EYE POINTER: A Real Time Cost Effective Computer Controlling System Using Eye and Head Movement

Conference Paper · April 2016

4 authors, including:

Shahed Anzarus Sabab
Northern University Bangladesh
8 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Sayed Rizban Hussain
Islamic University of Technology
2 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Hasan Mahmud
Islamic University of Technology
56 PUBLICATIONS **335** CITATIONS
SEE PROFILE

EYE POINTER: A Real Time Cost Effective Computer Controlling System Using Eye and Head Movement

Shahed Anzarus Sabab, Sayed Rizban Hussain, Hasan Mahmud, Md. Hasanul Kabir, Md. Kamrul Hasan
Systems and Software Lab (SSL), Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Dhaka, Bangladesh
Email: {sabab05, rizban, hasan, hasanul, hasank}@iut-dhaka.edu

Abstract— **In this paper, we are introducing a low cost and real time computer interaction technique with eye and head gestures along with voice commands, based on a standard webcam. Improving the accuracy of the cursor movement was one of the main challenges while building a system like this. We have applied Kalman filtering for smoothing the mouse cursor movement. The use of voice modality, along with head and eye gestures, made the computer operation more interactive. Calculations of our own defined heuristics were used to improve the accuracy of head and eye detection rate. We obtained remarkable results in eye blink detection, voice command accuracy, and cursor movement. This new multimodal technique, named "EYE POINTER" can be a useful system for the physically impaired people, such as amputees.**

Keywords—face and eye detection; image processing; eye focus point; voice commands; kalman filter; heuristics calculation; gestures; computer navigation; eye movement synchronization.

I. INTRODUCTION

Nowadays, human computer interaction is a vital concept for researchers. The computer mouse is arguably the most fundamental element that lets us interact with our computers. As the usage of computer has increased along with the number of users, the interaction devices should provide advanced and user friendly techniques. Rather than using traditional input devices which may not fit a varieties of users, researchers are working on multimodal input, such as eye gestures, voice commands and touch based interaction.

It has been shown that the Reaction Time (RT) latency of the hand is slower than the RT latencies of the eye and head when the subject had to make a button press response with either the index or middle finger of the right hand, dependent upon whether the stimulus occurred to the right or left of the control fixation point [1]. Figure 1 shows the response time for eye and finger in ms.

The number of computer users is increasing every day. There is a large number of disabled people in the world that need to be taken into account. Among them, amputees (people who lost hands in an accident) and acheiropodia (born without hands and feet) [18] exist in great numbers and cannot have the facilities of using a computer because of the lack of having a pair of working hands. For people who do not have hands, there is no efficient solution to interact with the computer. Elderly people also face problems to use current input devices. Considering all these facts, we are introducing another technique to interact with a computer. We are only using a webcam and a microphone. The webcam is used to take the head and eye gestures and to synchronize the mouse pointer with these gestures. The microphone is used to take voice commands, which are interpreted as commands for the computer. As eye and head gestures, along with voice commands, are used to interact with computers, our system is fast and comparable with real time interaction. For the physically disabled people, this can be a way of interacting with a computer.

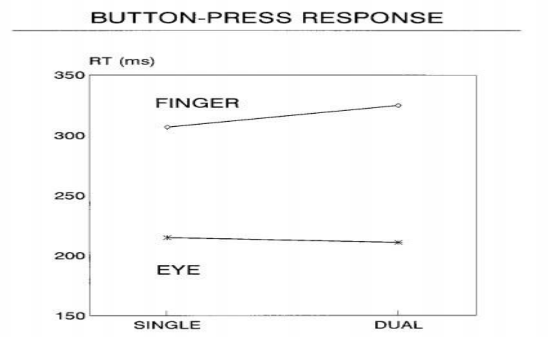

Figure 1. Mean RT latencies of eye and button in both single and dual task conditions

In our system, we have considered the standard resolution of the input image, so that the hardware cost is relatively low. The compensation of the lower resolution image is done using several filtering (median filter, Gaussian filter) and contrast stretching method. A former solution [2][3] was very expensive and did not provide all functionalities that we are providing in our system. We developed our own algorithm to detect eye and face more accurately only using a webcam. Our own developed voice commands were designed considering what human intend to perform through natural voice commands. We introduced Kalman filtering for getting better accuracy of mouse pointer movement. We developed several gestures of the head (e.g., tilting head to the left, right, up, down and movement of the head to any direction) and eye (e.g., eye blink) as navigation methods, which are new in the research area. Therefore, we believe that our interaction technique will bring a new contribution to Human Computer Interaction.

