MODELING VISUAL ATTENTION FOR GAZE-CONTINGENT VIDEO PROCESSING

Andrew T. Duchowski and Bruce H. McCormick, {andrewd|mccormick}@cs.tamu.edu

Department of Computer Science
Texas A&M University
College Station, TX, 77843-3112

ABSTRACT

We present a visuotopic model of visual attention for gaze-contingent video processing which incorporates dynamic Regions Of Interest (ROIs) to anticipate saccades and preserve preview benefit. The temporal nature of the model is founded on characteristics of two principle types of eye movements: saccades and fixations. Results are presented demonstrating the application of the model to a video sequence.

Keywords: Gaze-contingent Display, Multiresolution, ROI, Wavelets, Video Processing.

1. INTRODUCTION

Our model is based on the assumption that visual attention selects the next point of fixation in the periphery [1]. In the model, foveal ROIs match the information requirement for processing the "what" of visual attention. Regions outside ROIs are smoothly spatially-degraded to match the Human Visual System (HVS) acuity function. Anticipating saccades, peripheral ROIs (pROIs) are used for two purposes: (1) to preserve preview benefit [2], potential visual attractors are represented at high resolution for the "where" of visual preattention; and (2) to provide a high resolution inset so that a rapid change of fixation does not meet a low resolution region in a gaze-contingent application.

2. THE MODEL

Saccades typically complete in the range of 10-100ms [2]. Anticipating the destination of the saccade, the new fixation location must be fully represented at foveal resolution just prior to saccade completion. That is, pROIs must be fully represented within $.01 - 0.1 f_r$ frames, where f_r is the displayed frame rate (in fps). Since experiments have shown that a decision may be made to abort and reprogram the current saccade at the sudden onset of new stimulus during the saccade [2], pROIs must also be generated gradually so as not to interfere with the ballistic outcome of the saccade. The use of a temporal ramp is a suitable approach [3].

Fixations range from 150-600ms [2], implying that (foveal) ROIs should persist for $.15 - .6 f_r$ frames. Once fixation has been relocated, and the 600ms duration requirement has expired, the ROI should be extinguished gradually so as not to artificially induce the observer to refixate on the current target. A temporal ramp may again be used for this purpose.

3. RESULTS

Our ROI approach is similar to the work presented in [4], with wavelet coefficients decimated to synthesize a smoothly spatially-degraded image, relative to each ROI. ROIs are superimposed over each subband, with coefficients scaled according to a smooth mapping function matching HVS spatial acuity [5]. Unlike the work in [4], our goal is an unobtrusive representation where the observer does not differentiate between the original and processed sequences and furthermore is not induced to fixate on regions represented by ROIs.

Figure 1 shows processed frames from the 40-frame *tennis* sequence with ROIs determined a priori by a fictitious (but not implausible) visual scanpath [1], shown in Figure 2. Assuming a 30fps display rate, fixation durations (including pursuits) are $\leq$ 200ms (6 frames), saccade durations are $\leq$ 33ms (1 frame), and preview and fadeout of ROIs is accomplished in 66ms (2 frames) before and after fixation changes, currently without a temporal ramp. The difference images in Figure 1 show degradation of high frequency components outside the imposed ROIs, which according to the model, should be imperceptible given fixations centered on the ROIs.

[1] The term *scanpath* was originally coined by Noton and Stark [6].

We are currently installing an eye tracker to validate the model through subjective quality testing. Although ideally the model should be applied in real-time, with pROIs determined automatically and invoked just prior to fixation changes, we will first experiment with processing video off-line. A record of the subject's fixation points over individual frames will be made in preliminary viewing trials. Since we expect scanpaths to differ from trial to trial, this record will be used to encode a list of candidate frames, each with a pROI representing a possible future fixation. During display of the processed sequences, when a saccade is detected, the candidate frame containing the pROI closest to the projected fixation will be displayed next. We expect to show perceptual benefit of sequences with pROIs over sequences processed with only a single (foveal) ROI.

