Eye Pattern Analysis in Intelligent Virtual Agents

ABSTRACT

This research discusses a new approach for intelligent virtual agents (IVA) to use patterns of a user's eye motion to better understand the user. The analysis of eye motion on the pattern level can deliver three values to an IVA: speed, interaction reliability, and a more complete understanding of user attention. Current eye tracking interfaces use fixation as a means of target acquisition and/or selection. There are several problems with this approach concerning issues of speed, system reliability and the understanding of user attention. This research builds a system, called InVision, to demonstrate how the analysis of eye fixation at the pattern level can help provide solutions to these problems. First, interface speed is quick through the use of pattern identification as a means of selection. Second, pattern correlation can add reliability to an eye tracking interface. Finally, the ability to understand the context of a user's eye motion is provided through pattern interpretation. An IVA is built using this eye pattern analysis technique.

Author: Michael Li
Co-Author: Ted Selker, Associate Professor, MIT Media Laboratory

1. Introduction

The analysis of eye motion patterns provides a powerful capability for intelligent virtual agents (IVA) to more fully understand interaction with a human user. People say they want to meet to look into another persons eyes. What do they see? The analysis of eye motion patterns could be a powerful tool to understand a persons thoughts. Eye tracking provides a channel through which IVA's can intelligently observe a human. A pattern-centric approach to eye tracking can improve human computer interaction by making an IVA aware of such things as context and social interaction that can be determined from eye movement.

Analyzing patterns of eye motion can deliver three improvements to computer human interaction: interaction speed, reliability, and a more complete understanding of user attention. Eye fixation is the primary means of identifying user attention in present day eye tracking. Fixation is also commonly used for control purposes such as selection in an interface. several problems plauge the use of traditional eye fixation techniques in intelligent virtual agent interfaces including speed, interface reliability and the understanding of user attention. This work proposes the use of eye fixation patterns as solutions to these problems. Interaction speed is increased through the use of pattern identification as a means of selection. Pattern correlation can help improve the reliability of an eye-tracking interface. Such a system also can have the ability to understand the context of a user's eye-motion is provided through pattern interpretation.

For the purposes of this work, an eye tracking interface tool called InVision is built that uses an eye pattern analysis approach. Specifically, this interface uses patterns of eye fixations to analyze and interpret eye-motion data. Such an analysis moves beyond simple eye fixation identification and examines how the eye has moved around an image in the context of that image. Using the InVision interface tool, the three values proposed by this work that are offered to virtual agents through the analysis of eye patterns are investigated.

First, the performance of an eye pattern analysis approach in attentive selection is quantitatively evaluated. An experiment, the Eye Selection Test, is run for this purpose using the InVision system. The performance of the pattern approach is experimentally compared to that of an approach using simple fixation for selection. Two variables are measured: selection accuracy and selection time, reflecting system reliability and system speed respectively.

The second part of the work qualitatively studies how a virtual agent can use this proposed technique of examining a user's eye fixation patterns to reveal a more complete understanding of user attention. The Kitchen InVision project is a virtual agent that studies patterns of fixations for the purposes of identifying, interpreting and responding to user cognitive state. By understanding user attention on a more complete and contextual level through eye-motion pattern analysis, intelligent agents will be capable of y being able to better predict and accommodate a user's interests, tasks and questions.

2. Background

Eye tracking can be valuable as a channel through which IVA's can observe a human user on a contextual, emotional, and social level. The speed of interaction with an eye tracking interface is incredibly fast. Ware and Mikaelian (1987) show that simple target selection and cursor positioning operations can be performed twice as fast using an eye tracking interface than using a mouse.

Fixation is an important phenomenon that is commonly used by eye tracking systems to provide an indication of local user attention. Most eye tracking interfaces use fixation as the basis for *target acquisition* which is the task of identifying a particular object. Target acquisition is generally an aspect of selection for eye tracking interfaces as well. Some eye tracking interfaces use eye fixation for both target acquisition as well as *selection*, such as IBM's Suitor project (Blue eyes: Suitor). The Suitor project, also known as Interest Tracker, distinguishes between a "glance" and a "gaze" by the amount of time a user has fixated on a particular area and uses a gaze as a means of selection. Techniques using fixation duration, or *dwell-time*, as a means of selection generally use a threshold value between 250-1000ms (Edwards, 1998). If a fixation lasts longer than the chosen threshold value, a selection is initiated. Interfaces using fixations for either for target acquisition or selection are referred to in this paper as *fixation-based* interfaces.

