ARFLED: Ability Recognition Framework for Learning and Education

Shoya Ishimaru
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Shoya.Ishimaru@dfki.de

Andreas Dengel
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Andreas.Dengel@dfki.de

ABSTRACT
Learning is one of the vital behaviors of human beings. This paper demonstrates a framework to augment learning activities by packaging two key ideas: Eyetifact and HyperMind. Eyetifact is a system that converts data of eye movements beyond the difference of sensing devices to collect a large amount of training data for machine learning. HyperMind is a digital textbook that displays learning materials dynamically based on a learner's cognitive states as measured by several sensors. In order to implement these two ideas, we have conducted experiments related to eyewear computing, textbook reading behavior analysis, and stress sensing. The contributions of this research are to investigate approaches that recognize human abilities and to transfer them from experts to others.

ACM Classification Keywords
H.5.m Information interfaces and presentation (e.g., HCI): Miscellaneous

Author Keywords
Activity recognition; cognitive state; eyewear computing; eye tracking; learning; education; reading activity; quantified self

INTRODUCTION AND SPECIFIC PROBLEMS
Learning – the act of acquiring new knowledge, skills, abilities, and expertise – is one of the vital behaviors of human beings. In particular, people in the modern world are always required to learn situational new skills. The reason is that advances in technology are constantly changing their lifestyles and the ways they work. However, the computer is not only a requester of new skills; it can also be a partner in enhancing learning [Dengel 2016]. For instance, a system which senses learners' behaviors can help learners by providing personalized information based on their interests and problems. If learners' expertise can be recognized, it will be possible to transfer the secret of success from experts to novices. We focus on such a research domain and define it as *Ability Recognition*.

UbiComp/ISWC'17 Adjunct , September 11–15, 2017, Maui, HI, USA

ISBN 978-1-4503-5190-4/17/09…$15.00

DOI: https://doi.org/10.1145/3123024.3123200

Ability consists of physical movements, cognitive states, and contexts of the activities. Since Douglas Carl Engelbart proposed the framework of *Augmenting Human Intellect* [Engelbart 1962] and Mark Weiser created the terms *Ubiquitous Computing* [Weiser 1993] and *Calm Technologies* [Weiser and Brown 1997], many researchers have investigated recognition of human activities by using several sensors for giving proactive assistance. Compared to other two factors, recognizing cognitive states (e.g., attention, concentration, comprehension) is still a challenging task because it is hard to detect them by simple motion sensing approaches. Advanced sensors are required. For example, eye tracking is one of the most effective sensing approaches. But most of eye tracking devices are not designed for wearing regularly (e.g., the devices are expensive; battery lives are not enough to cover a whole day; cables on devices prevent a user from moving naturally). Therefore, the more the sensing approach is advanced, the more the environment of data recording is limited. To utilize modern machine learning techniques including deep learning, we need to come up with the idea how to record a huge amount of training data on such valuable sensors.

The specific problems addressed in this work are three-fold. (1) How can we record a large amount of natural behaviors in real environments? (2) How can we recognize problems and interests of learners by sensors? (3) How can we recognize and transfer the secret of success from experts to novices?

RELATED WORK
There is a large corpus of work focusing on understanding human behaviors from activity to context towards ability recognition. This section summarize them with highlights of two aspects: the target of recognition and the sensing modality.

Activity Recognition
The starting point of this field was to recognize "what" a user is doing. Human activities can be classified into two rough categories: physical activities (e.g. walking, running, cycling, sleeping) and cognitive activities (e.g. reading, writing, talking). Most physical activities can be recognized by motion sensors on the body [Foerster et al. 1999, Bao and Intille 2004] or a smartphone [Dernbach et al. 2012]. Recognizing cognitive activities is a more challenging task because body movements during these activities are limited. Additional sensors are required to recognize they are taking place.

Figure 1. Examples of sensors in this research field sorted by their pervasiveness (left: advanced but expensive; right: inexpensive)

One of the interesting approaches is to use eye movements. The relationship between cognitive activities and eye movements is well explored in cognitive science and psychology [Rayner 1998]. Bulling et al. have classified tasks including cognitive activities by using electrooculography (EOG) sensors [Bulling et al. 2011]. The trend of eye tracking research in activity recognition is going towards recognitions in a natural uncontrolled environment. A long-term eye tracking and egocentric vision dataset using Pupil (shown in Figure 1) has been proposed [Steil and Bulling 2015].

