Digitization of Hand-Written Notes Using a Wearable Camera

Eiji Watanabe, Takashi Ozeki, and Takeshi Kohama

***Abstract*—In taking hand-written note, the taking process plays important roles for recalling related information in addition of notations (words and figures) in the note. In this paper, we focus on the process through taking hand-written note. First, introduce an image processing method for the deduction of the position of pen head by using a wearable type camera. Next, we propose a method for the transformation of the physical position of the pen head into the logical position based on behaviors by students. Here, the logical position denotes the line and character numbers in the hand-written note. Moreover, we detect the face movement by using the block-matching method.**

***Index Terms*—Image processing, hand-written note, noting process, wearable camera.**

I. Introduction

In the progress of digital devices such as tablet type PC, we can easily input information. However, it is important for human to take notes and memorandum in everyday life such as learning, recording and communication. Especially, students have to receive information (characters, speech and images) from teachers and audio-visual devices in the educational opportunity. Moreover, hand-written notes and memorandum play important roles for understanding of contents and recalling related information in lectures. In this case, students have to process information efficiently and write summarized information on their notes. Moreover, Kishi et al. have analyzed the relations between note taking and scores on tests by students and discussed the importance of the process of note taking in lectures [1]. However, we have to consider analog type data (hand-written note) to digitize in consideration of the usage in computers and digital devices efficiently. Therefore, various types of digital pens have been developed and many systems for digital pens have been proposed and discussed [2], [3].

In this paper, we discuss the extraction of the taking process of hand-written note and discuss the digitization of hand-written notes by a wearable camera. Here, we propose an image processing method for the extraction of the process of taking the note by using a sunglasses type camera. Concretely, the position of pen head can be detected by image processing based on images recorded by a sunglasses type camera. Next, we propose an image processing method for the analysis of hand-written note [4]. Authors have been already proposed an image processing method for the detection of the pen head [5]. In this method, the position of the note corner is utilized for the compensation of the position of the pen head. However, it is difficult to get the position of the note corner in real case by a wearable camera. Therefore, in this paper, we transform the physical position of the pen head into the logical position by using the detection of the face movement. The face movement can be detected by the block-matching method. Here, the logical position denotes the line and character numbers in the hand-written note. Finally, we show the experimental results for a few hand-written notes.

Manuscript received May 6, 2014; revised July 12, 2014. This work was supported in part by JSPS KAKENHI Grant Number 25330428.

Eiji Watanabe is with the Faculty of Intelligence and Informatics, Konan University, Kobe, Japan (e-mail: e_wata@konan-u.ac.jp).

Takashi Ozeki is with the Faculty of Engineering, Fukuyama University, Fukuyama, Japan.

Takeshi Kohama is with the School of Biology-Oriented Science and Technology, Kinki University, Kinokawa, Japan.

II. Digitization of Hand-Written Notes

Fig. 1 shows the process for the detection of the position of the pen head and classification of behaviors by using wearable camera. First, the position of the pen head can be detection based on images recorded by a sunglasses type camera as one wearable camera [4]. Next, the behavior by the student can be classified into "writing", "new line" and others by using the position of the pen head. Moreover, we can determine the "logical" position of the pen head by using the above behaviors.

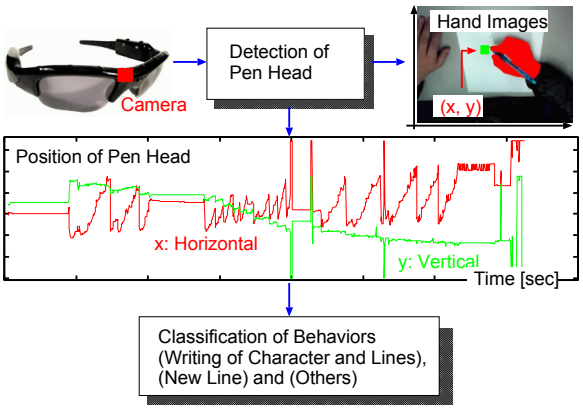


Fig. 1. Detection of the position of the pen head and classification of behaviors by using wearable camera

A. Detection of the Position of the Pen Head

The process of detection of the pen head by image processing is shown in Fig. 2. Roughly speaking, images recorded by the wearable camera consist of four regions (note, pen, finger, shadow and others). First, we extract the four regions based on the color information. Next, the two behaviors (writing and others) can be classified based on the

relation among these regions.

