made; the DigitalDesk [12] is a pioneering study. EnhancedDesk [13] also seeks the smooth integration of paper and digital information on a desk. Both of the above allow automatic retrieval and presentation of digital information by recognizing the contents [12] or the tag [13] printed on a page, and direct manipulation of digital information by gesture recognition. WikiTUI [14] is designed to provide bi-directional interaction with digital contents using the Wiki infrastructure. A reader can add and obtain digital annotations based on the page he/she is reading. They basically consist of augmentation of desk operations, and thus the working area is restricted to the desk location. On the contrary, our augmentation is done on the book side, which provides a user with a certain amount of freedom in the workplace.

Most closely related to our work regarding the application concept is eyeReader, which was proposed in the Augmented Text project [15]. The eyeReader system realizes gaze-based multimedia contents acquisition while reading an electronic book, which is analogous to traditional mouse-based web browser interaction; corresponding multimedia information is presented when an "eye-cursor" comes to a specific position in a text. Bahna and Jacob's work is also similar to ours; they have proposed an interaction technique, a *peripheral border display system*, for computer-based reading tasks [16]. The system conveys general awareness of the environment of a text with wall-projected abstract information. Peripheral images change automatically as the reader goes further in the document, where a user reads an electronic document presented between the screen and a user. They showed that the border display interaction technique successfully provides additional information without additional user effort and with no apparent negative effect on reading time or comprehension. We have tested not only an automatic full-screen presentation like theirs, but also a semi-automatic one based on negative feedback from an initial study utilizing such a fully automatic approach.

In terms of the method of identifying content to provide at a certain moment, recognizing a reading page is a relative approach [17, 18], where content is linked to one or a range of pages. On the other hand, embedding contents into a page allows direct identification using an appropriate detector [19, 20]. Back et al. augmented a book with RFID tags and a receiver to provide additional information based on a page [17]. Some projects have applied visual tags: a reader of the Magic Book [19] watches three-dimensional characters related to the page through a head mounted display. The Interactive Textbook system [20], an application of EnhancedDesk, literally provides an interactive experience with electronic contents linked to a page. Furthermore, a completely new material of paper, polymer conductive ink [18], has been investigated to detect a bending action; however, specially manufactured pages are required, i.e., redesign is needed. This makes the cost of a book high and prevents an existing book from adopting an augmented service. In WikiTUI, a user must tell the system the current page number every time she flips a page of a (physical) book. This is realized by pointing at the old number projected on the operating table, and the system increments an internal page counter. This means he/she flips both the actual page and the virtual page. The page number is correct as long as a reader follows the flipping rule; however, it requires the reader's attention. The Magic Book requires an extra device, a head mount display (HMD), which requires a reader to adapt to a new style of reading a book, which may be appropriate for educational and entertainment use. In contrast, our system aims at realizing page detection in a cost effective manner, allowing a reader to use almost the same actions as when reading an ordinary book. Attaching a device to a common part of a book enables this, i.e., a book cover and bookmark, rather than offering special pages or requiring a reader to wear special instruments or to *utilize a system*.

3. Sentient book

In this section, the design of a sentient book is described.

3.1. Object augmentation approach

An application that we consider and implement in a later section literally provides context dependent audio/visual contents to a reader while she is reading a paper-based book. In particular, we are interested in books such as novels and travelogues that are read in a sequential manner. A reader will see a slideshow of the Gion festival or hear the festival music when she is reading a page related to that festival, for example. To make the presentation effective, the timing is important: if the slideshow were to appear when she is reading a page that describes another festival, she would be confused. Considering the role of the (additional) material, the presented information should be easily linked to a word or a scene to be explained. We consider a system that keeps track of the reading position (tracking eye-movement) and that has the potential to minimize the time difference of presentation. A combination of eye tracking and optical character recognition (OCR) technologies would enable this very straightforward approach. However, due to the difficulty in deployment and the cost, it is not feasible in our daily life environments at present, although much effort has been applied to investigating practical eye-tracking [21]; the sensing system requires careful installation and calibration to some extent. Furthermore, a user needs to read a book at a specific location or wear special devices, e.g., a video camera or eye-tracker.

We have determined not to take such a “heavyweight” approach; rather we augment an object itself with sensors to measure the interaction of a user (reader) with a book. Here, the interaction is page-flipping. The accumulated number of page-flipping events indicates the two-page spread that is facing a reader (see Figure 1-(a)-(i), for example). This means that she is reading somewhere in that two-page spread at a certain moment. If a set of contents is mapped to pages, then flipping allows retrieval of the contents. As a consequence, we have developed add-on devices to detect page-flipping events in a post hoc manner. The fine-grained measurement of the reading position is sacrificed, but the lightweight nature of the approach should be attractive for free-style reading. Through application prototyping, we will investigate appropriate timing and presentation with such an approach. Note that a solution that utilizes specially manufactured pages with content IDs, e.g. RFID or visual tagging, and a dedicated detector, allows precise identification [17, 19]. However, to benefit from already published books, we have not taken this approach.

3.2. Add-on devices for page flipping detection

The requirements for the device are listed below:

1. **Natural extension of traditional reading style:** If a user is required to intentionally carry out one or more actions in addition to a traditional reading style, she might feel confused. On the other hand, if the additions are well-defined so that they could be “woven” into a sequence of page flipping, she will soon become familiar with the new style.

2. **Robustness with respect to various reading styles:** There are various styles of reading. Some people read a book in a horizontal manner on a desk, while others might hold it above the desk. Constraining readers to a particular style stresses them. Hence, the page-flipping detection mechanism should be robust with respect to reading style.
3. **Support for both soft and hard cover books:** As reported in [11], we often utilize the bendable nature of a soft cover to flip a page of a book. However, this is not applicable to a hardcover book, which indicates that flipping detection should not depend on the deformation of a book cover.