Fixation detection is an entire sub-field of eye tracking work While the physiological concept of a fixation is understood, many different algorithms for fixation detection (S. Zhai, personal communication, February 1, 2001) have been developed. A technique for fixation detection is used in this work that is similar to the fixation recognition approach described by Jacob (1995). It is believed that the specific choice of the fixation algorithm will not affect the results and conclusions of the work presented.

More recently, has been performed regarding the use of eye motion patterns to attempt to understand user cognitive state. One of the earliest uses of eye motion pattern to understand user attention is in a system created by Starker and Bolt (1990) that displays a planet from "The Little Prince," a book by Antoine de Saint Exupery. It uses patterns of natural eye movement and fixation to make inferences about the scope of a user's attention. Edwards (1998) uses eye movement pattern identification in the Eye Interpretation Engine, a tool that can recognize types of eye movement behavior associated with a user's task. Perhaps one of the works that is most relevant to this work is done by Salvucci (1999) who describes a technique called fixation tracing, a process that infers user intent by mapping observed actions to the sequential predictions of a process model. This technique translates patterns of eye movement to the most likely sequence of intended fixations. Several pilot interfaces were implemented in developing this work. These interfaces studied grouping, patterns of association and identification in search. The use of eye motion patterns is still a relatively new research area that will become more prevalent as eye tracking technology improves.

3. Approach

Eye motion patterns are at the center of this work. Patterns of eye movement are comprised of a sequence of points representing the locations of the eye fixation points[i] over time. While research has been performed on patterns of object selection to infer user intention and state, this work explores a new direction: the use of *patterns of eye fixations*. Several advantages can be gained from analyzing patterns of eye motion in eye tracking interfaces. The technique of analyzing eye movement on the pattern level can have three significant effects on current eye tracking systems that this section will propose and discuss. Such a technique can offer speed, reliability and a better understanding of user attention and each of these three effects are individually discussed below.

3.1 Speed through Pattern Identification

Pattern identification in eye motion data can increase selection speed. Identifying a pattern of eye motion can be much quicker than detecting a sequence of object selections. The task of using eye fixations as a means of selection can take an unpredictably long period of time depending on how good the system accuracy and system calibration is. This can greatly delay a system's response time.

Eye-gaze tends to be both too inaccurate and imprecise. A user's eye fixations, as recorded by an eye tracker, are often not centered over visual targets. Two things cause this inaccuracy: users can fixate anywhere within a one-degree area of the target and still perceive the object with the fovea (Jacobs 1995), and eye-trackers have a typically accuracy of approximately one-degree (Salvucci, 1999). Imprecision is introduced into eye-gaze data through the involuntary jittery motions produced by the eye. When the eyes appear to be looking at a single point, they are actually making a series of abrupt jumps. Research has demonstrated that impulses along the optic nerve occur only at the moment when the image changes on the retina (Babsky, Khodorov, Kositsky, & Zubkov, 1975). During constant application of light on visual receptors, impulses quickly cease along the corresponding optic nerve fibers and vision effectively disappears. For this reason, eye motion contains incessant imperceptible jumps that constantly displace the retinal image. This series of jumps stimulate new visual receptors and produce new impulses on the optic nerve, ultimately enabling the process of vision.Eye control is neither accurate nor precise enough for the level of control required to operate today's UI widgets such as scrollbars, buttons, hyperlinks and icons. Zhai, Morimoto, and Ihde (1999) show that the area represented by a user's fixation is approximately twice the size of a typical scrollbar, and much greater than the size of a character.

