A Standard for a Next Generation Over-head
Display as an Application Platform: Optimizing Bionic Vision

Pavleen Thukral, Nadia Abutaleb, Ria Banerjee, Mathew Bird
Lockheed Martin Explorer Post 1010, Poolesville High School SMCS

A Standard for a Next Generation Over-head Display as an Application Platform: Optimizing Bionic Vision

Abstract

The gap between man and machine can only be filled by two options: make the machine more human or make the human more machine. The next generation of eye-sight, named InSight in this paper, furthers 20/20 vision through the advancement of today's simple eyeglasses. A high-quality next-generation webcam, iris-tracking sensors, a computer chip managing the wireless connection to the user's cell phone, and a projection display will be placed within a pair of standard glasses. The device will track the user's two irises and use the curvature of his eye to calculate a vector intersection for line of sight, establishing what he is currently focusing his eyesight on. The object will then be analyzed by various computer imaging algorithms that will then display information about the object, such as tags associated with the object on the web, company names associated with a logo, etc. on the visual display of the glass's lenses. In addition, a zoom feature can be activated through a hand motion from the user. InSight is a new, convenient method of gathering information and takes 20/20 vision to a whole new level. The InSight hard ward and software solution will act as a platform for developers to make applications on a next generation user interface. The iPhone revolutionized the phone industry. When Apple first came out with the iPhone and started their chain of application/touch screen enabled devices, people were stunned and amazed because it was a new dawn of technology that made amazing technological possibilities available to the laymen population. However, as exciting as it is – the future is not in touch technology, but in sight. We will soon grow into a world where humans aren't the only creatures that see the world around us; computers will see it, too. They will read what we read, giving us further insight with the combination of AI. They will pull from a scene what we would like to know. The future of technology is not in the more amazing things possible on a screen, but in the fusion of interactions with the physical world with those of the virtual world. The future of commercial technology is an overhead display with which we can interface the world we see with the digital world of information and computing.

1 Present Technology

1.1 Eye Tracking

Most current eye tracking technologies focus on 2D analysis of the eye. The eye has three primary motions. The eye can move with horizontal rotation, vertical rotation, and rotation about a third axis, commonly referred to as the visual axis. In most current eye tracking systems 2D in which the third axis is not counted and the 3D position of the eye is not known. This makes the eye tracking system rely on the relative line of sight with respect to the user. When trying to measure where a human is looking, one must also consider the position of the head. Generally, a human's eyes reach a target faster than the head's turning motion. However, this continuous turning affects the position of the eye, while not drastically changing the position of the iris in two space, because the eye is not two dimensional and has curvature. This restricts current eye tracking systems as the user must remain steady relative to the monitoring device to achieve accuracy. In existing technologies, eye tracking systems employ uncomfortable mouth guard holdings or attachments to the head [5]. Systems that allow users to not have this uncomfortable interaction with the technology are called remote eye tracking systems. They allow for eye tracking without contact with the user, as well as free head movement.

1.2 Eye Tracking Synthesis

In current technology, both eye tracking and head-mounted displays are commercially available. However, the integration of these components together in a compact, portable, and robust system has not been developed. Integration results in high functionality in three dimensional scenes, but with the exchange of structure and usability.

1.3 Computer Vision

Computer vision today is a relatively new technology that has barely been experimented with for 10 years. Current technologies, although impressive, come nowhere close to our own natural vision. Identifying predefined objects through spot detection and mapping a scene is the most sophisticated technology available. Algorithms today are able to identify objects and contours based on shading and lines and are continuously modified in real time based on programming efforts. Canny edge detection is one of most common edge detection solutions and is showcased in the popular open sourced computer vision solution, OpenCV [4]. Canny edge detection works by converting the grayscale representation of an image into values based on the darkness of the grayscale. Local maximums and minimums are found based on the gradients of these values when represented as contours. The connections between the extrema points are the edges [4].

Specifically for eye tracking, current systems use a computer vision algorithm known as the Starburst Algorithm [7]. The idea of this algorithm is to synthesize a feature based approach and a model based approach. The algorithm begins with two steps. The first is noise reduction in the image gained of the eye. This is done by calculating a normalization factor based on average intensity of each horizontal line in the image. The second is reduction of any reflection from the cornea of the eye. This is important as these features increase error when examining the difference between contours of the eyes and can create local extrema that can trick eye tracking algorithms. An iterative feature based approach is applied to find pupil edges. This is done by mapping contours based on the time difference from the infrared rays. Then, local extrema of the contours is found by estimating gradients based on the data. The algorithm fits an ellipse to parts of the pupil using progressions of the points (edges or extrema) found using a random sample consensus paradigm (RANSAC). A random sample consensus paradigm is a database of sorts systematically organized under parameters of eye features. From the RANSAC, a set of feature points is acquired which is then used to fit an ellipse to subsets of the detected edge points [5]. In current systems, the estimation of a gaze is based off of infrared technology.