With these requirements in mind, we have analyzed page-flipping activities to specify the common parts of a book that characterize page flipping. Then, we have adopted a detection strategy, i.e. *obstacles are added so that a reader must remove them to complete one page flip*, and the movement of the obstacles is converted into a flipping event. As “sensitive obstacles”, we have augmented two types of book items with accelerometers: a book cover and a bookmark. As shown in Figure 1-(a), the book cover type is utilized by (1) picking up the

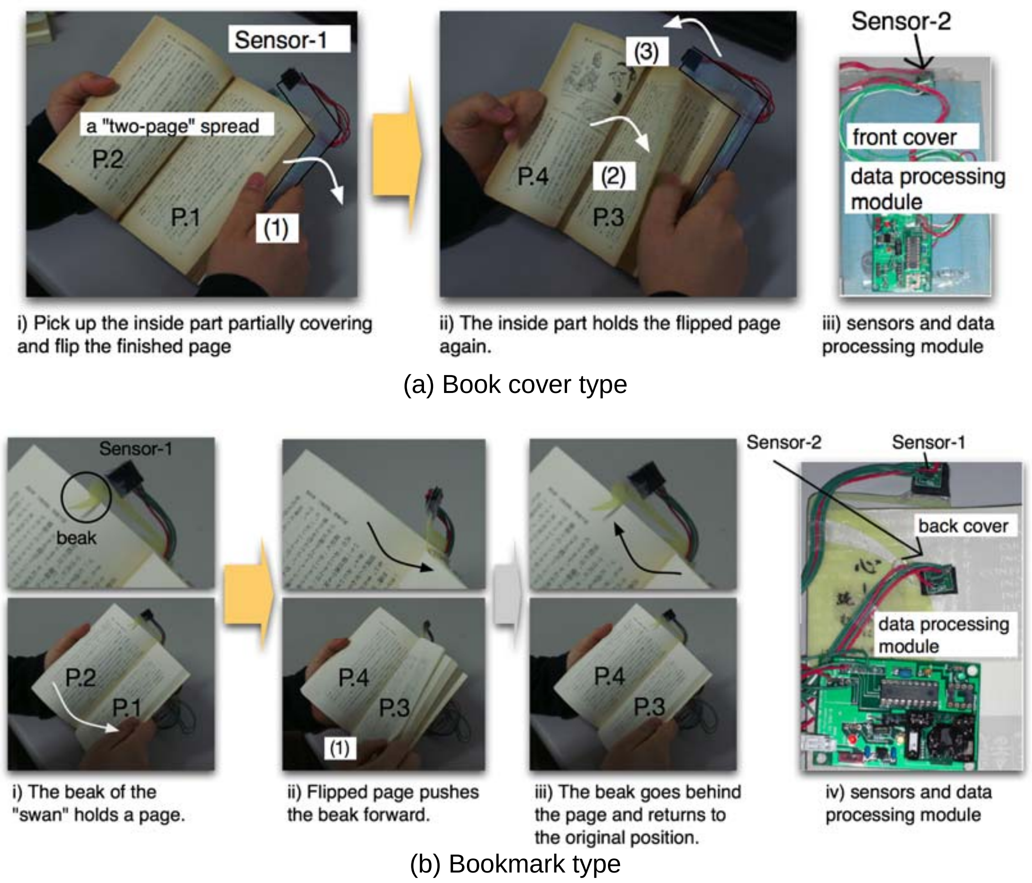

Figure 1. Page Flipping Detector Device: a) Book cover and b) Bookmark Type.
Note that the page is flipped from left to right; that is the traditional way of reading a Japanese book (reading a page vertically from top to bottom, and from the right side to the left).

inside part partially covering the reading surface when a reader flips the finished page (2). The inside part of the cover is translucent so that it does not interrupt the reader's view. Every time a page is flipped, the part is flipped accordingly to hold the page (3). The flipping detection by the bookmark type is realized by analyzing the movement of a wire prop that holds the pages to be read (Figure 1-(b)). We have augmented a commercial product called SwanTouch [22]. Whenever a reader flips a new page, the page pushes the beak of a “swan” forward (Figure 1-(b)-(ii)), and soon the beak goes behind the page due to its wire construction (iii). Therefore, the movement of the beak corresponds to flipping a page. The removal of these obstacles, i.e. the translucent inside cover and the beak of a swan, is naturally embedded into ordinal page flipping, which is a solution for the first requirement. Furthermore, the two methods are not affected by the change in shape of the book cover (the third requirement). The usability of the devices will be evaluated in Section 3.4.

We have developed more than 20 versions of the two devices to find a suitable design in terms of usability and performance of the detection mechanism, for which the size and the sensing position as well as the material are subject to test. To meet the second requirement, two accelerometers were attached to each device: one of them was utilized as a reference to avoid confusing the movement of the book itself. In the book cover version, a 3-axes accelerometer was attached inside the cover to detect the movement of the part (marked as Sensor-1 in Figure 1-(a)-(i)). Also on the front cover side, the other 3-axes accelerometer (Sensor-2) was attached. The material of the cover is polyethylene terephthalate (PET) of 0.3 mm thickness, and the “reverse gamma shape” of the interior part was finally determined to balance the usability (the readability of a text and the ease of manipulation) with the detection performance. Regarding the bookmark version, one accelerometer was attached on the beak (Sensor-1 in Figure 1-(b)-(i)), while the other was on the back cover (Sensor-2). SwanTouch is made of polypropylene; however, we also tested other materials, i.e., an acrylic sheet of 0.5 mm thickness and a 0.3 mm PET version. Among these, the polypropylene version performed best.

Sensing the change in thickness of the remaining (or already read) pages might also enable page-flipping detection. As proposed in [23], the thickness of sheets of paper could be measured by sensing the capacitance between the first and the last sheets. However, the characteristic is not linear, and it is saturated at about 20 pages thickness. Therefore, we do not consider it suitable for a book that has more pages.

3.2. Page-flipping detection algorithm

The algorithm is explained in Figure 2, where (a) and (b) indicate the sensor readings from Sensor-1 and Sensor-2, respectively. Additionally, column (I) corresponds to typical reading styles, while column (II) is provided as an example of non-flipping movement. Note that the vertical axes indicate values obtained from an A/D converter, and the vertical axis of (d) is represented with a log-scale. The sliding window size has been set to 20 samples (1 second) with no overlapping, and the threshold has been set to 5 in a heuristic manner. We have applied the same flipping detection technique in the two versions since they basically have similar round trip motions of the *sensitive obstacles* mentioned above.