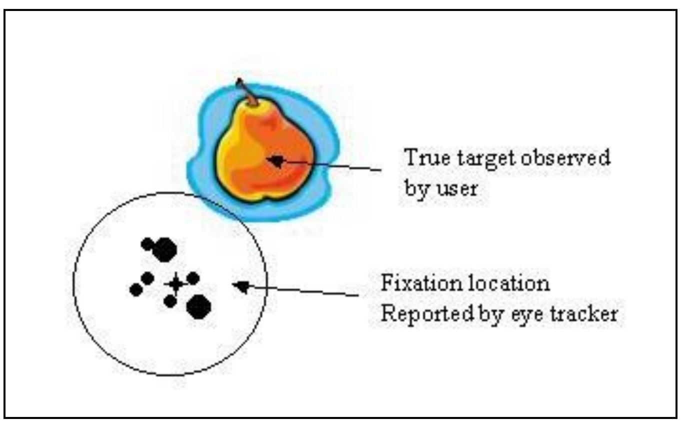

Figure 1. A sample of eye movement that shows the inaccuracy and imprecision related to eye-gaze.

Fixation-based interfaces are limited in their ability to handle and interpret eye tracking data because of this noise. By watching only for fixations, these interfaces adopt an easy technique to filter the noise, but at the same time end up ignoring important data as well. Consider the following problem: a user's eye gaze remains fixed on a point on the screen, but the eye tracking calibration is slightly off, resulting in a off-center fixation somewhere other than the location intended (see Figure 1). A fixation-based interface cannot effectively accommodate such noise.

Off-center fixations can undermine eye control reliability as well. An off-center fixation could fall on an object other than the one intended, producing an incorrect system response. Algorithms and techniques exist that attempt to map off-center fixations to intended targets. Similarly, cluster analysis methods can help determine likely target areas of attention (Goldberg & Schryver, 1995). These techniques do not always produce correct results however, especially as the number of target objects increase and the size of the target objects decrease.

The use of a pattern of fixations as a means of selection bypasses the need to select individual objects with fixations and thus can dramatically speed up selection time. Figure 2.a. displays a fixation path using a normal fixation-based method that requires a fixation to land on a target for a specified period of time. Figure 2.a. shows the need for several fixations per object selection when using fixation for selection. Figure 2.b. shows the use of eye fixation pattern for selection, which requires fewer fixation points to identify a selection. Because selection through pattern does not require many fixation points, selection speed for such a task is dramatically improved with little if no cost to accuracy.

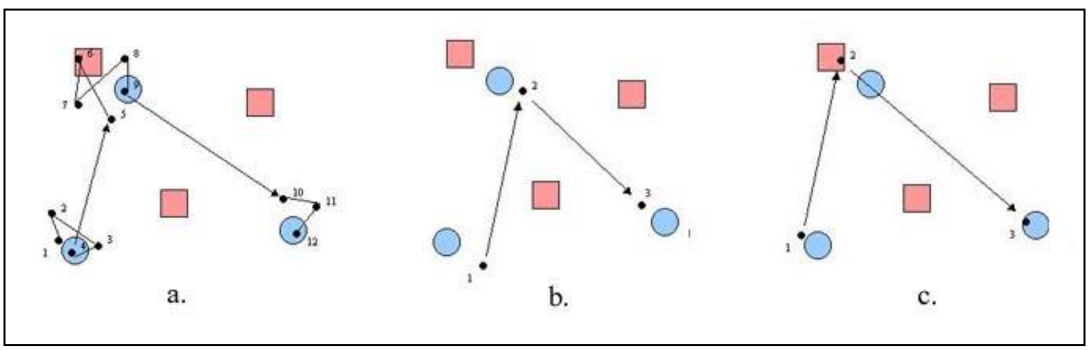

Figure 2. Three different eye fixation paths are shown for the same task of selecting blue circles in the image. *a.* selection of three objects by fixation *b.* selection of three objects by pattern of fixation *c.* selection reliability through pattern correlation

3.2 Reliability through Pattern Correlation

This work demostrates that the technique of analyzing eye movement on the pattern level can improve reliability in eye tracking interfaces by increasing accuracy in target acquisition. In order for a pattern to be recognized, a very characteristic path must be taken by the user’s eyes. Such a requirement can lead to a specificity that is hard to produce by either accident or luck. A system that is designed to listen for eye patterns waits for the user’s eyes to move in a certain pattern before an event or action is initiated. This offers a much more reliable response than a system that looks only at the location of the user’s gaze to initiate action.

