37.3: Dynamic Magnification of Video for People with Visual Impairment

Robert B. Goldstein, Henry Apfelbaum, Gang Luo and Eli Peli

The Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA

Abstract

Magnification is an effective aid for people with conditions causing resolution loss, but it inherently restricts the field of view. Therefore, magnification must be centered on the most important point of the scene. We determine this "point of regard" (POR) *by recording eye movements of normal subjects while they watch video.*

1. Introduction

People who suffer from loss of visual resolution due to eye diseases could benefit from modification in information displays. The most common modification used today is magnification. Magnification inherently restricts the field of view and thus may impede the acquisition of peripheral information attained in normal vision by the use of eye movements. This problem may be addressed by dynamic control of the display. Control may be automatic or under the user's control, or a combination of both. Dynamic control of text size for patients with central field loss was investigated by a number of labs[1-4] and is the basis for the success of electronic reading magnifiers and software screen magnifiers providing access to text in print and electronic formats. We propose to apply a similar approach to improve access to television and other video sources.

Magnifying television images using electronic or computational zoom [5] enables users to select the desired level of magnification using a remote control, and to vary the magnification used from time to time. Although manual "zoom and roam" devices are available in commercial television systems (such as video conferencing and DVD players), the rapid changes of scenes in most video movies does not allow for optimal manual control over the position of the magnified section of the image.

Only part of the magnified scene can be presented on the screen. Consequently, large parts of the scene become invisible. We propose pre-selecting the point in the image on which to center the magnified view (the point of regard: POR) and providing that position with each frame.

This selection should maintain the most relevant details in view, to the degree possible, when magnified. We used eye movement recordings from multiple normally-sighted observers watching the video program to determine the desired POR. Although other methods of determining the POR can be envisioned, this method is the most automatic and objective.

Together with DigiVision [6], we developed a computer controlled "zoom and roam" device for playback of the video clips. The computer plays a DVD and simultaneously reads the POR derived from eye movement recordings. These coordinates are sent to the zoom and roam device so that the magnified image is centered on the POR coordinates. Here we describe the process of recording and processing the POR information and its use in a prototype device.

2. Methods

19 normally-sighted subjects (7 men under 40 years old; 3 men over 45; 5 women under 40; and 4 women over 45) were seated at a distance of 74 inches from a 22-inch wide (27" diagonal) television screen, corresponding to a 16.9^ox12.7^o visual angle. Subjects viewed movie clips while eye movements were recorded with an ISCAN model RK726PCI Pupil/Corneal Reflection Tracking System equipped with an RK620-PC Autocalibration system. The ISCAN device had a nominal accuracy of 0.3^o over ± 20^o range [7] and a sampling rate of 60Hz (but was recorded at 30 Hz). This system tracks head movements within a narrow range, permitting gaze monitoring without head restraint, which makes for a more comfortable and natural viewing situation. The ISCAN device was calibrated using a 5-point calibration scheme [7]. The calibration was checked, and if necessary, repeated before each of the six video clips was viewed. Eye position coordinates and pupil diameters were transmitted to a PC. A Visual Basic program, using the MSWebDVD object of Microsoft DirectX 8.1, simultaneously controlled the DVD and read the ISCAN data from a serial port.

The video clips were selected to span a broad level of activity, from stationary newscasters to athletes in motion. The videos were selected to appeal to both younger and older audiences (Table 1). The movies were presented in 16x9 format (HDTV), but were displayed on a 4x3 screen, resulting in a 16.9^ox9.5^o image. A total of 707 minutes of eye movement recordings were collected during video viewing.

Table 1: Video clip categories and length used.

Category	Title	Time
Talk Show	Quiz Show (1994)	6:40
Romance	Shakespeare in Love (1998)	7:06
Sports	Any Given Sunday (1999)	4:12
Documentary	Blue Planet (2001)	8:14
News	Network (1976)	4:02
Comedy	Big (1988)	6:29
	Total (min:sec)	**37:29**

3. Analysis

The data files from the recordings consist of frame-number tagged records containing the (x, y) coordinates of the eye position along with pupil horizontal and vertical diameters. The analysis of these eye position data consisted of:

1. Preprocessing of the individual eye movement recordings to remove artifacts and blinks, and to identify saccades and fixation segments. (See [8] for descriptions of types of eye movements).
2. Merging of the recordings within age-gender groups to determine fixation overlaps in time and in position.
3. Filtering of the resulting POR file to remove jitter.