In Section II, the related works are described. Sections III contains some key terms that are defined throughout our work. In Section IV, the system overview is described. Section V has the major focus on the architecture of this system. Section VI contains evaluation and test analysis of the system.

II. RELATED WORKS

It has been found that musculoskeletal disorders due to computer work are directly related to ergonomic factors [4]. Traditional computer input devices have some difficulties. They affect the muscles in the human body in many ways. An inappropriate location of the computer can strain the shoulders. When the mouse and the keyboard are in such a place that someone needs to reach, then their shoulders extend forward and the shoulder blades rotate. Too much of these stretches of shoulder muscles causes spasm, fatigue, headaches and stiffness in the neck and shoulders. Long term effects may include severe shoulder pain and muscular imbalance [5][6]. Furthermore, any repetitive movement can result in a health problem called repetitive strain injury (RSI). Using a computer mouse for a long period of time is common cause of RSI in the wrist.

Many input techniques have been introduced as solutions of these aforementioned issues of interacting with a computer. Most of the techniques propose an external hardware. People need to wear external hardware, such as head mounted devices or extra glasses to give gaze information [3]. Tobi 1750 [2] eye tracking device takes eye gaze information using inferred camera. There are three classifications of the infrared wavelength. IR-A, IR-B and IR-C. The International Commission on Non-Ionizing Radiation Protection’s (ICNIRP) statement on *Far Infrared Radiation Exposure* have concluded that mainly IR in the IR-A band in high intensities are hazardous to the skin and eye [7]. Moreover, these devices are expensive for the general people to use.

Our solution is based on a simple webcam that does not use any infrared ray. It reduces the burden of carrying external devices. The cost of a webcam minimal compared to the other aforementioned interactive devices. Some computers, such as laptop, has a webcam integrated into them, so our system reduces the cost as well. The main challenge of this system was to work with a low resolution image that we can get from standard webcam. To fetch the required information from that input image, we used image processing techniques and our own algorithm. Another challenge was to design an interactive system that can interpret human intention. We have designed our voice commands in such a way so that it can predict human intention.

III. PROPOSED SYSTEM

Our system includes many interactive features which are very user friendly and easy to use. Our system provides whole computer navigation facility along with extensive features. We designed the system keeping in mind human cognition, intention and normal interaction techniques that a person follows while interacting in his or her daily life. Computer navigation is done by taking the movement information of the human head and eye and the commands are passed through voice input the same way a person expresses his or her intentions. The voice commands are also designed in a way such that a user need not memorize them, e.g., for moving mouse cursor the voice command is “move mouse cursor or move cursor”, which is very straight forward in regards to the human intention. Again, the cursor movement is smoothened using Kalman filtering [8], so that it gives the same kind of accuracy that a traditional mouse provides. We have proposed our very own heuristics calculation for face and eye detection purpose, which allowed the system to detect face and eye with a very high accuracy rate.

To understand our system, the following keywords descriptive knowledge is necessary:

Fixation or visual fixation is the maintaining of visual gaze on a single location. Whenever someone finds something of interest, his eyes fixate in that direction.

ROI (region of interest) [9] is the selected region of information among several regions of information. Here, the facial position and the possible eye position is the region of our interest. It is used to estimate visual interest information. We have calculated the ROI using our own heuristics.