Several features of an eye movement sequence can make pattern identification more reliable than fixation identification. The distance and vector angle between two points in an eye movement sequence are both attributes that can be used to help validate a pattern. Even the information concerning the location of where the eyes start and end can be used to help confirm a pattern. Identifying these features can provide data redundancy and correlation in noisy data from eye tracking systems. Figure 2.c. shows how pattern correlation can improve the reliability of a selection. This figure shows how the task of selecting blue circles can be distinguished from the task of selecting red squares based on the pattern of the eye fixation path. Using a pattern-based approach, a system can examine several validating factors to establish consistency in interpreting a user's intent, which ultimately improves the reliability of the interface.

3.3 Understanding User Attention through Pattern Interpretation

A system using an eye pattern based approach can better understand the concept of user attention. Several problems can be identified with the way fixation-based eye tracking systems determine user attention. Without looking at sequences of eye motion, it is difficult to appreciate attention on a complete level. Through examination of eye motion at the pattern level, the scope of a user's interest/attention can be better determined and identified.

The scope of user interest/attention is not something that is adequately addressed in current eye tracking systems. A traditional eye tracking system generally approaches the task of identifying user attention based on eye fixations. While the direction of the gaze usually points to the object of interest, this is not always the case. Several fixations within a particular area might indicate the user's interest in a single object in that location, or it could also be indicative of interest in a couple smaller objects. A system that has knowledge of the objects in the image and uses a pattern-based approach can better determine if the user is interested in a face, or specifically the nose on that face. By looking at eye tracking data in aggregate patterns, the data can be processed at a higher semantic level. The points of eye positions while giving little meaning themselves, can be grouped into patterns that have relevance to what the user is really looking at and what the user is concerned with.

Eye patterns can also give a better indication that a user has in fact given attention to a particular object. A system that can combine the history of a user's gaze with information about the objects in an image can build a better model of attention. The current location of a user's eye-gaze alone has proven insufficient for determining attention but the analysis of how the user's eyes moved in a given time period gives a much more complete picture. With this technique, the problem of distinguishing meaningful attentive vision from idle looking will be easier to approach.

Patterns of eye fixation can directly reflect user task. This represents a new area not emphasized by current eye tracking work which has primarily focused on patterns of object selections. Salvucci (1999) proposes similar work using fixation tracing to facilitate eye movement analysis to infer user intent at the fixation level. This work helps to better infer intended fixation location from recorded eye movement. Work performed by Edwards (1998) distinguishes eye movement into three mutually exclusive categories of behavior: searching, knowledge movement, and prolonged searching. From these characteristic patterns of eye movement, inferences can be made about user intent. This work investigates how aggregations of eye motion patterns can correlated to contextual user attention, such as user task. A computer user's eyes move in a specific way across the screen that is characteristic in part of the type of task, whether writing a paper, browsing the web, searching for a file, checking email or launching an application.

Patterns promise the potential of helping eye tracking systems begin to understand the user on a much higher-level.

4. Results

4.1 The Eye Selection Test Experiment

The Eye Selection Test demonstrates how a pattern-based approach can improve speed and reliability in an eye tracking interface in the task of identifying user attention. An experiment is performed to evaluate the pattern-based InVision interface in comparison to a fixation-based interface. The objective of this experiment is to address the first part of this work's hypothesis: to quantitatively investigate whether a pattern-based analysis can improve the reliability and speed of an eye tracking interface. Object size, while not the only measurable independent variable, is one of the biggest factors influencing selection performance. For this reason, selection speed and accuracy is measured for each interface over the size of the objects being selected. This provides a level means of comparison across different interfaces. The following sections outline the experiment performed, present the experiment results, and finally discusses the analysis of the results.

This experiment displays a sequence of trials each consisting of circular targets on the subject's screen. Targets appear three at a time in random locations on the screen (see Figure 3). Target size is randomized across trials but all the objects in the same trial are of equal size. The subject is instructed to select each of the three targets as rapidly as possible when the targets appear at the beginning of the trial. Selection, defined in the context of this interface, is equivalent to target acquisition. When an individual target has been selected, it disappears, and after the three targets in the trial have all been selected, a new trial is displayed. If a trial is not completed in a designated time, it is skipped. The trial begins when the user selects the first target and ends when the last target selection in a trial is completed. The experiment records whether the targets have been successfully selected along with the time duration of the trial. The number of off-center fixations, or fixations that don't select a target, are recorded and are used to reflect the relative inaccuracy of the selection process. After the subject has completed the series of trials, data analysis is available. The data from each trial with the same object sizes is combined for the purposes of analysis.

