To appear in the Proceedings of the ACM SIGCHI symposium on engineering interactive computing systems (EICS 2009)

Input Evaluation of an Eye-Gaze-Guided Interface: Kalman Filter vs. Velocity Threshold Eye Movement Identification

Do Hyong Koh Sandeep Munikrishne Gowda Oleg V. Komogortsev
Department of Computer Science, Human Computer Interaction Lab,
Texas State University-San Marcos, USA
+1 512 245 0349
dk1132@txstate.edu sm1499@txstate.edu ok11@txstate.edu

ABSTRACT
This paper evaluates the input performance capabilities of Velocity Threshold (I-VT) and Kalman Filter (I-KF) eye movement detection models when employed for eye-gaze-guided interface control. I-VT is a common eye movement identification model employed by the eye tracking community, but it is neither robust nor capable of handling high levels of noise present in the eye position data. Previous research implies that use of a Kalman filter reduces the noise in the eye movement signal and predicts the signal during brief eye movement failures, but the actual performance of I-KF was never evaluated. We evaluated the performance of I-VT and I-KF models using guidelines for ISO 9241 Part 9 standard, which is designed for evaluation of non keyboard/mouse input devices with emphasis on performance, comfort, and effort. Two applications were implemented for the experiment: 1) an accuracy test 2) a photo viewing application specifically designed for eye-gaze-guided control. Twenty-one subjects participated in the evaluation of both models completing a series of tasks. The results indicates that I-KF allowed participants to complete more tasks with shorter completion time while providing higher general comfort, accuracy and operation speeds with easier target selection than the I-VT model. We feel that these results are especially important to the engineers of new assistive technologies and interfaces that employ eye-tracking technology in their design.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, input devices and strategies, interaction styles.*

General Terms
Algorithms, Measurement, Performance, Design, Experimentation, Human Factors.

Keywords
Human Computer Interaction, Kalman Filter, Pointing Device Evaluation, Eye Tracker.

1. INTRODUCTION
An eye-tracking device can be used as an interactive input modality for users with disabilities or as an additional interaction channel for normal users [2]. Engineering challenges behind using this type of input are 1) eye-tracking failures due to eye squinting, eye moisture, and blinks and 2) noise due to eye-tracking hardware inaccuracies and micro eye movements. These challenges can be solved by processing raw eye position signal and classifying it into meaningful components such as fixations (eye movements that occur when gaze is dwelling on objects), saccades (eye movements between two separate fixations), and pursuits (eye movements that occur when eyes are tracking moving objects). Fixations (dwell-time) are the most common modality for an eye-gaze-guided computer interaction [9, 13, 17]. This modality of interaction assumes that a fixated object is selected when duration of a fixation reaches a predefined threshold. Very few works exist that employ saccade based interaction [15]. Pursuit-based interaction is an unexplored topic in the Human Computer Interaction (HCI) community to the best of our knowledge. In this paper we consider only fixation-based mode of interaction.

Several models exist for eye movement classification, including the mostly used Velocity-based Threshold (I-VT) model described by Salvucci and Goldberg [11]. I-VT model is usually selected because of the ease of its implementation and its low computational cost, but the model is not robust and is not capable of handling high levels of noise present in the eye position data.

The Kalman filter is a recursive estimator that computes a future estimate of the dynamic system state from a series of incomplete and noisy measurements. Eye trackers frequently fail to report eye position data and the reported data is susceptible to noise due to the individual anatomical properties of users and limited spatial resolution of the equipment. Therefore, Kalman filter framework can be applied to process raw eye position data to provide more accurate and robust estimation of the eye position signal. At the same time, the Kalman filter is capable of classifying eye movements [12]. The use of a Kalman filter in a real-time eye-gaze-guided computer interface was first discussed by Komogortsev and Khan [6] where researches have indicated that the filter can be successfully used during eye-tracking failures. Kumar et al. [8] presented the case where a Kalman filter provided smoothing to raw eye position signal, thereby increasing the stability of the input. Unfortunately, both research groups did not provide a comprehensive evaluation of the interface performance driven by the Kalman filter.