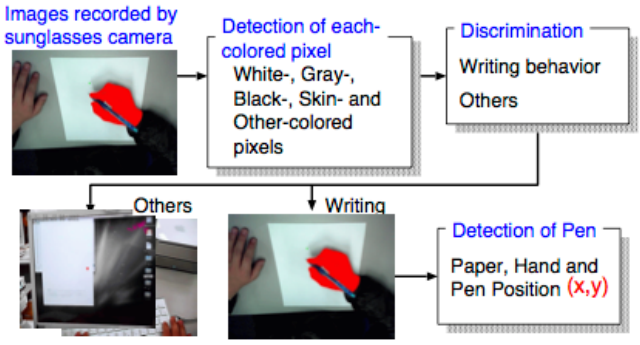


Fig. 2. Detection of the pen head by image processing.

B. Discrimination of Writing and Other Behaviors

We propose a discrimination method between writing and other behaviors based on the above features (the number of each-colored pixels).

First, we extract the four regions (note, pen, finger and shadow) by using the color information. As shown in Fig. 3, the above regions have white-, black-, skin- and gray-colored pixels respectively. Therefore, we adopt the number of each-colored pixels in the four regions as features for the objective image.

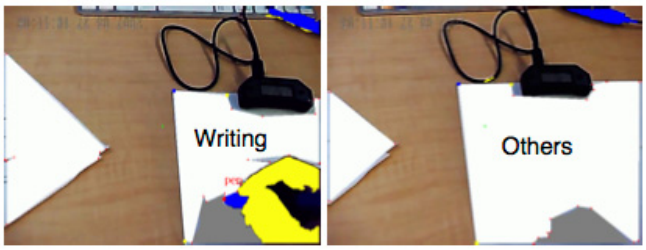


Fig. 3. Discrimination of the two behaviors (writing and other behaviors).

Next, we propose a discrimination method between writing and other behaviors based on the above features. As shown in Fig. 2, the image for writing behavior includes the skin region. On the other hand, the image for other behavior includes no skin region. Therefore, we focus on the difference between the two images for writing and other behaviors and can discriminate between the two behaviors by using the following index:

$$R_{Skin} = \frac{n_{Sk}}{n_B + n_{Sh} + n_{Sk} + n_W + n_O}, \tag{1}$$

where n_B, n_W, n_{Sk}, ns and n_O denote the numbers of black-colored pixels, white-colored pixels, skin-colored pixels, shadow-colored pixels and other-colored pixels respectively. Thus, if $R_{Skin} > \varepsilon_{Skin}$, then the image can be discriminated as writing behavior. Here, ε_{Skin} denotes a threshold.

C. Discrimination of Writing and Other Behaviors

As shown in Fig. 4, the objective images for writing behavior include paper, finger, pen, and shadow regions. First, we select images including regions with black-colored pixels as candidate regions for pen position. We determine the left- upper point (x_{Pen}, y_{Pen}) as the pen head. Next, we calculate the two indices R_{Note} and R_{Shadow} for the region surrounded by the red-colored region in Fig. 4 as follows;

$$R_{Note} = \frac{n_W}{n_B + n_{Sh} + n_{Sk} + n_W + n_O}, \tag{2}$$

$$R_{Shadow} = \frac{n_{Sh}}{n_B + n_{Sh} + n_{Sk} + n_W + n_O}. \tag{3}$$

Since images in taking note include pen head and fingers, we evaluate the indices $0 < \varepsilon_s < \varepsilon_n < \varepsilon_t$ and R_{Shadow}. Therefore, if the indices R_{Note} and R_{Shadow} satisfy the following conditions $R_{Note} > \varepsilon_{Note}, R_{Shadow} > \varepsilon_{Shadow}$ respectively, the position (x_{Pen}, y_{Pen}) can be determined as the pen head.