To emphasize the distinction between moving and stable states, the variance within a certain window is utilized. A variance is calculated for each axis (x, y and z), and then they are averaged to obtain a representative value for a sensor (c), i.e. Sensor-1 or Sensor-2. Next, the ratio of the value of Sensor-1 to that of Sensor-2 is calculated (d). Here, Sensor-2 acts as a baseline. We have adopted the ratio since we found it difficult to distinguish the actual flipping from mere movement of a book itself when only one sensor (Sensor-1) was utilized. As can be seen in the right column, the two sensors show very similar waveforms when a book is just carried; however in a page flipping situation, only the inside and the beak parts are actually affected, and the body of a book moves independently (compare (a) and (b) in both column (I) and (II)). Therefore, it is our opinion that the ratio performs well. Finally, to obtain a result for the *detection of flipping*, a threshold has been applied. In (d), three flipping events are detected at f_1, f_2, and f_3. Robustness with respect to various reading styles (the second requirement) has also been achieved. In Figure 2, flipping pages with three reading styles were shown: horizontal, tilted, and suspended. They look different from each other in the original sensor readings (a), but the ratios have a very similar form in (d). Thus, the proposed algorithm distinguishes page-flipping from mere movement of a book itself in a robust manner. A performance evaluation of the two devices is presented in the next section

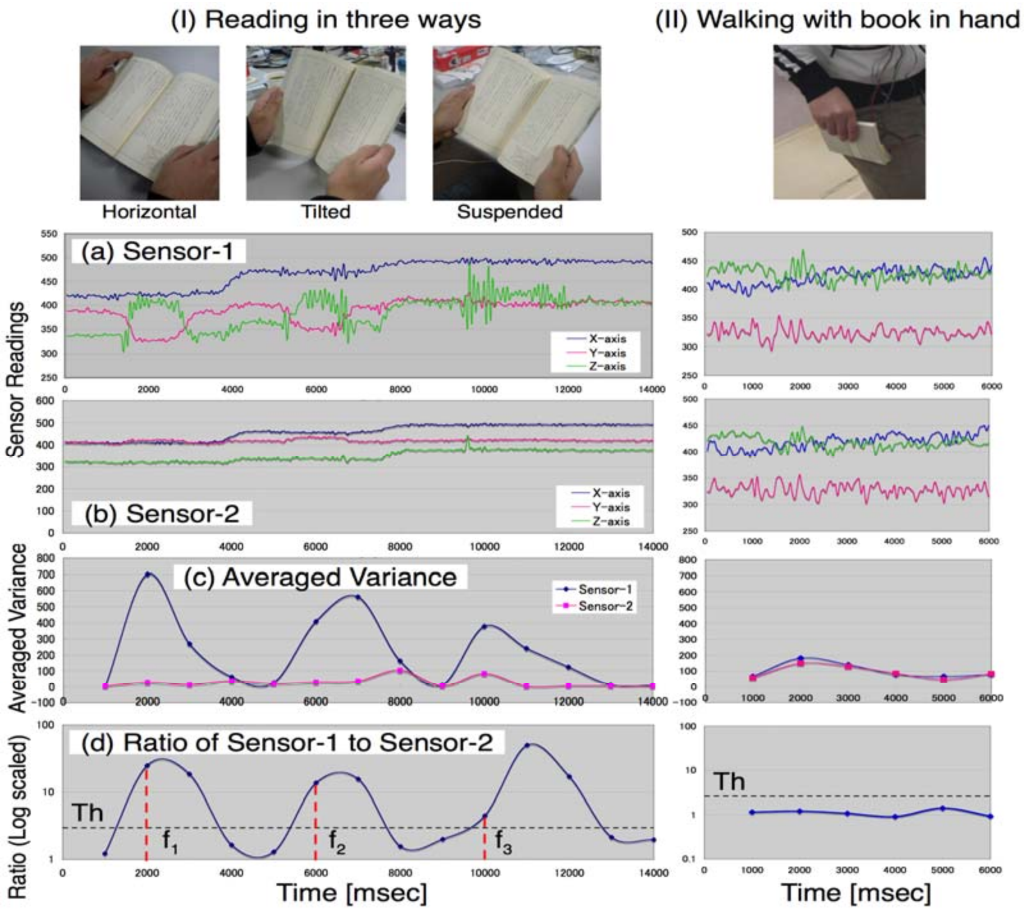

Figure 2. Data Plotting of Sensor Readings (a) and (b), Averaged Variance (c), and Log-scaled Ratio of Sensor-1 to 2. “Th” in (d) indicates the threshold for flipping detection. Column (I) indicates the data from typical reading styles, while column (II) represents non-flipping movements, i.e. walking with a book in hand.

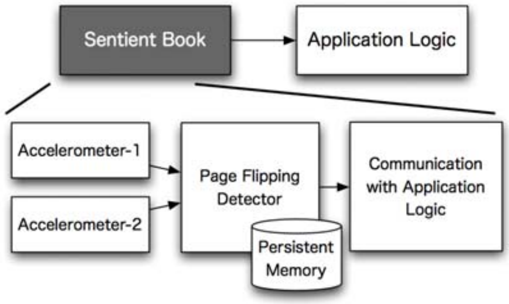

Figure 3. Architecture of the Sentient Book Application and Internal Components of the Sentient Book

3.3. Architecture and current implementation

Figure 3 illustrates the architecture of the Sentient Book application, where a Sentient Book device detects page-flipping events using two accelerometers, and communicates the accumulated number of events to the application logic. The accumulated number of page flipping events is maintained in persistent memory, which allows a user to resume reading without flipping pages from the beginning.

The two prototypes have been constructed using our general-purpose wireless sensor/actuator node named Tonttu, which consists of two 3-axes accelerometers, a PIC16F88 microcontroller, a ZigBee-based wireless communication module and generic I/O ports. In the current implementation, the page-flipping event is remotely detected, which means the add-on devices wirelessly transmit sensor readings every 50 msec to a PC, and the detection is performed there. However, to reduce energy consumption of the node and to hide the complexity of data analysis, our future implementation will detect the flipping locally (on the Tonttu side), and only an event will be communicated.

3.4. Performance and usability evaluation of a sentient book

We conducted a performance evaluation on page flipping event detection, and a usability test of the two types of add-on devices.

3.4.1. Methodology: Twelve subjects (10 undergraduates and an adult couple in their 50's) participated in the test. To assess the intuitiveness of the augmentation, we did not inform them as to the manner of use in the first trial (0). Then, they were instructed as to the correct manner of use and the principles of page identification. Three types of usage were tested to determine the robustness of the algorithm: (1) horizontal on a desk, (2) tilted on a desk, and (3) held suspended in the air. The subjects were told to flip 20 times (40 pages) for each type. The memorability was also tested; two of them had the same test a week to ten days later. They were the subjects who could not work out the methods without instruction in the first trial.

