International Journal of Engineering Education Vol. 29, No. 3, pp. 568–577, 2013
Printed in Great Britain

0949-149X/91 $3.00+0.00

Smart E-learning: Enhancement of Human-Computer Interactions Using Head Posture Images*

YÜCEL UĞURLU
Department of Integrated Information Technology, Aoyama Gakuin University, Sagamihara City, Kanagawa 252-5258, Japan.
E-mail: yucel.ugurlu@ni.com

This paper proposes a novel e-learning system that incorporates human-computer interaction data to build a smart e-learning system. A supervised image segmentation algorithm is used to detect the face and hair of students in head posture images. A simple and effective human presence detection and gaze direction estimation method is then developed based on changes in the face and hair information. First, the proposed algorithm is tested using 10 different students with seven different head postures each and 92% of the head postures are identified accurately. Second, the method is applied to real time video sequences containing 80 frames that lasted 400 seconds, which are acquired using an integrated web camera, and similar results are obtained. Finally, human-computer interaction data, which is an indicator of student attention, is calculated based on the human presence and gaze direction over time. The experimental results show that the proposed approach enhances human-computer interactions for e-learning systems and helps us to evaluate student performance.

Keywords: engineering education; e-learning; human-computer interaction; machine vision

1. Introduction

Teaching and learning methods have changed significantly due to recent trends in computer technology. E-learning is one of the most popular approaches because of its tremendous potential and future engineering education trends. E-learning is defined as learning and teaching online via network technologies, and in the last decade, it has emerged as one of the most powerful approaches for addressing the growing needs for education [1, 2]. At present, many classes are supplemented with virtual labs or e-learning environments. The biggest advantages of e-learning systems over traditional classes are their on-demand functionality, because they can be accessed anytime or anywhere, and their rich content such as videos, graphics and texts, and their fully personalized structures. The e-learning approach is learner-centred, and its design involves a system that is interactive, self-paced, repetitious, and customizable [3]. The educational advantages of this approach are manifold. For instance, students might become more engaged in their learning because they make active choices as they navigate through the material, while they can move at their own speed because they work with personalized content.

On the other hand, current e-learning systems are far from perfect. It is clear that e-learning differs from conventional educational, mainly because e-learning separates teachers from students and students from students [4]. Unlike conventional education, this quasi-permanent separation means that teachers and students are unable to engage in face-to-face communication during e-learning, which results in an absence of interaction. Some of the biggest challenges of e-learning systems are measuring student responses, adaptively changing the teaching speed or content, and the lack of student emotion because of the limits of human-computer interactions [5]. Therefore, the next generation of e-learning should consider human interactions. As a result, it is very important that the learning styles of individuals and cultural differences should be understood to develop better e-learning systems [6].

Efficient human-computer interactions must be designed to make the use of e-learning environments more effective [7]. Three main types of user interfaces are used: sensor-based, visual-based, and audio-based. In most e-learning systems, human-computer interaction is very primitive and mainly limited to the use of a mouse or keyboard to progress to the next topic or to answer simple quizzes. An important requirement for the new generation of interfaces is the differentiation between using intelligence to make the interface and the way that the interface interacts with users. Thus, intelligent human-computer interaction designs require interfaces that incorporate some form of intelligence in the perception of users and the responses by them [8, 9]. Examples include speech-enabled interfaces that use natural language to interact with users and devices that visually track the user movements or gaze to respond accordingly [10–12].

In the last few years, we have developed an e-learning system for teaching graphical programming in higher education [13, 14]. However, the

* Accepted 15 January 2013.

system is based on slides and programming videos, which only requires mouse and keyboard interactions. Therefore, a new approach is proposed in this study for understanding and evaluating student interactions during the learning process. Mouse and keyboard entries provide very limited information about user interactivity, which makes personal assessments difficult. In contrast, visual information provides continuous data about individuals, which can be utilized to improve the learning content or build smart e-learning environments.

In this study, we propose a smart e-learning system that incorporates head posture detection to enhance the human-computer interaction level with students. This remainder of this paper is organized as follows. In Section 2, the system outline is presented and the proposed approach is explained. Subsequently, this method is applied to a head posture database and video data in Section 3. The paper closes with some concluding remarks and implications for future work.

2. Vision-based human-computer interactions

Human-computer interaction is a crucial component of e-learning systems that affects the efficiency and quality of their usage and provides means of communication between the user and the virtual environment [15]. At present, many e-learning systems have been developed with very advanced graphical user interfaces, although the role of the human-computer interaction remains at classical levels. Most current human-computer interaction methods are based on mouse clicks and keyboard entries. However, e-learning systems should be smart enough to incorporate the behaviour of students and adapt to the learning styles of individual students. Thus, e-learning systems ensure high standards of accessibility and usability by making learner interactions with virtual systems as natural and intuitive as possible. Various sensors, cameras, and innovative human-computer interface approaches are required to understand the behaviour of learners [16, 17].