To provide an objective and subjective performance analysis of I-KF and I-VT models, we have designed an accuracy test and a standalone real-time eye-gaze-driven photo viewing application (iGaze).

Performance of I-VT and I-KF was conducted following several evaluation guidelines of the ISO 9241 Part 9 standard. This standard allows conducting an evaluation of non keyboard/mouse based input devices. Previously, Zhang and MacKenzie [18] successfully applied the ISO 9241 Part 9 standard to the eye-gaze guided task for a simple target selection task. In our research, we applied the standard guidelines to a full standalone eye-gaze-guided application.

Twenty one subjects participated in the evaluation where eye movement identification was done by either I-VT or I-KF model. In addition to the accuracy test, each subject had to complete a series of five tasks using the iGaze application. The results allow us to conclude that participants with I-KF were able to complete more tasks on average along with a shorter completion time while providing smoother, higher accuracy, higher operation speed, and easier target selection than the I-VT model.

The paper is organized as following: first, the description of the I-VT and I-KF classification models, followed by input evaluation description including an accuracy test, the ISO 9241-9 standard, and the design principles employed in the creation of iGaze photo viewing application. The paper ends with results, discussion, and conclusion sections.

2. REAL-TIME EYE MOVEMENT CLASSIFICATION

2.1 Velocity Threshold Classification (I-VT)

I-VT distinguishes fixations and saccades based on the observation of velocities between two separate eye positions. If the sampled velocity is greater than the threshold then the corresponding eye position sample is marked to be a part of a saccade; otherwise the eye position sample is marked to be a part of a fixation [11]. Consecutive eye position points, classified as fixation, are collapsed into a single fixation with a center coordinates computed as a centroid of all points in the fixation. Classified fixations are subsequently merged into larger fixations by the criteria based on two parameters: latency and distance between two subsequent fixation points. Velocity separation threshold is the main parameter responsible for the correct performance of the I-VT model. Different literature sources indicate different values for this threshold. Salvucci and Goldberg [11] indicate values above 300°/s for saccades and below 100°/s for fixations. Komogortsev and Khan [6] suggest 30°/s and 5°/s. Practical approach necessitates the empirical selection of such parameter with the value of 75°/s selected in our system.

I-VT classification is simple to implement, but the model has low tolerance to noise caused by equipment failures and/or micro eye movements.

2.2 Kalman Filter Classification (I-KF)

The Kalman filter is a recursive estimator that computes a future estimate of the dynamic system state from a series of incomplete and noisy measurements. A Kalman Filter minimizes the error between the estimation of the system's state and the actual system's state. Only the estimated state from the previous time step and the new measurements are needed to compute the new state estimate. Many real dynamic systems do not exactly fit this model; however, because the Kalman filter is designed to operate in the presence of noise, an approximate fit is often adequate for the filter to be quite useful [1].

The general mathematical framework of the Kalman filter is described by Brown and Hwang [1]. In our implementation, I-KF models an eye as a system with two states: position and velocity. The acceleration of the eye movement is considered as white noise with known maximum acceleration. The details of the Kalman filter parameterization that we have employed in our work were presented by Komogortsev and Khan [6].

The use of Kalman filter allows generating predicted position and velocity signal. Velocity prediction can be applied as a part of chi-square test for eye movement classification [12] and eye position signal prediction can provide the data during eye-tracking failures. The predicted position signal is employed for fixation parameters calculation, providing signal during tracking failures (data loss).

Chi-square test monitors the difference between predicted and observed eye-velocity:

$$\chi^2 = \sum_{i=1}^{p} \frac{(\dot{\theta}_i^- - \dot{\theta}_i)^2}{\delta^2} \qquad (1)$$

In Equation (1), $\dot{\theta}_i^-$ is the predicted eye-velocity by Kalman filter. $\dot{\theta}_i$ is the computed eye-velocity based on the measured eye position signal. δ is the standard deviation of the measured eye-velocity during the sampling interval and p is the sampling window size. If χ^2 is smaller than the threshold the corresponding eye position sample is marked to be a part of a fixation; otherwise the eye position sample is marked to be a part of a saccade. Consecutive eye position points that are classified as fixation are collapsed into a single fixation following the same criteria as for I-VT model. I-KF provides stable fixation detection if δ is selected to be a constant. In our model, we empirically selected $\delta^2 = 1000$ and $p = 5$. χ^2 threshold was empirically selected to be 50.