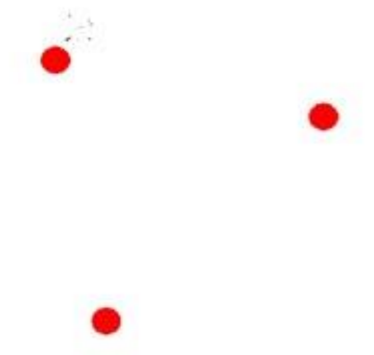

Figure 3. The Eye Test Selection Experiment

The InVision system's pattern-based approach is compared to a system that uses simple fixation as a means of target acquisition, an approach that looks simply at whether a fixation falls within an object. While better fixation algorithms exist to recognize the intended location of off-center fixations (such as one that chooses the closest target to the fixation), this fixation-based approach is chosen for use as a base level comparison. For the actual experiment, 5 tests each composed of 500 trials were run on each interface. Each selection test used a random target object size between 5 and 150 pixels and placed the target at a random location on the screen. Through this experiment, selection accuracy is compared across object size between a fixation-based interface and the pattern-based InVision system.

4.2 Results, Speed Comparison

Data regarding trial times is collected from the experimental runs, combined and then summarized. In order to compare the two different sets of data, the time recorded for each trial is divided by three, the number of targets per trial. This gives a representation of the selection time per object for a certain object size using a particular interface. The selection times recorded per experimental run reflect an average of the selection times across all trials of a particular object size. The selection times for each experimental run is plotted across object size for both interfaces and the results are summarized in Figure 4. A best-fit line is drawn through the two data samples.

Figure 4. Selection Time vs. Target Object Size for Fixation (blue) and Pattern-Based (red) Approaches

It is apparent that the data set representing the fixation-based approach requires a rather large number of fixations compared to the pattern-based approach. For the fixation-based technique, the number of fixations need to select an object is non-linear and reflects the calibration and inaccuracies in the eye tracking system. In a fixation-based approach, several fixations are required before a fixation falls on the target object. At small object sizes, the fixation-based approach requires a large selection time. This is due to the fact that as object sizes decrease, objects becomes harder to select. However as object sizes increase, the selection time approaches a constant. The significance of these results is that the pattern-based approach is able to remove the unpredictability and non-linearity of selection by not selecting through fixation but through pattern.

4.3 Results, Accuracy Comparison

Selection accuracy is used as a measure for system reliability. Data is taken from the same experimental runs used in the speed comparison analysis performed above. Selection accuracy for a particular object size is measured by dividing the number of target objects within the trial by the total number of fixations recorded in the trial. Similar to the speed comparison analysis performed above, the data collected from each experiment for each object size is averaged across trials. The percent selection accuracy for each experimental run is plotted across object size for both interfaces and is displayed in Figure 5. A best-fit line is drawn through the two data samples.

