Mouse Activity by Facial Expressions Using Ensemble Method

Anandhi.P[1], Gayathri.V[2]

[1]ME-Student, [2]Assistant Professor, Department of CSE, Srinivasan Engineering College, Perambalur, Tamil Nadu, India,

Abstract: *This graduation project aims to present an application that is able of replacing the traditional mouse with the human face as a new way to interact with the computer. Human activity recognition has become an important application area for pattern recognition. Here we focus on design of vision-based perceptual user interfaces. The concept of the second order change detection is implemented which sets the base for designing complete face-operated control systems ,where the mouse is replaced by 'nose' as pointer and 'blinking eyes' for clicking option. To select a robust facial feature, we use the pattern recognition paradigm of treating features. Using a support vector machine (SVM) classifier and person-independent (leave-one person-out) training, we obtain an average precision of 76.1 percent and recall of 70.5 percent over all classes and participants. The work demonstrates the promise of eye based activity recognition (EAR) and opens up discussion on the wider applicability of EAR to other activities that are difficult, or even impossible, to detect using common sensing modalities. The visual based human computer interaction is probably the most widespread area in HCI research. Different aspects of human responses can be recognized as a visual signal. One of the main research areas in this section is through Facial Expression Analysis and Gaze Detection (Eyes Movement Tracking).*

Keywords—Ubiquitous computing, feature evaluation and selection, pattern analysis, signal processing.

I. Introduction

The applications of any computer vision system require face tracking to be fast, affordable and, most importantly, precise and robust. In particular, the precision should be sufficient to control a cursor, while the robustness should be high enough to allow a user the convenience and the flexibility of head motion. A few hardware companies have developed hands-free mouse replacements. In particular, in accessibility community, several companies developed products which can track a head both accurately and reliably. These products however either use dedicated software or use structured environment (e.g. markings on the user's face) to simplify the tracking process. At the same time, recent advances in hardware, invention of fast USB and USB2 interfaces, falling camera prices, an increase of computer power brought a lot of attention to the real-time face tracking problem from the computer vision community.

The approaches to vision-based face tracking can be divided into two classes: image-based and feature-based approaches. Image-based approaches use global facial cues such as skin color, head geometry and motion. They are robust to head rotation and scale and do not require high quality images. In order to achieve precise and smooth face tracking, feature-based approaches are used. These approaches are based on tracking individual facial features. These features can be tracked with pixel accuracy, which allows one to convert their positions to the cursor position. This is the reason why vision-based games and interfaces are still not common.

Excellent pupil localization and blinking detection performance is reported for systems which use structured infrared light. These systems register the light reflection in the users' eyes in order to locate the pupils. Good results in eye blinking detection are also reported for the systems based on high-resolution video cameras which can capture eye pupils with pixel accuracy. It is intensively used by mechanisms of visual attention employed in biological vision systems. A common approach to detecting moving objects in video is based on detecting the intensity change between two consecutive frames caused by the object motion. The simplest way of detecting such a change, which will refer to as the first order change, is to use the binaries threshold absolute difference between two consecutive video frames. This is what has been used so far by other systems to detect blinks.

Computer vision is the science and technology of machines that see. As a scientific discipline, computer vision is concerned with the theory and technology for building artificial systems that obtain information from images or multi-dimensional data. Examples of applications of computer vision systems include systems for

i. Controlling processes (e.g. an industrial robot or an autonomous vehicle).
ii. Detecting events (e.g. for visual surveillance).

The organization of a computer vision system is highly application dependent. Some systems are stand-alone applications which solve a specific measurement or detection problem, while other constitute a sub-system of a larger design which, for example, also contains sub-systems for control of mechanical actuators, planning, information databases, man-machine interfaces, etc. The specific implementation of a computer vision system

also depends on if its functionality is pre-specified or if some part of it can be learned or modified during operation.

Perceptual Vision Technology is the technology for designing systems, referred to as Perceptual Vision Systems (PVS) that use visual cues of the user, such as the motion of the face, to control a program. The main application of this technology is seen in designing intelligent hands-free Perceptual User Interfaces to supplement the conventional input devices such as mouse, joystick, track ball etc.