This is used instead of the visual spectrum due to the ability to control the light used. The estimation of gaze is determined by using a physical model of an eye defined by dependent parameters of cornea radius, offset of gaze from optical axis (the difference between the physical center of your pupil and where your mind differentiates gaze) and then the largest constraint on the technology the distance between the user and the two dimensional display for which the gaze is tracked. This means that in current systems, the gaze can only be accurately measured knowing not only the distance from the target to the user, but also that the frame of the scene in two dimensional. Current technologies accomplish gaze recognition with an average accuracy of 1.2 degrees [7].

As an application platform this concept would need to have vision systems built in to be able to recognize objects with respect to your gaze. A major limitation of the time is that algorithms are only proficient for predefined object parameters. They know what they are looking for beforehand. Gaze is different. In a real life situation you could be looking at any object. No system exists that is able to separate the different objects in a scene without predefined knowledge of the scene.

2 Future Technology

2.1 Eye Tracking

Our technology will focus on using near infrared illumination to determine the 3D position of the eye. By determining the 3D position of the eyeball, the system will be able to provide the complete 3D representation of the line of sight. This will allow for all head movement and will be less restrictive in structure. Current near infrared illumination technology reflect controlled infrared light and use the starburst algorithm to detect contours and use this to ellipse the pupil. However, there are many restrictions to the current technology that must be overcome.

The first limitation is that current NIR technology requires the knowledge of the distance between the target and the user. This is not possible for our device as the scene through our head-mounted display will be three dimensional. Gaze on a three dimensional scene contrasting with a two dimensional scene is very different due to depth. Current technologies provide an error of 1.2 degrees when determining gaze. For objects that are farther away, the degree of error is much more significant. Also, with a three dimensional scene, the distance between the target and the user is varied. Current systems that overcome this problem are bulky and have flawed accuracy when integrated with displays. This means that coupled with the three dimensional nature of an eye ball, the three dimensional depth of the targets the gaze estimation has to have a decimal error; otherwise, eye tracking will not provide accurate measurements, and the user interface for the device will be lacking in sophistication. Integration will also require the shrinking of all components.

Improvements to the Starburst Algorithm

The starburst algorithm optimizes the time used to find counters in the eye by focusing on the central portion of the image. An even more optimized procedure would be to focus not only on the central part of the eye, but on the ring where the ellipse of the pupil would be found. Converting this ring structure and mathematically unfolding and translating coordinates into a rectangular pseudo coordinate system would provide for easily defining contours as the contour line would now be linear. After finding the equation of the contour, it would be run through a mathematical formula to translate coordinates back into the real coordinate system, theoretically rolling the ring back up. This would be much more efficient and would take less time as RANSAC would not be required. Only the relative distances of the points measured in the pseudo coordinate system would need to be recorded to map the curvature of the ellipse. Employing a complicated formula would require a lot less time than a search through a database of ellipse parameter values.

2.2 Eye Tracking Synthesis

Current technology is not able to integrate head-mounted display technology with eye tracking components together in a compact, portable, and robust system that are comfortable for everyday use. InSight will incorporate the technology such that all components will fit into common current thick framed glasses described in the website component of the submission. This will incorporate nano-circuitry not possible today, including a solution to jumping electrons in nano-circuits. In addition, high exposure to NIR illumination at close range from our device would result in damage to the iris. However, our device would allow for minimal output while still getting reliable readings. The solution to this would yield procedures to increase signal intensity in close range of the device during input while minimizing signal during output.

2.3 Computer Vision

The future of computer vision relies in future developments in artificial intelligence technology. The ability for a computer system to be able to make decisions about what in view is an individual object that has not been predefined. Features such as facial features, buildings, and text can be predefined; however, advancements must be made so that any kind of object – an exotic snow globe, a hockey stick, a Christmas gift of abstract shape, or even bags, all in any orientation in three space—can be recognized as an individual object and identified or searched for based on visual information. The most rapid increase in this field is based on mathematical advancements with research into shape spaces. Shape spaces is a mapping process with which one takes a shape in two space and translates it into a point in shape space. Shape spaces are often used in gesture recognition as when shapes change over time they map a trajectory in shape space[4]. The same technology can be applied to picking out one object from a bundle of multiple objects in a scene. The obvious complications to this technology are the amount of undefined parameters. We do not know orientation, color, or truly anything about what the object will be; however, the technology must learn to identify objects in a 2D array of pixels. In a sense, the device will either need to learn and catalog objects as it goes along or be advanced enough to identify objects without prior knowledge about them and simply focus on advanced heuristics with shadows and contrast by identifying empty space.

2.4 Processing

Portability is extremely important to InSight. Because all computation will be done through an off-board wireless computing unit (probably a cell phone), processing technology will have to be significantly advanced. The fastest portable processing unit on current devices is up to 1.6 GHz. For artificial intelligence computation, much more processing power will have to be added. The IBM 5.2Ghz processor is the size of a common face. High power processing units must be compacted and reduced to a pocket-sized device. For portability and sufficient smoothness in software portable processing, units must advance to a standard above 2Ghz.