3.4.2. Performance of page flipping detection: Table 1 shows the detection accuracy. The accuracy of flipping detection is defined by the ratio of the number of counted pages to the actual number of pages that the subjects flipped. The standard deviation (SD) was also calculated to see the variation in individuals. The averaged accuracy for the book cover and the bookmark version was 88.1% (sd=11.4%) and 92.5% (sd=8.3%), respectively.

Table 1. The Performance of Page-Flipping Detection

	Book cover (N=820)		Bookmark (N=960)	
	Average [%]	SD [%]	Average [%]	SD [%]
(0) Without explanation	84.4*	20.3	21.3	18.5
(1) Horizontal on a desk	88.7	8.5	92.1	8.5
(2) Tilted on a desk	85.2	15.4	91.3	9.2
(3) Suspended (above a desk)	90.4	11.4	94.2	8.3

This indicates that the bookmark version detected page-flipping extremely well with small variation among individuals. We consider the reason for the difference is that the book cover has a wide range of motion, which could lead to variation of utilization among individuals, and the heterogeneity makes it difficult to detect page-flipping with a simple threshold. On the contrary, the algorithm itself is robust in the sense that the differences among the three styles are small.

The case without explanation (1)-(3) shows low accuracy and large deviation. In the case of the book cover version, seven subjects could not find out about picking up the cover when they were not told anything. They just put the flipped pages on the cover. We consider the gesture of putting the flipped page under the inside part of the cover was far from a natural reading gesture. On the contrary, the accuracy of the bookmark version was extremely low. This is because none knew the way of flipping, but they continued to read the book anyway. However, once they were instructed, the accuracy and the deviation were improved. The methods were easy to learn due to the simplicity and the seamless integration into ordinal book reading. Regarding memorability, the result shows the accuracy of the three cases were almost the same as before. We consider this is because they were told not only the method itself, but the principle of page-flipping detection. Additionally, the physical appearance of the devices reminded them of the usage.

Misdetection with the bookmark version generally arose from the situation where the flipped page went through the beak part without a large movement. If we made the material of the bookmark more rigid, it could more firmly hold the page to be flipped next. Then, detection might be more accurate due to the larger bouncing acceleration of the beak. However, at the same time, it would become more difficult to flip and sometimes the page might be damaged.

The common limitation for both methods is that they do not support random access to the contents. This comes from the method of identifying the current page. The advantage of random access is that it not only allows the system to know the page number if a reader opens the book suddenly, but also it can eliminate accumulating errors in page-flipping detection. We basically read a novel and a travelogue in a sequential manner, which indicates that only rarely do we need to access a page randomly. However, the effect of accumulating errors might not be negligible for a thick book. For example, a bookmark system attached to a book with 200 pages can misidentify 15 pages (= 100 flips × 2 page/flip × 7.5% error) by the end. To address this issue, an extension of the current device that has error correction or page number adjustment functionality needs to be investigated while taking into account the impact on a traditional reading style. Alternatively, a body-mounted camera and sophisticated page number recognition would allow random access. This could expand the types of books that might use the system; however, robust camera control against free body motion and the manner of reading a book would be required.

3.4.3. Usability of add-on devices: The proposed devices have been designed to retain the traditional reading style as much as possible. From the interviews, we found that the subjects felt that the bookmark version was less obtrusive than the book cover version. The bookmark version requests a user to pay attention to the beak part to some extent; however, it has a great advantage over the book cover version. The beak holds the next page to be read (Figure 1-(b)-(i)), whereas the flipped page is held by the book cover (Figure 1-(a)-(i)). This means a reader of the book cover version needs to open the cover when he/she wants to check the flipped pages again. This causes misdetection of flipping since the movement of the inside cover is the same. In contrast, the bookmark version has no such limitation. The information is provided based on the incremented page number determined by the movement of the beak. A reader might be presented with different information while he/she is checking a page that he/she flipped before. We consider this is not so big a problem in a novel or a travelogue since the activity is basically confirming an unclear point. However, a pause button added to the beak or at another location would contribute to an improvement in usability.

4. Application prototype: a Virtual Illustration System

The aim of this prototyping is to investigate the applicability of the lightweight, i.e. augmented object, approach. In addition, the effectiveness of such an augmented reading experience is evaluated to find a new class of implicit interactive applications. We have developed a Virtual Illustration System for these purposes. The system provides a reader with multimedia information according to the page that he/she is reading. This, as the name indicates, is an extension of a traditional illustration printed on paper. The information could be a still image, a movie or a sound clip. Furthermore, more ambient information such as the color of the room light, vibration of a chair, or an aroma in the room, could be provided if an appropriate actuator were available. The information would contribute to enlivening a scene or supporting a reader to enable them to understand the description in a book. Note that we recognize there are also people who want to enjoy creating images in their minds without system support, especially for a novel. We do not intend to change their minds; the proposed application is for those who are curious about such augmented experiences.

4.1. Designing the virtual illustration system

The key design issues and decisions are as follows:

1. **Mapping of multimedia contents to one or a range of pages:** It is obvious that providing a meaningful set of multimedia contents is crucial to successful service. This prototyping does not focus on the creation of the contents: we assume a "page-contents mapping table" is provided by someone such as a book editor, an author or a hobbyist. The table maps the accumulated number of page-flipping events to multimedia files or actuators. We have defined an XML-based language to specify the content file names, corresponding page numbers, and the duration to show/play one unit of content.

   Alternatively, if an electronic counterpart of the target book is provided, unsupervised methods as proposed in [24, 25] might be applied, which identify keywords in a lyric to provide suitable images or colors.