II.   RELATED WORKS

It analysis the eye movements of people in transmit in an everyday environment using a wearable electrooculographic (EOG) system. It compares three approaches for continuous recognition of reading activities: a string matching algorithm which exploits typical characteristics of reading signals, such as saccades and fixations.

It can store data locally for long-term recordings or stream processed EOG signals to a remote device over Bluetooth. It describes how eye gestures can be efficiently recognized from EOG signals for HCI purposes. In an experiment conducted with 11 subjects playing a computer game we show that 8 eye gestures of varying complexity can be continuously recognized with equal performance to a state-of-the art video-based system. Physical activity leads to the EOG signal. It describes how these EOG signal can be removed using an adaptive filtering scheme and characterize this approach on a 5-subject dataset. In addition to HCI, it discusses how this paves the way for EOG-based context-awareness, and eventually to the assessment of cognitive processes.

It proposes a new biased discriminate analysis (BDA) using composite vectors for eye detection. A composite vector consists of several pixels inside a window on an image. The covariance of composite vectors is obtained from their inner product and can be considered as a generalization of the covariance of pixels.

The design, implementation and evaluation of a novel eye tracker for context-awareness and mobile HCI applications. In contrast to common systems using video cameras, this compact device relies on Electrooculography(EOG). It consists of goggles with dry electrodes integrated into the frame and a small pocket-worn component with a DSP for real-time EOG signal processing. It describes how eye gestures can be efficiently recognized from EOG signals for HCI purposes.

Eye tracking research in human-computer interaction and experimental psychology traditionally focuses on stationary devices and a small number of common eye movements. The advent of pervasive eye tracking promises new applications, such as eye-based mental health monitoring or eye based activity and context recognition.

It might require further research on additional eye movement types such as smooth pursuits and the vestibule-ocular reflex as these movements have not been studied as extensively as saccades, fixations and blinks. It develop a set of basic signal features that we extract from the collected eye movement data and show that a feature-based approach has the potential to discriminate between saccades, smooth pursuits, and vestibule-ocular reflex movements.

It presents a user-independent emotion recognition method with the goal of recovering affective tags for videos using electroencephalogram (EEG), pupillary response and gaze distance. It selected 20 video clips with extrinsic emotional content from movies and online resources. Then, EEG responses and eye gaze data were recorded from 24 participants while watching emotional video clips.

The analysis of human activities in videos is an area with increasingly important consequences from security and surveillance to entertainment and personal archiving. Several challenges at various levels of processing robustness against errors in low-level processing, view and rate invariant representations at mid-level processing and semantic representation of human activities at higher-level processing make this problem hard to solve.

Gesture recognition pertains to recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head, and/or body. It is of utmost importance in designing an intelligent and efficient human computer interface. The applications of gesture recognition are manifold, ranging from sign language through medical rehabilitation to virtual reality.

It provides a survey on gesture recognition with particular emphasis on hand gestures and facial expressions. Applications involving hidden Markov models, particle filtering and condensation, finite-state machines, optical flow, skin color, and connectionist models are discussed in detail. Existing challenges and future research possibilities are also highlighted.

The introduction of eye movement analysis as a new sensing modality for activity recognition. The development and characterization of new algorithms for detecting three basic eye movement types from EOG signals (saccades, fixations, and blinks) and a method to assess repetitive eye movement patterns. The development and evaluation of 90 features derived from these eye movement types, and the implementation of a method for continuous EAR and its evaluation using a multi participant EOG data set involving a study of five real-world office activities.

All parameters of the saccade, fixation and blink detection algorithms were fixed to values common to all participants; the same applies to the parameters of the feature selection and classification algorithms. Despite person-independent training, six out of the eight participants returned best average precision and recall values of between 69% and 93% using the SVM classifier. However, two participants returned results that were lower than 50%. Participant four had zero correct classifications for both reading and copying, and close to zero recall for writing; participant five had close to zero recall for reading and browsing. On closer inspection of the raw EOG data, it turned out that in both cases the signal quality was much worse compared to the others. The signal amplitude changes for saccades and blinks - upon which feature extraction and thus classification performance heavily depend - were not distinctive enough to be reliably detected.