Context Recognition

The more sensors developed, the more interested researchers became in recognizing the context of human activities (i.e., "when", "where", "by whom", and "why" the activity is performed). For example, first-person vision is utilized in the context recognition. Through an egocentric camera attached to the head or the body, activities and the contexts can be estimated from objects in front of the person [Ma et al. 2016]. To recognize social interactions is also a key factor to understand the context of talking activities [Fathi et al. 2012]. Instead of the on-body sensors mentioned above, remote sensors (e.g., fixed cameras [Dimakis et al. 2008], microphones [Stager et al. 2004]) have also been employed to recognize contexts because they can record the interactions between humans and the environment.

Ability Recognition

The most abstract subject of recognition is "how" the activity is performed (e.g., comparing performances between participants in the same task). Since it is close to the cognitive processes, an obvious approach to tracking it is to monitor brain activities directly. Electroencephalography (EEG) [Gevins et al. 1998], magnetic resonance imaging (MRI) [Cox and Savoy 2003], and near-infrared spectroscopy (NIRS) [Ishimaru et al. 2014a] can be candidate, if we can accept the limitation of the recording environment. Physiological sensing including measuring the autonomic nervous system (ANS) [Sapolsky 1994] is more realistic. Cognitive states (e.g., concentrations, mental workload, boredom) can be measured by the changing of pupil diameter [Kahneman and Beatty 1966] and nose temperature [Kunze et al. 2015]. Eye tracking is on the balance of the pervasiveness and the potential. In the field of sensing reading actiivties, for instance, unknown words for a reader [Okoso et al. 2015] and his/her English level [Yoshimura et al. 2015] can be predicted from eye movements. Recently eyewear computers became more and more affordable [Amft et al. 2015, Kuhn et al. 2016].

METHODOLOGICAL APPROACHES AND THE KEY IDEAS

In order to solve three problems mentioned in the Introduction, we propose methods and tools for ability recognition, and package them as a framework. The framework mainly consists two ideas: *Eyetifact* and *HyperMind*. Eyetifact is related to the first problem and HyperMind is for the others.

Eyeticact

AI technologies including deep learning have rapidly accelerated research in the field of image, audio and natural language processing. However, ability recognition is not sufficiently based on the benefits technology offers. Our hypothesis of the reason is that collecting a large amount of data for training is more difficult than it is in other research fields.

To solve this problem, we propose the system that converts data beyond the difference of sensors. Figure 2 represents an overview of this idea. It converts eye movements recorded by a perversive sensor into artificial eye movements which are equivalent to the data recorded by an advanced sensor by learning the relation of data from two devices. By utilizing this system, researchers can record a lot of natural eye movements with commercial eyewear computers in real life, convert them to valuable data, and apply deep learning techniques.

HyperMind

Regarding the second and the third problem, it is difficult to cover many scenarios in the limited time frame of the doctoral project. Thus we treat a single case in the thesis.

Still today, various school subjects are aligned with contents that are captured in textbooks. Although curiosity is an important factor for learning, every student has a different way of learning based on individual speed and preferences, textbooks have traditionally been found to be static and consistently dull for a variety of learners. Therefore, students sometimes avert their eyes from reading a textbook because it is boring.

One of the solutions to this problem is to develop a digital textbook which can make learning instractions adaptive and dynamic on display. The system recognizes students' cognitive states by using several sensors including an eye tracker. Then it changes the content and the layout dynamically so as to improve a student's motivation and understanding. For instance, playing a video instead of showing a static picture should attract their interests. Since students prefer different representations depending on their skill levels [Mozaffari et al. 2016], the system displays the adapted representation based on cognitive state analysis. By tracking the level of understanding of the content, the system selects or generates exercises they should solve so as to correct misunderstandings.

