Utilizing Human-Computer Interaction Data to Extract User Interests from Web-based Learning Systems

Yücel Uğurlu

Department of Creative Informatics, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: yucel.ugurlu@ni.com

Received 7 July 2013

Accepted 13 August 2013

Abstract

We propose a novel approach by utilizing human-computer interaction data to extract user interests from web-based learning systems. The proposed approach is based on user access and a wearable context recognition system. Under our approach, a series of screen images are captured by an imaging device worn by users engaged in web-based learning. These images help in detecting specific vendor logo information, which is then used to deduce the web-based learning context. The compiled history of recognized context and learning access is finally compared to extract user interests. Experimental results show that the proposed approach is robust, and can identify the relevant context in 96% of cases. The proposed method was successfully applied to 16 users for 1 h of learning time to extract high and low interest topics from web-based learning systems.

Keywords: Web-based learning, context recognition, graphical programming, human-computer interaction.

1. Introduction

In recent years, web-based learning systems have attracted considerable attention in higher education. Moreover, web-based learning, in general, has become an integral method for delivering knowledge. The primary advantage of web-based learning is the hyper-availability of rich digital content capable of engaging students, both in and out of classrooms. With the supply of this digital content continuing to grow, web-based learning systems will need to filter and adapt content to the needs and interests of an individual user [1–3].

In particular, an automated adaptation process, which incorporates smart human-computer interaction and context recognition strategies, would ensure an improved web-based learning experience [4–7]. Indeed, the goals of next-generation web-based learning systems align well with the goals of exemplary instruction: delivering the right content, to the right person, at the proper time, in the most appropriate way—any time, any place, any path, any pace [8, 9].

To deliver on these goals, we have focused our research on the analysis of human-computer interaction and a context recognition for web-based learning system based on extraction of visible information using wearable imaging devices. We believe this approach satisfies most requirements of next-generation web-based learning systems in an automated and non-intrusive way, thus ensuring more adaptive and user-friendly web-based learning systems [10].

This paper is organized as follows: In Section 2, the proposed approach is explained, and a test implementation is described; in Section 3, the system is applied to a sequence of screen images for single and multiple users; and finally, in Section 4, the concluding remarks are given.

Published by Atlantis Press

2. Context Recognition

Enabling systems to act smartly and independently is the most obvious use for context awareness technology. In web-based learning, particularly, context awareness can be used to decide what, when, and how learning resources are delivered to users [11]. The primary aim of our work is to discover and develop methods for recognizing a user's context from sensor data using pattern recognition.

In a previous study, we developed a method for recognizing individuals' behavior using cameras embedded on users' PCs by analyzing their head postures [12]. However, understanding individuals' interests based on this approach is a challenging task, and the accuracy of the system varies significantly depending on a user's race and hair color. By comparison, wearable imaging devices provide much a wider coverage, and are easy to integrate into many applications [13].

In this study, a wearable imaging device, composed of a portable camera mounted on eyeglasses, was used to collect user interaction data from web-based learning systems. A general schema is provided in Fig. 1. As shown, context recognition data provides additional information about user behavior and actual usage of the web-based learning system. Moreover, these additional sources of information can be easily integrated into the final evaluation process.

This context recognition module is tasked with determining user behavior based on image processing and geometric feature matching. Using wearable imaging devices to capture images can significantly alter the brightness, angle, and distance of the images, as a user may move his/her head, and the captured images may also include a significant amount of noise.

Thus, the context recognition system is decomposed into three ordered processes: image enhancement, feature detection, and evaluation of user interests.

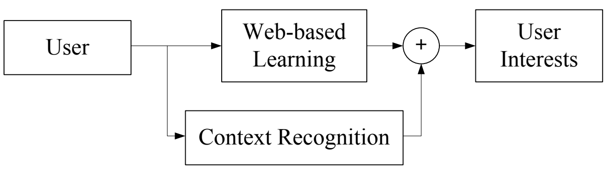


Fig. 1. System overview: integration of context recognition with web-based learning system to extract user interests

2.1. *Image enhancement*

By enhancing image information and restraining noise and brightness, we can significantly improve the accuracy of feature detection. The flowchart for our image enhancement process is shown in Fig. 2. The input image is used here to obtain the wavelet transform images. The decomposition of images into different frequency ranges helps to isolate the frequency components introduced by intrinsic deformations or extrinsic factors. For this reason, discrete wavelet transform (DWT) is the key component of image enhancement system [14].