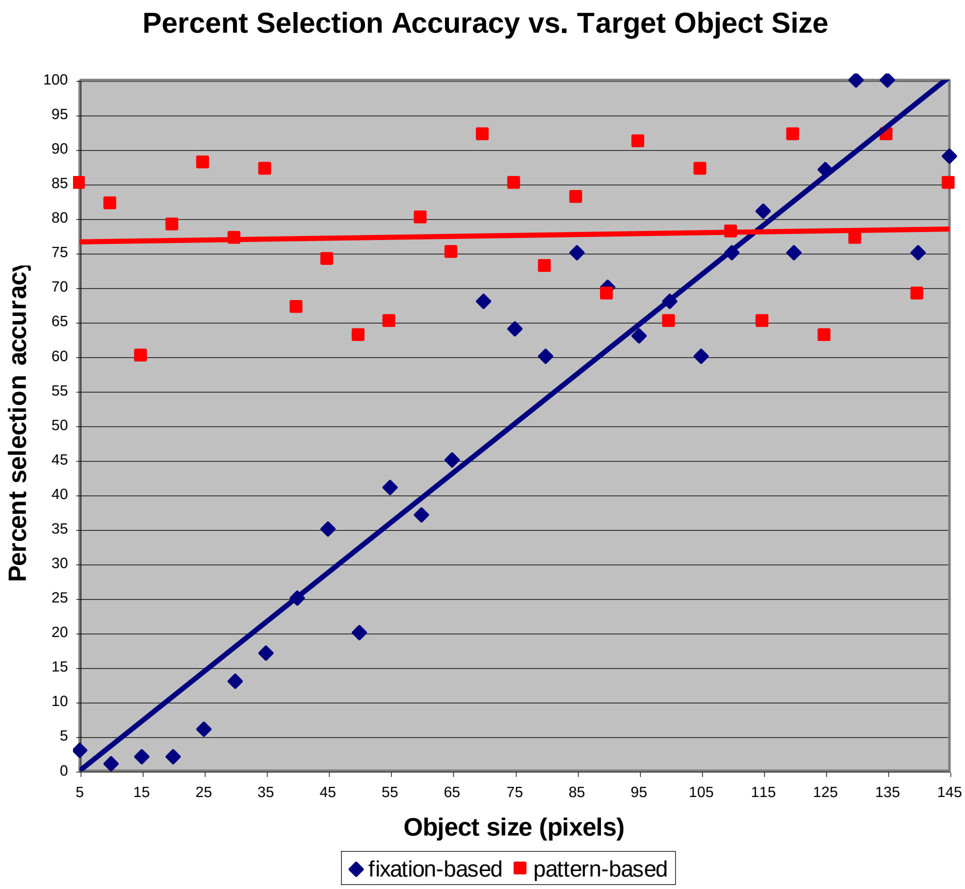

Figure 5. Percent Selection Accuracy vs. Target Object Size for Fixation (blue) and Pattern-Based (red) Approaches

The results for the accuracy comparison show that the InVision interface performs at a much higher selection accuracy than the fixation-based interface that is used for the experiment. Figure 5 shows a graph of the selection accuracy of the interfaces while object size varies. This graph shows that InVision performs better across all object sizes and performs significantly better when the object size is small, maintaining a high level of selection accuracy where the fixation-based system becomes very inaccurate.

The InVision system is also more adept at distinguishing between close objects than other systems. InVision correctly distinguishes the target object from two objects that are relatively close on the screen at a higher frequency than does the fixation-based technique that is examined.

4.4 Kitchen InVision

The Kitchen InVision is an intelligent virtual agent that can interact with a user by *listening* to patterns of eye fixation. An image of a kitchen is displayed along with several items commonly found in a kitchen. Interaction in this non-command (Nielson 1990) interface is achieved by watching a user's eye motion, interpreting patterns of eye fixation and delivering a visual response reflecting a change in the state of the kitchen or one of the items.

The interaction across this interface from user to system is not necessarily a direct one. Eye movements reflect thought process, and indeed a person's thought may be followed to some extent from eye movement analysis (Yarbus, 1967). The project described does not employ direct control and manipulation but rather builds a non-command interface in which the system *responds* to the interpreted eye pattern rather than being controlled. Therefor no command instructions need be provided on how to use the system. Providing instructions specific to the interface control implies a direct control mode of interaction between user and system, which as stated earlier, is not the intent of the project.

The sysetm can recognize the following tasks: cooking a turkey, washing dishes, and unpacking food from a grocery bag. Each task is made up of a sequence of independent actions. For example, the task of cooking a turkey involves opening a cabinet, putting a tray on the counter, taking the turkey out of the fridge, putting it on the tray, sticking the turkey in the oven, cooking the turkey and taking it out at the appropriate time. Active objects in the kitchen function in much the same way as one would expect them to in real life; the fridge, oven, and cabinets can open, food items can be cooked or be placed in the fridge, dishes can be washed, etc (see Figure 6).

Figure 6. The Kitchen InVision Project

The Kitchen project depends on InVision to recognize several types of fixation patterns. Simple patterns such as detecting user attention on one or more objects are recognized by the system, as well as complex patterns involving higher-level analysis and interpretation of aggregate patterns.