3. INPUT EVALUATION

To test input performance capabilities of the I-VT and I-KF models, we employ two methods in evaluation. The first method is an accuracy test that provides an objective measurement of accuracy performance. The second method uses an eye-gaze-guided application with performance evaluated through a series of tasks and a questionnaire suggested by the ISO 9241-9 standard. The use of real application allows to test for objective measurements such as task completion time and the number of tasks completed and the questionnaire provides a subjective evaluation in terms of performance, comfort, and effort.

3.1 Accuracy test

This procedure involves participants looking at 17 sequentially presented points that are uniformly distributed on the computer screen. When a subject fixates at each point, the raw eye position signal is processed by either I-VT or I-KF and corresponding fixation parameters such as location coordinates, the onset time, and the duration are determined. The coordinates of the eye position within the detected fixations are compared to the center of presented stimulus. This allows for the computation of error between reported location of the gaze and the actual gaze point. At the end of the recording, the error values are averaged between all points and presented on the computer screen. Additionally, an accuracy test computes and presents data loss parameter that indicates the amount of erroneous (not detected) eye position samples provided by an eye tracker for the participant.

3.2 ISO 9241-9 standard

The ISO 9241-9 standard [4] is used for the evaluation of computer pointing devices with suggested measurement of performance and comfort. The standard provides a questionnaire for evaluating performance, comfort, and effort of a given input modality.

3.3 Eye gaze-guided system design

The design challenge of any eye-gaze-guided computer system can be separated into three categories: layout design, individual component size selection, and visual feedback. We suggest the following guidelines to address the issues presented in each category.

3.3.1 Layout

The minimum spacing between each individual component of the interface should not be less that the reported eye tracker accuracy. Usually the accuracy value is around 0.5° [3]. The accuracy of the eye-tracker equipment degrades closer to the periphery of the computer screen; therefore, the layout of the individual interface components should provide an increased spacing at the boundary of the computer screen.

3.3.2 Individual component size

The size of the individual component should not be less than reported eye-tracker's accuracy to provide an accurate selection of this component by a fixation detection algorithm. An ideal component size would be around 2° of the visual angle due to the similar size of the human fovea – region in our eye providing the highest acuity of vision [5]. We recommend keeping individual component size between 0.5-2°.

3.3.3 Visual Feedback

Eye-gaze-guided systems do not have a mouse cursor following the eye gaze because of the chasing effect that appears when the reported gaze location does not match with the actual gaze position on the screen [3]. Nevertheless, the eye-gaze-guided system must indicate to the user that the interface component is currently being attended to. We recommend highlighting the boundary of individual interface component with different colors when user's gaze is directed toward this component in order to indicate the longevity of attention. Users successfully use this mechanism for the interface control and validation.

4. iGAZE INTERFACE

We designed an iGaze interface as an image viewing application capable of navigating through a set of grouped pictures (albums) using eye movements as the primary input. The iGaze provides functionality of scrolling through a list of albums, selecting and expanding the desired album into a viewable set of pictures, and finally enlarging the selected picture providing a full screen view to the user. Component selection was done by dwell time selection method, where the selection occurs based on the specified fixation duration. (200 ms.). As reported by Tien and Atkins [14] the value of 200 ms. provides an adequate balance between speed of interaction and accuracy.

The iGaze application was designed according to the eye-gaze-guided system design, i.e., all individual components called gidgets (term indicating the specificity of widgets designed for eye-gaze selection.) had spacing greater than 0.5° with individual gidget size larger than 1°. The size of gidgets was further increased at the screen boundaries to compensate for hardware inaccuracies.

4.1 Modes of interaction

4.1.1 Album mode

In this mode, the interface displays list of albums on the left side and grid of thumbnails of the currently selected albums on the right side. Also, both the album list and the thumbnails can be scrolled up and down using the "UP" and "DOWN" buttons.