First, the image to be enhanced is decomposed into four frequency sub-bands using wavelet transform [15]. Second, 2D image interpolation and subtraction are applied to adjust overall image brightness and eliminate the lighting variations. Finally, the enhanced image is obtained through inverse wavelet transform of the sub-band images.

The resulting image has been decomposed into two

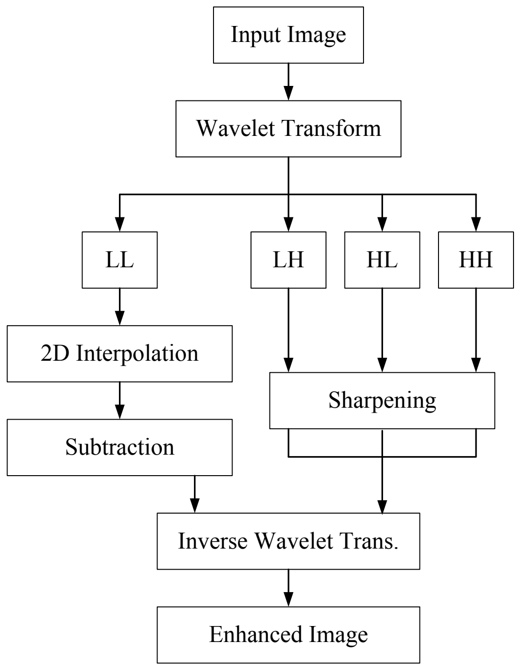


Fig. 2. Block diagram of image enhancement processes

parts: a low frequency part, which is a kind of average of the original signal, and a high frequency part, which is what remains after the low frequency part is subtracted from the original signal. Since the low frequency image is expected to have significant brightness variations, it should be corrected before pattern matching.

In this study, Haar wavelet is used because it is fast and can detect sharp objects. Haar wavelet is a sequence of rescaled "square-shaped" functions, and represents the simplest possible wavelet.

Haar wavelet's mother wavelet function is described as:

$$\psi(\mathrm{t}) = \begin{cases} 1 & 0 \le t < 1/2, \\ -1 & 1/2 \le t < 1, \\ 0 & \textit{otherwise.} \end{cases} \tag{1}$$

Its scaling function is described as:

$$\varnothing(\mathrm{t}) = \begin{cases} 1 & 0 \le t < 1, \\ 0 & \textit{otherwise.} \end{cases} \tag{2}$$

The 2D wavelet decomposition of an image is performed by first applying a 1D DWT along the rows of the image, and then decomposing the results column-wise. This operation results in four decomposed sub-band images, referred to as low–low (LL), low–high (LH), high–low (HL), and high–high (HH). The resolution of these sub-band images is half that of the original image.

In order to detect brightness variations in the input image, a 2D surface-fitting algorithm is employed, leveraging a bicubic spline interpolation method [16]. Bicubic splines are used for obtaining smoothness in two-dimensional interpolation, and they guarantee that the first and second partial derivatives of cubic interpolating polynomials are continuous, even at the data points.

In image processing, bicubic interpolation is often chosen over bilinear or nearest neighbor interpolation for image resampling. In contrast to bilinear interpolation, which only takes 4 pixels (2×2) into account, bicubic interpolation considers 16 pixels (4×4). Images resampled with bicubic interpolation are smoother and have fewer interpolation artifacts. For our purposes, the LL image is divided into a rectangular grid, and the maximum brightness value for each pixel area is calculated. The final 2D image is obtained by performing bicubic interpolation on the LL image, subtracting the original LL image from the interpolated image, and eliminating the brightness variations from the difference.

Meanwhile, the high frequency components, LH, HL, and HH, are sharpened using a 3×3 Laplacian kernel, and used to obtain an enhanced image with the inverse wavelet transform. (Note that the resolution of the enhanced image is exactly the same as the input image.)

2.2. *Geometric feature matching*

Context recognition methods, proposed in a previous study, are very similar to the techniques used in pattern recognition [17]. The aim of pattern recognition is to classify data on the basis of its properties and available a priori knowledge. In our case, classification involves association of the various vendor or school logos, unique icons, and other characters that are consistent and visible throughout the web-based learning systems.