Visual-based human-computer interaction, which includes facial expression analysis, body movement tracking, gesture recognition, and gaze detection, is probably the most widespread approach used in research. The principles of detecting eyes, noses and lips and using their geometric combinations are well established for understanding facial expressions [18, 19]. Component-based approaches have produced better results than global approaches because the individual components vary little whereas the variations related to pose changes are mainly geometric. Most previous component-based methods have used both grey and colour images to recognize faces [20, 21]. Consequently, most of these methods are computationally expensive and some can only deal with frontal faces with few variations in their size and orientation. Recently, a robust component-based face detection algorithm was proposed with fewer computational demands [22].

The current research replicates the classroom environment in virtual space by including visual checks. For example, individuals look at the PC screen while learning is in progress and they interact appropriately, while paying attention to the e-learning content. However, indications that no learning is taking place include looking away from the PC, falling asleep, or being outside the camera's field of view, which all suggest disinterest.

On the other hand, typical face and hair recognition approaches are insufficient for understanding a person's level of interaction with a computer [23]. Therefore, in this research, a head posture detection algorithm is configured in several steps: image segmentation, human presence detection, and gaze direction estimation. It is assumed that the lighting conditions and background textures do not change significantly during the learning period. An overview of the vision-enabled human-computer interactive e-learning system is shown in Fig. 1.

2.1 Colour image segmentation

In computer vision, segmentation is the process that partitions a digital image into multiple segments. The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyse. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Several

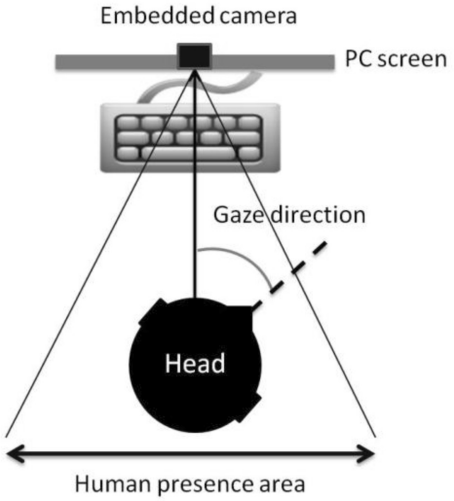

Fig. 1. Vision-enabled human-computer interactive e-learning system (top view).

general-purpose algorithms and techniques have been developed for image segmentation, such as thresholding, clustering, histogram-based methods and split-and-merge methods [24, 25]. There is no general solution to the image segmentation problem, so these techniques often have to be combined with domain knowledge to solve an image segmentation problem effectively.

In this study, image segmentation is based on a supervised training and classification process using colour images. In this approach, the classification function is learned from the training images. Colour segmentation compares the colour features of each pixel with those of the surrounding pixels or a trained colour classifier before segmenting an image into colour regions.

First, the RGB colour space is transformed into the HSL (Hue, Saturation, Lightness) colour space. The RGB colour space is the most common colour space, but R, G, and B are dependent on the illumination [26]. Thus, hair and face detection may be unsuccessful in the RGB colour space when the illumination conditions change. In the HSL colour space, the hue is generally related to the wavelength of the light, so it allows significant discrimination between face and hair regions. The conversion between the RGB and HSL colour spaces is conducted as follows:

$$H = \cos^{-1}\left[\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right], G \geq B \tag{1}$$

If $B > G$, then $H = 30 - H$, since H is an angle in degrees.

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)] \tag{2}$$

$$L = \frac{1}{2}\max(R,G,B) + \frac{1}{2}\min(R,G,B) \tag{3}$$

During the feature extraction and training phase, known samples are provided manually to develop the rules. A know sample consists of a region in the image that contains the colour you want the classifier to learn and a label for the colour. For every sample that is added during the training phase, the colour classifier calculates a colour feature and assigns the associated class label to the feature. Eventually, all the trained samples (colour feature with the label) added to the classifier are saved in a file that represents a trained colour classifier. The training process consists of the following steps:

(1) Identification of the categories/classes to be used.
(2) Selection and definition of the training data.
(3) Choosing the statistical classifier and conducting classification.
(4) Developing the result display.
(5) Evaluating the classification performance.

Four different pattern categories are defined for the head posture images: the background, clothing, face, and hair. Several patterns are manually selected to represent each category from the head posture images. The training patterns are manually selected to represent the full characteristics of the head images because the colour variations differ, even in the same category.

After the training process, the colour samples can be classified into their corresponding classes for colour identification. During the classification phase, the classification engine calculates the colour feature of the sample that we want to identify and classifies it according to the trained sample using one of the existing classification algorithms. Among the various supervised statistical pattern recognition methods, the nearest neighbour rule achieves consistently high performance; hence, a new sample is classified by calculating the distance to the nearest training case. Finally, a new greyscale image is generated based on the classification results, which is used for further image analysis.

The colour classifier uses the HSL colour space to calculate the colour feature for each sample that needs to be trained or classified. The colour feature represents the three-dimensional colour information of the sample in a one-dimensional format. The colour classifier calculates the colour features according to the following steps.

