COMMUNICATION AID FOR PARALYZED PEOPLE BY EYELID TRACKING

[1]Abijith Vignesh T, [2]Aishwarya P K, [3]Dr.S. Sobitha Ahila
[1]Student, [2]Student, [3]Associate Professor
[1] Computer Science and Engineering Department,
[1] Easwari Engineering College, Chennai, India

ABSTRACT: One of the major form of paralysis that exists is the Locked-in syndrome or pseudo coma where a patient is aware but cannot move or communicate verbally, due to complete paralysis of nearly all voluntary muscles in the body except for vertical eye movements and blinking. The individual is conscious and sufficiently intact cognitively to be able to communicate with eye movements. These people once trained in the international Morse code with the blinking of their eyes, the system would be able to interpret their blinks into respective letters of the English alphabets converting them into text and synthesising them into voice and alert messages. The system monitors their eyelid movements by detecting their face and eyes. These eye blinks are tracked by a camera in real time and converted to text. This acts as a boon to the people affected by pseudo coma in helping them communicate important information to others around them.

Keywords: Morse code, Pseudo coma, eye detection speech synthesizer.

1 INTRODUCTION

Artificial Intelligence have come a long way since it is inception and machine learning are one of the major subset belonging to it. Machine learning helps the system to learn on its own with data, not requiring explicit programming. Machine learning have several applications in the modern world, some including virtual assistants, prediction, social media services, spam filtering, online customer support and product recommendation.

Machine learning is a subfield of artificial intelligence. The goal of machine learning is to understand the data which can fit into the modules and can be understood and utilized by the users. Machine learning differs from traditional computational approaches as the algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve whereas in machine learning, the algorithms instead allow for computers to train on data inputs and use statistical analysis to output the values that fall within a specific range. As the result of this, machine learning facilitates computers in building models from sample data to automate decision-making processes based on data inputs. Machine learning is widely used in all the fields like facial recognition technology allows social media platforms to help users tag and share photos of friends, Optical Character Recognition (OCR) technology converts images of text into movable type, recommendation engines empowered by machine learning suggests what movies or television shows to watch next based on user preferences ,self-driving cars that rely on machine learning to navigate may soon be available to consumers.

In supervised learning, the computer is provided with example inputs that are labelled with their desired outputs. The purpose of this method is for the algorithm to be able to “learn” by comparing its actual output with the “taught” outputs to find errors and modify the model accordingly. Supervised learning therefore uses patterns to predict label values on additional unlabelled data. We will be using this approach to train our module for the eyes detection and for mapping of the Morse code with the eye blinks.

Locked-In Syndrome otherwise known as pseudo coma, is a condition in which a person there is complete paralysis of all voluntary muscles except for the ones that control the movement of the eyes i.e. the person is aware but cannot communicate verbally due to complete paralysis of nearly all voluntary muscles in the body except for vertical eye movements. The people affected by this are conscious and aware, with no loss of cognitive function. Locked-in syndrome is caused by damaged to the pons, a part of the brainstem that contains nerve fibres that relay information to other areas of the brain. People affected by this often lose their quality of life in interacting with other people often leading them to lead life in loneliness.

It is very difficult for people affected by locked in syndrome to communicate with their helpers leading them to not receiving the necessary attention they require. Our paper aims to provide a suitable system to bridge the communication gap between the pseudo coma people and normal people. Since the affected people can move their eyes to blink, our system will be able to detect their eyelid movements through the webcam available in the system. The International Morse code can be used as the standard of communication through their blinks. Every series of blink is captured and mapped corresponding to associated letter present in the Morse code which is then synthesized into text and voice.

To detect the face and eyes of a person using the webcam, we use the Histogram of Oriented Gradients which is a feature descriptor and SVM classifiers. With the help of OpenCV we can detect the faces and eyes of people using these facial descriptors. Support vector machine is formulated to solve the individuals pattern reorganization problem. The face reorganization is done by modifying the interpretation of the output SVM classifier and representation of the facial images. This algorithm is also used for identification and verification of the face in the image which estimates the identity of the persons face from the database.

Morse code is a method of transmitting text information as a series of on-off tones, lights, or clicks. In our system we’ll be using blinks to transmit it. Each alphabet in the English language has a series of dots and dashes to represent each letter as shown in fig (1). The dot represents one unit while the dash represents 3 units, the space between letter is 3 units. We use the same principle in our system where a dot is represented by a small eye blink, a dash is represented by a long eye blink and each unit represents a second.