Pattern matching can be extremely challenging, as many factors can alter the way an object appears to a vision system. Traditional pattern matching technology relies upon a pixel-grid analysis process, commonly known as normalized correlation. This method looks for statistical similarity between a gray-level model—or reference image—of an object and certain portions of the image to determine the object's position.

Though effective in certain situations, this approach may fail to find objects, or to locate them accurately, under variations in appearance common to the production or situation of those objects, such as changes in object angle, size, and shading.

To overcome these limitations, a geometric pattern matching approach is preferred. This approach involves "learning" an object's geometry, or *template*, using a set of boundary curves that are not bound to a pixel grid, and then scanning for similar shapes in the image without relying on specific grey-scale levels.

In geometric pattern matching, an object template can be found across varying lighting conditions, noise, and geometric transformations such as shifting, rotation, or scaling [18–20]. This makes it possible to extract the contours of the image, by assigning each match a score that indicates how closely the template resembles the located match.

This match score is given as follows:

$$r = \frac{\textit{matched pixels}}{\textit{total pixels in ROI}} \tag{3}$$

where r is the matching score and ROI is the region of interest. The output indicator of the found number of matches indicates the number of exact matches of the template found in the database.

The result is a revolutionary improvement in the ability to accurately find objects despite changes in angle, size, and shading. The correlation coefficient has the value $r = 1$ if the two images are absolutely identical, $r = 0$ if they are completely uncorrelated, and $r = -1$ if they are antithetically uncorrelated, such as when one image is the negative of the other. In this study, scores greater than 0.5 ($r \geq 0.5$) are considered as significant correlation, and the corresponding images are treated as successful matches.

2.3. *Extracting user interests*

Integration of context recognition data into web-based learning systems is essential for evaluating user interests and adapting web-based learning systems to those interests. Utilizing only user access information and time is not enough to build intelligent and adaptive systems. The relationship between learning systems and human-computer interaction can be measured and used for the further analysis. For this purpose, personal access history and context recognition results are converted to numerical values such as 1 and 0, where 1 represents valid web-based learning usage and context recognition, and 0 represents failure.

Personal access history can be represented numerically using

$$b_t = \begin{cases} 1 & access\,confirmed, \\ 0 & otherwise. \end{cases} \tag{4}$$

where t is time, and b_t returns to 1 if user access is confirmed.

The context recognition data of a user is calculated using

$$d_t = \begin{cases} 1 & r_1 \geq 0.5\, or\, r_2 \geq 0.5, \\ 0 & otherwise. \end{cases} \tag{5}$$

If $b_t = d_t$, web-based learning access is correlated with context recognition data for the specified access time, indicating that web-based learning is taking place at that moment.

Low interest topics can be extracted using the following equation:

$$m_t = 1 - (b_t - d_t) \tag{6}$$

When $m_t = 0$, there is a good match between web-based learning usage and actual context awareness. However, when $m_t = 1$, the user likely has low interest in the presented content, or has lost motivation for other reasons.

Note that $m_t = 2$ is an invalid value, because it can only occur when $b_t = 0$ and $d_t = 1$, i.e., when no web-based learning history has been recorded in the system.

$$n_t = mode(m_{t-2}, m_{t-1}, m_t, m_{t+1}, m_{t+2}) \tag{7}$$

Since our pattern recognition method is not perfect, false detections can occur in d_t data points, yielding misleading results. To eliminate these false detections, sequential image information is used, as per Eq. (7). In Eq. (7), "*mode*" refers to the most frequently occurring value in the neighborhood. Thus, false detections up to 2 sequential images are corrected using their neighboring data points.

In case of $n_t = 0$, web-based learning access history and context recognition results do not match each other, and this implies a loss of user interest. In other words, even though a user has access to a web-based learning system, context recognition fails to detect the logo information on the screen images.

Eq. (7) provides us information in an acquired and analyzed image level. However, web-based learning systems are generally constructed by learning categories or topics. Therefore, actual user interest should be identified in terms of topic name or number.

For this purpose, the images are categorized according to corresponding topics, and the total learning time is extracted in addition to n_t values. User interest level for a specific topic can then be calculated using the following equation:

$$L_j = \sum_{i=1}^{k} A_i \sum_{i=1}^{k} (n_t) / k \tag{8}$$

where L is user interest level, j is topic number, A is total access time, and k is the number of images corresponding to one specific topic. Using Eq. (8), user interest level can be represented numerically based on the access history and context recognition analysis.