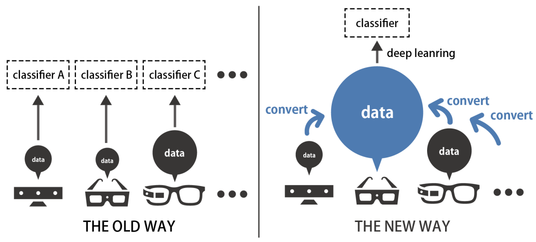

Figure 2. Eyetifact

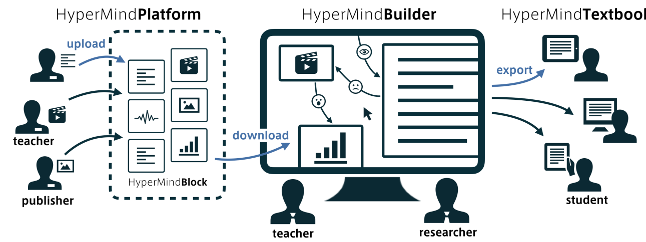

Figure 3. HyperMind

We also provide the platform for researchers, teachers, textbook publishers, and students as shown in Figure 3. Publishers and teachers upload learning materials as blocks to our open platform. Teachers and researchers build an intelligent textbook by connecting blocks with triggers (i.e., determining when/how a new block appears) on the basis of students' cognitive states. Students enjoy reading the intelligent textbook. Their reading behaviors are recognized by eye tracking.

RESEARCH CONDUCTED AND PLANED AHEAD

Eyewear Computing

As pilot studies, we have investigated the potential of pervasive eyewear computers. We proposed a method to detect eye blink frequencies and head motions by using the sensors built on Google Glass to recognize daily activities [Ishimaru et al. 2014b]. The method was evaluated on a data set containing five activity classes (reading a book, watching a video, solving mathematical tasks, sawing cardboard and talking) of eight participants. The classification accuracy on user-dependent training was 67% by utilizing features from eye blink only and increased to 82% when extended with head motion patterns.

We have also proposed activity recognition method on JINS MEME [Ishimaru et al. 2014]. JINS MEME are Electrooculography (EOG) glasses that are equipped with three electrodes to detect eye movements and a 6-axis internal measurement unit (IMU) to detect head movements (see Figure 1). The device is more inexpensive than Google Glass and has enough long battery to cover daily activities. We have demonstrated interactions using the glasses as well [Ishimaru et al. 2015].

There is a relationship between cognitive abilities (e.g., vocabulary skills, critical thinking, academic scores) and reading habits especially daily reading volumes [Stanovich and Cunningham 1998]. As people can be motivated to be physically fit by monitoring step counts [Michie et al. 2009], we believe that tracking the number of words they read can help them improve their daily reading volumes. Therefore we have implemented the Wordometer 2.0: a system to quantify daily reading volume by estimating the number of words a user read from EOG signals measured on JINS MEME [Ishimaru et al. 2016b]. The estimation algorithm was evaluated with a dataset involving five participants read 38 documents (minimum: 27 words; maximum: 120 words; average: 60 words in one document). It estimated the number of read words with 11% error rate with user-independent training and 3.0% with user-dependent training.

Reading Behavior Analysis on a Textbook

In order to develop the intelligent textbook, we started from investigating reading behavior on a textbook [Ishimaru et al. 2016a]. In this study, we proposed methods to extract attentions and to predict comprehensions by utilizing mobile eye tracking glasses. As preprocessing, raw data from an eye tracking glasses are converted to gaze points on a document with a projection function based on SIFT features [Lowe 1999] and classified into fixations and saccades [Buscher et al. 2008]. In the attention extraction method, we define area of interest (AOI) on the textbook beforehand based on the role of contents (e.g., introduction, definition, application). Then the attention was calculated as normalized sum values of fixation durations in each AOI. In the comprehension prediction method, we proposed two types of approaches: AOI based and features in subsequence based. Both approaches utilize SVM to classify students into three classes (low, middle, high). For the classification, attentions in AOIs calculated above are used as feature in AOI based approach, and means and standard deviations of fixation durations and saccade lengths in one minute window, the size was selected experimentally, were used in subsequence based approach.

For evaluations of the methods, we asked eight 11-12 years old students to wear an eye tracking glasses, to read a textbook in Physics and to solve respective exercises. Ground truth of their comprehension was calculated from scores of the exercises. Experimental results revealed that students' reading behaviors represent their comprehensions. For example, students with high comprehensions tend to pay attention on the definition part to understand the content. The classification accuracy of the AOI based complehension prediction was 100% although it is not suitable for the realtime application because it requires the reading from the beginning to the end. Features from a window of gaze data in one minute was enough able to classify students' completion into three classes with 70% accuracy.