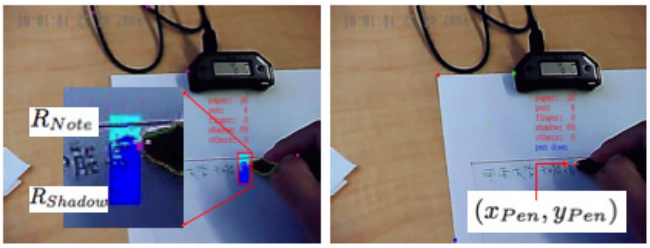
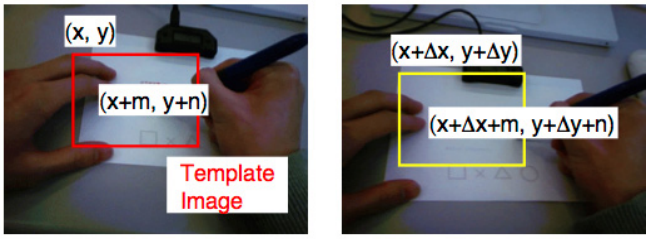


Fig. 4. Detection of the position of the pen head.

D. Detection of Face Movement

When students move their heads, the image recorded by a wearable camera has different viewpoints. If the corner of the note can be detected, the position of the pen head can be compensated by the corner [4]. However, it is difficult to get the position of the note corner in real case by a wearable camera. Therefore, we use the block-matching method for detection of the head movement as shown in Fig. 5.



(a) $(i-1)$-th frame (b) i-th frame

Fig. 5. Detection of the face movement by the block-matching method.

In Fig. 5 (a), the red-colored rectangle shows ROI (region of interest) in the template image (the (i-1)-th frame). By reference to the template image (the red-colored region) in Fig. 5 (b), the position of the template image in the i-th frame can be detected. Concretely, the difference $D(\Delta_x, \Delta_y)$ between the template image $f_{i-1}(x+m, y+n)$ and the i-th image $f_i(x+\Delta_x+m, y+\Delta_y+n)$ can be calculated by Eq. (4).

$$D(\Delta_x, \Delta_y) = \sum_m \sum_n (f_i(x+\Delta_x+m, y+\Delta_y+n) - f_{i-1}(x+m, y+n))^2, \tag{4}$$

where (x, y) denotes the left-upper position of the template image in the i-th frame and Δ_x and Δ_y denote the movement quantity of the template image. Moreover, (m, n) denotes the coordinate in the template image. By searching the position (Δ_x, Δ_y) so as to minimize the difference $D(\Delta_x, \Delta_y)$, the movement of the object can be detected. Here, we define the movement quantity as $(\Delta_x(i), \Delta_y(i))$. Furthermore, we define the accumulated movement quantity $T_x(i)$ and $T_y(i)$ as the face movement as follows;

$$T_x(i) = \sum_{j=1}^{i} \Delta_x(j), \quad T_y(i) = \sum_{j=1}^{i} \Delta_y(j)$$

III. Classification of Behaviors in Taking Note

In this section, we describe a method for classification of behaviors (writing characters, new-line and others).

A. References

Fig. 6 shows the horizontal and vertical positions of the pen head and sentences on note. If sentences are written from left to right, the horizontal position of pen head changes from to left similarly. On the other hand, the vertical position of pen head changes from top to bottom. Moreover, in beginning on a new line, the vertical position changes rapidly. As shown in Fig. 6, the position of the pen head has sharp and dull fluctuations according to writing behaviors. Here, we approximate the changes of the horizontal and vertical positions of the pen head by using piecewise lines.