High interest values indicate that there is high correlation between user access history and human-computer interaction data. This information can be utilized to build adaptive and intelligent frameworks for web-based learning systems. In case of low interest topics, the learning content should be revisited or improved upon.

3. Experimental Results

To measure the performance of the proposed system, we tested it on the users of National Instruments' e-learning portal. This portal consists of more than 200 topics, equivalent to 30 h of learning time, all dedicated to teaching graphical programming in LabVIEW to engineering students. The web-based learning content includes various one-topic videos, each from 5 to 10 minutes in length, readings, and short quizzes. The courses covered are as follows:

- LabVIEW programming I
- LabVIEW programming II
- Data acquisition and analysis
- Field-programmable-gate-arrays (FPGA)
- Real time programming
- Image acquisition and processing

For the sake of simplicity, we mainly focused on the first course, which is LabVIEW programming I.

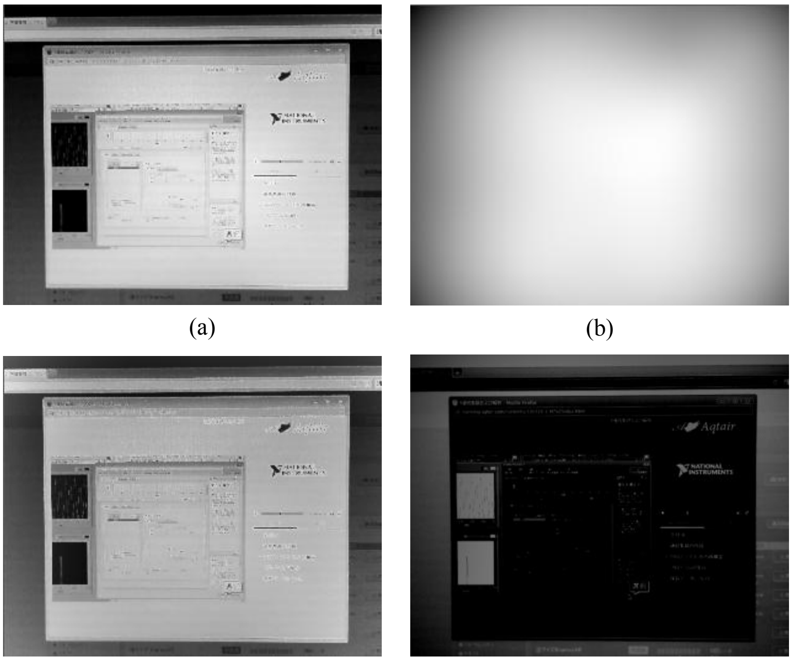
In practice, a unique user ID and password are provided to access all web-based learning functions. Since users are logged into the system using their unique user IDs, it is easy to acquire their access history, as formalized in Eq. (4).

After login, we captured a series of web-based learning screen images using the imaging device worn by the users. Each image was saved to the system with a time stamp for use in further evaluations. A total of 76 images were captured at a resolution of 1280×1024 pixels.

3.1. *Application to single user*

The image enhancement and geometric feature matching processes were applied to all 76 screen images, among which there were 64 images of displayed web-based learning content, and 12 images of the computer display without web-based learning content.

All input images used in this study are gray scale images. For image processing, NI LabVIEW and IMAQ Vision Development software were used to analyze the screen images [21, 22]. In LabVIEW, the *Multiresolution Analysis 2D.vi* function was used to acquire the wavelet transform images.



(a) (b) (c) (d)

Fig. 3. Image enhancement results: (a) input image, (b) 2D interpolated image, (c) enhanced image, (d) inversed image

(a) (b)

Fig. 4. Vendor logos and their contour detections: (a) Aqtair logo, (b) NI logo

Since the images were divided into different frequency bands, the LL images included the brightness variations, and needed to be adjusted prior to geometric pattern matching.

Fig. 3 shows the image enhancement results using the *2D Interpolate.vi* function. For this, the original images were divided into a 5×4 grid, and the maximum brightness level for each rectangular area was found. The *2D interpolate.vi* function accepts tabulated X, Y, and Z values, and provides interpolated values z_i that correspond to each x_i, y_i location.