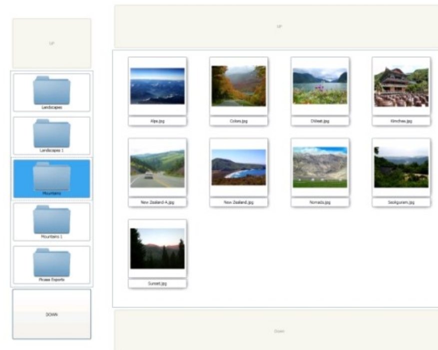

Figure 1. iGaze – album mode

4.1.2 Picture mode

In this mode, the thumbnail that gets selected in the album mode is displayed with enlarged view at the top right corner on a screen. At the bottom, thumbnails of other pictures in the album are listed horizontally across the screen with two buttons at each corner that enable scrolling left or right. Located on the left is a button that enables to switch back to album mode.

Figure 2. iGaze – picture mode

Visual feedback is provided for every gidget when an eye-gaze is directed toward the component. Specifically, all gidgets have a border that starts to glow blue as soon as the eye-gaze enters the gidget area. The border's glow color turns from blue to red as the fixation duration increases. The red border glow is designed to provide an indication that the gidget is about to get selected.

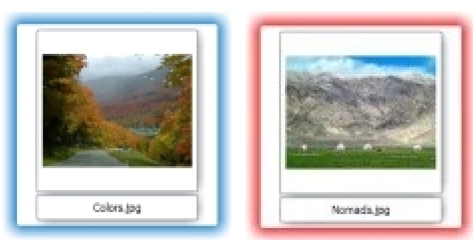

Figure 3. Visual feedback provided by gidgets.

5. METHODOLOGY

5.1 Participants

A total of 21 participants volunteered for the evaluation test. Participants' ages were from 18 to 35 (*mean* = 22.3). None of the participants had prior experience with eye tracking. Among these participants, 11 had normal vision and 10 wore glasses or contacts.

5.2 Apparatus

The experiments were conducted with Tobii x120 eye tracker, which is represented by a standalone unit connected to a 19 inch flat panel screen with resolution of 1280x1024. The eye tracker performs binocular tracking with the following characteristics: accuracy 0.5°, spatial resolution 0.2°, drift 0.3° with eye position sampling frequency of 120Hz. Tobii x120 model allows 300x220x300 mm freedom of the head movement. Nevertheless, a chin rest was employed for higher accuracy and stability.

5.3 Procedure

An accuracy test was applied to every participant at the beginning of each experiment. Next, each participant was asked to complete a sequence of five tasks using an iGaze application. Prior to each task, a subject was presented with an image cropped from one of the pictures stored in the iGaze application. The goal of each task was to find the original picture within three minutes. After each task, completion time was recorded. If the participant failed to find the picture, we marked it as time out. Half of the subjects were assigned to complete the tasks using the I-KF model and the remaining half were assigned to complete the tasks using the I-VT model. Such subject assignment was employed to negate possible interface learning effects that would occur if a subject would perform series of tasks using both models sequentially.

At the end of the procedure, subjects were asked to complete a survey containing a Device Assessment Questionnaire suggested by ISO 9241-9 guidelines, with some questions modified to be more task related. All questions were rated on a 5-point Lickert scale.

1. *Smoothness during operation was(very rough to very smooth)*
2. *The mental effort required for operation was (too low to too high)*
3. *The physical effort required for operation was (too low to too high)*
4. *Accurate pointing was (easy to difficult)*
5. *Operation speed was(too fast to too slow)*
6. *Eye fatigue (none to very high)*
7. *Target Selection(easy to very difficult)*
8. *Which will you prefer the eye gaze interface or interface with keyboard and mouse? (iGaze to Keyboard & Mouse)*
9. *Shoulder fatigue (none to very high)*
10. *Neck fatigue (none to very high)*
11. *General comfort (very uncomfortable to very comfortable)*
12. *Overall, the interface was (very difficult to use to very easy to use)*

6. RESULTS

6.1 Accuracy Test

The average fixation position detection error was 0.14° (SD=0.011) for I-VT and 0.126° (SD=0.007) for I-KF, indicating an improvement of accuracy of approximately 10% achieved by employing the I-KF model. The result was statistically significant (F(1,35067)=168.86,p<0.001).