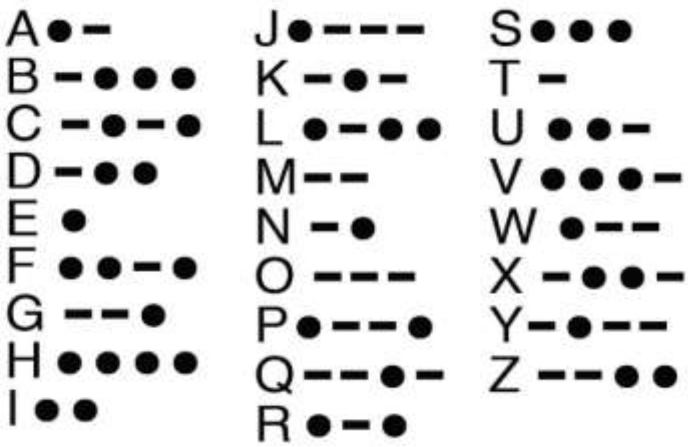


Fig.1: Morse code

2 LITERATURE SURVEY

"Biometric Recognition via Eye Movements: Saccadic Vigor and Acceleration Cues", IOANNIS RIGAS and OLEG KOMOGORTSEV [1], this paper is based on the concepts of Saccadic Vigor and Acceleration Cues for tracking the eye movements. Saccadic eye movements are executed via the combined application of agonist–antagonist forces from the oculomotor muscles. The research field of eye movement-driven biometrics explores the extraction of individual-specific characteristics from eye movements and their employment for recognition purposes. Thus, inspecting the acceleration profile of a saccade can provide valuable information regarding the underlying sources of the applied forces. The descriptive characteristics from the acceleration profiles of saccades and modeled the asymmetries of the acceleration and deceleration phases were extracted. The biometric performance that can be achieved from the integration of the saccadic vigor and acceleration features in the original CEM-B framework was evaluated by ANOVA analyses across selected fusion schemes using the data from the 20 random partitions .The experiments were conducted with 322 subjects (171 male/151 female), aged 18–46 (M = 22, SD = 4.2). Of those subjects, 163 had corrected vision (67 glasses/96 contact lenses). The statistical significance of these differences were examined by, the subjects performed two recording sessions each, separated by a time interval of approximately 20 minutes for each visual stimulus, leading to the collection of 1,866 unique recordings. They achieved a accuracy of 36% due to the limited number of data sets. Other issues include variable lighting conditions and environmental noise thus requiring a high-grade eye tracking device. The findings of this system support that the incorporation of the proposed features can lead to an increase in biometric accuracy and improve robustness when using different types of visual stimulus.

"Eye PACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays", MOHAMED KHAMIS, DANIEL BUSCHEK, TOBIAS THIERON, FLORIAN ALT, ANDREAS BULLING, [2] this paper presents an Eye Pact method that compensates the input error caused by parallax on public displays. This method uses a display mounted depth camera to detect the uses 3D eye position in front of the display. Eye PACT determines which pair of eyes to correct for depending on the closest arm to the display at the time of the touch event. This is done by utilizing skeletal tracking which provides the position of users' joints in the 3D space in front of the display. The user's pupils are continuously tracked by employing the approach by Smolyansky et al., which augments the Active Appearance Model (AAM) algorithm by using depth-based face segmentation detected by an RGBD camera. EyePACT determines the pair of eyes to correct for depending on the closest arm to the display at the time of the touch event. This is done by utilizing skeletal tracking, which provides the position of users' joints in the 3D space in front of the display. The positions of the user's hand, the joints are then used to determine which user is closest to the touchscreen at the time of the touch action. Once the user is determined, EyePACT corrects for parallax from that user's perspective. The speed and direction of the moving target is randomly decided. EyePACT can be exploited to support interaction.