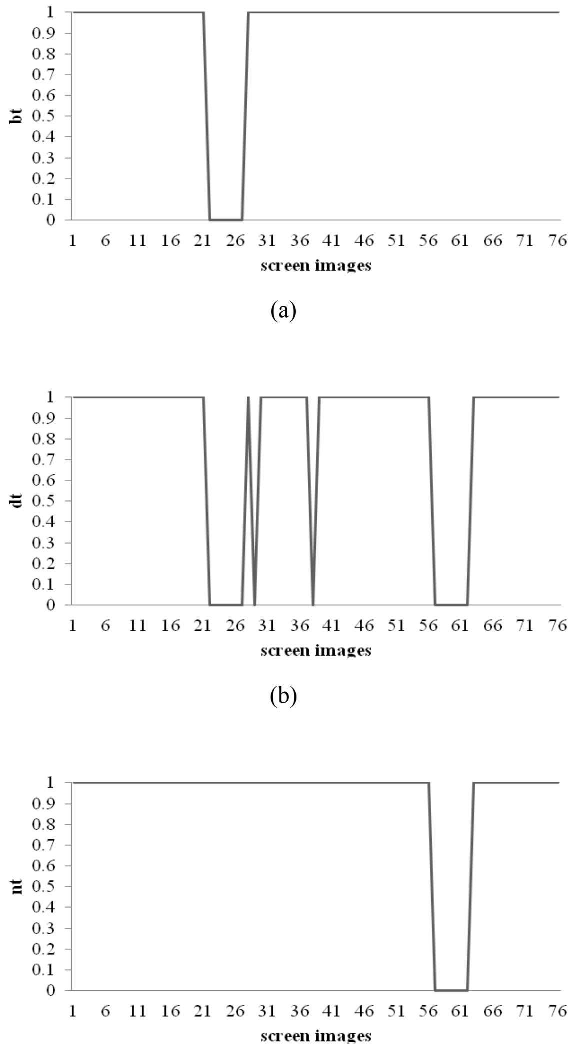
Next, vendor logos were manually selected and saved on the system. Fig. 4 shows the vendor logos of two companies, Aqtair and National Instruments Corporation, for which contour information was automatically detected and highlighted in the template images.

The *Geometric Matching.vi* function is used to find templates in LabVIEW. For pattern matching, scaling and rotation values were set within the ranges of 50%–150% and 0°–360°, respectively. All 76 images were processed using the same image enhancement and pattern matching processes.

Table 1 shows the analysis results of the proposed system for all screen images: 97% of context images were accurately recognized using vendor logo information, and the total accuracy of the system, including other contexts, was 96%. We consider this a very satisfactory result.

Table 1. The results of feature matching for a wearable context recognition system

Types of Images	Actual Number	Detected Number	Accuracy (%)
Web-based learning images	64	62	97
Other context images	12	11	92
Total images	76	73	96



(c)

Fig. 5. Practical usage of context recognition for web-based learning systems: (a) web-based learning access history, (b) context recognition data, (c) detection of high and low interest topics from screen images

A further evaluation of web-based learning access and context recognition is provided by the graphs in Fig. 5. Fig. 5(a) shows the web-based learning access history of a user throughout the 76 screen images, indicating that access history alone is not enough to understand actual usage of web-based learning systems. Fig. 5(b) shows the context recognition results based on geometric feature matching. This data provides us additional information on how the web-based learning system was used, even though there are several false detection points.

Finally, Fig. 5(c) shows the correlation of web-based learning access and real usage, with false detections eliminated using sequential image information, as formalized in Eq. (7). From this, it is clear that images 56–61 are uncorrelated with access history, implying weak human-computer interaction during that interval.

Based on this information, the corresponding images and access time is used to calculate the user interest level for each topic as given by Eq. (8). Therefore, based on the human-computer interaction data, we found that user had the highest interest for "DAQ assistant" and lowest interest for "state machine" topics.

3.2. *Application to multiple users*

One of the main benefits of a web-based learning system is to understand user behavior and customized learning structure of individuals based on the human-computer interaction data. For this purpose, we extend our approach to multiple users, and extract high and low interest topics, as given by Eq. (8).