Eye and head movement refers to voluntary or involuntary movement of the eye and head, helping in acquiring, fixating and tracking visual stimuli. The path followed during the movement of eye and head is the movement path. We are using the eye and head movement calculation to synchronize the mouse cursor movement with it.

Image enhancement technique includes several image preprocessing methodologies. This is a very important step for image processing and noise removal. This enhancement technique usually varies from application to application. For image enhancement, some key processing includes, i.e., filtering (Median filter, Laplacian filter, Gaussian filter, Weighted average filter) [10]. As we worked with low resolution images, we needed several enhancement techniques to compensate for the noises.

Kalman filter [8] is a recursive process that works in real time. It works on noisy observation of data like Gaussian noise. It is largely dependable on feedback. Kalman filtering is used to improve targets tracking, robot localization etc. It takes previous data and the current actual measurement and, based on the previous data, it predicts an optimal solution. Here, prediction and correction is determined by the gain factor.

If Measurement accurate > gain high (K_k) > observation dominate the filter response.

Or if Measurement noisy > gain low (K_k) > observed position considering estimated / predicted position is considered for filter response.

ISBN: 978-1-61208-468-8

Some basic equations of Kalman filter are shown in Figure2.

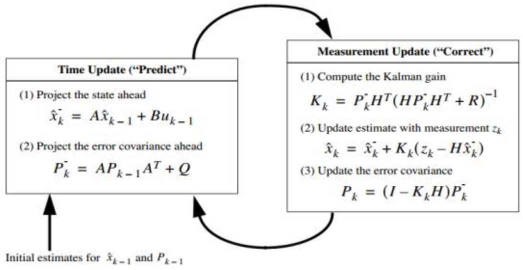
Time Update ("Predict")
(1) Project the state ahead
$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$
(2) Project the error covariance ahead
$P_k^- = AP_{k-1}A^T + Q$
Measurement Update ("Correct")
(1) Compute the Kalman gain
$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$
(2) Update estimate with measurement z_k
$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$
(3) Update the error covariance
$P_k = (I - K_k H)P_k^-$
Initial estimates for $\hat{x}_{k-1}$ and P_{k-1}

Figure 2. Equations of Kalman Filter

The time update equations can also be mentioned as predictor equations, while the measurement update equations can be mentioned as corrector equations. The final estimation algorithm resembles that of a predictor-corrector algorithm.

IV. SYSTEM OVERVIEW

The whole system is divided into two modes, i.e., "cursor control mode" (for navigation purpose) and "multi-mode" (for performing multipurpose operation).

The first mode is named the **Cursor control mode**. In this mode, the user can simply navigate throughout the computer. The user can move the cursor with eye and head gestures. Here, basic voice commands are implemented to do click options. The user can single click or double click or right click in this mode. They can perform click events with voice commands, as well as with eye blinks. Eye blink detection algorithm is used along with certain time threshold to do the click option. There is a specific voice command to activate the cursor control mode. If a user activates this mode, an activation confirmation feedback (e.g., cursor control mode is activated) will be returned from the system. In this mode, several operations can be performed, such as movement of mouse pointer, maximize or minimize windows, navigate throughout different folders, open and close programs/applications, control mouse pointer synchronization.

The second mode is named **Multi mode**. This mode is mainly introduced for multimedia purposes. After activating this mode (through voice command), voice confirmation feedback will be generated. Different voice commands are introduced here for different multimedia purpose. Here, we can watch video clips with gestures. We can fast forward the video or rewind the video using facial gestures. Other voice commands are introduced to do different tasks. There are voice commands for adjusting the volume or for watching full screen. There are commands to stop, play and pause the video. So, now, the users can enjoy and express their desire in a more flexible way. In this mode, the user can also view photo albums using head gestures. The user will be able to read document files by sliding it up and down using gestures. So, it makes the reading a more flexible and interesting technique.

V. SYSTEM FUNCTIONALITIES

Here, we describe our system functionalities and methodologies to implement them. The common functionalities are divided into 3 parts: 1) Navigation through eye and head movement, 2) Reading and multimedia control through different head and eye gestures, 3) Different operations through voice commands. Figure 3 shows the entire architecture.