2.5 Necessary Breakthroughs

To create InSight, several technological advancements are necessary. First, a next generation mini-web camera that will be able to zoom in while still retaining high image quality must be developed. Current web cameras will not retain the image quality necessary to make InSight a reality when shrunk down to the necessary size. Also, computer vision algorithms must reach a level of sophistication at which a computer can quickly analyze, distinguish, and identify objects in a field of view. Depth information from focus cues in 3D is generally incorrect due to the fact that "light comes from a planar display surface" [9]. This causes distortion in perceived depth. The iris-tracking sensors used must utilize a three-dimensional analysis of the eye and where it is gazing rather than the two-dimensional analysis that current iris-tracking algorithms use. In other words, the issue of depth creating differences between two-dimensional and three-dimensional images must be resolved by the iris-tracking algorithms.

Currently, methods for developing 3D displays with near correct focus cues are being researched, such as the volumetric display. This method attempts to render accurate focus cues by displaying millions of voxels, or volumetric pixels, within a physical volume [8]. However, methods such as this are often inefficient in their computations, have limited rendering volume, and face difficulty in "correctly rendering view-dependent lighting effects such as occlusion, specular reflection, and shading" [8]. These difficulties must be overcome for InSight to effectively display information to the user.

The third breakthrough is in optical display technology. When using lenses to present an image on a display, there is an overlap of the lenses in the field of vision. This is the same in one's eyes – there is an overlap in the fields of view of the right eye and the left eye. Algorithms to modify displays so that the displayed characters appear normal to the user are required to create InSight. Otherwise, there will be a repeat of the overlapped area in the display and the image will be distorted.

The final breakthrough is in nano-processing technology. For the device to function, it must have a nano-processor that is capable of wirelessly transmitting and receiving information to and from another processor that would do most of the computation. At the same time, the processor must be small enough to fit inside the rim of a typical pair of glasses. However, currently, multiple difficulties exist with shrinking down electrical components to such a small size. One such difficulty is that when the components are shrunk and packed into smaller spaces, the heat generated by the electrical components could easily become high enough to begin damaging the chips themselves [6]. In addition, at such small sizes, research has shown that there is an unwanted transfer of electrons between components [6]. To create InSight, current electronic components must be made small enough to fit in glasses while being kept far enough from each other to prevent illegitimate transfer of electrons in different parts of the circuitry and to prevent overheating. Apart from the nano-processor, the second processor that will be doing computation for the computer vision solutions must be powerful. For continuous real time processing of the amount of information and calculation required for the type of artificial intelligence necessary to identify objects without predefined parameters the processor must be very powerful. The fastest electric processor available today is IBM's 5.2 GHz processor chip released in September of 2010. This unit is the fastest non over clocked electric microprocessor invented. This technology is not commercially available and is made for use with IBM's z 196 series computers (industry use only) [Cnet current news]. It is also very big. The technology must improve and shrink.

Conclusions

InSight poses an extremely good standard for future over-head display technologies. For an overhead display to be usable and worth the cost, it must achieve a level of sophistication where it knows where we are looking and determines the object that we are looking at. From this level of base systems processes, the product will run third party applications made by developers. In a futuristic world, it is necessary for us to move away from making robots more bionic and give us the ability to advance our own senses to match machines. Modern platforms, such as Google's Android and Apple's iOS, have given us a taste of the power of an application platform open to developers. With the implementation of over-head display hardware and base software, the potential for application development for the improvement of our access to information and computing is limitless. If you enjoyed this outline for a standard of the over-head display concept, please read the papers listed in the Reference Section.

References

[1] Böhme, M., Meyer, A., Martinetz, T., & Barth, E. (2006, September 5). Remote eye tracking: State of the art and directions for future development. Retrieved from Institute for Neuro- and Bioinformatics, Germany website: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.1747&rep=rep1&type=pdf

[2] Carlson, W. (2003). A critical history of computer graphics and animation. Retrieved from The Ohio State University website: http://design.osu.edu/carlson/history/lesson17.html

[3] Fitzpatrick, P. (n.d.). Head pose estimation without manual initialization. Retrieved from Massachusetts Institute of Technology website: http://groups.csail.mit.edu/lbr/hrg/2001/paulfitz-headpose.pdf

[4] Hua, H., Krishnaswamy, P., & Rolland, J. P. (2006, May 15). Video-based eyetracking methods and algorithms in head-mounted displays. Optics Express, 14(10), 4328-50. Retrieved from http://www.opticsinfobase.org/abstract.cfm?URI=oe-14-10-4328

[5] Li, D., & Parkhurst, D. J. (2005, September 5). Starburst: A robust algorithm for video-based eye tracking. Retrieved from Human Computer Interaction Program, Iowa State University website: http://thirtysixthspan.com/openEyes/starburst.pdf

[6] Liu, S., & Hua, H. (2009, June 1). Time-multiplexed dual-focal plane head-mounted display with a liquid lens. Optics Letters, 34(11), 1642-4. Retrieved from http://www.u.arizona.edu/~liusheng/research/reference/09%20OL.pdf

[7] Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005, September 1). Focus cues affect perceived depth. Journal of Vision, 5(10). Abstract retrieved from http://www.journalofvision.org/content/5/10/7