 ISSN/0003-0966X/03/3402-1152-$1.00+.00

3.1 Preprocessing of individual eye files

The individual subject's recordings were processed to remove eye movement recording artifacts, caused by blinks and other failures. The recording could fail if the subject moved his head too fast for the tracking to stay locked or when other specular reflections, such as tear film menisci, were erroneously detected by the ISCAN as a cornea reflection. The remaining data were analyzed to detect and keep fixation and smooth pursuit segments. The DVD only interrupts the processor every 0.4 to 1 second [9] with frame number information. Frame numbers between such interrupts were calculated from the elapsed time, assuming a 30 frame per second rate. This procedure resulted, on occasion, in non-monotonic or duplicate frame numbers.

Blinks and loss of tracking are filtered from the file by removal of frames containing zero data or frames where the pupil diameter was very small or too large. Removal of these data resulted in regions of apparent "frame dropouts". Although we expected the rejection rate due to pupil criteria would be small, they contributed substantially to the rejection rate shown in Table 2. This is most likely an indication that there was a problem with the pupil data that decreased our yield of data, but did not invalidate the remaining eye position information.

The data were converted to distance and velocity in units of degrees and degrees/second. A 30^{o}/sec or higher velocity threshold was used to reject saccades, dividing the remaining data into sequences of fixations or pursuits.

A fixation or pursuit sequence was started when we were in a region of no frame dropouts, (a skip of one frame was allowed), AND when the velocity was below threshold. A sequence was ended when the velocity exceeded this threshold. In addition there had to be >5 frames (equivalent to 150ms) in any accepted sequence.

For all these segments, the mean and standard deviation of the x and y coordinates and their correlation (r_{xy}) was calculated. A sequence was characterized as a fixation sequence if the range of x and y values were less than 1^{o} or if r_{xy} was < 0.5. A sequence that has r_{xy} >0.5 is more likely to represent a line (and therefore be a pursuit). Smooth pursuit segments were not used in the merge process that follows. Future processing will include the pursuit segments as well.

3.2 Merging of eye recordings of multiple eye files to find time and position overlaps

The filtered data from different individual observers are then merged to detect time and position overlaps. The files were traversed and the fixation segments were compared. A segment with the lowest frame number was chosen as the "reference" segment. The segments of the other files were tested against the "frame range" of the reference segment. To be considered an overlap, there had to be at least 2 frames overlapping the reference segment. The time overlap region information, including a count of how many fixation segments there were that overlapped the reference segment, was recorded. Note that it is possible to have more fixation segments in a time overlap region than there were observers if the reference segment was long. Once a fixation segment was used to detect a time overlap region, that fixation segment was not used again as part of other overlaps.

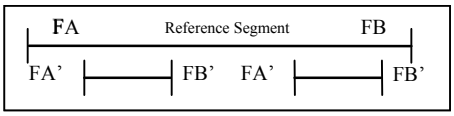

Figure 1. A time line showing a reference fixation segment and candidate overlapping segments with their starting and ending frame numbers (FA, FA', and FB, FB' respectively).

Then, across all segments being considered, the combined time overlap region statistics, $\langle x_S \rangle$, $\langle y_S \rangle$, σ_{SX}, σ_{SY} were calculated. The individual fixation segment averages were then compared to the time overlap region average to exclude any outlier that differed by greater than 2 standard deviations from the means. The means were then recalculated with these outliers excluded (Figure 2).

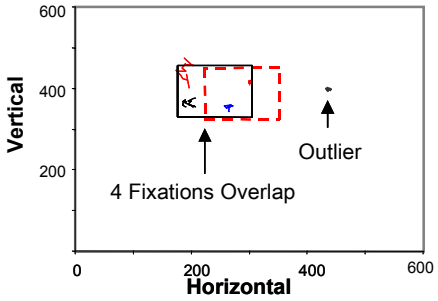

Figure 2. Eye position coordinates for five fixation segments within one time overlap region for three observers. The outlier is not used to calculate the POR for that time overlap region. The rectangles show the region of position overlap before (dashed) and after (solid) the outlier is removed.

To determine if there was a position overlap in the time overlap region, we set a criterion that the fixation segment averages must be within an area that was one quarter of the screen size. We rarely expect to need to magnify the image by greater than a factor of 4. Doing so might cause too much loss of context. The combined overlap region averages are again recalculated with only those fixation segments that had a position overlap. The final POR file for magnified playback of the video is based on these final time and position overlap region averages.