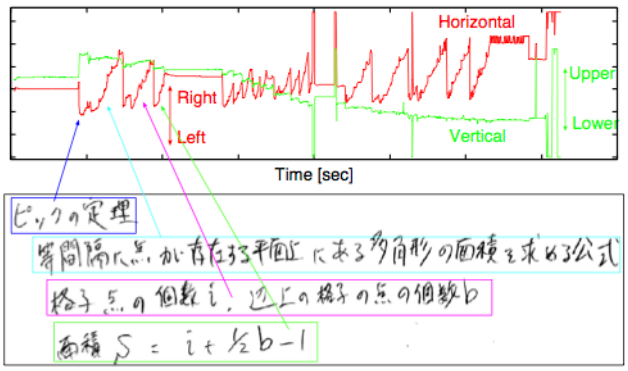


Fig. 6. Changes of the position of the pen head and sentences on note.

Fig. 7 shows the outline for the approximation of the position of the pen head. First, we define the position of pen head as $P_i(x_i, y_i)$ in the objective section. Here, x_i and y_i denote the frame number and the horizontal position of the pen head respectively. Also, we defined the both ends of the objective section as $P_s(x_s, y_s)$ and $P_e(x_e, y_e)$. The line ℓ_T between two points P_s and P_e can be represented by

$$\ell_T : y = m_T x + n_T, \tag{5}$$

where $n_R = y_s - m_T x_s$ and $m_T = \dfrac{y_e - y_s}{x_e - x_s}$, $x_s \neq x_e$.

The perpendicular distance d_i from the position $P_i(x_i, y_i)$ of pen head to the line ℓ_T is given by

$$d_i = \frac{|m_T x_i - y_i + n_T|}{\sqrt{m_T^2 + 1}}. \tag{6}$$

Next, the maximum $d_{\max}$ of the perpendicular distance is defined by $d_{\max} = \max_i d_i$. If $d_{\max} > \varepsilon_d$, then the line l_t is divided on the position $P_i(x_i, y_i)$ of pen head having the maximum perpendicular distance. On the other hand, if $d_{\max} \leq \varepsilon_d$, then the line l_t is adopted as an approximated line between the two points $P_s(x_s, y_s)$ and $P_e(x_e, y_e)$.

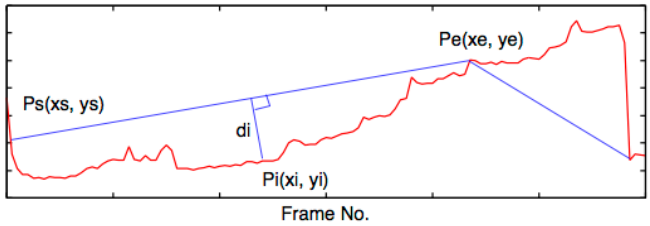


Fig. 7. Approximation of the horizontal position of the pen head by piecewise lines.

B. Classification of Behaviors in Taking Note

Based on the slope m_T of the approximated line, the behaviors in taking note can be classified into the four behaviors "stopping", "writing character and figures" and "writing a new line" and "others" [6].

1) *Stopping behavior*: If the slope m_T is satisfied with the following condition, it means that the position of pen head do not move within the given time. Therefore, the behavior within the given time can be classified as "stopping" behavior.

$$-\varepsilon_S \leq m_T \leq \varepsilon_S, \tag{7}$$

where ε_S is a threshold for the determination of "stopping" behavior.

2) *Writing behavior*: In writing behaviors, the horizontal coordinates x_i of the position $P_i(x_i, y_i)$ of pen head increases according to the frame no i. Thus, when the horizontal position x_i of the position satisfies the following condition, the objective behavior belongs to the writing one.

$$\varepsilon_S < m_T \leq \varepsilon_t, \tag{8}$$

where ε_t is a threshold for the determination of "writing" behavior.