These problems may be solved, in part, by an annotation process that uses video and precise gaze tracking. Activities from the current scenario could be redefined at a smaller time scale, breaking web-browsing into smaller activities such as "use scrollbar", "read", "look at image", "type", and so on. This would also allow us to investigate more complicated activities outside the office. An alternative route would be to study activities at larger time scales, to perform situation analysis rather than recognition of specific activities. Longer term eye movement features, for example the average eye movement velocity and blink rate over one hour, might be useful in revealing whether a person is walking along an empty or busy street, whether they are at their desk working, or whether they are at home watching television. Again, annotation will be an issue, but one that may be alleviated using unsupervised or self-labeling methods.

III. TECHNIQUES USED

i) Support Vector Machine algorithm
SVM are a new type of maximum margin classifiers: In "learning theory" there is a theorem stating that in order to achieve minimal classification error the hyper plane which separates positive samples from negative ones should be with the maximal margin of the training sample and this is what the SVM is all about.
The data samples that are closest to the hyper plane are called support vectors. The hyper plane is defined by balancing its distance between positive and negative support vectors in order to get the maximal margin of the training data set.
Support vector machine are supervised learning models with associated learning algorithms that analyze data and recognize patterns used for classification and regression analysis. The basic SVM takes a set of input data and predicts for each input.

ii)Ensemble Method

Ensemble method is the one of data mining concept. It is used to separate the relevant data and irrelevant data. Ensemble method is mainly used for retrieving the relevant data. Here ensemble method is mainly used for taking the nose and eye expressions from the Facial actions.
The eyes are tracked to detect their blinks, where the blink becomes the mouse click. The tracking process is based on predicting the place of the feature in the current frame based on its location in previous ones; template matching and some heuristics are applied to locate the feature's new coordinates.

iii) Face Detection Algorithm

It applies blink detection in the eye's ROI before finding the eye's new exact location. The blink detection process is run only if the eye is not moving because when a person uses the mouse and wants to click, he moves the pointer to the desired location, stops, and then clicks; so basically the same for using the face: the user moves the pointer with the tip of the nose, stops, then blinks. To detect a blink we apply motion detection in the eye ROI; if the number of motion pixels in the ROI is larger than a certain threshold we consider that a blink was detected because if the eye is still, and we are detecting a motion in the eye ROI, that means that the eyelid is moving which means a blink. In order to avoid multiple blinks detection while they are a single blink the user can set the blink length, so all blinks which are detected in the period of the first detected blink are omitted.

Architecture Diagram
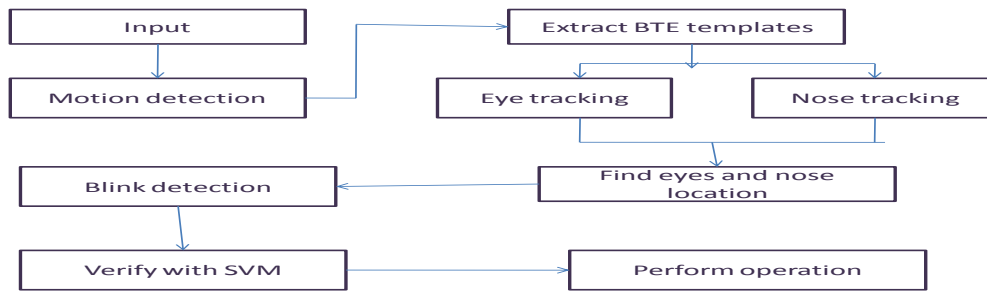
To detect motion in a certain region we subtract the pixels in that region from the same pixels of the previous frame, and at a given location (x,y); if the absolute value of the subtraction was larger than a certain threshold, we consider a motion at that pixel .If a left/right blink was detected, the tracking process of the left/right eye will be skipped and its location will be considered as the same one from the previous frame (because blink detection is applied only when the eye is still). Eyes are tracked in a bit different way from tracking the nose tip and the BTE, because these features have a steady state while the eyes are not (e.g. opening, closing, and blinking) To achieve better eyes tracking results we will be using the BTE (a steady feature that is well tracked) as our reference point; at each frame after locating the BTE and the eyes, we calculate the relative positions of the eyes to the BTE; in the next frame after locating the BTE we assume that the eyes have kept their relative locations to it, so we place the eyes‟ ROIs at the sample relative positions to the new BTE. To find the eye new template in the ROI we combined two methods: the first used template matching, the second searched in the ROI for the darkest5*5 regions (because the eye pupil is black), then we used the mean between the two found coordinates as the new eye location.