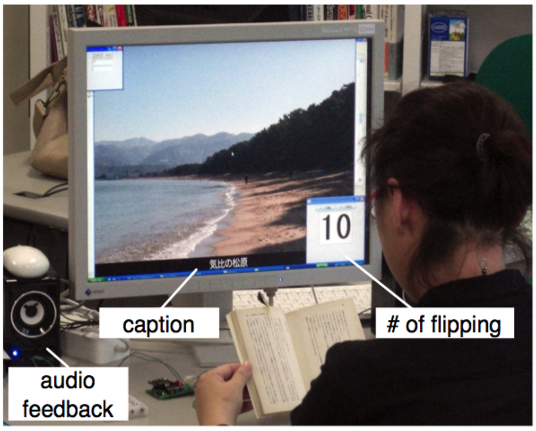

Figure 4. Typical Snapshot of the System

2. **Presenting the contents without additional user effort:** We took a push-based approach, where corresponding contents are automatically presented based on the table and flipping event. Here, a reader passively receives the information with her peripheral sense. In our earlier experience with an augmented mirror [26] that showed personalized information while a person was brushing their teeth, half of the subjects in usability tests preferred automatic activation of the presentation rather than an explicit indication of the start. Furthermore, Bahna and Jacob's work on a peripheral border display interaction technique [3] encouraged us to take such an automatic presentation approach, where images are projected on the wall while reading an electronic document. They showed the effectiveness of using this technique as a basis to communicate additional information without additional user effort.

   Figure 4 represents typical usage of the system. One full screen image is presented at a time. If more than two pages are assigned to one page-flipping event, they appear in turn with an interval specified by the creator. A black screen is shown if the contents creator does not want to show any image; otherwise, the same image is presented until the next number-content mapping appears.

3. **Supporting a reader to understand presented visual contents:** If a reader fails to find the relationship between the presented content and the description that she is currently reading, she might be confused. The lightweight nature of the reading position identification is likely to cause this problem: the resolution is to present material from the viewpoint of the two pages, not from a single word. So, as can be seen in Figure 4, a caption is shown at the bottom of an image to fill the gap between the timing of the presentation and the actual appearance of the corresponding word(s).

   Furthermore, the estimated page number is presented on the display in terms of feedback from the page-flipping action. As was the case with a sensor-based system, the page-flipping detection is not 100% accurate, and therefore the system might misidentify the content to be provided. Presenting the number, as well as playing a sound that indicates page-flipping, provides an indicator of the internal state of the system.

Figure 5 illustrates the system architecture. The major components of the system are the Sentient Book, the Virtual Illustration Controller and the Presentation Manager. The Virtual Illustration Controller is comprised of a contents definition module, an interpreter of definitions and local/remote contents. Contents definition is represented as a file that could be downloaded from the Internet or provided locally via external storage, e.g. an SD card. The International Standard Book Number (ISBN) could identify a contents definition. The interpreter responds to an event issued by the Sentient Book, where the event is the estimated page number calculated from the Sentient Book side. Once the interpreter identifies a content assigned to a two-page spread in the mapping table, it requests the Presentation Manager to handle it appropriately. An individual content unit can also be represented as a file: an image, a movie clip, or a sound clip. Even control commands for an actuator could be put into the file. The association of a Sentient Book device with a paper-based book is out of the design focus of the prototype. However, a terminal that runs Virtual Illustration Controller could realize this by scanning ISBNs, for example.

4.2. Initial user evaluation

We have conducted a user test on the virtual illustration system, and validated the design of a sentient book.

4.2.1. Methodology: Three subjects (undergraduates; two new subjects and one from the previous performance test) actually utilized the system with the bookmark version. Here, a travelogue “Kaido-wo Yuku” by Ryotaro Shiba was selected, where the author's experiences and random thoughts during travel along a country trail are described. We provided contents to help subjects imagine the scenery, which included old and contemporary maps, illustrations of an old Samurai-battle, landscapes, and novelty animals. Additional sound clips included a war whoop, the lapping of waves against a shore, and the sound of a rain shower. After reading some sections (34 pages), we had a semi-structured interview session. Note that none of the subjects had any prior knowledge about the places and events described in the book.

4.2.1. User feedback and implications: Regarding the readability of the book, two subjects felt rather annoyed because of two types of interruption: 1) change of images in a short time and 2) a large gap between an actual description and the presented information. The first interruption was caused when more than two pages were assigned to one page-flipping (a two-page spread) and changed within a short time. The subjects tried to see an image every time it appeared. For the first two subjects, the time to change was set to 3 to 7 seconds, and the third subject who did not feel annoyed was assigned an interval of 10 to 40 seconds. We consider that too short an interval prevented concentration on reading a text, and thus made those subjects feel that the extra material was intrusive. However, the interval is not the only reason for the perceived interruption; the appropriateness of the timing of presentation is important. In particular, premature presentation bothered them, because they had to determine the relationship to the part they were currently reading. This sometimes happened for contents related to the left page (latter part). As described in section 3, our approach of page number identification is based on an accumulation of page-flipping events, which limits the resolution of the reading position in a book to two pages. To improve the resolution, the speed of reading, i.e., words per minute, and the amount of text in every two-page spread are realized while utilizing the current setting.

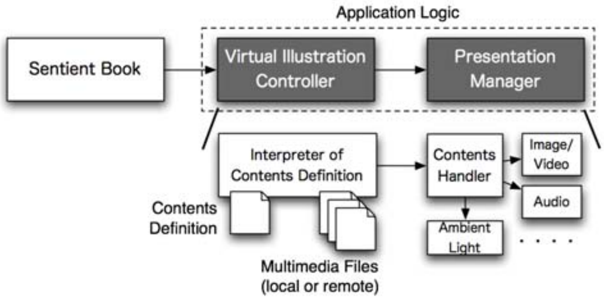

Figure 5. Architecture of the Virtual Illustration System and Internal Components

However, throughout the interview session, we noticed that the subjects needed information to resolve ambiguity. In particular, the subjects preferred images of maps. Although there was a hand-written map at the beginning of the book that illustrated the positional relationship of typical places, it was too abstract and unrelated for them to understand. The positional relation of many other places remained unclear, which was critical in the book since it was a travelogue. Other images that the subjects preferred were pictures of a novel animal, i.e., an ermine, and a mountain path because they were informative. On the other hand, the audio contents were not preferred since they were played at the wrong time. We have considered that a *semi-automatic* presentation would address this issue; not all the contents were interesting to the reader, and so such a "forced presentation" was rather annoying.