"Training and Testing Object Detectors with Virtual Images", Yonglin Tian, Xuan Li, Kunfeng Wang, Fei-Yue Wang [3], this paper presents a way to design artificial scenes and automatically generate virtual images with precise annotations. All the images and annotations were stored in the form of PASCAL VOC. Deformable parts model was one of the most effective object detectors based on histogram of oriented gradient (HOG) before the resurgence of deep learning. Faster R-CNN is currently a state of the art approach and widely used in object detection. Faster R-CNN is done by initializing the with ImageNet pre-trained weights. Performance of the model is worse on all sub-datasets while a bigger rate of descent of AP occurred on ParallelEye in the testing phase, which resulted from the smaller average area of object of ParallelEye. In this work, a flexible approach to building artificial scenes is proposed. The main intention is to set up virtual datasets with specific features and diverse annotations to train and test the object detectors. The experiment can work on 8 – 12 frames per second to produce different virtual images with different annotations of work station which is efficient when compared with manual work. They have achieved a result that the virtual datasets are viable to improve the performance of the object detector trained with the real data sets. Occlusion was not intentionally introduced to get a better understanding of the effect of pose change on the trained models. And it witnessed a huger drop of the rate of AP after the deletion of occluded objects from the training set because it has a higher occlusion rate.

"Learning Discriminative Subspace Models for Weakly Supervised Face Detection", Qiaoying Huang, Chris Kui Jia, Xiaofeng Zhang, and Yunming Ye [4], In this module, a subspace-based generative model is proposed to select positive instances by minimizing rank of the coefficient matrix associated with the subspace models of object instances in images. Based on the subspace assumption of object categories an approach is proposed to learn generative subspace models of object categories in the weakly supervised setting. The discriminative ability is improved by an enable strategy which is proposed by employing multiple subspace models. This paper focuses on object detection learning using only image-level weak labels. To reduce the learning difficulty, block coordinate descent method is being used. A set of training images including both positive and negative images is given in the setting of weakly supervised object class learning. The task is to learn a model for target object category so that it can be used for classification, i.e., checking whether a test image contains any positive instance and detecting their locations. The empirical evaluations are divided into single subspace model and multiple subspace model, but it utilizes only one subspace model to directly label testing instances. The proposed Single+Subspace model can detect the instances that are comparably close to the true instances. To evaluate the model performance, it plots several false positive bounding boxes labeled by Single+Subspace. It is observed that when the obstacles block more important area of human face, the proposed Single+Subspace fails to label and localize the correct instance also fails to label correct instances when there exists more than one person in the same image. The average precision of Single+Subspace 59.5% is stable and the average accuracy of Single+Subspace is 60.7%.

"Detecting Gaze Towards Eyes in Natural Social Interactions and Its Use in Child Assessment", EUNJI CHONG and KATHA CHANDA, ZHEFAN YE, AUDREY SOUTHERLAND and NATANIEL RUIZ, REBECCA M. JONES, AGATA ROZGA∗ and JAMES M. REHG∗ [5], This paper proposes approach to eye contact detection during adult-child social interactions in which the adult wears a point-of-view camera which captures an egocentric view of the child's behavior. The system is fully automated for eye contact detection to solve the sub problems of end to end feature learning and pose estimation. This wearable eye tracker gives the location of point of regard of POV image but does not precisely define what gaze target is present at that location. This difficulty increases the complexity of the automated behavior measurement of the system in wearable eye tracker. The detectors output gives an idea about the rough scope of what the current training set deals with the face detection is done with supervised Descent Method for tracking facial landmarks. The facial landmarks are then used to estimate the head pose with three degrees of freedom- yaw, pitch and roll. The results of the facial landmark identification are used to crop the eye regions and extract Histogram of Oriented Gradients (HOG) features from the eye patches. The Pose-Implicit CNN, deep learning architecture that predicts eye contact while implicitly estimating the head pose is used. A deep learning system that detects moments of eye contact from egocentric videos of adult-child interactions while also implicitly estimating the head pose is presented. A basic problem with using any form of eye tracking to analyze face-to-face gaze is the need to identify the gaze target given the estimated gaze direction Landmark detection is successful only for 75.56% of frames, in comparison to 97.94% for face detection and . The precision-recall results at Intersection over Union (IoU) thresholds of 0.5 and 0.75. The main drawback of this module is that the examiner must wear glasses.