3.3 Post processing filtering to remove jitter

The resulting eye files frequently had POR coordinates that differed only slightly from one time overlap region to the next. When these were used to control the magnified video playback, the result was perceived as "jitter", or small jumps occurring at intervals on the order of a second. Therefore, a "jitter filter" step was implemented that smoothed these short-term fluctuations.

The filtering was done by finding "jump boundaries", where the respective x or y coordinates of two successive fixation points differed by more than 1/8th the maximum extent of the screen (half of the fraction used in the position overlap criteria). Once the "jump boundaries" were determined, the average x and y values were obtained for all fixation segments within the region for which there were no jumps. The fixation position for this multi-segments region was then set to this average and used in producing the jitter-filtered eye file.

4. Results

There were about 10 times as many fixation segments as there were smooth pursuit segments (Table 2). Since the preprocessing stage employed very strict criteria it produced a large rejection rate of data (Table 2). Although only 30 to 50 percent of the recorded data produced usable fixation and pursuit segments, we are confident that those that were accepted represent good quality fixations around bona-fide centers of interest. The redundancy built into having multiple subjects view all segments permitted us to obtain reliable multi-subject based POR despite the low yield for each individual observer.

Table 2. Percent of eye recordings rejected for each group of observers and the number of fixation segments used in this study as well as pursuit segments that will be used in the future. Despite the large rejection rate, the large number of accepted segments and redundancy across subjects permitted successful determination of POR.

Group	Rejected	# Fixations	# Pursuits
M<40	51.7%	16032	1404
F<40	67.7%	9644	752
F>45	68.5%	6777	471
M>45	70.4%	4914	465

If most people were found to fixate on the screen center, then it would be just as effective to always magnify around the screen center. Figure 3 shows the POR coordinates on the screen for two fully-processed clips. This shows that although clustered around the screen center, there was wide variation in the POR coordinates.

Histograms of the number of fixation segments that determined each time overlap region were computed (Figure 4). For those fixations that overlapped in time, Figure 5 shows how many overlapped in position. Of note is that of over 3000 fixations that overlap in time, only 280 did not overlap in position. This illustrates that the POR is the same about 90% of the time across observers from the same age and gender group.

5. Discussion

We have demonstrated that it is possible to determine the coordinates of the most important part of a scene by recording the eye movements of normally-sighted observers while they watch a movie. Over 90% of the time, different people look at the same objects.

The above analysis discarded "smooth pursuit" sequences of eye movements in the merge process. Smooth pursuits serve the same function as fixation and should be included in future analysis.

On occasion there may be more than one POR in a scene. This would happen if there were two reasonable possible points of attention in a scene, such as two people facing each other. One observer might look at person A first and then shift his/her attention to person B. Another observer might do the reverse.

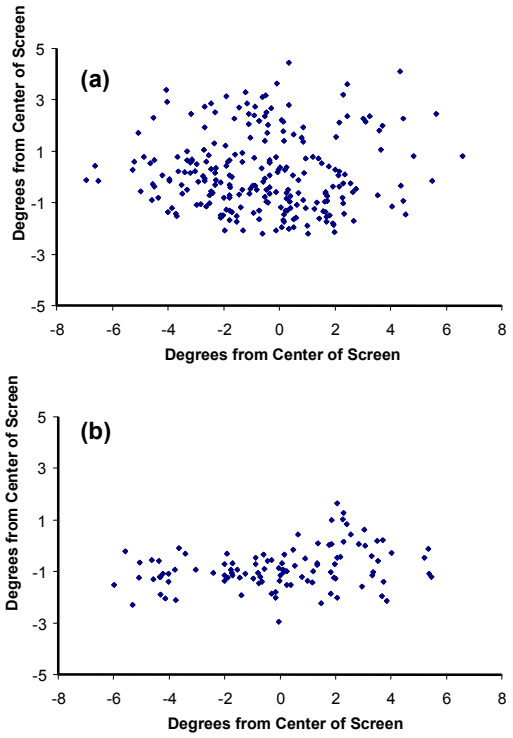

Figure 3. Fixation coordinates found for two video clips a. "Big" and b. "Network", determined from eye recordings of seven M<40 subjects. In both cases there are many instances where the POR is significantly off center.

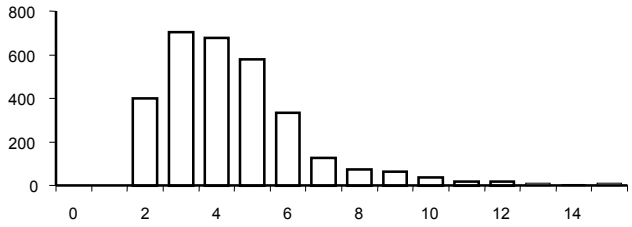

Figure 4. Number of overlapping (in time) fixations for seven M<40 subjects watching 37 minutes of video. The number of fixations is greater than the number of subjects in the cases where there were long fixation durations.