4. CONCLUSION AND FUTURE WORK

We have presented a model for gaze-contingent video processing utilizing ROIs to anticipate saccades and preserve preview benefit. We are currently formulating a framework for dynamic ROI representation which includes the following studies: (1) incorporating gradual onset and fading of ROIs via a temporal ramp; (2) considering other perceptual limitations of the HVS such as contrast sensitivity; and (3) developing methods to automatically detect potential visual attractors in the periphery. We are working on extending our algorithm to include temporal analysis utilizing the 3D wavelet transform, and generalizing the model to express dynamic ROIs as space-time *Volumes Of Interest (VOIs)*.

5. REFERENCES

[1] A. H. C. Van der Heijden. *Selective Attention in Vision*. Routledge, 1992.

[2] Keith Rayner, editor. *Eye Movements and Visual Cognition: Scene Perception and Reading*. Springer-Verlag, 1992. Springer Series in Neuropsychology.

[3] Lew B. Stelmach and Wa James Tam. Processing Image Sequences Based on Eye Movements. In Bernice E. Rogowitz and Jan P. Allebach, editors, *Human Vision, Visual Processing, and Digital Display V*, pages 90–98, San Jose, CA, February 8-10 1994. SPIE.

[4] E. Nguyen, C. Labit, and J-M. Odobez. A *ROI* Approach for Hybrid Image Sequence Coding. In *Intl. Conference on Image Processing (ICIP)*, pages 245–249, Austin, TX, November 13-16 1994. IEEE.

[5] Andrew T. Duchowski and Bruce H. McCormick. Representing multiple ROIs with wavelets. In *Human Vision and Electronic Imaging*, San Jose, CA, January 28-February 2 1996. SPIE. In review.

[6] David Noton and Lawrence Stark. Eye Movements and Visual Perception. *Scientific American*, 224:34–43, 1971.

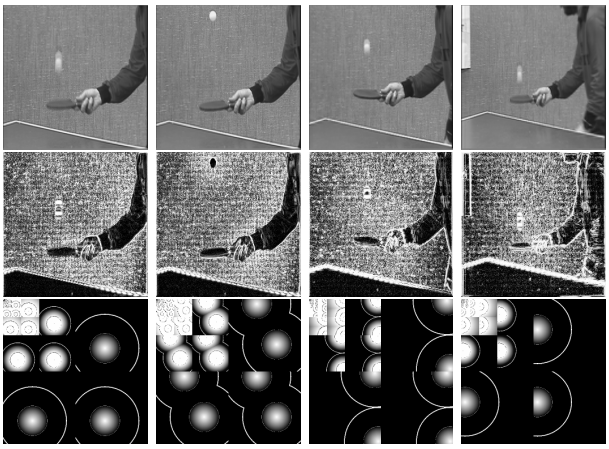
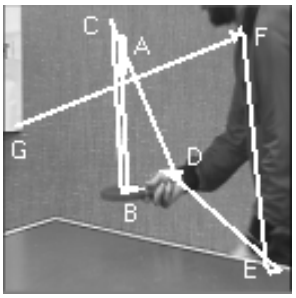


Figure 1: Frames {10,20,30,40} of the *tennis* sequence. Top: images processed with Daubechies-6 wavelets. Middle: difference frames (enhanced through histogram equalization for clarity). Bottom: wavelet scaling function superimposed over wavelet pyramid. White regions represent coefficients scaled by constant 1, black regions by constant 0. Intermediate regions are scaled by a linearly interpolated value on the interval [0,1]. White rings are artificially superimposed resolution level boundaries, interspersed according to the HVS acuity function.



A: $f_1 \rightarrow f_6$, fixation (pursuing ball), B: $f_7 \rightarrow f_{12}$, saccade to paddle, fixation, C: $f_{13} \rightarrow f_{18}$, (ball hits paddle) fixation (pursues ball), D: $f_{19} \rightarrow f_{24}$, saccade to wrist (contrast edge), fixation, E: $f_{25} \rightarrow f_{30}$, saccade to left hand (onset object), fixation, F: $f_{31} \rightarrow f_{36}$, saccade to shirt collar, fixation, G: $f_{37} \rightarrow f_{40}$, saccade to picture (onset edge), fixation.

Figure 2: Example of possible scanpath (superimposed over frame 40).