The Kitchen InVision project demonstrates how eye patterns can be used to interpret high-level user attention. The success of this demonstration is not entirely unexpected; it is logical that patterns of eye movement preserve data relating to a user's intention and attention. Where the eye comes from, what the eye has seen and when the eyes have moved all are factors that help understand user attention on more than just a localized scope. It should be stated that the concept of attention is a complicated idea that cannot be adequately identified by the eyes alone. However, as defined earlier, the scope of this work focuses on the eye's contribution to the state of user attention and attempts to better understand what user attention is, and how to help identify it using data collected from eye motion.

A preliminary example of a non-command interface, is kitchen InVision; it both human computer interaction and intelligent virtual agents. It serves as an example of an interface that uses a combination of eye fixation pattern and contextual information as the means of identifying user task and intention. Previous endeavors to understand patterns of eye motion have emphasized the use of patterns of selected objects to make inferences about user intention, rather than patterns of fixations. A pattern-based approach can offer speed and reliability in the research of using patterns to explain cognitive intent, especially in the area of non-command interfaces and context-sensitive environments.

5. Conclusion

This work has proposed the use of interpreting eye motion data through patterns of aggregate eye movement and has discussed how the field of intelligent virtual agents can benefit from eye pattern analysis. A system called InVision is built which adopts a pattern-based approach to eye motion interpretation. InVision provides a platform on which interfaces using eye pattern analysis can be built. The abilities of a pattern-based approach are tested and evaluated by using the interface structure provided by InVision. Next, comparison benchmarking is performed between a pattern-based and a fixation-based approach. Finally an intelligent virtual agent, called Kitchen InVision, is created to demonstrate how patterns of eye fixation can be used to infer context-specific user intent. Results point to several advantages gained through the use of patterns, confirming the benefits of a pattern-based approach proposed earlier in this paper. The three benefits that are demonstrated by this research to be gained from the use of eye pattern analysis are:

1. Speed through pattern identification
2. Reliability through pattern correlation
3. Understanding through pattern interpretation

To conclude, the analysis of fixations on the pattern-level has been identified as a new approach for understanding the human user that can offer both better ability as well as new capability to eye tracking interfaces. It is hoped that the research outlined in this paper will give encouragement to future research and development of eye fixation patterns within the space of intelligent virtual agents.

References

Babsky, E., Khodorov, B., Kositsky, G., & Zubkov, A. (1975). Human physiology. Moscow: Mir Publishers, 182-214.

Blue eyes: Suitor [WWW Document]. URL http://www.almaden.ibm.com/cs/blueeyes/suitor.html (visited 2001, February 2).

Edwards, G. (1998). A tool for creating eye-aware applications that adapt to changes in User Behaviors. International ACM Conference on Assitive Technologies 1998. 67-74.

Goldberg, J. H. & Schryver, J. C. (1995). Eye-gaze determination of user intent at the computer interface. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), Eye movement research: Mechanisms, processes, and applications (491-502). New York: Elsevier Science Publishing.

Jacob, R. J. K. (1995). Eye tracking in advanced interface design, *in* W. Barfield & T. Furness, eds, `Advanced Interface Design and Virtual Environments'. Oxford: Oxford University Press, 258-288.

Nielsen, J. (1993, April). Noncommand user interfaces [Communications of the ACM, volume 36, no. 4]. 83-99.

Salvucci, D. D. (1999). Inferring intent in eye-based interfaces: Tracing eye movements with process models. Proc. CHI ’99, 254-261.

Starker, I. & Bolt, R. A. (1990). A gaze-responsive self-disclosing display. Proc. ACM CHI ’90 Human Factors in Computing Systems Conference, Addison-Wesley/ACM Press, 1990. 3-9.

Yarbus, A. L. (1967). Eye movements during perception of complex objects, in L. A. Riggs, ed., `Eye Movements and Vision'. New York: Plenum Press, chapter 7, 171-196.

Ware, C. & Mikaelian, H. T. (1987). An evaluation of an eye tracker as a device for computer input. Proc. ACM CHI, GI ’87 Human Factors in Computing Systems Conference. 183-188.

Zhai, S., Morimoto, C. & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. Proc. ACM CHI ’99 Human Factors in Computing Systems Conference, Addison-Wesley/ACM Press, 1999. 15-20.

[i] In this research we look at patterns of fixation points, although eye movement patterns are not necessarily limited to fixation points alone.