IV.   EXPERIMENTAL EVOLUTION

We designed a study to establish the feasibility of EAR in a real-world setting. Our scenario involved five office-based activities—copying a text, reading a printed paper, taking handwritten notes, watching a video, and browsing the Web—and periods during which participants took a rest (the NULL class).We chose these activities for three reasons. First, they are all commonly performed during a typical working day. Second, they exhibit interesting eye movement patterns that are both structurally diverse and have varying levels of complexity. We believe they represent the much broader range of activities observable in daily life. Finally, being able to detect these activities using on-body sensors such as EOG may enable novel attentive user interfaces that take into account cognitive aspects of interaction such as user interruptibility or level of task engagement.

CLASSIFICATION PERFORMANCE

SVM classification was scored using a frame-by-frame comparison with the annotated ground truth. For specific results on each participant or on each activity, class-relative precision and recall were used.Figure 10 shows the average precision and recall, and the corresponding number of features selected for each participant. The number of features used varied from only nine features (P8) up to 81 features (P1). The mean performance over all participants was 76.1 percent precision and 70.5 percent recall. P4 reported the worst result, with both precision and recall below 50 percent. In contrast, P7 achieved the best result, indicated by recognition performance in the 80s and 90s and using a moderate-sized feature set.
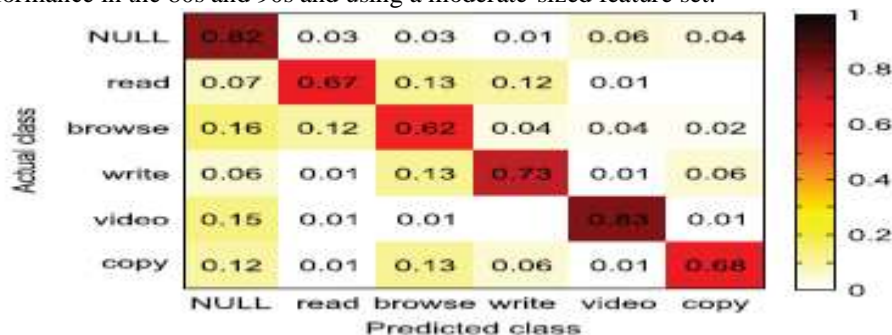

Figure 1: Summed confusion matrix from all participants, normalized across ground truth rows.

All of the remaining eight participants (two females and six males), aged between 23 and 31 years (mean ¼ 26:1, sd ¼ 2:4) were daily computer users, reporting 6 to 14 hours of use per day (mean ¼ 9:5, sd ¼ 2:7). They were asked to follow two continuous sequences, each composed of five different, randomly ordered activities, and a period of rest. For these, no activity was required of the participants but they were asked not to engage in any of the other activities. Each activity (including NULL) lasted about five minutes, resulting in a total data set of about eight hours.