3) *New line*: In the beginning on a new line, the horizontal coordinates x_i of the position $P_i(x_i, y_i)$ return to the beginning of a line. Thus, when the horizontal position x_i of the position satisfies the following condition, the objective behavior belongs to the behavior for the beginning on a new line.

$$m_T \pounds e_{NL}, \tag{9}$$

where ε_{NL} is a threshold for the determination of the

behavior for the beginning of a new line.

4) *Other behaviors*: As other behaviors, we can enumerate the following behaviors; manipulating the keyboard, erasing the note and looking the display. In these cases, the sunglasses type camera can not detect the pen and the fingers of the writer. Thus, these behaviors belong to "other" behaviors. Here, the above threshold values have the relation $0 < \varepsilon_s < \varepsilon_n < \varepsilon_t$.

IV. Experimental Results

We used a sunglasses type camera as shown in Fig. 8 (a) as one wearable camera. This camera (resolution: 640×480 [pixel], frame rate: 30 [fps]) has CCD sensor fixed on the center of bridge. Moreover, Fig. 8 (b) shows a sample of hand-written note and the time for writing is about 60 [sec]. Here, the number of students 3.

(a) Wearable camera

(b) Hand-written note

Fig. 8. Sample of hand-written note and wearable camera.

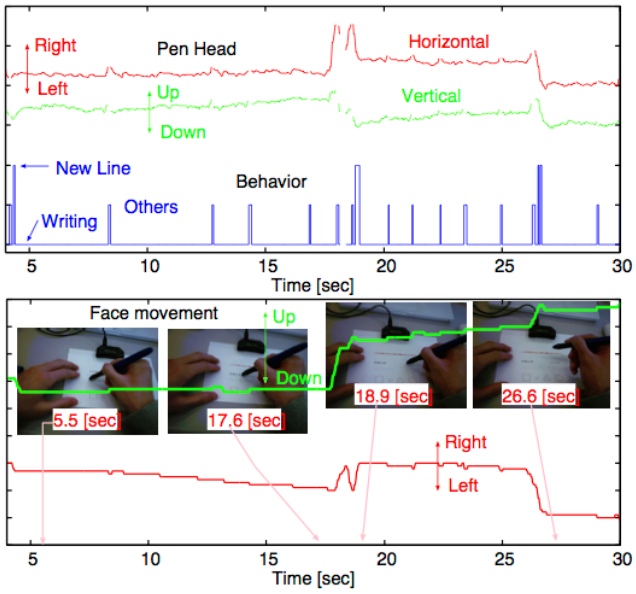


Fig. 9. Pen position and face movement detected by a wearable camera.

Fig. 9 shows the position of the pen head and the face movement detected by the proposed method. In the pen position, blue-colored lines show the behaviors ("writing", "new line" and "others") classified by the proposed method. From Fig. 9, we can summarize the transformation of the physical position of the pen head into the logical position as follows;

- In the range of 5 to 18 [sec], the number of other behaviors is 4 and these are generated by the pen-up behavior. Therefore, the first string consists of four Japanese characters and the above other behaviors correspond to the writing behavior for four Japanese characters.
- Similarly, in the range of 19 to 26 [sec], the number of other behaviors is 6 and these are generated by the pen-up behavior. Therefore, the second string consists of 6 numerals (012345) and the above other behaviors correspond to the writing behavior for six numerals.
- Moreover, in the range of 19 to 26 [sec], the vertical position of the pen head is lower than the first string. This is the cause that the vertical position of the face is higher than the first string as shown in Fig. 9.