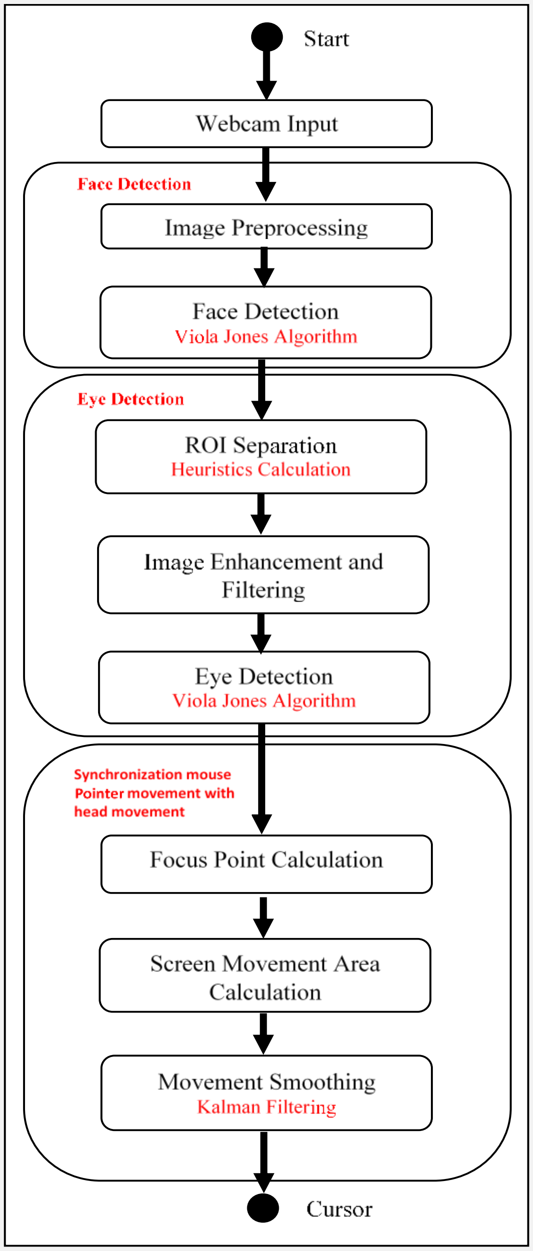

Figure 3. System Architecture

The process begins with taking an input image from an ordinary webcam. As the main challenge in this research is to work with standard resolution image, we have considered the image resolution as 800x600. The frame rate is set to the standard 30 fps. The webcam is set on top of the monitor so that the face of the user can be recorded.

1) Navigation through eye and head movement:

To implement this functionality, face detection methodology is applied by using Viola Jones algorithm [11][12]. It is a machine learning based approach. This technique employs a Haar-features (digital image features used in object recognition) based approach, which makes the rapid and accurate object detection possible. In our case, we have used a sample dataset [13] which includes thousands of negative and positive facial image information. For enhancing the poor image, input taken by the webcam is converted into gray. After that, binary thresholding is applied on the input image. Then, transformation of the image is done using image pyramid. An image pyramid is a collection of images - all arising from a single original image - that are successively down sampled until some desired stopping point is reached. There are two kinds of image pyramids. We have used Gaussian pyramid [14]. Next, contrast stretching Histogram Equalization [10] methodology is applied to enhance the contrast of the image. Face detection is done on the input image afterwards. To detect the eyes, a heuristic calculation is applied to determine the possible area of the eye within the detected face region. This area is named as Region of Interest (ROI) (orange rectangle in Figure 4).