3 PROPOSED WORK

The proposed system consists of the following modules:

- Face/eye detection
- Conversion of eye blinks to text using Morse code
- Conversion of text to speech
- Alert module

3.1 FACE/EYE DETECTION

The face detection is used to find the facial landmarks which are used to localize and represent salient regions of the face like nose ,eyes ,ears ,mouth ,eye brows .The skin color detection is done to the image obtained and on which the noise reduction is done to minimize the grain structure of the medium .After which the face in the image is localized and the key facial structures are detected . Images are *manually labeled*, specifying **specific** *(x, y)*-coordinates of regions surrounding each facial structure. The facial landmark detector is used to estimate the location of ***68 (x, y)-coordinates*** that map to facial structures on the face. Once the face is identified, the index are marked. Histogram of Gradient(HOG) + Linear SVM object detector used to train highly accurate object classifiers — eyes. The linear support vector machine is trained on the positive and the negative samples so that we can detect and extract HOG descriptors for these samples. With that the eye coordinates are identified from the face index.

Fig.2: Visualizing the facial landmark coordinates.

3.2 CONVERSION OF EYE BLINKS TO TEXT USING MORSE CODE

From the output of the previous module 6(x, y) coordinates are marked around the identified eye (fig 3.). The Eye Aspect Ratio (EAR) is then found with the help of the following formula.

$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

The eye aspect ratio is then continuously monitored, and a blink is registered whenever it becomes zero (i.e. when the person closes their eyelids). The user is required to follow the Morse code chart to register a sequence of short and a long blink for the respective English alphabets. These blinks are stored until a 5 second pause is encountered after which it is mapped to the corresponding letter. The letters are then grouped together into a word whenever a 7 second pause is encountered.

From the output of the previous module 6(x, y) coordinates are marked around the identified eye (fig 3.). The Eye Aspect Ratio (EAR) is then found with the help of the following formula.

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

The eye aspect ratio is then continuously monitored, and a blink is registered whenever it becomes zero (i.e. when the person closes their eyelids). The user is required to follow the Morse code chart to register a sequence of short and a long blink for the respective English alphabets. These blinks are stored until a 5 second pause is encountered after which it is mapped to the corresponding letter. The letters are then grouped together into a word whenever a 7 second pause is encountered.

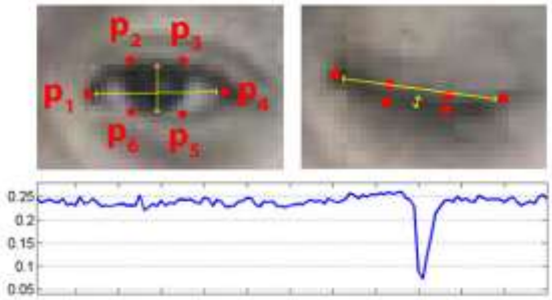


Fig.3: The open and closed eye coordinates ratio is calculated and plotted in the graph.

3.2 CONVERSION OF TEXT TO SPEECH

A text-to-speech (TTS) system converts plain text into speech. The plain text which is obtained from the previous output after the text normalization. The phonetic transcription is done to each word and divides and marks the words into prosodic units like phrases, clauses and sentence. This process is known as text to phoneme. Phonetic transcriptions and prosody information together make up the linguistic representation and the synthesizer converts the linguistic representation into sound.

3.3 ALERT MODULE

When the helper for the paralyzed person is not nearby, the paralyzed person can alert him through the system by sending an alert text message. The message which is generated by mapping with the Morse code from the paralyzed person is sent as a text message to the helper's mobile phone using the Twilio API for text messages in case the internet is not available.

4 FUNCTIONAL WORKFLOW

The functional workflow represents the functions of each module in a graphical manner. The eyes and face of the person is detected when they investigate the camera. The number of blinks is counted using the eye aspect ratio and stored until a pause is encountered. These set of blinks are then mapped to the corresponding alphabet using the Morse code. These letters are then grouped to form a word. Once the message is in the text form, a text message is sent to the caretaker on his/her mobile phone. The same message is also outputted as voice using the speech to text convertor, which is eventually sent as a text message to the helper.

