Guidance of Eye Movements on a Gaze-Contingent Display

M Dorr, T Martinetz, K Gegenfurtner, and E Barth

Institute for Neuro- and Bioinformatics, University of Lübeck
Abteilung Allgemeine Psychologie, Universität Giessen
{dorr, martinetz, barth}@inb.uni-luebeck.de, Karl.R.Gegenfurtner@psychol.uni-giessen.de

Abstract. We present a system that allows to display and manipulate high resolution image sequences as a function of gaze in real time. An eye tracker was connected to a video workstation and software was developed with an emphasis on low latency, so that the whole system is capable of reacting to an eye movement within less than 30 ms. In first experiments with this system, we measured the effect of presenting peripheral visual patterns on the subjects' scan paths when viewing 20 s long video clips of natural outdoor environments. Small red squares or looming patterns were overlaid on the video for 120 ms at locations about 12º away from the position of gaze. The actual location was chosen randomly from candidate locations that had some minimal amount of saliency (spatio-temporal curvature). Results show that for some subjects, about 70% of the peripheral stimuli triggered a saccade to the desired location, while in the worst case this rate was about 20%. Our research aims at creating new forms of visual communication and vision-based interaction [1].

Introduction

We perceive our visual world as colourful and rich in detail across the whole visual field. The fact that we actually perceive only very little detail at a time remains mainly unconscious because we constantly scan the visual world by moving our eyes. These eye movements are not only a necessary prerequisite for successful visual perception, but their exact order does also determine to a large extent *how* we perceive the visual world. For specific tasks, the human visual system deploys specific patterns of eye movements, or scan paths. This also means that a given scan path can be unsuitable for unexpected situations or tasks, for example change detection [2, 3]. We therefore believe that the scan path should become part of visual communication systems as are the classical image attributes, luminance and colour. Thus, the scan path needs to be recorded, processed, and “displayed”. We plan to display a desired scan path by guiding the gaze of the observer.

We will present here a prototype system that can affect the eye movements subjects make while looking at video clips of natural dynamic scenes. To this end, these video clips are manipulated dependent on the current fixation point and the spatio-temporal curvature of the image sequence to create stimuli that are designed to attract the observers' gaze.

These experiments will allow us to continue towards systems that are able to guide the scan path of an observer. Such a recommended scan path could improve human-machine communication, maximize information content of a display, or could function as an aid to disabled people.

Theory

With an image sequence defined by its image-intensity function $f(x, y, t)$ and partial derivatives denoted f_x, f_y, f_t, the structure tensor is

$$J = w * \begin{pmatrix} f_x f_x & f_x f_y & f_x f_t \\ f_x f_y & f_y f_y & f_y f_t \\ f_x f_t & f_y f_t & f_t f_t \end{pmatrix}$$

with w as a spatial Gaussian smoothing kernel. The intrinsic dimension (inD) of f is n if n eigenvalues λ_i of J are non-zero, thus i2D features are those where the spatio-temporal curvature

$$S = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3$$

is non-zero. Image sequences are completely defined by their i2D regions [4].

System

Because of the computational complexity of video-based eye-tracking and real-time image processing on high resolution video, the system consisted of two high-end personal computers, one for tracking of gaze, the other one for display of actual stimuli. The first workstation was connected to a commercially available SensoMotoric Instruments iViewX Highspeed eye tracking system that has a temporal resolution of 240 Hz and a spatial resolution of approximately 0.5-1º. Gaze data was sent via a dedicated gigabit UDP network connection to the display workstation, to which a 22" video display unit (Iiyama HM204DT) was attached that span a visual field of approximately 44º x 33º at a viewing distance of 50 cm. The VDU ran at 150 Hz with a spatial resolution of 1280 x 720 pixels (the video resolution).

For on-line saccade detection, a velocity-based algorithm was used with two different thresholds for saccade onset and offset.

Videos

Videos of 20 s duration were taken in and around Lübeck with a digital JVC JY-HD10 video camera and stored in MPEG-2 format. One movie shows a bumblebee flying around flowers, two show street scenes at an intersection and a roundabout, and the fourth movie was taken on a sunny day at the beach. Spatial resolution was 1280 x 720 pixels (HDTV standard) and temporal resolution was 30 Hz.

Stimuli

Two different kinds of stimuli were used. The first stimulus type was a simple red square with a size of 1º x 1º. Its luminance was proportional to local spatial contrast as defined in [5]. The second type was based on the assumption that peripheral movement towards an observer should have a high perceptual saliency. Therefore, we designed a looming pattern that magnified the contents of a square of 2º size with an increasing factor from 1 to 3 over 60 ms, mimicking the optical expansion of an approaching object.

Candidate Points

We considered only those locations in the video to be suitable for stimulation that belonged to the 50 points of a particular frame with the highest amount of spatio-temporal curvature. This measure has also been used to predict saccade targets [6]. Using these points excluded constant image regions where the looming pattern would have had no visual effect at all. Because of the strong preference of the human visual system for horizontal saccades, another constraint was that the angle between the horizontal axis and a line between the candidate point and gaze was less than 45º.

Timing

While freely watching videos, subjects make about 2-3 saccadic eye movements every second. To ensure that stimulation would not occur too close to one of these regular saccades, when stimulation could not possibly have an effect because of the afferent latency of the visuomotor system, the point in time of the next saccade was predicted by adding the median of all previous intersaccadic intervals to the last saccade. Stimulation was then switched on 250 ms before the estimated next saccade. Because of the variance of fixation duration, stimulation was switched off after 120 ms or when the subject made a saccade.