6.2 Tasks Completed & Completion Time

The average number of tasks completed was 3.22 for I-VT (SD=1.48) and 3.8 (SD =1.03) for I-KF, indicating that participants were able to complete approximately 18% more tasks using I-KF model. The result was statistically significant (F(1,17)=7.15,p<0.01).

The average completion time was 92.59 seconds (SD=39.26) for I-VT and 88.06 seconds (SD =18.05) in I-KF indicating that participants were able to complete tasks on average 5% faster using I-KF. The result was statistically significant (F(1,17)=4.86,p<0.04).

6.3 Completion Time vs. Accuracy

The accuracy error indicates the difference between the actual eye-gaze position and position reported by the eye-tracker. Larger accuracy error results in erroneous or unsuccessful selections therefore hindering interaction performance. Following plot provides the task completion times given the specific accuracy error.

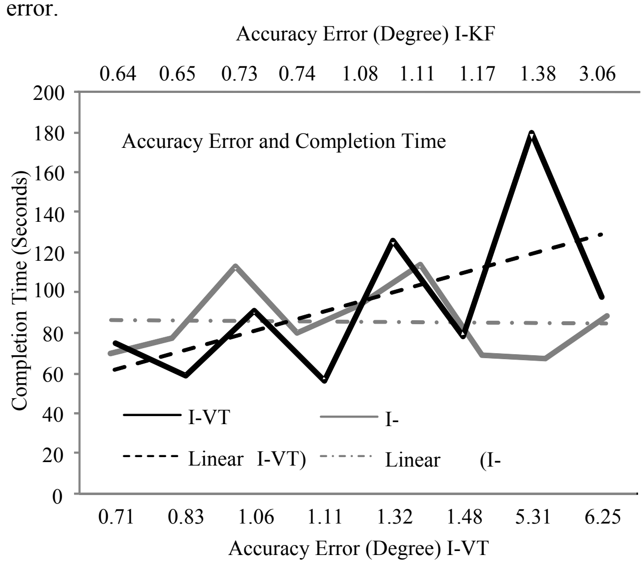

Figure 4. Completion time vs. Accuracy. Upper x-axis represents recorded range for I-KF model; lower x-axis represents recorded range for the I-VT model

The results indicate that lower accuracy provided a higher negative impact for the I-VT model, increasing the amount of time required to complete a task. It is apparent that completion time increased with lower accuracy value. In terms of the I-KF model, the accuracy error did not have such a significant effect, *e.g.*, completion time at 0.71° accuracy was approximately the same as at 3° of accuracy.

6.4 Completion Time vs. Data Loss

Data loss indicates the quantity of invalid data from the eye tracker for each subject. Data loss was computed as a part of the accuracy test. Data loss can have a substantial negative effect on the eye-gaze-guided system performance causing delays and incorrect fixation detections. As a result, it is important to evaluate the performance of the eye-gaze-guided system given the specific level of data loss. Figure 5 presents the results.

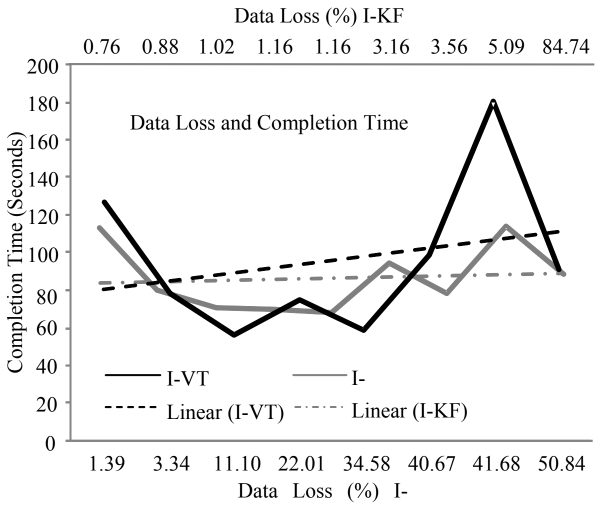

Figure 5. Completion Time vs. Data Loss. Upper x-axis represents recorded range for I-KF model; lower x-axis represents recorded range for the I-VT model

The performance I-KF results indicate very slow linear growth of average completion times with increased data loss. Remarkably, participants were able to perform well with the I-KF model even when data loss reached levels of more than 80%. With I-VT model, the data loss impact was much greater. Completion times were higher for the same level of data loss. It was not possible to interact with iGaze interface using the I-VT model when data loss exceeded 50% of the eye tracking data.