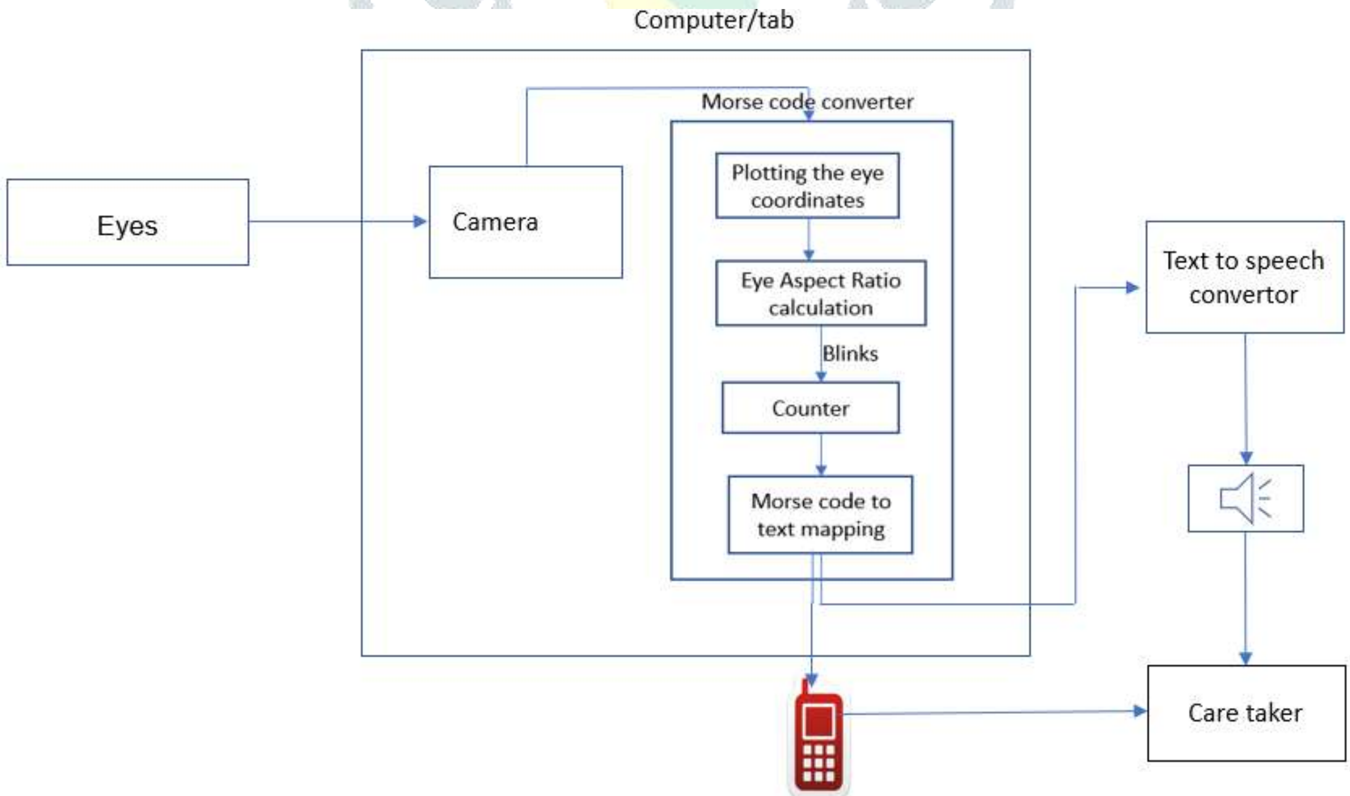
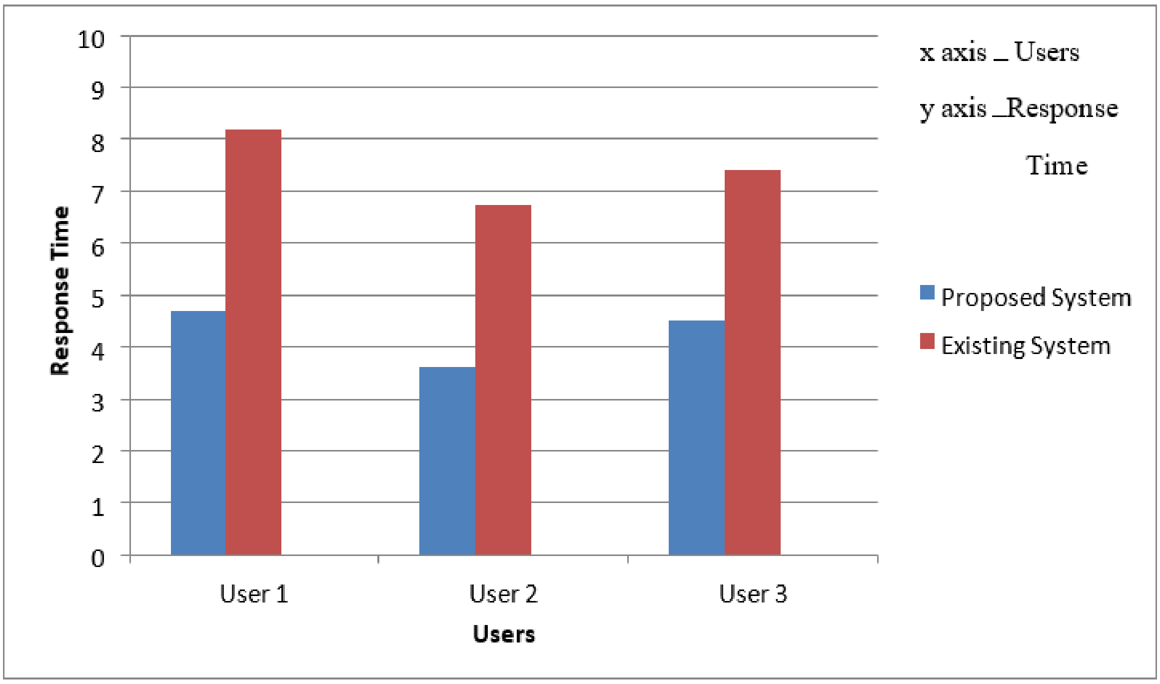