Procedure

5 subjects were shown 4 movies each. The red squares stimulation was overlaid the first and last movie, while the two other movies were shown with the looming pattern stimulation. The instruction to the subjects was to just watch the videos. Because of the need to stabilize their heads in a chin-rest, subjects were told that their eye movements were to be recorded.

Results

We obtained data for 5 subjects. As examples, we present data for 2 subjects in Fig. 1 and 2. Because the different video clips gave similar results, these data are pooled over all four movies, and thus pooled over both types of stimulation as well. Stimulation had the strongest effect on subject 2 (shown to left) and the weakest effect on subject 3 (shown to the right). Subject 1 reported to have consciously fixated, therefore rendering stimulation ineffective. Nevertheless, some stimulus-directed saccades were made. Data for subject 4 are qualitatively similar to those of subject 2, while subject 5 showed intermediate results.

In Fig. 1, lines with an almost vertical downward slope indicate stimulus-directed saccades. Note that for both subjects, there is a high number of such saccades around 200 ms, which coincides well with typical saccade latency.

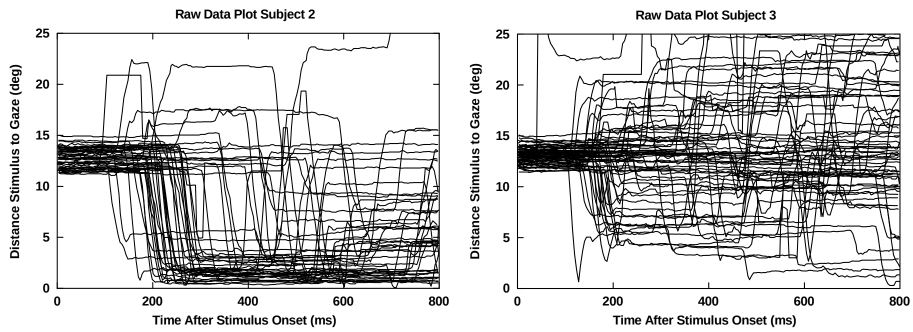

Fig. 1. Data for two subjects, pooled over all four movies. Plotted is how the distance between point of gaze and stimulus location develops over time after stimulus onset. Each line shows one stimulus presentation. Almost vertical slopes indicate saccades, with downward slopes showing eye movements towards the stimulus location.

Discussion

The following conclusions seem appropriate despite the small number of experimental trials. The strength of the guidance effect depended more on the subjects than on the movies and the type of stimulation. None of the subjects reported to be following the stimuli with their gaze, except for one subject in one quiet scene. Nevertheless, subjects noticed that red dots appeared in the movies. The looming pattern was less visible and was perceived as a video artifact. In future experiments, however, we will attempt to create stimuli that remain below the detection threshold.

We have implemented the first gaze-contingent system that can operate with low latencies on a high-resolution video stream and we conclude that scan path guidance seems to be possible in principle.

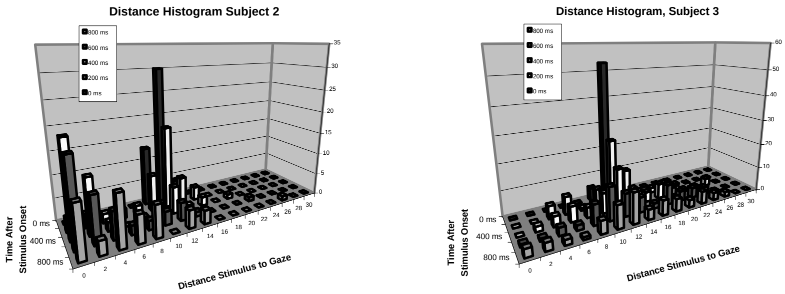

Fig. 2. These plots show the distribution of distances between point of gaze and stimulus location for several points of time after stimulus onset. Each colour shows a histogram of a cross-section of the plots in Fig. 1. The peaks around 12º at 0 ms are due to stimulus placement. Later peaks around 0º show the number of saccades towards the stimulus location.

Acknowledgements

Research is supported by the German Ministry of Education and Research (BMBF) under grant number 01IBC01B. We thank SensoMotoric Instruments GmbH for their eye-tracking support; data were obtained using their iViewX Highspeed system.

References

1. http://www.inb.uni-luebeck.de/Itap
2. J K O'Regan, H Deubel, J J Clark, and R A Rensink, Picture changes during blinks: looking without seeing and seeing without looking. Visual Cognition, 7:191-212, 2000
3. D J Simons and C F Chabris, Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception, 28:1059-74, 1999
4. C Mota and E Barth, On the uniqueness of curvature features. In G Baratoff, H Neumann (eds), Dynamische Perzeption, vol. 9 of Proceedings in Artificial Intelligence, 175-8, 2000.
5. P Reinagel and A M Zador, Natural scene statistics at the centre of gaze. Network: Computation in Neural Systems, 1, 341-350, 1999.
6. E Barth, J Drewes, and T Martinetz, Dynamic predictions of tracked gaze. In Seventh International Symposium on Signal Processing and its Applications, Special Session on Foveated Vision in Image and Video Processing, 2003.