(1) Convert the colour sample into the HSL colour space.
(2) Calculate the hue, saturation, and lightness histograms for the colour sample. The hue and saturation histograms contain 256 pixel levels each.
(3) Reduce the luminance histogram to eight different levels, which are suppressed by 12.5%. The National Instruments (NI) Colour Classifier accentuates the colour information in the sample by suppressing the luminance histogram.
(4) Combine the hue, saturation and luminance values, which include 520 (256 + 256 + 8) components, to produce a high resolution colour feature.
(5) Obtain medium and low-resolution colour features by applying a dynamic mask to the high-resolution colour feature. The medium- and low-resolution colour features are subsets of the high-resolution colour feature.

The NI Colour Classifier supports the entire closed region of interests; therefore, any closed shape of the colour regions can be used. Additionally, no pre-processing algorithms are associated with the NI Colour Classifier.

2.2 Human presence detection

Detecting human presence is comparatively simple with e-learning systems because a specific amount of face information is sufficient for detection. Presence detection is defined here as the human presence detected by imaging systems, which is an important step in determining the learner's interaction level with a computer. The image segmentation based approach is very fast and flexible because it is independent of the location of facial features such as lip movements and closed eyes.

Face and hair pixel information are used to identify the human presence level, as follows:

$$P_i = \frac{(f_i + h_i)}{(f_0 + h_0)} 100 \tag{4}$$

where f_0 and h_0 are the total number of face and hair pixels in the initial view, respectively, and f_i and h_i represent images in different head postures. In the initial view, $f_i = f_0$ and P_i equals 100.

These values are affected whenever a learner moves their head, such as by changing the distance or moving outside the camera's field of view. For example, P_i is zero when no-one is present. However, the presence rate changes significantly with partial or near/distant views.

2.3 Gaze direction estimation

Gaze direction estimation is another criterion used to evaluate human-computer interaction. Gaze detection is generally an indirect form of interaction between the user and the machine, which is used mainly to understand the user's attention or intention in computer vision. If the learner's gaze direction changes, the hair-to-face ratios are also affected in the segmented images. The following equation is used to estimate the gaze direction:

$$D_i = \left| 100 - \frac{\frac{h_i}{f_i + h_i}}{\frac{h_0}{f_0 + h_0}} 100 \right| \frac{P_i}{P_0} m \tag{5}$$

where m is the multiplier, which is calculated by

$$m = \frac{h_0}{I} n \tag{6}$$

where n is a gender constant that is 10 for women and 20 for men, and I is the total number of image pixels. Human gender is determined automatically by testing the hair to face pixel ratio in the initial image. If the hair to face ratio is greater than 0.7, it is assumed that the person is a women. In this study, all of the constants were obtained experimentally using a face recognition database.

The initial value of D_0 is equal to 0, i.e., the learner is looking directly at the computer screen. If the learner's face and hair ratio does not change, the value of D_i remains close to zero. If the learner looks in a different direction, however, this ratio changes significantly. More hair information is available in most situations, especially in side or back views of subjects. Obviously, there are personal hair visibility differences that depend on the gender and hairstyle, although there are some exceptions such as people with no hair.

2.4 Human-computer interaction check

To identify the student interaction with a computer and an e-learning system, a decision-making algorithm is developed based on the human presence ratio and the gaze direction estimate.

The 'if' case

$$R_i = \begin{cases} True, & P_i \geq 40 \text{ and } D_i \leq 10 \\ False, & otherwise \end{cases} \tag{7}$$

is used.

Experimental results show that the presence ratio can be changed if the learner moves closer or outside the camera's field of view during the learning period. Therefore, a presence change of up to 40% is considered acceptable; it implies that the person exists and that they are interacting appropriately with the computer. However, the gaze direction algorithm also has an error range. Therefore, to be safe, a directional change of more than 10° is considered to be a sign of human-computer interaction failure.

After applying this decision-making algorithm, *True* implies that a student is engaging in a human-computer interaction and that learning is in progress. However, *False* shows that learning is not occurring because the learner has a very weak computer interaction or they have lost interest due to the presence ratio and the gaze direction failure. Finally, the *True* and *False* information can be easily integrated with e-learning systems to improve the content, evaluate individual performance, or analyse a student's interest levels.

3. Experimental results

3.1 Application to a face recognition database

A series of experiments were conducted using the MIT-CBCL face recognition database based on the algorithms introduced above [27]. The dimensions

of the images are 768 × 576 pixels and they comprise images of three women and seven men taken from different angles. Some of the images were regenerated from the original view; subsequently, seven image files were acquired for each person: the initial view, distant view, near view, slight right, full right, slight left, and full left head postures.

The NI Vision Assistant is used for image segmentation because of its user-friendly and interactive environment. A *Colour Segmentation* function is available for training and classification in the NI Vision Assistant. First, several patterns were carefully selected from the initial view to represent the background, clothing, face, and hair. The pattern sizes of the regions were not identical because of the manual operation of the PC mouse, although they were generally approximately 30 × 30 pixels. The classification results were confirmed visually based on the segmentation images. Some image segmentation failures occurred due to the similar colour of the learner's hair and clothing, so additional samples were included in the training process. Fig. 2 shows one of the head posture images and its image segmentation results. Four different grey levels are displayed in the images, which represent the face, hair, clothing and background textures.