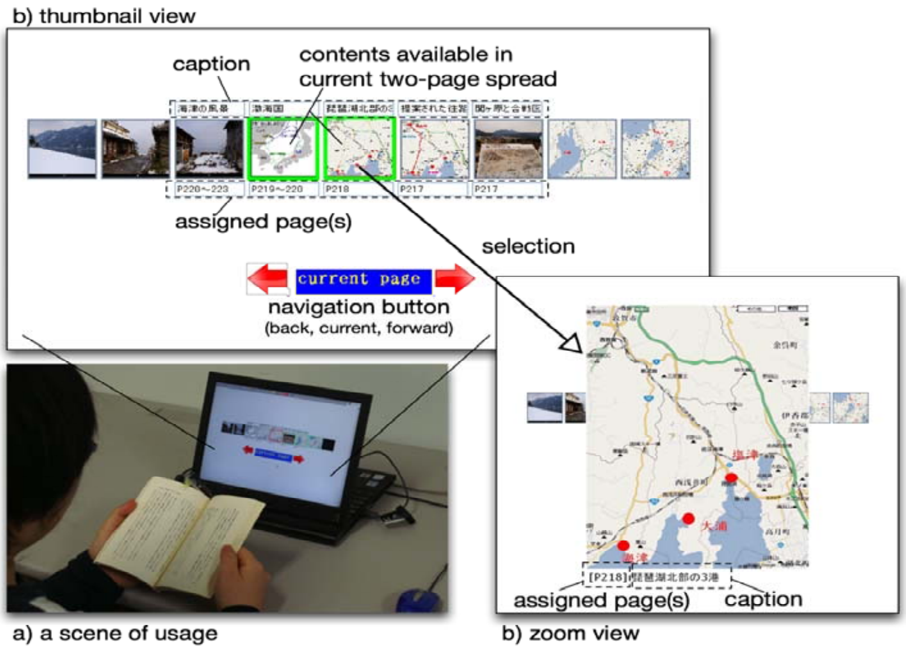

Figure 6. The components of and the interaction with the improved

4.3. Improvement on the presentation

Based on the analysis of the initial user test, we have designed a new version of the presentation as a process of iterative design. Figure 6 illustrates the components of and the interaction with the new presentation. The new one consists of an overview and a *zoom-up*, where an overview is presented automatically based on a page-flipping event, while zoom-up is initiated manually. An overview is constructed with a set of thumbnails, corresponding keywords, assigned page numbers and navigation buttons. On detecting a flipping event, the thumbnails assigned to the newly opened two-page spread are appended to the tail of the thumbnail queue. If the queue has space, older thumbnails could still be presented. To help a reader identify the new contents, the border of the thumbnail or icon image is highlighted with a thick and colored line. The navigation buttons, i.e. back, forward and jumping to the current page, allows a reader to check thumbnails that are not in the queue at any time. Note that each thumbnail is clickable to obtain a zoom-up view. For non-visual contents, a clickable icon invokes a dedicated actuator.

4.4. Evaluation of the interruption by the automatic/semi-automatic presentations

We have conducted an evaluation of the interruption with two presentations, which aims at investigating how much the presentation interrupts a reader and how she actually perceived the interruption. The evaluation has been conducted from both a qualitative and quantitative point of view.

4.4.1. Methodology: Twelve males (ages 22 to 24) volunteered for this test. Six were assigned to the automatic version, while the remaining six were assigned to the semi-automatic version. Both groups were assigned the same book and image-based contents for comparison, where the same travelogue as was used for the initial study was utilized (but the length was shortened to 26 pages to reduce the workload). All of them were masters' students in the department of computer science, but not part of the authors' research group. Eleven subjects hardly read at all (less than one book per half year). Note that none of the subjects had any prior knowledge about the places and the events described in the book, and also that the subjects of the initial user study were not included.

In contrast to the previous test, an experimenter simulated the page-flipping detector device, i.e., an experimenter controlled the timing of presentation. This helped to avoid the effect of the usability of the device, thus allowing them to concentrate on the presentation. The subjects were asked to read the book; they were told that they would be videotaped during the reading task for later analysis, that they would be interviewed at the end, and that they could safely ignore information that was presented in a 15-inch tablet PC (see Figure 6-(a) as an example). A web-camera was attached to the display frame, which was utilized to approximately detect a switch of the gaze of the subjects in an economical manner. The subjects of the semi-automatic version could choose the method of interaction with the display: touching by a stylus or mouse clicking.

The subjects were also told that they would have a memory test afterwards, which was to make them concentrate on the primary task, *reading*, and to make the presentation effective and realistic as a source of interruption. Furthermore, as described in section 4.5, the result of the memory test was utilized to evaluate the performance of the virtual illustration system itself and the two presentations with respect to comprehension of the book. Therefore, we made an inter-group comparison, rather than applying the two versions to the same group.

Table 2. Summary of evaluation on interruptive nature of the presentations

	Automatic		Semi-automatic	
	Average	SD	Average	SD
Score of perceived annoyance (1=definitely annoying, 5=not annoying at all)	2.2	0.8	4.2	0.4
Number of gaze switches	55.3	12.6	43.7	19.4
Time taken for a "look" [sec]	2.3	0.4	5.5	2.8

4.4.2. Results and implications: Table 2 summarizes the results: 1) the feedback from the subjects about perceived annoyance on a 5-point Likert scale (1 = definitely annoying, and 5 = not annoying at all, 2) the results of the video analysis on the mean number of *gaze-switches,* and 3) the mean time taken for a look. The results shows that the perceived annoyance was improved in the semi-automatic version; the number of gaze-switches was reduced and the time taken to parse an image increased once they decided to look at the image.

The new presentation, i.e., the semi-automatic version, showed all thumbnails that were assigned to a two-page spread at the time that a user flipped a page. In contrast, the old version automatically showed images one-by-one at a specific interval if more than two images were there. In the experiment, when there was an image to present during one page-flip, an additional 1.88 images appeared in this manner. This indicates the members of the "automatic" group might be interrupted 1.88 times. Here, the first image that appeared just after page-flipping was excluded. On the other hand, there was no such interruption in the semi-automatic version; the video analysis and our observation revealed that the majority of subjects in the group glanced at the newly appended thumbnail, but checked a zoomed image later. They remembered the keywords and utilized one if desired.

As described above, the subjects of the semi-automatic version interacted with the system to check interesting contents either by tapping with a stylus of a tablet PC or with a mouse click. Here, eleven subjects utilized a mouse from the beginning, and the one who initially selected a stylus finally changed to a mouse. The reason for such a dominant preference of the mouse interaction is that they thought or found it convenient to control a mouse near the book, rather than move their hands to the surface of the tablet PC that was tilted in front of them (see Figure 6-(a)). We consider this is also an interruption that was caused by the presentation.