EOG signals were picked up using an array of five 24 mm Ag/AgCl wet electrodes from Tyco Healthcare placed around the right eye. The horizontal signal was collected using one electrode on the nose and another directly across from this on the edge of the right eye socket. The vertical signal was collected using one electrode above the right eyebrow and another on the lower edge of the right eye socket. The fifth electrode, the signal reference, was placed in the middle of the forehead. Five participants (two females and three males) wore spectacles during the experiment. For these participants, the nose electrode was moved to the side of the left eye to avoid interference with the spectacles.
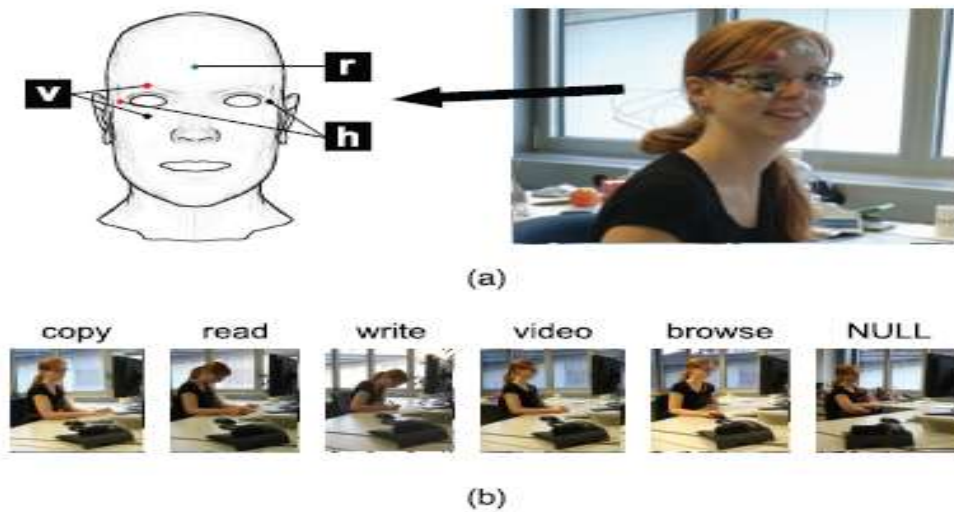
Figure 2:(a) Electrode placement for EOG data collection (h: horizontal, v: vertical, and r: reference). (b) Continuous sequence of five typical office activities: copying a text, reading a printed paper, taking handwritten notes, watching a video, browsing the Web, and periods of no specific activity (the NULL class).

The experiment was carried out in an office during regular working hours. Participants were seated in front of two adjacent 17 inch flat screens with a resolution of 1;280 ,1;024 pixels on which a browser, a video player, a word processor, and text for copying were on-screen and ready for use. Free movement of the head and upper body was possible throughout the experiment. Classification and feature selection were evaluated using a leave-one-person-out scheme: We combined the data sets of all but one participant and used this for training; testing was done using both data sets of the remaining participant. This was repeated for each participant. The resulting train and test sets were standardized to have zero mean and a standard deviation of one. Feature selection was always performed solely on the training set. The two main Parameters of the SVM algorithm, the cost C and the tolerance of termination criterion were fixed to C ¼ 1 and _ ¼ 0:1. For each leave-one-person-out iteration, the prediction vector returned by the SVM classifier was smoothed using a sliding majority window. Its main parameter, the window size W_{sm}, was obtained using a parameter sweep and fixed at 2.4 s.

Segmentation the task of spotting individual activity instances in continuous data remains an open challenge in activity recognition. We found that eye movements can be used for activity segmentation on different levels depending on the timescale of the activities. The lowest level of segmentation is that of individual saccades that define eye movements in different directions "left," "right," and so on. An example for this is the end-of-line "carriage return" eye movement performed during reading. The next level includes more complex activities that involve sequences composed of a small number of saccades. For these activities, the wordbook analysis proposed in this work may prove suitable.

This is indicative of a task where the user might be concentrated on a relatively small field of view, but follows a typically unstructured path. Similar examples outside the current study might include interacting with a graphical user interface or watching television at home. Writing is similar to reading in that the eyes follow a structured path, albeit at a slower rate. Writing involves more eye "distractions" when the person looks up to think for example. Browsing is recognized less well over all participants (average precision 79 percent and