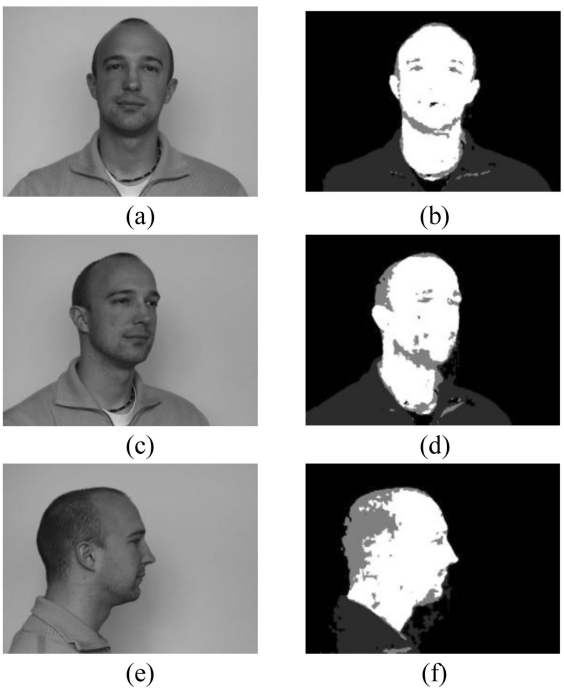

Fig. 2. Head images and segmentation results: (a) initial view, (c) slight right view, (e) full right view, whereas (b), (d), and (f) are segmentation results for (a), (c), and (e), respectively.

Second, the NI Vision Assistant segmentation script was imported into the LabVIEW environment for further head posture analysis. LabVIEW is a graphical programming environment that includes various image and signal processing functions. Some important LabVIEW steps include calculating the pixel numbers for the face and hair regions and adding some mathematical and logical calculations to represent the human-computer interaction.

Equations (4) and (5) were used to calculate the human presence level and to estimate the gaze direction for seven head postures. Finally, the results were evaluated using Equation (7). Table 1 shows the analysis results for the person shown in Fig. 2, where P_0 and D_0 were 100% and 0° for the initial view, respectively. However, the presence value and gaze directions changed accordingly when the head postures changed. The initial view, distant view, and near view postures indicated that the students were interacting appropriately with the computer. However, a gaze direction greater than 10° indicated interaction failure. P_i and D_i values are also plotted in Fig. 3.

Finally, the same image segmentation and analysis approach was applied to seven men and three women, each of whom had seven different head postures. Each person had a unique hairstyle and skin tone, and they wore unique clothing, as shown in Fig. 4. The initial images were used for training purposes. Thus, 60 head posture images were evaluated using the proposed algorithm.

The initial head posture of each student was used for image segmentation training purposes. The remaining six images (near, far, slight right, full right, slight left, and full left) were used in the final evaluation process. The results showed that only 8% were misidentified, mainly due to similarities in the hair and clothing colour. However, 92% of the head postures were correctly identified. The results are summarized for all students in Table 2. The pro-

Table 1. Head posture analysis for the seven different situations.

	Face pixels	Hair pixels	Human presence ratio (%)	Gaze direction estimate (degrees)	Human computer interaction
Initial view	82195	12573	100	0	True
Distant view	52604	8870	65	3	True
Near view	114581	15179	137	9	True
Slight right	76530	18794	101	28	False
Full right	75859	30645	112	75	False
Slight left	79889	18390	104	24	False
Full left	79245	31791	117	77	False

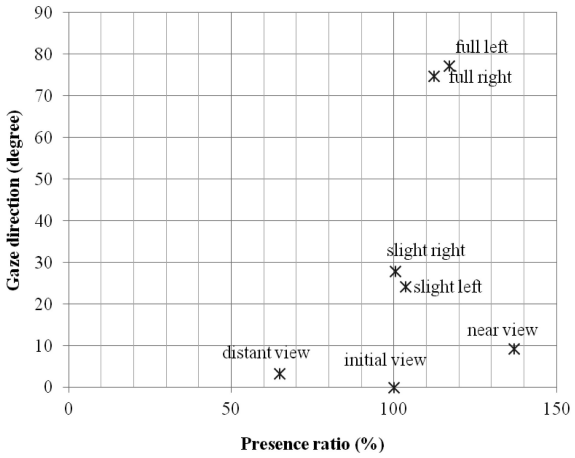

Fig. 3. Analysis results for seven different head posture images.

posed algorithm was effective with specific human types and races, especially when the skin and hair colours were different. However, the failure rate might increase if the person does not have hair, if they were working in an environment with a very complex background, or if they had very similar skin and hair colours such as Africans and Indians.