A web-based learning system is constructed using several main and sub-topics. The main learning topics are as follows:

1. Express VI
2. LabVIEW introduction
3. File input/outputs
4. Debugging
5. LV basic
6. Graphs
7. Subroutines
8. Parallel programming
9. Measurement and automation explorer
10. Customization
11. Design patterns
12. Reading texts

Sub-topics are not listed here; nevertheless, the number of sub-topics changes from 2 to 10 for each main topic, and includes 5 to 10 min of videos. In addition, there are short quizzes and some reading texts in the sub-topic sections.

In the experiment, we applied our approach to 16 users who were new to LabVIEW programming. We allocated 1 h of learning time, and acquired 60 images per person during the entire learning time. The average access rate was 8 topics out of the 64 topics in the *LabVIEW Programming I* course.

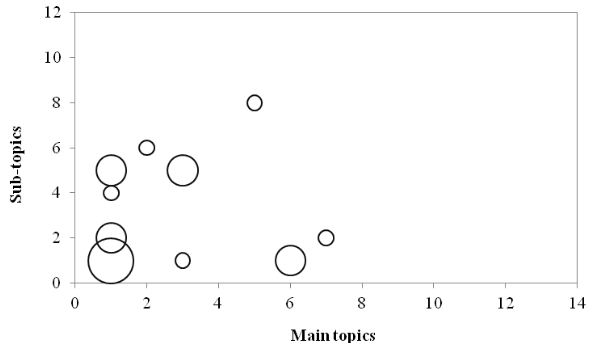
Context recognition method was utilized to understand the user interest for each topic, as explained in Section 2. We calculated user interest level for each viewed topic based on the learning time and context recognition analysis. Highest and lowest interest topics of multiple users are summarized in Table 2.

Table 2. Highest and lowest interest topics of multiple users

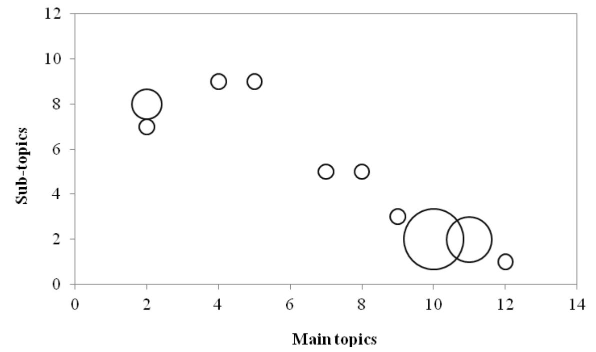
Users	Highest Interest Topics	Lowest Interest Topics
1	(1–1)	(10–2)
2	(1–5)	(2–8)
3	(2–6)	(11–2)
4	(1–4)	(7–5)
5	(1–1)	(10-2)
6	(3–1)	(8–5)
7	(6–1)	(12–1)
8	(3–5)	(10–2)
9	(5–8)	(9–3)
10	(1–1)	(2–8)
11	(1–2)	(11–2)
12	(7–1)	(4–9)
13	(1–5)	(2–7)
14	(1–2)	(11–2)
15	(3–5)	(10–2)
16	(6–1)	(5–9)

For the sake of simplicity, main and sub-topic numbers are used in Table 2 to represent highest and lowest user interests. For example, (1−5) indicates 1st main topic and 5th sub-topic, which corresponds, respectively, to "express VI" and "continuous loop" function related video content in the web-based learning system.

The same results are plotted in the bubble chart too. The graphs are given in Fig. 6 for the highest and lowest interest topics corresponding to each user. The size of bubbles represents the frequency of the user interest topics. Common interest topics are indicated by bigger bubbles. The size (width) of the bubbles is proportional to the frequency.



(a)



(b)

Fig. 6. Extracted user interests for multiple users: (a) highest interest topics, (b) lowest interest topics

Fig. 6(a) shows the highest interest topics corresponding to each of the 16 users. High interest topics are related to fundamental concepts such as “express VI,” “LabVIEW introduction,” and “file inputs/outputs.” On the other hand, users have low interest in advanced topics such as “customization” and “design patterns,” most probably because they are new to LabVIEW programming, as shown in Fig. 6(b).

In our experimental studies, target audiences were selected from our teaching course, and the duration of the experiment was limited to 1 h. Nevertheless, preliminary results show that the proposed approach can be easily applied to larger groups to extract user interests. In addition, it allows us to have a better understanding of web-based learning usage. Thus, we can build a more adaptive and user-friendly learning environment.