recall 63 percent) but with a large spread between people. A likely reason for this is that it is not only unstructured, but also it involves a variety of sub activities—including reading—that may need to be modeled. The copy activity, with an average precision of 76 percent and a recall of 66 percent, is representative of activities with a small field of view that include regular shifts in attention (in this case, to another screen). A comparable activity outside the chosen office scenario might be driving, where the eyes are on the road ahead with occasional checks to the side mirrors. Finally, the NULL class returns a high recall of 81 percent. However, there are many false returns (activity false negatives) for half of the participants, resulting in a precision of only 66 %.
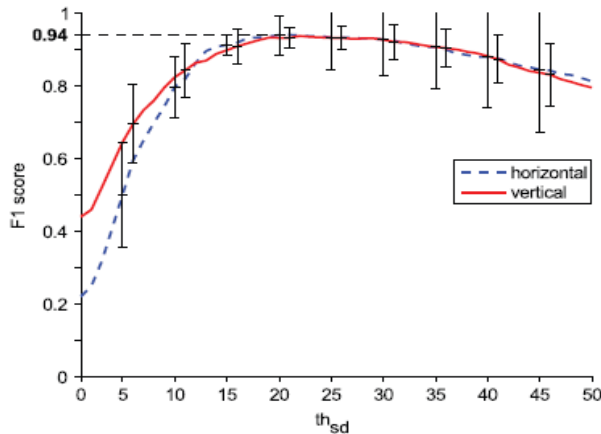

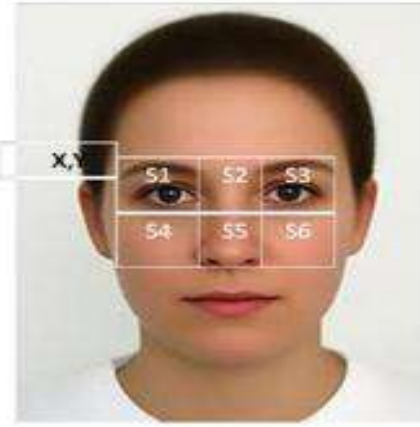**Figure 3: Comparison of threshold th$_{sd}$,F1 score**   **Figure 4:SSR Filter**

Two participants, however, returned results that were lower than 50 percent. On closer inspection of the raw eye movement data, it turned out that for both the EOG, signal quality was poor. Changes in signal amplitude for saccades and blinks—upon which feature extraction and thus recognition performance directly depend—were not distinctive enough to be reliably detected. As was found in an earlier study [6], dry skin or poor electrode placement are the most likely culprits. Still, the achieved recognition performance is promising for eye movement analysis to be implemented in real-world applications, for example, as part of a reading assistant, or for monitoring workload to assess the risk of burnout syndrome. For such applications, recognition performance may be further increased by combining eye movement analysis with additional sensing modalities.

RESULTS

These problems may be solved, in part, by using video and gaze tracking for annotation. Activities from the current Scenario could be redefined at a smaller timescale, breaking browsing into smaller activities such as "use scroll bar," "read," "look at image," or "type." This would also allow us to investigate more complicated activities outside the office. An alternative route is to study activities at larger timescales, to perform situation analysis rather than recognition of specific activities. Long-term eye movement features, e.g., the average eye movement velocity and blink rate over one hour, might reveal whether a person is walking along an empty or busy street, whether they are at their desk working, or whether they are at home watching television. Annotation will still be an issue, but one that maybe alleviated using unsupervised or self-labeling methods.

V.   Conclusion

The project designed to match the mouse operations with facial expressions was implemented with first few modules like "Frame Grabber" module which is used to take video inputs converts them into the frames and those frames sent into the modules like "Six-Segmented Rectangular Filter" and "Support Vector Machine" to detect the regions of the face. The exact operation like the mouse movement and mouse clicks may match with eye blinks in the feature work. Some trackers used in human-computer interfaces for people with disabilities require the user to wear special transmitters, sensors, or markers. Such systems have the disadvantage of potentially being perceived as a conspicuous advertisement of the individual's disability. Since the eye Blinks uses only a camera placed on the computer monitor, it is completely non intrusive. The absence of any accessories on the user make the system easier to configure and therefore more user-friendly in a clinical or academic environment.

References

[1]   S. Mitra and T. Acharya, "Gesture Recognition: A Survey," IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev., vol. 37, no. 3, pp. 311-324, May 2007.