4.5. Evaluation on the effects of the virtual illustration

We also conducted another evaluation that aimed at investigating the effect of the additional information, i.e. virtual illustrations.

4.5.1. Methodology: In addition to the twelve subjects in the above-mentioned test, six males (ages 22) were added as a control group; they were assigned to the same book *without* any device.

Table 3. Summary of evaluation on effects of virtual illustration

		Control group		Automatic		Semi-automatic	
		Average	SD	Average	SD	Average	SD
Score of perceived burden (1 = not burdensome at all, 5 = definitely burdensome)		3.4	1.1	3.7	0.9	3.5	0.8
Correctly Answered [%]	Text-question	40.5	22.5	66.7	17.8	66.7	15.8
	Image-question	16.7	13.6	55.6	20.8	50.0	21.5
	Map-question	3.3	7.5	16.7	21.3	40.0	23.1

After reading the book, the eighteen subjects had a written memory test, where three types of questions were given: 1) seven text questions (the subjects filled in or marked text-based answers, which appeared both in the book and in the caption of the presented contents), 2) six image-questions (the subjects identified images that were answers of questions, where the subjects could answer the questions either by recalling the images or by imagination from the written text, and finally 3) eight map-questions (the subjects were asked to identify the place of events, fill in the names of places or trace the route traveled).

4.5.2. Results and Implications: Table 3 summarizes the results: 1) the feedback from the subjects about the perceived extra workload on a 5-point Likert scale (1 = not burdensome at all, and 5 = definitely burdensome), and 2) the percentage of the three types of the memory test questions that were correctly answered; the text, image, and map questions. There was no obvious difference in terms of the perceived burden among the three groups; the slight difference between two presentation groups came from the perceived interruption according to the interview with the subjects. Regarding comprehension of the book, the subjects who utilized either presentation showed a better performance compared to the control group. This seems obvious; however, taking into account the similar levels of the perceived burden, we argue that both presentations were effective in adding some value while requiring a similar amount of extra work from the readers.

The difference between the groups with respect to the two presentations is relatively large in the “map-question” category. Answering the questions in the category required the subjects to look at the map images carefully in order to construct positional models in their minds. Our lightweight approach of inferring the position of reading sacrifices high precision of inference, and therefore, the information could be presented at unwanted times. This is the issue of the automatic version as pointed out in section 4.2. We consider that the semi-automatic version performed well due to the subjects’ intentional acquisition of the map information, i.e., they needed to confirm the positional relationship of relevant places. We also consider that contents such as an animal or scenery were simple enough to understand at a glance, which might make the correct answer ratio high even in the “automatic group”. On the other hand, members of the semi-automatic version might miss some information due to their disinterest, which might cause them to answer incorrectly; however, we have not confirmed this.

5. Discussion and conclusion

In this article, we have proposed an augmented book cover and a bookmark as means for implicit interaction with a book, referred to as Sentient Books. Implicit interaction is a style of human-computer interaction, in which a user's explicit command is not required to control a system. The augmented items sense their state of use by accelerometers and detect page-flipping events. Accumulation of the events is utilized by an application, the Virtual Illustration System, to estimate the current two-page spread and then to infer a reader's "context". In the application, corresponding multimedia contents are identified by referring to a predefined page-contents mapping table. Thus, a user reads the book in an ordinal way, and at the same time, her context is implicitly captured and given to an application as an input. Here, we assumed that the class of books that are read in a sequential manner, e.g. a novel book or a travelogue, is the target for the application.

The devices were evaluated in the performance of page-flipping event detection and their usability. The detection accuracy of the book cover and the bookmark versions were 88.1% and 92.5%, respectively. The subjects felt that the bookmark version was less obtrusive than the book cover version; they needed to pay greater attention to part of the bookmark. We are planning to investigate a mechanism to correct or adjust the page estimation determined by the user to be applicable to a thicker book that has a large page estimation error. Currently, the Sentient Book device extracts "page-flipping" as a context, which is obtained by combining the sensor readings from two accelerometers. Another application that utilizes the context is, for example, a smart notification system that determines the timing of interruption by taking into account the page currently being read; an interval to the next chapter could be a chance to notify via e-mail. We can also imagine an application in a classroom, where the progress of a problem-solving exercise is implicitly captured to motivate pupils to do more in a competitive manner as well as to help a teacher to dynamically design a class. Furthermore, to make the device a more generic platform for book-based implicit interaction, we intend to identify other context elements for a book.

We initially provided the application with automatic contents presentation. However, the initial user feedback showed us the need for on-demand contents selection to minimize interruption to a reader. So, a new presentation with a thumbnail-based overview and detail view was proposed. A reader can perceive the presence of some information with peripheral vision when it appears, and she can obtain as much of the actual contents as she wants. A proactive change of the screen happens once when a page is flipped, although there may be more than two units of content to present. Additionally, selecting an interesting unit of content from an overview fills the gap between a presented unit of content and the description. User feedback indicates that the new approach was less annoying, and that the perceived burden throughout reading was not so much different from ordinal reading. In the semi-automatic presentation, page-flipping was utilized as a means of pre-selection of possible contents in a particular two-page spread. We consider this to be a practical approach to cope with low resolution and ambiguous contextual information when there are multiple candidates to select.

We also evaluated the effects of virtual illustration; a comparison among the automatic, semi-automatic and "no presentation" (control group) versions was made. The results showed that the semi-automatic version was most effective for understanding complex information, e.g., a map that illustrates positional relationships among relevant places in a travelogue. We recognize that performance strongly depends on many factors: the book itself, the type of the book, the contents, the individuals, etc. Therefore, we will test with other factors to obtain more generic design principles.

Regarding the display device, a PC display or a tablet is not the only choice to provide the Virtual Illustration service. This restricts the location for reading a book. We can utilize a wearable display for "free space reading". Actually, we have informally tested a single eye HMD [27], and we have the impression that the interruption caused by automatic presentation of an image is not as great as is normal for a PC display. We will investigate a suitable interaction method for such a display, which includes presentation for small and low-resolution display areas and selection without a mouse or a stylus.