6.5 Questionnaire

According to the subjective evaluation, iGaze performance with the I-VT was perceived as smoother, induced less physical, eye and neck fatigue with more smoothness during operation than the iGaze with the I-KF. However, the iGaze with the I-VT caused greater shoulder fatigue.

On the other hand, the iGaze with I-KF required less mental effort for the operation In addition, I-KF provided more accurate pointing, higher operation speed, and easier target selection. Besides, participants gave higher preference score to I-KF model when they compared the eye-gaze-driven interface to the interface controlled by keyboard/ mouse. I-KF scored higher in general comfort evaluation. Participants also answered that the iGaze interface driven by the I-KF was easier to use when compared to the I-VT. The differences in the evaluation of above mentioned categories were not statistically significant.

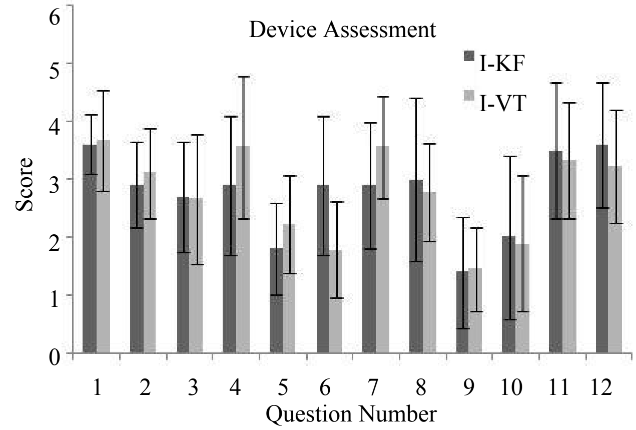

Figure 6. Device Assessment Questionnaire results.

7. DISCUSSION

Based on our observations, participants who were wearing eye correcting devices did not perform well in cases when iGaze interface was controlled by the I-VT model. Very frequently, such participants have lower accuracy and higher data loss. However, I-KF performed much better in those cases due to predictive and accuracy improvement capability of the filter.

In the current implementation of I-KF model, parameters such as threshold value for the chi-square test, sampling window size, and initial values for the system covariance matrix were empirically selected. In theory, such parameters depend on eye tracker sampling frequency and individual calibration accuracy. In the future, we would like to create a formula that would allow computing I-KF setup parameters based on experiment parameters.

The difference between subjective score values was not statistically significant with exception on question on general comfort and the difficulty of the interface use. The increase of subject pool should address this issue.

8. CONLUSION

This paper has evaluated subjectively and the objectively the performance of the Velocity Threshold (I-VT) and the Kalman Filter (I-KF) eye movement detection models. Both eye movement classification models were implemented with a standalone eye-gaze-driven photo viewing application (iGaze) and additionally tested with an accuracy test. The results of the evaluation indicate that on an average I-KF allowed participants to complete more tasks with shorter completion time while providing higher general comfort, accuracy, operation speed and easier target selection than the I-VT model. The scores on the general comfort and the overall interface evaluation were also higher when Kalman filter was employed.

The task completion time for the I-KF model did not decrease substantially in cases of decreased accuracy and high data loss. For the I-VT model, the task completion time increased when higher data loss and/or low accuracy of the equipment was reported. In two cases of extreme data loss (>50%), the iGaze application was unusable with the I-VT model, but when control was switched to I-KF it was possible to complete assigned tasks.

Very importantly, accuracy verification test indicated that I-KF provided a 10% of the accuracy increase when reporting locations of the detected fixations. This finding suggests that by employing I-KF model in a design of an eye-gaze-guided systems it would be possible to improve the accuracy of the eye-tracker equipment itself, which is especially potent in cases of emerging low-cost eye trackers.