From Figure 4, the Y coordinate of point A(X, Y) is calculated by adding top of the face rectangle (yellow box) with face rectangle height multiplied by a scale factor. Here, the scale factor is user dependent. We have found that our chosen scale factor (3/11) gives the best result for any user sitting in front of the camera. Therefore, point A(X, Y) is determined when the X coordinate is equal to the X coordinate of the face rectangle and the coordinate is previously calculated. Accordingly, the X coordinate of point B(X, Y) is the addition of X coordinate of the face rectangle (yellow box) and face rectangle width, and the Y coordinate of point B(X, Y) is previously calculated. The search area size is approximated by height: height of the face rectangle multiplied by the scale factor and width: face rectangle width. Finally, the possible ROI area (orange box) is approximated from the value of point A(X, Y) and the search area size. The calculations are as follows:

```
YCoordSearchEye = FaceRectangleTop +
(FaceRectangleHeight x 3/11)
StartingEyeSearchPoint = new Point (XCoordFaceRectangle,
YCoordSearchEye)
SearchEyeAreaSize = new Size (FaceRectangleWidth,
(FaceRectangleHeight x 3/11))
PossibleROIEyes = new Rectangle (StartingEyeSearchPoint,
SearchEyeAreaSize)
```

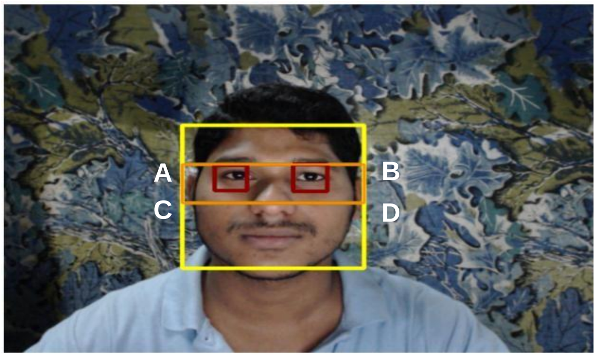

Figure 4. Face and Eye detection

For eye detection inside the ROI region (orange box), Viola Jones algorithm [11][12] is applied using a sample dataset matching of eye [15]. After completion of the detection process, synchronizing the mouse pointer movement with the head movement is needed. For that, we have considered a focus point calculation mechanism.

From Figure 5, the focus point F(X, Y) is calculated using a heuristic value. The X coordinate of F(X, Y) is the addition of the X coordinate of face rectangle (yellow box) and half of the search area size width and the Y coordinate is the addition of the Y coordinate of A(from Figure 4) and half of the search area size height.

```
FocusPoint = new Point (XCoordFaceRectangle +
SearchEyeAreaSizeWidth / 2, YCoordSearchEye +
SearchEyeAreaSizeHeight / 2)
```

The considered focus point is synchronized later as the mouse pointer. Figure 5 shows the white dot point in between the two eyes which we have considered as focus point.

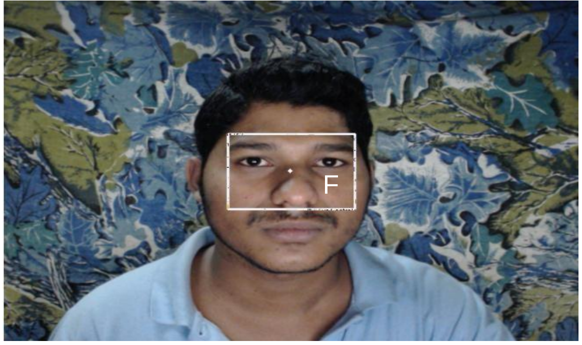

Figure 5. Focus point

Screen movement area calculation is the base region, inside where we have taken the relative focus point movement. This area is the subdivision of the total screen resolution. The movement of the focus point inside this region is multiplied by a factor to interpret that point throughout the total screen resolution. As a result, the little movement of the focus point inside the region is interpreted as the cursor movement within the monitor screen resolution. Though this rapid detection and calculation process in real time is very much effective, there were some mechanical noises which resulted in some disturbance in the movement path of the cursor. We found a blinking effect of the cursor, which is the presence and absence of the mouse cursor due to noises. Therefore, the accuracy in the navigation process was negatively affected. To compensate for this we have implemented Kalman filter [8]: a set of mathematical equations that provides an efficient computational means to estimate the state of the movement. It supports estimation of past, present and future states and it can do so even when the precise nature of the modeled system is unknown. Finally, after applying Kalman filtering we obtained a very smooth movement path resulting in greater accuracy in the navigation process.