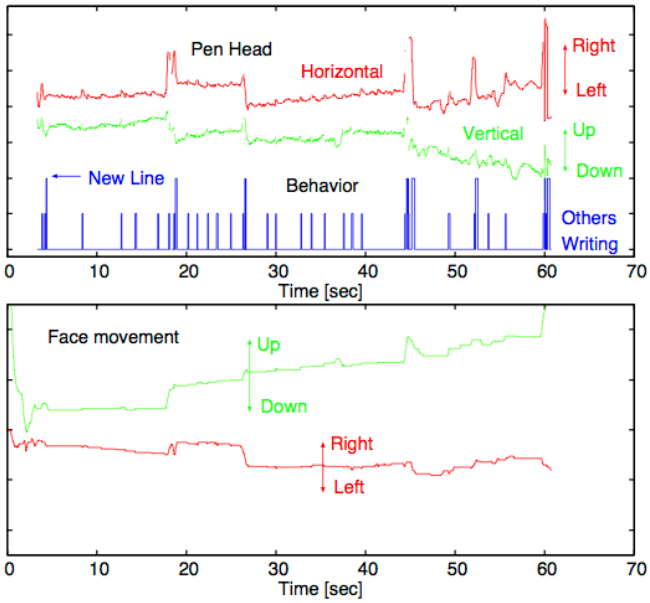


Fig. 10. Pen position and face movement detected by wearable camera for Student-A.

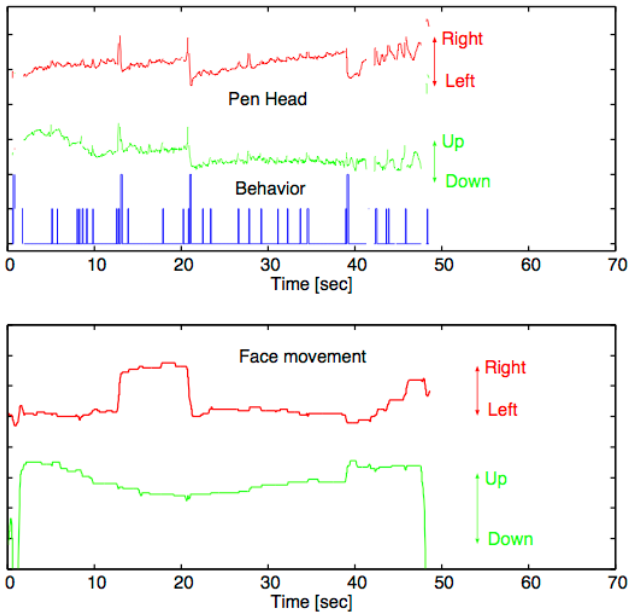


Fig. 11. Pen position and face movement detected by wearable camera for Student-B.

Fig. 10 shows the position of the pen head and the face movement detected by a wearable camera for Student-A. By focusing on the horizontal position of the pen head, we can see that four strings are written for Student-A. Moreover, the face movement in horizontal is changing according to the new line. However, the face movement in vertical is not detected correctly. We have to compensate the pen position and the face position in vertical.

Furthermore, Fig. 11 and Fig. 12 show the position of the pen head and the face movement detected by a wearable camera for Student-B and Student-C respectively. In Fig. 11, we can see that an important behavior (three times of "new line") and the logical process in taking a note can be extracted. On the other hand, in Fig. 12, we can see that excessive behaviors concerning on "new line" are occurred. This is the cause that the pen movement for Student-C is larger compared with Student-B.

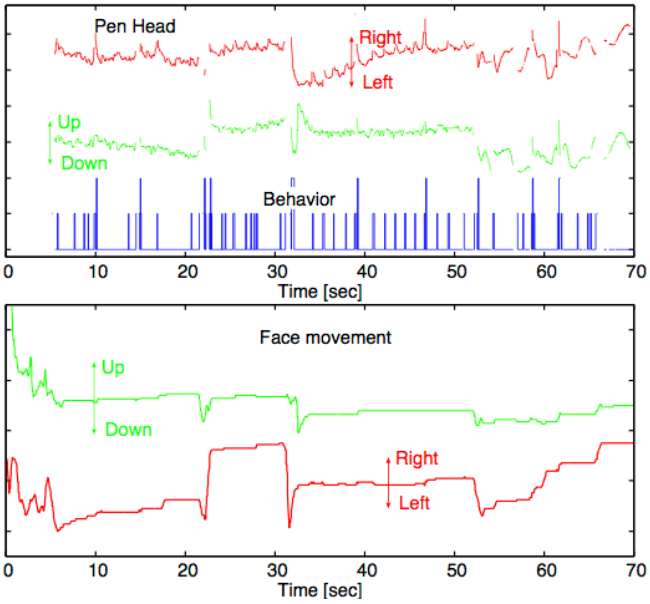


Fig. 12. Pen position and face movement detected by wearable camera for Student-C.