Acknowledgments

This work was supported by The Ministry of Education, Culture, Sports, Science and Technology in Japan under a Grant-in-Aid for Division of Young Researchers. We would also like to thank the subjects who took part in our evaluation.

References

[1] G.Chen and D.Kotz, "A Survey of Context-Aware Mobile Computing Research", Technical Report TR2000-381, Department of Computer Science, Dartmouth College, 2000.
[2] B.Brumitt, B.Meyers, J.Krumm, A.Kern and S.Shafer, "EasyLiving: Technologies for Intelligent Environments", In Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing (HUC2K), pp. 12-19, 2000.
[3] A.Harter, A.Hopper, P.Steggles, A.Ward, and P.Webster, "The Anatomy of a Context-Aware Application", In Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '99), pp.59-68, 1999.
[4] A.Schmidt, "Implicit Human Computer Interaction Through Context", Personal and Ubiquitous Computing, Vol. 4, No. 2-3, pp. 191-199, Springer London, 2000.
[5] K.Hinckley, J.Pierce, E.Horvitz, and M.Sinclair, "Foreground and Background Interaction with Sensor-Enhanced Mobile Devices", ACM Transaction on Computer-Human Interaction (TOCHI), Vol. 12, No. 1, pp. 1-22, 2005.
[6] K.Fujinami, F.Kawsar and T.Nakajima, "AwareMirror: A Personalized Display using a Mirror", In Proceedings of International Conference on Pervasive Computing (Pervasive2005), LNCS3468, pp. 315-332, 2005.
[7] K.Fujinami and T.Nakajima, "Sentient Artefact: Acquiring User's Context Through Daily Objects", In Proceedings of the 2nd International Symposium on Ubiquitous Intelligence and Smart Worlds (UISW2005), LNCS 3823, pp. 335-344, 2005.
[8] Althi Inc., Image Novel, http://www.althi.co.jp/tat.html (in Japanese, accessed: 5 January, 2009)
[9] CELSYS Inc., Book Surfing, http://www.celsys.co.jp/en/solution/booksurfing (accessed: 5 January, 2009)
[10] E Ink Co., Electronic paper displays, http://www.eink.com (accessed: 5 January, 2009)
[11] J.Watanabe, A.Mochizuki, and Y.Horry, "Bookisheet: Bendable Device for Browsing Content Using the Metaphor of Leafing Through the Pages", In Proceedings of the 10th International Conference on Ubiquitous Computing (Ubicomp2008), pp. 360-369, 2008.
[12] P.Wellner, "Interacting with paper on the DigitalDesk", Communications of the ACM, Vol. 36, No. 7, pp. 87-96, 1993.
[13] H.Koike, Y.Sato, and Y.Kobayashi, "Integrating Paper and Digital Information on EnhancedDesk: A Method for Realtime Finger Tracking on an Augmented Desk System", ACM Transaction on Computer-Human Interaction (TOCHI), Vol. 8, No. 4, pp. 307-322, 2001.
[14] C.S.Wu, S.J.Robinson, and A.Mazalek, "Turning a page on the digital annotation of physical books", In Proc. of the 2nd International Conference on Tangible and Embedded Interaction (TEI'08), pp. 109-116, 2008.
[15] German Research Center for Artificial Intelligence (DFKI), Augmented Text Project, http://www.augmentedtext.net/ (accessed 10 Jan. 2009).
[16] E.Bahna and R.Jacob, "Augmented Reading: Presenting Additional Information Without Penalty", In CHI'05 extended abstracts on Human factors in computing systems, pp. 1909-1912, 2005.

[17] M.J.Back and J.Cohen, "Page detection using embedded tags", In Proc. of the 13th annual ACM symposium on User Interface and Software Technology (UIST2000), 2000.
[18] K.W.May, "Conductive ink based page detection for linking digital and physical pages", In extended abstract of Human Factor in Computing Systems (CHI'01), pp. 275-276, 2001.
[19] M.Billinghurst, H.Kato and I.Poupyrev, "Magic Book: transitioning between reality and virtuality", In extended abstract of Human Factor in Computing Systems (CHI'01), 2001.
[20] H.Koike, Y.Sato, Y.Kobayashi, H.Tobita, and M.Kobayashi, "Interactive Textbook and Interactive Venn diagram: Natural andIntuitive Interfaces on Augmented Desk System", In Proc. of the SIGCHI conference on human factors in computing systems (CHI'00), pp. 121-128, 2000.
[21] T.Ohno and M.Naoki, "A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction", In Proceedings on the 2004 symposium on Eye tracking research & applications (ETRA'04), pp.115-122, 2004.
[22] SwanTouch, http://swanmark.hp.infoseek.co.jp/index.html (in Japanese, accessed: 5 January, 2009)
[23] C.Decker, et al., "DigiClip: Activating Physical Documents" in Proc. of 4th International Workshop on Smart Appliances and Wearable Computing (IWSAWC 2004), pp. 388-393, 2004.
[24] D.Sekulovski, Gijs Geleijnses, B. Kater, J.Korst, S.Pauws, and R.Clout, "Enriching text with images and colored light", Multimedia Content Access: Algorithms and Systems II, Eds. T.Gevers, R.Jain, and S.Santini, Vol. 6820, pp. 4-15, 2008
[25] D. Shamma, B. Pardo, and K.Hammond, "Music Story: a Personalized Music Video Creator", In Proceedings of the 13th annual ACM International Conference on Multimedia, pp. 563-566, 2005.
[26] K.Fujinami and F.Kawsar, "An Experience with Augmenting a Mirror as a Personal Ambient Display", In Proceedings of the 8th Asia-Pacific Conference on Human-Computer Interaction (APCHI 2008), pp. 183-192, 2008.
[27] Scalar Co., Teleglass, http://www.scalar.co.jp/english/products/teleglass.html (accessed: 1 May, 2009)

Authors

Kaori Fujinami is an associate professor in the Computer, Information and Communication Sciences Department at the Tokyo University of Agriculture and Technology. He received his MS in Electrical Engineering and Ph.D. in Computer Science from Waseda University in 1995 and 2005, respectively. His research interests include smart object systems, human-computer interaction and activity recognition.

Nobuhiro Inagawa is a MS student of the Computer, Information and Communication Sciences Department at the Tokyo University of Agriculture and Technology. He received his B.Eng. in 2008. His research interests include image analysis, smart object systems and human-computer interaction.