2) Reading and multimedia control through different head and eye gestures

For reading and multimedia control, we have defined several head and eye gestures. Different operations are introduced using very effective and easy to learn head gestures. The common operations are: scrolling documents and reading documents, viewing images and switching images, forwarding and rewinding videos, switching through folders etc.

From different gestures, we have used head up, head down, head tilt left and head tilt right. Figure 6 shows the different head gestures used in "EYE POINTER".

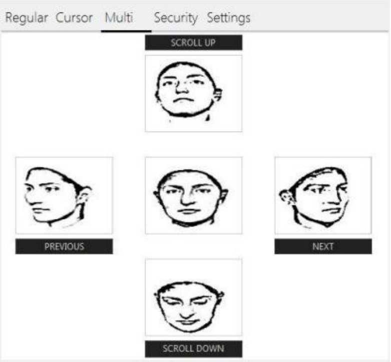

Figure 6. Different Gestures used in EYE POINTER

Mouse click events were defined through eye blinks, whereas click events can also be triggered by voice commands. With proper setup and lighting conditions, eye blink activation triggers the click event.

3) Different operations through voice commands

A number of voice commands are predefined for different operations on the computer. The operations are: maximize window, minimize window, increase /decrease volume, close any window/program, zoom in/out, play/pause video, copy/paste/cut, single click, double click, right click, show previous window, eye blink detect activation on/off, cursor movement on/off, pointer sensitivity increase/decrease, pointer speed increase/decrease, show cursor coordinate, switch through different modes, scrolling up/down, show next/previous picture, full screen on/off and many more.

VI. EVALUATION AND RESULT ANALYSIS

Evaluating a system is a continuous process. We evaluated the system on a group of twenty people of different ages. Among them, 5 were aged between 15 and 20, 9 were between 21 and 40 and the others were above 40. They were given a particular task to do using our system. Their performance was calculated based on the following parameters:

- Time to complete that task.
- Number of errors made in a particular time.
- Number of users making a particular error.
- Number of people completing the task successfully.

In real life, it was difficult to have real amputees as users to evaluate the system. So, during the experimental setup, we have chosen random people, who were restricted to use their hands for computer interaction. The task was to create a folder into a drive and copy a file from another folder which is located in another drive and paste it into that folder. In the experiment, we asked the users to do the task using the traditional mouse and then by using our system.

At the very first stage of the evaluation, we did not use the Kalman filter in our system. For that reason the evaluation results were not satisfactory although we obtained a good accuracy rate of face (100%) and eye region detection (95%).However, the pointer movement accuracy was low. The blink detection rate was moderate. Users were facing problems in performing the tasks. Our system takes voice commands with a very high accuracy rate. So, they faced no problem while providing voice commands. The main concern was the less accurate movement of the pointer. The initial accuracy was approximately 60%.

We faced an important problem while synchronizing the mouse pointer with the eye and head movements, as there was some noise in the movement path of the pointer. To compensate, we used the Kalman filter. This filter provides estimated values based on the previous values. So, this increased the accuracy of the pointer movement when some frames were still missing at the time of tracking. Issues due to missing frames were compensated by the estimated value calculation. So, the pointer movement became smoother. Accuracy increased up to 90%. Table I shows the accuracy rates of the different functionalities of our system.

TABLE I: ACCURACY RATE

Feature	**Accuracy**
Blink Detection	80%
Voice Commands	95%
Pointer Movement (Without Kalman)	60%
Pointer Movement (With Kalman)	90%

Previously, when the Kalman filter was not used, the average number of errors made by the users while completing the task was very high. However, after using the Kalman filter the error rate has decreased remarkably. This result is presented in Figure 7.