[2]   P. Turaga, R. Chellappa, V.S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," IEEETrans. Circuits and Systems for Video Technology, vol. 18, no. 11, pp. 1473-1488, Nov. 2008.

[3]   B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C.J. Bula, and P. Robert, "Ambulatory System for Human Motion Analysis Using a Kinematic Sensor: Monitoring of Daily Physical Activity in the Elderly," IEEE Trans. Biomedical Eng., vol. 50, no. 6, pp. 711-723, June 2003.

[4]   J.A. Ward, P. Lukowicz, G. Tro¨ster, and T.E. Starner, "Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1553-1567, Oct. 2006.

[5]   N. Kern, B. Schiele, and A. Schmidt, "Recognizing Context for Annotating a Live Life Recording," Personal and Ubiquitous Computing, vol. 11, no. 4, pp. 251-263, 2007.

[6]   A. Bulling, J.A. Ward, H. Gellersen, and G. Tro¨ster, "Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography," Proc. Sixth Int'l Conf. Pervasive Computing, pp. 19-37, 2008.

[7]   S.P. Liversedge and J.M. Findlay, "Saccadic Eye Movements and Cognition," Trends in Cognitive Sciences, vol. 4, no. 1, pp. 6-14, 2000.

[8]   J.M. Henderson, "Human Gaze Control during Real-World Scene Perception," Trends in Cognitive Sciences, vol. 7, no. 11, pp. 498-504, 2003.

[9]   A. Bulling, D. Roggen, and G. Tro¨ster, "Wearable EOG Goggles: Seamless Sensing and Context-Awareness in Everyday Environments," J. Ambient Intelligence and Smart Environments, vol. 1, no. 2,pp. 157-171, 2009.

[10]   A. Bulling, J.A. Ward, H. Gellersen, and G. Tro¨ster, "Eye Movement Analysis for Activity Recognition," Proc. 11th Int'lConf. Ubiquitous Computing, pp. 41-50, 2009.

[11]   Q. Ding, K. Tong, and G. Li, "Development of an EOG (Electro- Oculography) Based Human-Computer Interface," Proc. 27th Int'l Conf. Eng. in Medicine and Biology Soc., pp. 6829-6831, 2005.

[12]   Y. Chen and W.S. Newman, "A Human-Robot Interface Based on Electrooculography," Proc. IEEE Int'l Conf. Robotics and Automation, vol. 1, pp. 243-248, 2004.

[13]   W.S. Wijesoma, K.S. Wee, O.C. Wee, A.P. Balasuriya, K.T. San, and K.K. Soon, "EOG Based Control of Mobile Assistive Platforms for the Severely Disabled," Proc. IEEE Int'l Conf. Robotics and Biomimetics, pp. 490-494, 2005.

[14]   R. Barea, L. Boquete, M. Mazo, and E. Lopez, "System for Assisted Mobility Using Eye Movements Based on Electrooculography," IEEE Trans. Neural Systems and Rehabilitation Eng., vol. 10, no. 4,pp. 209-218, Dec. 2002.

[15]   M.M. Hayhoe and D.H. Ballard, "Eye Movements in Natural Behavior," Trends in Cognitive Sciences, vol. 9, pp. 188-194, 2005.

[16]   L. Dempere-Marco, X. Hu, S.L.S. MacDonald, S.M. Ellis, D.M. Hansell, and G.-Z. Yang, "The Use of Visual Search for Knowledge Gathering in Image Decision Support," IEEE Trans. Medical Imaging, vol. 21, no. 7, pp. 741-754, July 2002.

AUTHORS PROFILE

P.Anandhi received the B.E Degree CSE and now he is an M.E student in the Department Of Computer Science & Engineering, Srinivasan Engineering College – Dhanalakshmi Srinivasan Group of Institutions, Perambalur,TN, India.
Her research interest includes Data mining, Wireless Networks and Image Proceesing.

V.Gayathri is working as Assistant Professor at Srinivasan Engineering College – Dhanalakshmi Srinivasan Group of Institutions, Perambalur,TN, India.
Her main research interest includes Data Mining and Neural Networks.