One of our motivations to create the iGaze application was to provide an access to a computer to people with disabilities. Based on the findings provided in this paper, we can recommend the I-KF model to the designers of the eye-gaze-guided computer interfaces. I-KF is a real-time detection algorithm based on matrix multiplication, it is clearly improves systems performance by increasing the accuracy of the eye-tracking device and providing eye-position data during eye tracker failures.

In our future work, we plan to create and evaluate more interactive applications and improve the robustness of the eye-gaze-guided computer applications by further improving the accuracy of the eye-gaze input.

9. ACKNOWLEDGMENTS

We would like to thank following people for their help in developing iGaze interface: Ukwatta Jayarathna, Bushra Iftikhar, Jason Ramsey, Liam Feldman, Stephen Gross, Joseph Valdez, Prathima Boddu, Shruthi Anantharamaiah, Han Loo. This work was in part supported by Sigma Xi grant **G200810150639.**

10. REFERENCES

[1] Brown, R. G., and Hwang, P. Y. C. (1997) Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions. New York: Wiley.

[2] Douglas, S. A., Kirkpatrick, A. E., and MacKenzie, I. S. (1999). Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. The ACM Conference on Human Factors in Computing Systems - CHI '99 (pp. 215-222). New York: ACM.

[3] Duchowski, A. T. (2007). Eye Tracking Methodology: Theory and Practice. London, UK: Springer-Verlag.

[4] Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. International Standard, International Organization for Standardization (2000) ISO. ISO/DIS 9241-9.

[5] Irwin, D. (1992). Visual Memory Within and Across Fixations. Eye Movements and Visual Cognition, Springer Series in Neuropsychology (pp. 146-165). New York NY: Springer.

[6] Komogortsev, O. V., and Khan, J. (2007). Kalman filtering in the design of eye-gaze-guided computer interfaces. Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, 679–689.

[7] Komogortsev, O. V., and Khan, J. (2008). Eye movement prediction by Kalman filter with integrated oculomotor plant mechanical model. The 2008 symposium on Eye tracking research & applications (pp. 229-236). New York: ACM Press.

[8] Kumar, M., Klingner, J., Puranik, R., Winograd, T., and Paepcke, A. (2008). Improving the accuracy of gaze input for interaction. The 2008 symposium on Eye tracking research & application (pp. 65-68). New York: ACM Press.

[9] Kumar, M., Paepcke, A., and Winograd, T. (2007). Eye point: practical pointing and selection using gaze and keyboard. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, CA.

[10] Leigh, R. J., and Zee, D. S. (2006). The Neurology of Eye Movement. UK: Oxford University Press.

[11] Salvucci, D. D., and Goldberg, J. H. (2000). Identifying fixations and saccades in eye tracking protocols. The Eye Tracking Research and Applications Symposium (pp.71-78). New York: ACM Press.

[12] Sauter, D., J., Martin B., Di Renzo, N., and Vomscheid, C. (1991). Analysis of eye tracking movements using innovations generated by a Kalman filter. Med Biol Eng Comput, 29(1), 63–69.

[13] Sibert, L. E., and Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. The SIGCHI conference on Human factors in computing systems (pp. 281-288). New York: ACM Press.

[14] Tien, G. and Atkins M. S. (2008), Improving hands-free menu selection using eyegaze glances and fixations. In proceedings of the 2008 symposium on Eye tracking research and applications. Savannah, Georgia: ACM, 2008.

[15] Urbina, M. H., and Huckauf, A. (2007). Dwell time free eye typing approaches. The 3rd Conference on Communication by Gaze Interaction (COGAIN 2007) (pp.65-70). Leicester, UK.

[16] Yarbus, L. (1967). Eye movements and vision. Moscow: Institute for Problems of Information Transmission Academy of Sciences of the USSR.

[17] Zhai, S., Morimoto, C., and Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. The SIGCHI conference on Human factors in computing systems: the CHI is the limit (pp.246-253). New York: ACM Press.

[18] Zhang, X., and MacKenzie, I. S. (2007). Evaluating eye tracking with ISO 9241 - Part 9. Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, 779-788.