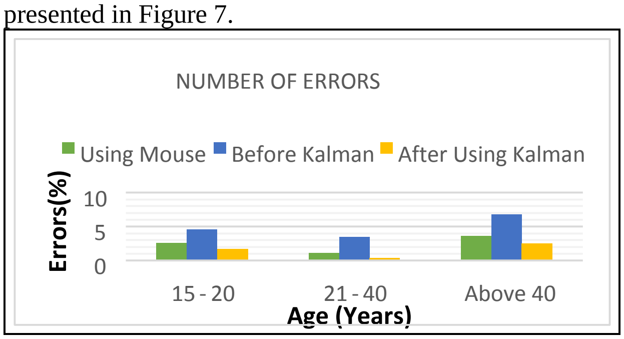

Figure 7. Average Number of errors made by the users.

Next, we repeated the experiment with the same group of people. This time the success rate was very high. Users were able to complete the task with a very high accuracy. It was found that the average time taken to complete the task was best for people aged between 21 and 40. Figure 8 shows the average time information.

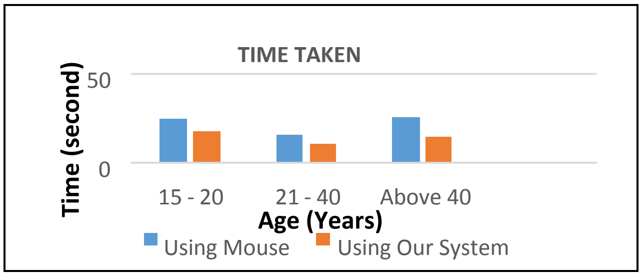

Figure 8. Average Time taken by the user

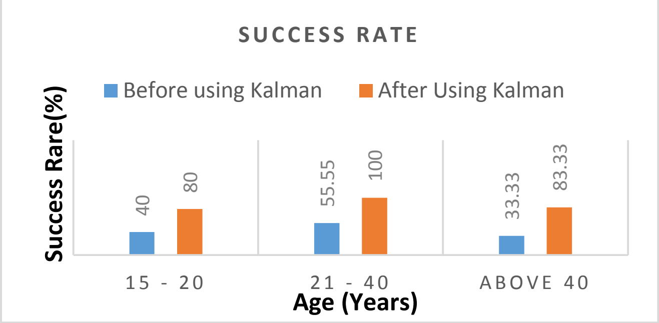

Figure 9. Success Rate.

In the final experiment, the number of people completing the task was very high. Almost everyone completed the task successfully. However, it is clear from Figure 9 that, before using Kalman, the success rate was really poor.

User overview and feedback after using our system was remarkable. People are always interested in using new technologies. Our unique system got the attention of everyone. They felt great interest in using our system.

The voice commands were preferred by the users in our experiment. The commands are so simple to learn and remember. It feels like communicating with another person. There is appropriate feedback for the commands, so the users will get a real time experience of computer interaction. There is no extra burden of using external devices to pass commands. According to the user feedback we received, we proposed this user interactive system where, there is a scope of change if needed.

VII. Conclusion

Trying to build a system like “Eye Pointer” is not new Researchers were trying to build this kind of system for decades [16][17]. But now, the advancement and the availability of technology gives us an advantage. The availability of the latest dataset helped us build a low-cost system that is very effective. ”Eye Pointer” is an innovative approach while using these materials. This is a generic navigation system which makes the computer navigation technique a real time interactive process. The main focus point of this paper is to build a navigation system that is completely hands free with the cheapest cost possible. We have overcome many challenges in designing this interactive system such as: designing voice commands considering human cognition, increasing the mouse pointer movement accuracy rate, enhancing the detection capability of eyes and face, designing a system so that it can fit standard goals of a user. Our system is designed with the focus on disabled people. We have tried to give them a solution mechanism to interact with the computer. However, there are some limitations. Our blink detection can only work in the proper lighting condition. In case of voice input we have found that, in an environment of loud noises, our system takes wrong input. Finally, our future plan is to update our research so that our system can be an effective approach for any end user.