The human presence detection algorithm performed robustly for both near and distant views of the ten students, as can be inferred from Table 2 and Fig. 5. The calculated values for the far and near views were very close to the average values; therefore, a standard deviation of less than 8% was ensured. However, the gaze direction estimation was more sensitive to personal differences such as hairstyle, gender, or lighting conditions. Therefore,

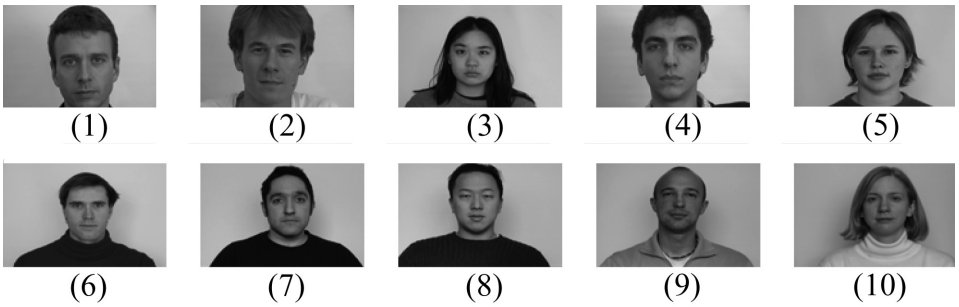

Fig. 4. MIT face recognition database images.

Table 2. Comparison of the average values for the actual and estimated head posture images of all students

	Distant view (presence ratio)	Near view (presence ratio)	Slight turn view (gaze direction)	Full turn view (gaze direction)
Actual head posture value	60%	130%	30°	90°
Estimated head posture value	64%	134%	28°	81°
Standard deviation	1.7%	7.3%	12.8°	16.8°

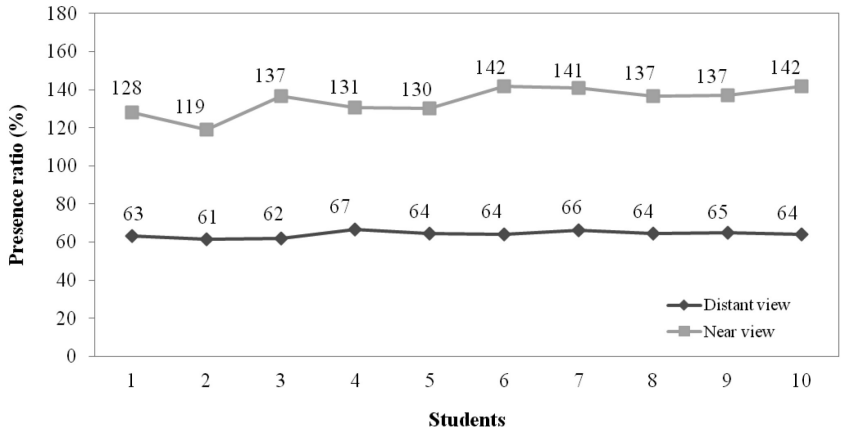

Fig. 5. Presence ratio changes for all students.

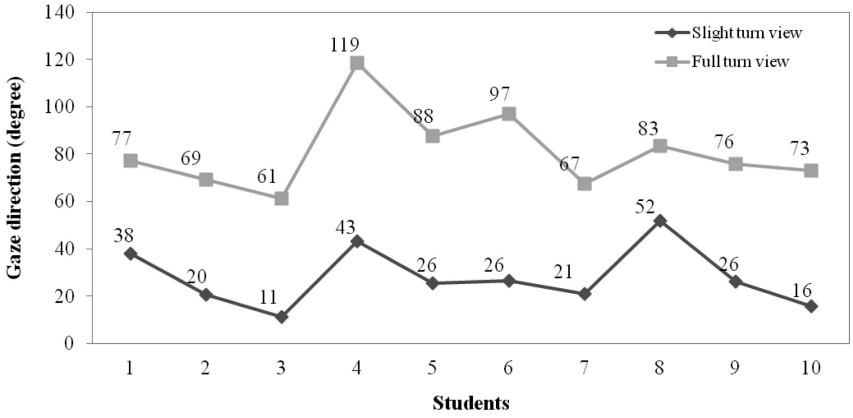

Fig. 6. Estimated gaze direction changes for all students.

the gaze direction values fluctuated more throughout the student face images, as shown in Table 2 and Fig. 6. Despite these errors, the calculated average slight right/left view was 28° and the calculated full right/turn view was 81°, which are very close to the actual values. In addition, the average values for the presence ratio and gaze direction estimates were close to the actual head positions, although a limited number of head posture images were used.

The proposed approach has two main advantages over traditional face and gaze detection techniques. (i) it is independent from facial features and locations; therefore, it is more robustness against eye, lip, and mouth movements and facial occlusions; (ii) the algorithm is based on image segmentation; hence, no intensive mathematical computations are required to identify the human presence or to estimate the gaze direction. However, there will always be exceptions and challenges during wider application of the proposed algorithm.

3.2 Application to video sequence

At present, most personal computers have integrated cameras for video conferencing or chatting. Using our system, we acquire the video data from an embedded PC camera that captures 1280 × 720 pixels and we apply head posture analysis to understand the human-computer interaction with an e-learning system. The PC camera is very easy to use but the video quality is not perfect. Therefore, the video frames were pre-processed using a median filter before colour image segmentation. The median filter is a nonlinear digital filtering technique, often used to remove noise. Using this method, a significant amount of noise was eliminated and the video frames were smoothed.