Stress Sensing

Stress management is a key to keep be motivated in learning. In particular, external stresses including thermal heat decrease a learner's performance and productivity. Thus we have proposed methods to estimate heat stresses in a working environment [Hoffmann et al. 2016]. We succeeded to extract face temperature by combining sensor signals of FLIR One (see Figure 1) and a facial landmark tracking.

Mental illness, especially depression is one of the most pressing concerns all over the world. We also have demonstrated the idea of "Thermometer for the Mind": the mental state estimating system by a user's activity log derived from wearable devices [Ishimaru and Kise 2015]. In this study, we investigated how information about physical activity from a smartphone and social activity from a Web service can be used to estimate a user's mental state. We recorded one participant's step counts as one of the measurements of physical activities and Twitter post counts as one of the measurements of social activities with his self-assessment ground truth for 5 months. The classification accuracy into three classes (the mood is low, middle, or high) was 60%.

Future work

The implementation of Eyetifact is scheduled as the next step. On the basis of the pilot studies, we are designing the method to convert data from several devices. Since some of the devices can be worn at the same time, we start from recording two sensor's data and treat data from the advanced sensor as ground truth. Then the relation of two signals is trained by LSTM [Hochreiter and Schmidhuber 1997].

In the context of HyperMind, we investigated "when" the dynamic changing should be appeared by measuring a student's comprehension. We plan to implement dynamic changing on textbooks, and investigate three research questions. (1) What kind of additional information (e.g., text, image, video) can improve students' learning abilities? (2) What is the best way to display additional information (e.g., popup, using a white space) overcoming the split attention effect? (3) How to control showing/dismissing the additional learning materials?

CONTRIBUTION IN THE FIELD OF UBICOMP

Expected contributions of this research are to investigate sensing approaches that recognize human abilities and to provide the framework. Specifically, Eyetifact can accelerate work around eyewear computing. HyperMind supports researchers who want to apply sensing approach in the field of ubiquitous computing to the educational field. In the long term, by using the recognitions as the base research, we would like to demonstrate the work of transferring expert abilities to others.

ACKNOWLEDGEMENTS

This work is supported by JSPS KAKENHI and JST CREST (Grant Numbers: 17K12728, JPMJCR16E1).

REFERENCES

Oliver Amft, Florian Wahl, Shoya Ishimaru, and Kai Kunze. 2015. Making regular eyeglasses smart. *Pervasive Computing, IEEE* 14, 3 (2015), 32–43.

Ling Bao and Stephen S Intille. 2004. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*. Springer, 1–17.

Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Troster. 2011. Eye movement analysis for activity recognition using electrooculography. *IEEE transactions on pattern analysis and machine intelligence* 33, 4 (2011), 741–753.

Georg Buscher, Andreas Dengel, and Ludger van Elst. 2008. Eye movements as implicit relevance feedback. In *CHI'08 extended abstracts on Human factors in computing systems*. ACM, 2991–2996.

David D Cox and Robert L Savoy. 2003. Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19, 2 (2003), 261–270.

Andreas Dengel. 2016. Digital Co-Creation and Augmented Learning. In *Proceedings of the 11th International Knowledge Management in Organizations Conference on The changing face of Knowledge Management Impacting Society*. ACM, 3.

Stefan Dernbach, Barnan Das, Narayanan C Krishnan, Brian L Thomas, and Diane J Cook. 2012. Simple and complex activity recognition through smart phones. In *Intelligent Environments (IE), 2012 8th International Conference on*. IEEE, 214–221.

Nikolaos Dimakis, John K Soldatos, Lazaros Polymenakos, Pascal Fleury, Jan Curín, and Jan Kleindienst. 2008. Integrated development of context-aware applications in smart spaces. *IEEE Pervasive Computing* 7, 4 (2008), 71–79.

Douglas C Engelbart. 1962. Augmenting human intellect: a conceptual framework. (1962).

Alircza Fathi, Jessica K Hodgins, and James M Rehg. 2012. Social interactions: A first-person perspective. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1226–1233.

F Foerster, M Smeja, and J Fahrenberg. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior* 15, 5 (1999), 571–583.