4. Conclusion

This study proposed a new approach to extract user interests from web-based learning systems using human-computer interaction data. The approach involved image capture by wearable devices, image enhancement, and geometric pattern matching techniques. The accuracy of the context recognition system was found to be quite high, even though the images were acquired at different times, under different lighting, orientation, and distance conditions. Finally, the algorithm was tested on single and multiple users to extract high and low interest topics from web-based learning systems.

The real benefit of context recognition system and utilizing human-computer interaction data in web-based learning systems is to understand a user’s learning behavior so as to extract his/her choice of popular topics. As we know, the biggest challenge in web-based learning system is lack of face-to-face interaction between an instructor and a user. In addition, each user is unique, has varying levels of knowledge, and learns differently. Therefore, extracting user interests is very important to improve web-based learning content, and to build a more intelligent and adaptive learning environment.

In the future, we would like to focus on understanding the main characteristics of highest and lowest interest topics as well as their relationship with personal needs. For this purpose, a larger dataset comprising multiple users would be collected and analyzed to identify similarities and differences between users and their learning behavior.

Acknowledgement

This work was supported by National Instruments Japan Corporation.

References

1. R. Ubell, Engineers turn to e-learning, *IEEE Spectrum* **37**(10) (2000) 59–63.
2. R.C. Clark and R.E. Mayer, *E-learning and the Science of Instruction*, (Pfeiffer, 2007) pp. 7–48.
3. C. Virginio, et al., Perspectives and challenges in e-learning: towards natural interaction paradigms, *J. Visual Languages and Computing* 15 (2004) pp. 333–345.
4. F. Karray, et al., Human-computer Interaction: Overview on state of the art, *Int. J. on Smart Sensing and Intelligent Systems* **1**(1) (2008) 137–159.

5. J.S. Greenstein, Pointing devices, *Handbook of Human-Computer Interaction*, (Elsevier Science, Amsterdam, 1997).
6. B.A. Myers, A brief history of human-computer interaction technology, *ACM Interactions*, **5**(2) (1998) 44−54.
7. C.M. Karat, et al., Conversational interface technologies, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application*, eds. J.A. Jacko and A. Sears, (Lawrence Erlbaum Associates, Mahwah, 2003).
8. R. Benson and C. Brack. Developing the scholarship of teaching: What is the role of e-teaching and learning? *Teaching in Higher Education*, **14**(1) (2009) 71−80.
9. A. Bosom, et al., Excellence in virtual education: The tutor online approach, *J. Cases on Information Technology*, **9**(2) (2007) 61−74.
10. V. Shute and B. Towle, Adaptive e-learning, *Educational Psychologist*, **38**(2) (2003) 105−112.
11. M.L. Blum, *Real-time context recognition*, (MIT Media Lab, 2005).
12. Y. Ugurlu, Smart e-learning: Enhancement of human-computer interactions using head posture images, *Int. J. Engineering Education*, **29**(3) (2013) 568−577.
13. Y. Ugurlu, Wearable context recognition for distance learning systems: A case study in extracting user interests, *Studies in Computational Intelligence*, (Springer, London, 2013) pp. 65−75.
14. H. Demirel and G. Anbarjafari, Image resolution enhancement by using discrete and stationary wavelet decomposition, *IEEE Trans. on Image Processing* **20**(4) (2011) 1458–1460.
15. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding* (Prentice Hall, 1995).
16. A. Kouibia and M. Pasadas, Approximation of surfaces by fairness bicubic splines, *Advances in Computational Mathematics* 20 (2004) pp. 87–103.
17. R. Duda, et al., *Pattern Classification*, (Wiley, New York, 2001) pp. 9−13.
18. G. Sharma, et al., Image recognition systems using geometric matching and contour detection, *Int. J. Computer Applications* **51**(17) (2012) 49–53.
19. M.T. Goodrich and J.S.B. Mitchell, Approximate Geometric Pattern Matching under Rigid Motions, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **21**(4) (1999) 371−379.
20. S.H. Chang, et al., Fast algorithm for point pattern matching: Invariant to translations, rotations and scale changes, *Pattern Recognition*, **30**(2) (1997) 311−320.
21. T. Klinger, *Image Processing with LabVIEW and IMAQ Vision*, (Prentice Hall PTC, 2003).
22. C.G. Relf, *Image Acquisition and Processing with LabVIEW*, (Taylor and Francis CRC ebook, 2007) pp. 96−155.