Fig.4: Functional workflow

5 RESULT

We are comparing the results of our system along with an earlier system which consists of a wearable glass fitted with infrared camera to detect the blinks. The analysis showed that the system proposed in this paper was much faster than the previous one. It also didn't have the need to use any kind of intrusive wearable devices like the glasses mentioned before.



Graph 1: Performance Analysis of Response time.

Trial(s)	Blinks detected (yes/no)	Letters mapped (yes/no)
1	Yes	Yes
2	Yes	Yes
3	No	No
4	Yes	Yes
5	Yes	Yes
6	Yes	Yes
7	Yes	No
8	Yes	Yes
9	Yes	Yes
10	Yes	Yes
11	Yes	Yes
Accuracy(%)	90.9	81.8

Table 1: Accuracy of the proposed system

The experiment was carried out using a paralyzed person, aged 52. The above table was formulated based the success of detecting a short blink and a long blink, successfully converting them to the corresponding letter in the Morse code. We managed to achieve an accuracy of about 81.8 %.

6 CONCLUSION AND FUTURE SCOPE

The communication aid is provided for the physically paralyzed people and is designed to make it easier and understandable for others. The paralyzed person is trained with the Morse code eye blinks to generate the corresponding text and voice with an additional feature of enabling a text message and alert to the care-taker. This enables the affected person to communicate with normal people. This system doesn't require the affected used to wear any intrusive wearable or any expensive sensors. It is cheaper and easy to use.

The project can be further enhanced by:

- We can use the same concept in much smaller handheld device to improve portability
- The response time of the system to could be further improved.

7 REFERENCES

[1] Ioannis Rigas and Oleg Komogortsev, Texas State University Reza Shadmehr, Johns Hopkins University," Biometric Recognition via Eye Movements: Saccadic Vigor and Acceleration Cues", ACM Transactions on Applied Perception, Vol. 13, No. 2, Article 6, January 2016.

[2] Mohamed Khamis, Ubiquitous Interactive Systems Group, LMU Munich, Germany Daniel Buschek, Ubiquitous Interactive Systems Group, LMU Munich, Germany Tobias Thieron, Ubiquitous Interactive Systems Group, LMU Munich, Germany Florian ALT, Ubiquitous Interactive Systems Group, LMU Munich, Germany A Bulling, Max Planck Institute for Informatics, Saarland Informatics Campus, Germany,"EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays ", ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies, Vol. 1, No. 4, Article 146. December 2017.

[3] Yonglin Tian, Xuan Li, Kunfeng Wang, Member, IEEE, and Fei-Yue Wang, Fellow, IEEE, "Training and Testing Object Detectors with Virtual Images", IEEE/CAA Journal of Automatica Sinica, Vol. 5, No. 2, March 2018.

[4] Qiaoying Huang, Chris Kui Jia, Xiaofeng Zhang, and Yunming Ye, "Learning Discriminative Subspace Models for Weakly Supervised Face Detection", IEEE Transactions on Industrial Informatics, Vol. 13, No. 6, December 2017.