Alan Gevins, Michael E Smith, Harrison Leong, Linda McEvoy, Susan Whitfield, Robert Du, and Georgia Rush. 1998. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40, 1 (1998), 79–91.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

Sabine Hoffmann, Helga Tauscher, Andreas Dengel, Shoya Ishimaru, Sheraz Ahmed, Jochen Kuhn, Carina Heisel, and Yutaka Arakawa. 2016. Sensing thermal stress at office workplaces. In *Proceedings of the 5th International Conference on Human-Environment Systems ICHES 2016 Nagoya*.

Shoya Ishimaru, Syed Saqib Bukhari, Carina Heisel, Jochen Kuhn, and Andreas Dengel. 2016a. Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 1041–1045.

Shoya Ishimaru and Koichi Kise. 2015. Quantifying the mental state on the basis of physical and social activities. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 1217–1220.

Shoya Ishimaru, Kai Kunze, Koichi Kise, and Andreas Dengel. 2016b. The wordometer 2.0: estimating the number of words you read in real life using commercial EOG glasses. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 293–296.

Shoya Ishimaru, Kai Kunze, Koichi Kise, and Masahiko Inami. 2014a. Position Paper: Brain Teasers - Toward Wearable Computing That Engages Our Mind. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 1405–1408.

Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014b. In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. In *Proceedings of the 5th Augmented Human International Conference*. ACM, Article 15, 4 pages.

Shoya Ishimaru, Kai Kunze, Katsuma Tanaka, Yuji Uema, Koichi Kise, and Masahiko Inami. 2015. Smart Eyewear for Interaction and Activity Recognition. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 307–310.

Shoya Ishimaru, Kai Kunze, Yuji Uema, Koichi Kise, Masahiko Inami, and Katsuma Tanaka. 2014. Smarter Eyewear: Using Commercial EOG Glasses for Activity Recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 239–242.

Daniel Kahneman and Jackson Beatty. 1966. Pupil diameter and load on memory. *Science* 154, 3756 (1966), 1583–1585.

Jochen Kuhn, Paul Lukowicz, Michael Hirth, Andreas Poxrucker, Jens Weppner, and Junaid Younas. 2016. gPhysics–Using Smart Glasses for Head-Centered, Context-Aware Learning in Physics Experiments. *Transactions on Learning Technologies* (2016).

Kai Kunze, Susana Sanchez, Tilman Dingler, Olivier Augereau, Koichi Kise, Masahiko Inami, and Terada Tsutomu. 2015. The augmented narrative: toward estimating reader engagement. In *Proceedings of the 6th Augmented Human International Conference*. ACM, 163–164.

David G Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on Computer vision*, Vol. 2. Ieee, 1150–1157.

Minghuang Ma, Haoqi Fan, and Kris M Kitani. 2016. Going deeper into first-person activity recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. 1894–1903.

Susan Michie, Charles Abraham, Craig Whittington, John McAteer, and Sunjai Gupta. 2009. Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology* 28, 6 (2009), 690–701.

Saleh Mozaffari, Saqib Bukhari, Andreas Dengel, Pascal Klein, and Jochen Kuhn. 2016. A Study on Representational Competence in Physics Using Mobile Remote Eye Tracking Systems, Proceedings. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*.

Ayano Okoso, Takumi Toyama, Kai Kunze, Joachim Folz, Marcus Liwicki, and Koichi Kise. 2015. Towards Extraction of Subjective Reading Incomprehension: Analysis of Eye Gaze Features. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1325–1330.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372.

Robert M Sapolsky. 1994. *Why zebras don't get ulcers*. WH Freeman New York.

Mathias Stager, Paul Lukowicz, and Gerhard Troster. 2004. Implementation and evaluation of a low-power sound-based user activity recognition system. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, Vol. 1. IEEE, 138–141.

KE Stanovich and AE Cunningham. 1998. What reading does for the mind. *American Education Journal* (1998).

Julian Steil and Andreas Bulling. 2015. Discovery of everyday human activities from long-term visual behaviour using topic models. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 75–85.

Mark Weiser. 1993. Some computer science issues in ubiquitous computing. *Commun. ACM* 36, 7 (1993), 75–84.

Mark Weiser and John Seely Brown. 1997. The coming age of calm technology. In *Beyond calculation*. Springer, 75–85.

Kazuyo Yoshimura, Kai Kunze, and Koichi Kise. 2015. The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents. In *Proceedings of the 13th International Conference on Document Analysis and Recognition*. 251–255.