V. Conclusions

In this paper, we have extracted the taking process of hand-written note and discuss the digitization of hand-written notes by a wearable camera. Here, we have proposed an image processing method for the extraction of the position of pen head and the face movement by using the block-matching method. From experimental results, we have obtained the following remarks;

- The logical process in taking a hand-written note can be extracted by the proposed method.
- It is efficient for the compensation of the pen head to utilize the face movement. However, we have some problems concerning on the accuracy of the compensation of the position of the pen head. As future works, we would like to discuss the analysis of the structure of note based on the relations among character regions [7] and the recognition for hand-written characters [8].

References

[1] T. Kishi, H. Tsukada, and E. Nojima, "Relation analysis of presence of note-taking and test score after the lecture," *Educational Technology Research*, vol. 28, pp. 265–268, 2004.

[2] K. Ito, S. Uchida, I. Masakazu, S. Omachi, and K. Kise, "Recovering handwritings via pen-tip camera," *IEICE Tech. Rep.*, vol. 108, no. 432, pp. 19-24, 2009.

[3] T. Kazuhiro, U. Seichi, I. Masakazu, O. Shinichiro, and K. Koichi, "Stroke recovery for data embeddment and extraction into handwriting pattern," *IEICE Tech. Rep.*, vol. 105, no. 615, pp. 61-66, 2006.

[4] E. Watanabe, "A study on analysis of handwritten notes," *IEICE Tech. Rep.*, vol. 108, no. 432, pp. 13-18, 2009.

[5] E. Watanabe, T. Ozeki, and T. Kohama, "Embedding time tag in hand-written memorandum by using sunglasses type camera," in *Proc. IWAIT 2012*, in CD-ROM, p. 5, 2012.

[6] E. Watanabe, "Analysis of movements of students in e-learning — Analysis based on movements in taking notes," *IEICE Tech. Rep.*, vol. 109, no. 39, pp. 1-6, 2009.

[7] S. Marinai, M. Gori, and G. Soda, "Artificial neural networks for document analysis and recognition," *IEEE Trans. on PAMI*, vol. 27, no. 1, pp. 23–35, 2005.

[8] K. Kise, S. Omachi, S. Uchida, and M. Iwamura, "Current status and future prospects of camera-based character recognition and document image analysis," *IEICE Tech. Rep.*, vol. 104, no. 742, pp. 85-90, 2005.

Eiji Watanabe received the B.E. and M.E. degrees from Tokushima University, Japan, in 1983 and 1985 respectively. He received the Ph.D. from Osaka Prefecture University, Japan, 1997. He is now a professor of the Faculty of Intelligence and Informatics, Konan University, Japan. His research interests include learning algorithms for multi-layered neural networks and their applications to image processing. He is a member of IEICE and IEEE.

Takashi Ozeki received the B.S. degree from Kyoto University in 1988. During 1988-1992, he worked for Brother Corporation. He received M. and Dr. Info. Sci. from JAIST in 1994 and 1997, respectively. He is now a professor of the Faculty of Engineering, Fukuyama University. His research interests include image processing, optimization theory and numerical analysis. He is a member of IEICE, ITE and SIAM.

Takeshi Kohama received the M.E. in knowledge-based information engineering and the Ph.D. in system information engineering from Toyohashi University of Technology, Toyohashi, Japan, in 1997. Since 2010, he is in the Department of Computational Systems Biology, Faculty of Biology-Oriented Science and Technology, Kinki University. His research interests include human vision system, visual perception, and neuroinfor- matics of vision. Dr. Kohama is a member of the JNSS, the JNNS and the ITE in Japan.