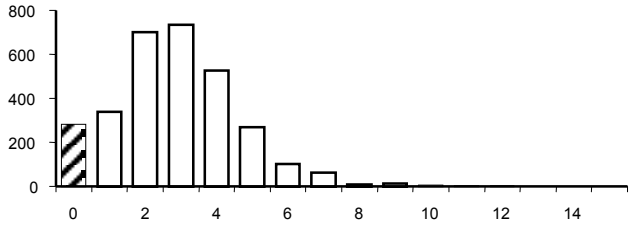

Figure 5. For those fixation segments where there were time overlaps, the number of fixations where the positions of the fixation overlapped within a 4.2° (horizontal) visual angle (3.2° vertically). Note that in only 280 out of 3000 fixations did the 7 M<40 subjects NOT look in the same position.

Algorithms need to be developed to detect this situation and deal with it. One way to deal with this situation is inclusion of the coordinates of several PORs with each frame for differing age-gender groups of viewers. This will only be useful if we determine that different groups indeed view consistently different objects. Another solution might be to change the magnification so that the magnified region includes both PORs.

The loss of context issue can be addressed in two ways. One way is for visually impaired viewers, using a remote control, to quickly toggle between magnification and no-magnification (Temporal Multiplexing) [10] Another way, which we are actively pursuing, is to superimpose an "edge-detected", high contrast, non-magnified image (a 'cartoon') over the magnified image (Spatial Multiplexing) [10]. We have developed a device that allows this (Figure 6). The user will be able to toggle this superposition on and off.

Figure 6. A 2.0X magnified image with an edge-detected cartoon of the full image superimposed.

When magnification is selected, the system magnifies the image as required and shifts the part specified by the POR to the screen center. It is possible for the user to override this function such that other parts of the magnified image may be scrolled onto the screen and viewed. The override or roaming function is likely to be useful only in static situations and scenes. In fast-moving scenery (as in a typical movie), there is no time to scan the scene before it is changed.

However, there are many television programs, varying from game shows to talk shows, where such a manual override may be useful.

Although formal evaluation of the process is underway with a population of visually impaired subjects, we have anecdotal evidence that this is an effective technique.

6. Impact

The effective use of electronic magnification as an aid to watching videos has never been realized. Although current television systems can do this in a low-cost and efficient manner, the scene changes that are inherent to video prevent using that capability effectively. Measuring and providing the coordinates of the most important part of the scene along with each frame, may realize magnification's full potential as a low vision aid. The eye movement method presented here is a natural and efficient way of determining these PORs. We envision that, just as programs are now being provided in "closed captioned" format, videos can be provided with these PORs encoded. In addition to its use for television viewing, the same system can be used for any videotape, DVD or other method of presenting motion videos.

7. Acknowledgements

NIH Grants EY05975 and EY12890

Rick Hier, DigiVision Inc. collaborated in developing the system

8. References

[1] Rubin, G. and K. Turano, Reading without saccadic eye movements. Vision Research, 1992. **32**: p. 895-902.
[2] Legge, G., et al., Psychophysics of reading II. Low vision. Vision Research, 1985. **25**: p. 253-66.
[3] Fine, E. and E. Peli. Computer display of dynamic text. in Vision 96: International Conference on Low Vision. 1996. Madrid, Spain.
[4] Fine, E. and E. Peli, Benefits of rapid serial visual presentation (RSVP) over scrolled text vary with letter size. Optometry and Vision Science, 1998. **75**: p. 191-6.
[5] Kazuo, S. and J. Shimizu. Image scaling at rational ratios for high-resolution LCD monitors. in SID 00 Digest of Technical Papers. 2000. San Jose, CA: Society for Information Display.
[6] Digivision, DZ-1. 2001: San Diego, CA.
[7] ISCAN, Raw Eye Movement Data Acquisition Software Operating Instructions. 3.1 ed. 1996, Burlington, MA: ISCAN, Inc.
[8] Stark, L. and D. Norton, Eye movements and visual perception. Sci Am, 1971. **224**(6): p. 35-43.
[9] Microsoft, DirectX. 2001, Microsoft: Redmond, WA.
[10] Peli, E., Vision multiplexing - an engineering approach to vision rehabilitation device development. Optometry and Vision Science, 2001. **78**: p. 304-315.