Feature extraction and training is implemented using NI Vision Assistant and the classification files are obtained for the initial view, as described in Section 2.1. If the segmentation results were not satisfactory, we increased the number of training

Fig. 7. Front panel of LabVIEW when analysing video frames.

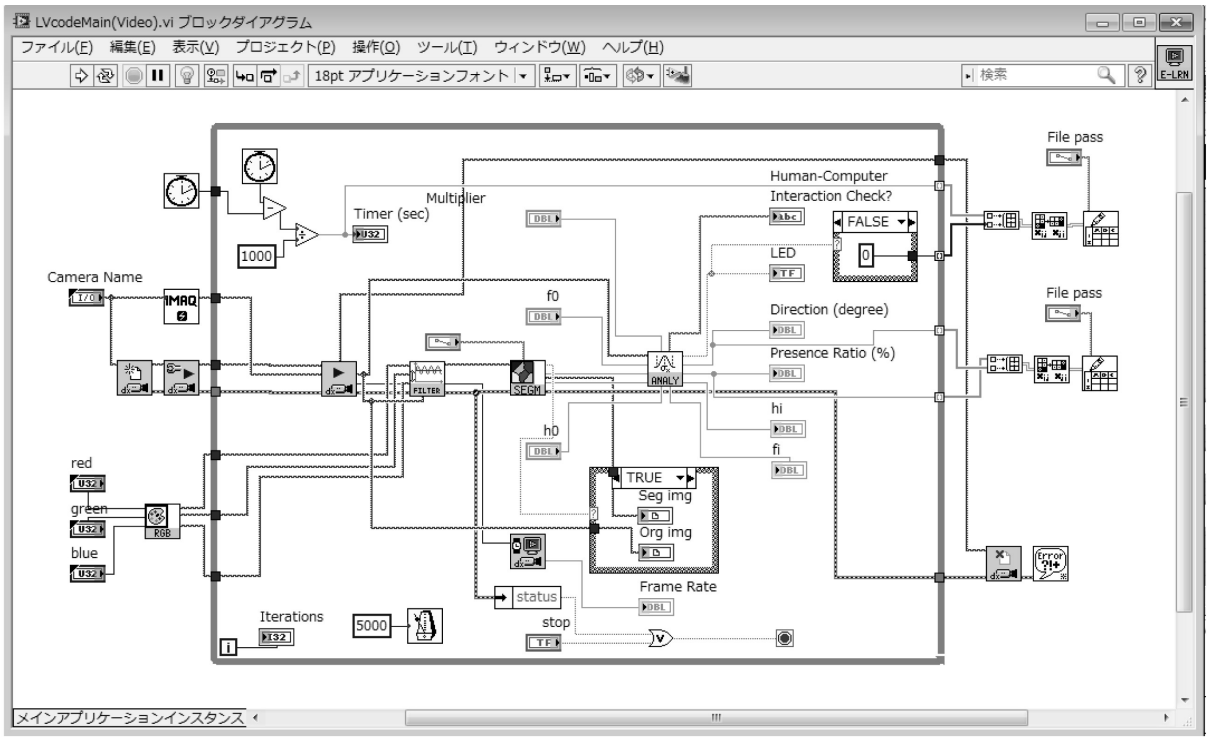

Fig. 8. Block diagram of the main LabVIEW VI code.

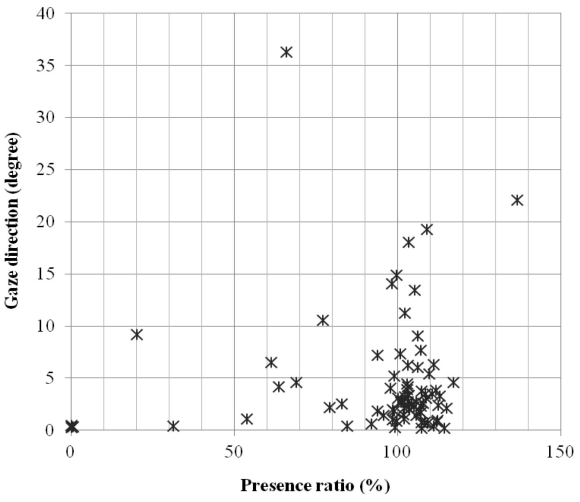

Fig. 9. Analysis results of the video sequence.

patterns. The RGB colour information is very sensitive to the lighting conditions. Therefore, we assume that the lighting is stable during the e-learning period. On the other hand, this system does not require fast video acquisition and processing to track human-computer interactions. Hence, video acquisition rate was set to 12 frames per minute and the images were processed as described in Section 2. The actual program used by the LabVIEW front panel and its block diagram are shown in Fig. 7 and Fig. 8. The original video frame and segmentation results, the human-computer interaction check, and the video acquisition time are displayed in the front panel.