References

[1] Harold Bekkering, Jos J. Adam, Herman Jingma, A. Huson and H. T. A Whiting, “Reaction time latencies of eye and hand movements in single- and dual-task conditions”,in Experimental Brain Research, Springer-Verlag 1994,7 September 1993, 97(3):471-6.

[2] Young-Min Jang, Rammohan Mallipeddi, Sangil Lee, Ho-Wan Kwak and Minho Lee, “Human intention recognition based on eyeball movement pattern and pupil size vibration,” School of Electrical Engineerin, Kyungpool National University, in Neurocomputing 2013 Ngip-khean Chuan and Ashok Sivaji, “Combining Eye Gaze and Hand Tracking for Pointer Control in HCI”, in IEEE Colloquium on Humanities, Science & Engineering Research, 2012.

[3] Michal Kowalik, “How to build low cost eye tracking glasses for head mounted system,” martin tall on gaze interaction, A blog on research and developments in eye tracking and gaze interaction, 2010.

[4] Jens Wahlström, “Ergonomics, musculoskeletal disorders and computer work,” Department of Occupational and Environmental Medicine, Sundsvall Hospital, SE-851 86 Sundsvall, Sweden, 2005.

[5] Annabel Cooper and Leon Straker, “Mouse versus keyboard use: A comparison of shoulder muscle load”, International Journal of Industrial Ergonomics, 1998; 22:351-357.

[6] B. M Blatter and P. M Bongers, “Duration of computer use and mouse use in relation to musculoskeletal disorders of neck or upper limb”, International Journal of Industrial Ergonomics, Volume 30, Issues 4–5, October–November 2002; 295–306.

[7] Soyun Cho et al., “Effects of Infrared Radiation and Heat on Human Skin Aging “, Journal of Investigative Dermatology Symposium Proceedings (2009) 14; 15–19.

ISBN: 978-1-61208-468-8

[8] Greg Welch and Gary Bishop, “An Introduction to Kalman Filter,” Department of Computer Science, University of North Carolina, Technical Report ,TR 95-041, July 24, 2006.

[9] Jing Zhang, Li Zhuo, Zhenwei Li and Yingdi Zhao, “An Approach of Region of Interest Detection Based on Visual Attention and Gaze Tracking,” Signal and Information Processing Laboratory, Beijing University of Technology, Communication and Computing (ICSPCC), 2012 IEEE International Conference ; 228-233.

[10] R. Gonzalez and R. Woods, Digital Image Processing, 3rd ed.: Pearson Education, Reference book, 2009.

[11] Paul Viola and Michael Jones, “Rapid Object Detection using a Boosted Cascade of Sample Features,” in computer vision and pattern recognition 2001.

[12] Paul Viola and Michael J. Jones, “Robust Real-Time Face Detection,” International Journal of Computer Vision, 57(2), 2004; 137–154.

[13] https://github.com/Itseez/opencv/blob/master/data/haarcascades/haarcascade_frontalface_alt_tree.xml , retrieved: October, 2015.

[14] http://docs.opencv.org/2.4/doc/tutorials/imgproc/pyramids/pyramids.html, retrieved: October, 2015

[15] https://github.com/Itseez/opencv/blob/master/data/haarcascades/haarcascade_eye.xml, retrieved: October, 2015.

[16] Mohammand Usman Ghani, Sarah Chaudhry, Maryam Sohail and Muhammad Nafees Geelani, “GazePointer: A Real Time Mouse Control Implementation Based On Eye Gaze Tracking,” Department of Electrical Engineering, COMSATS Institute of Information Technology, Lahore, Pakistan , Multi Topic Conference (INMIC), 2013 16th International; 154-159.

[17] Yingbo Li, David S. Monaghan and Noel E. O’Connor, “Real-Time Gaze Estimation using Kinect and a HD Webcam,” INSIGHT Centre for Data Analysis, Dublin City University, Ireland, 2014.

[18] Stefan Mundlos and Denise Horn, Limb Malfunctions: An atlas of genetic disorders of limb development, Berlin; London: Springer, 2011, P: 62.