[5] Eunji Chong and Katha Chanda, Georgia Institute of Technology Zhefan YE, University of Michigan Audrey Southerland and Nataniel Ruiz, Georgia Institute of Technology Rebecca M. Jones, Weill Cornell Medicine Agata Rozga∗ and James M. Rehg∗, Georgia Institute of Technology, "Detecting Gaze Towards Eyes in Natural Social Interactions and Its Use in Child Assessment", ACM On Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 3, Article 43. September 2017.

[6] Eunji Chong and Katha Chanda, Georgia Institute of Technology Zhefan YE, University of Michigan Audrey Southerland and Nataniel Ruiz, Georgia Institute of Technology Rebecca M. Jones, Weill Cornell Medicine Agata Rozga∗ and James M. Rehg∗, Georgia Institute of Technology, "Detecting Gaze Towards Eyes in Natural Social Interactions and Its Use in Child Assessment", ACM On Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 3, Article 43. September 2017.

[7] Shouyi Yin, Peng Ouyang, Xu Dai, Leibo Liu, and Shaojun Wei," An AdaBoost-Based Face Detection System Using Parallel Configurable Architecture with Optimized Computation", IEEE Systems Journal.

[8] Janthakal Rajeshwari, Kwadiki Karibasappa, Madigondanahalli Thimmaiah Gopalkrishna," Adaboost modular tensor locality preservative projection: face detection in video using Adaboost modular-based tensor locality preservative projections", IET COMPUT. VIS., 2016, VOL. 10 ISS. 7, PP. 670-678,2016.

[9] Artem Dementyev, Microsoft Research, Redmond & MIT Media Lab Christian Holz, Microsoft Research, Redmond," DualBlink: A Wearable Device to Continuously Detect, Track, and Actuate Blinking for Alleviating Dry Eyes and Computer Vision Syndrome", ACM On Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 1, Article 1. March 2017.

[10] Marc Tonsen, Germany Julian steil, Max Planck Institute for Informatics, Saarland Informatics Campus, Germany Yusuke Sugano, Osaka University, Japan Andreas bulling, Max Planck Institute for Informatics, Saarland Informatics Campus, Germany," InvisibleEye: Mobile Eye Tracking Using Multiple LowResolution Cameras and Learning-Based Gaze Estimation", ACM On Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol 1 No 3 Article 106, September 2017.

[11] T. Xie, Q. Zheng, W. Zhang, and H. Qu, ''Modeling and predicting the active video-viewing time in a large-scale E- learning system,'' IEEE Access, vol. 5, pp. 11490–11504, 2017.

[12] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, ''Dynamic image networks for action recognition,'' in Proc. IEEE Conf. comput. Vis. Pattern Recognit., Jun. 2016, pp. 3034–3042.

[13] S. Herath, M. Harandi, and F. Porikli. (2016). ''Going deeper into action recognition: A survey.'' [Online]. Available: https://arxiv. org/abs/1605.04988.

[14] S. Tippaya, S. Sitjongsataporn, T. Tan, M. M. Khan, and K. Chamnongthai, ''Multi-modal visual features-based video shot boundary detection,'' IEEE Access, vol. 5, pp. 12563–12575, 2017.

[15] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, ''Survey of spatiotemporal interest point detection algorithms in video,'' IEEE Access, vol. 5, pp. 10323–10331, 2017.

[16] B. Fernando, E. Gavves, J. O. M, A. Ghodrati, and T. Tuytelaars, ''Rank pooling for action recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 773–787, Apr. 2017.

[17] Christopher Ackad, Martin Tomitsch, and Judy Kay. 2016. Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 2343–2347.

[18] Florian Alt, Andreas Bulling, Lukas Mecke, and Daniel Buschek. 2016. Attention, Please!: Comparing Features for Measuring Audience Attention Towards Pervasive Displays. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16). ACM, New York, NY, USA, 823–828.

[19] Double Glazing Info. 2017. Webpage. (2017). http://www.double-glazing-info.com/Choosing-your-windows/Types-of-glass/ Security-glass Retrieved February 11, 2017.

[20] InterelectroniX. 2017. Webpage. (2017). https://www.interelectronix.com/en/products/touch-screen/custom/laminated-glass.html Retrieved February 11, 2017.

[21] David Lindlbauer, Klemen Lilija, Robert Walter, and Jörg Müller. 2016. Influence of Display Transparency on Background Awareness and Task Performance. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 1705–1716.

[22] Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16). ACM, New York, NY, USA, 12.