In the experiment, 80 video frames are acquired during 400 seconds of e-learning time. We simulated common human behaviours during the e-learning process such as changes in the head position, face occlusion, gaze directions, and moving outside the camera's field of view. During the first 150 seconds, the user changed their head positions, occluded their face with their hand, and moved closer to and further from the camera, while keep track of the e-learning content. The user moved outside the camera's field of view for around 200–220 seconds. He moved their head in different direction for between 280–330 seconds, which indicated a loss of attention. Finally, they returned to their original position. The presence ratio and gaze direction distribution for each frame are shown in Fig. 9. As seen in Fig. 9, most of the video frames are located around the 100% presence ratio and less than 10° of the gaze direction during the actual learning period.

Finally, the human-computer interactions are calculated for each frame using Equation (7). This showed that 89% of the video frames, i.e., 71 frames out of 80, were identified correctly and the proposed algorithm successfully recognized human-computer interaction despite the occluded face images and significant head movements. Most misidentifications are caused by poor image quality and the subsequent effects on the segmentation results. Fig. 10 shows a comparison of the actual head

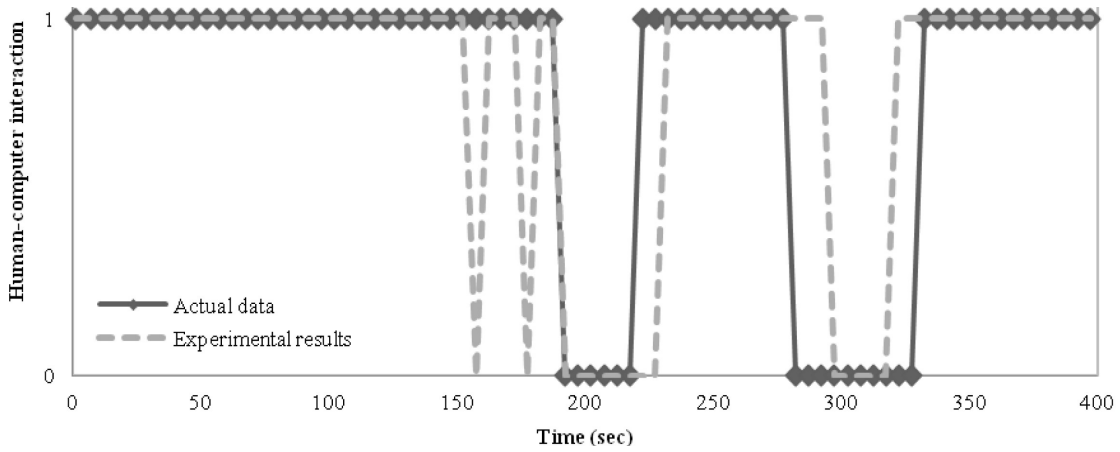

Fig. 10. Comparison of actual and experimental results of the video sequence. 1 and 0 correspond to *True* and *False* values, respectively, for human-computer interactions on the vertical axis.

posture data and the corresponding experimental results throughout the 400 seconds of the e-learning period. For the sake of simplicity, the human-computer interaction values are represented as 1 and 0 on the vertical axis instead of *True* and *False*.

In summary, the analysis of 7 minutes of the experimentation period demonstrated that the proposed algorithm can be applied to e-learning systems to recognize human presence and to estimate the gaze direction. However, more studies should be conducted under different scenarios for robust recognition, such as to detect closed eyes and sleeping students.

4. Conclusion and future work

In this paper, we introduced a new visual approach for enhancing human-computer interactions, especially in virtual environments and e-learning systems, which may supplement the analysis of mouse and keyboard strokes. Instead of using facial feature detection techniques, a simple and user-interactive approach was proposed for detecting a human presence and estimating the gaze direction. The main advantages of the proposed algorithm are its computational efficiency and the fact that it is independent of facial features, such as eye, lip, and mouth movement. The results showed that this technique can be applied in e-learning systems to track the interest of students. The proposed approach was also applied to a short video sequence to demonstrate the practical usage and the accuracy of the algorithm.

In general, human-computer interaction data can be used for the following purposes in e-learning systems: (i) calculating the overall student learning time; (ii) understanding the needs of students and their study topics; (iii) improving the e-learning content based on data from multiple students, common disinteresting topics, and other variables. Therefore, human-computer interaction data has huge potential for understanding learning in virtual environments. A fundamental aim of human-computer interaction is to improve the interactions between users and computers by making computers more usable and receptive to the user's needs. A long-term goal of human-computer interaction is to design systems that minimize the barrier between the human cognitive models of what they want to accomplish and how the computer understands the user's task and behaviour, in order to build adaptive and smart e-learning systems.

In this study, a limited number of human races were represented. However, additional experiments will be implemented to include different human races, such as users with a dark skin tone, in order to understand their emotions and expressions. Finally, future research will be dedicated to human-computer interaction integration methods using images, voice, and some sensor data to build smart and adaptive e-learning systems. In this manner, we might even recognize sleeping students, which is a very common human-computer interaction failure in remote learning systems. The integration of human-computer interaction data into the system is still an open topic and needs further elaboration and experimental validation. Therefore, integration and assessment methods will be considered in future work.

References

1. M. Rosenberg, *E-Learning: Strategies for Delivering Knowledge in the Digital Age*, McGraw Hill, 2001.
2. S. Alexander, E-learning developments and experiences, *Education + Training*, **43**(4/5), 2001, pp. 240–248.
3. C. Twigg, Quality, cost and access: The case for redesign, *The Wired Tower*, Prentice-Hall, New Jersey, 2002.
4. D. Keegan, *Theoretical Principles of Distance Education*, Routledge, New Fetter Lane, 1993.

5. R. W. Picard and J. Klein, Computers that recognize and respond to user emotions, *Theoretical and Practical Implications*, MIT Media Lab Tech Report 538, 2002.
6. L. Shen, M. Wang and R. Shen, Affective e-learning: Using emotional data to improve learning in pervasive learning environment, *Educational Technology & Society*, 2009, **12**(2), pp. 176–189.
7. B. Shneiderman, The future of interactive systems and the emergence of direct manipulation, *Behavior and Information Technology*, 1982, pp. 237–256.
8. M. T. Maybury and W. Wahlster, *Readings in Intelligent User Interfaces*, Morgan Kaufmann Press, San Francisco, 1998.
9. A. Kirlik, *Adaptive Perspectives on Human-Technology Interaction*, Oxford University Press, Oxford, 2006.
10. S. L. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson and D. Ferro, Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions, *Human-computer Interaction*, **15**, 2000, pp. 263–322.
11. D. M. Gavrila, The visual analysis of human movement: a survey, *Computer Vision and Image Understanding*, **73**(1), 1999, pp. 82–98.
12. L. E. Sibert and R. J. K. Jacob, Evaluation of eye gaze interaction, *Human Factors in Computing Systems*, 2000, pp. 281–288.
13. Y. Ugurlu, Measuring the impact of virtual instrumentation for teaching and research, *Proceedings of IEEE Global Engineering Education Conference*, Amman, Jordan, April 4–6, 2011, pp. 152–158.
14. Y. Ugurlu and H. Sakuta, E-learning for graphical system design courses: A case study, *Proceedings of IEEE International Conference on Technology Enhanced Education*, Kerala, India, January 3–5, 2012, pp. 1–5.
15. M. Eisenhauer, B. Hoffman and D. Kretschmer, State of the art human-computer interaction, 2002, GigaMobile/D2.7.1.
16. F. Karray, M. Alemzadeh, J. A. Saleh and M. N. Arab, Human-computer interaction: Overview on state of the art, *International Journal on Smart Sensing and Intelligent Systems*, 2008, **1**(1), pp. 137–159.
17. H. Ekanayake, D. D. Karunarathna and K. P. Hewagamage, Cognitive architecture for affective e-learning, *Proceedings of the International Conference on eLearning for Knowledge-Based Society*, Bangkok, Thailand, August 3–4, 2006, pp. 1–8.
18. R. Brunelli and T. Poggio, Face recognition: Features versus templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 1993, pp. 1042–1052.
19. S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, Springer, 2004.
20. R. L. Hsu, M. Abdel-Mottaleb and A. K. Jain, Face detection in color images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2002, **24**(5), pp. 696–706.
21. M. J. Jones and J. M. Rehg, Statistical color models with application to skin detection, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, USA, June 23–25, 1999, pp. 274–280.
22. P. Viola and M. Jones, Robust real-time face detection, *International Journal on Computer Vision*, 2004, **57**(2), pp. 137–154.
23. G. J. Edwards, C. J. Taylor and T. F. Cootes, Learning to identify and track faces in image sequences, *Proceedings of IEEE International Conference on Computer Vision*, Washington DC, USA, January 4–7, 1998, pp. 317–322.
24. S. L. Phung, A. Bouzerdoum and D. Chai, Skin segmentation using color pixel classification: Analysis and comparison, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2005, **27**(1), pp.148–154.
25. D. L. Pham, C. Xu and J. L. Prince, Current methods in medical image segmentation, *Annual Review of Biomedical Engineering*, **2**, 2000, pp. 315–337.
26. A. Atharifard and S. Ghofrani, Robust component-based face detection using color feature, *Proceedings of The World Congress on Engineering*, London, U.K., July 6–8, 2011, pp. 3781–791.
27. B. Weyrauch, B. Heisele, J. Huang and V. Blanz, Component-based face recognition with 3D morphable models, *Proceedings of IEEE Workshop on Face Processing in Video*, Washington DC, USA, June 27–02, 2004, pp. 202–207.

Yücel Uğurlu is a research associate and lecturer in the Department of Integrated Information Technology at Aoyama Gakuin University. He received B.Sc. and M.Sc. degrees in Electronics Engineering in 1993 and 1995, respectively, from Ankara University and a Ph.D. in Information Science from the Tokyo Institute of Technology in 1999. He is a member of IEEE, the Japanese Society for Engineering Education, and the Japanese Society of Applied Science. His research interests are engineering education, learning tools and environments, e-learning, human-computer interaction, machine vision, and image processing.