Probability models for Visual Search

Literature Survey

Umesh Rajashekar

EE381K - Multidimensional Digital Signal Processing

FALL 2000

The University of Texas at Austin

Abstract

The development of efficient artificial machine vision systems depends on the ability to mimic aspects of the human visual system. Humans scan the world using a high-resolution central region called the fovea and a low resolution surrounding area to guide the search. A direct consequence of this non-uniform sampling is the *active* nature in which human visual system gathers data in the real world using fixations and saccades. In this report, we look at a few techniques that attempt to mimic the visual search strategies of the human visual system.

1.INTRODUCTION

The human visual system uses a dynamic process of actively scanning the visual environment. The active nature of scanning is reflected in the eye scanpath pattern. These sequence of fixations and saccades (constituting the scanpaths) are attributed to the distribution of the photoreceptors on the retina. The photoreceptors are packed densely at the point of focus on the retina - fovea and the sampling rate drops almost exponentially from the fovea. Fig 1 shows a typical retinal sampling grid. As a result, humans see with very high resolution at the fixation point and the resolution falls away from the fixation point. Fig 2 shows a typical image on the retina - the fixation point being the middle of the image. In order to build a detailed representation of the image, the eye scans the scene with a series of fixations and jumps (saccades) to new fixation points. Information is gathered by the eye during fixations while no information is gathered during the saccades. The fixation duration is about 200ms. Fig 3 shows a typical scan path of the human eye while looking at the image [1].

The active nature of looking has its advantages in terms of speed and reduced storage requirements (due to the non-uniform resolution across the image) in building artificial vision systems. It also have significant applications in the area of video compression where the region around the fixation point in the video sequence is transmitted with high resolution and regions away from the fixation point are blurred. The development of foveation based artificial vision systems and video compression schemes depends on the ability to determine the fixation points/area of interest regions in the image. However, in general, we cannot predict a person’s scanpath while viewing a scene in a realistic way. One common solution to determine the eye scan path is the use of eye trackers. An alternative solution is to develop models for the fixation problem. Since deterministic solution to the fixation point prediction problem is impossible (different people look at the same image using different scan paths based on the motive), I propose to investigate the possibility of building a probabilistic model for eye fixations in a visual search environment.

Besides the applications already mentioned, the development of such a fixation model has significant applications in computer vision applications such as pictorial image database query and image understanding.

2. Previous models for fixation point selection

The primary goal of many machine vision systems has been the development of algorithms that interpret visual data from cameras to help computers to *see*. Most of the active vision systems developed are developed for a specific task and hence perform only in constrained scenarios. In this section we will briefly go over three such techniques.

1. Image features and fixations

Privetera and Stark [2] propose a computational model for human scanpaths based on intelligent image processing of digital images. The basic idea is to define algorithmic regions of interest (aROI) generated by the image processing algorithms and compare the result with human regions of interest (hROI). The comparison of the aROI and hROI is accomplished by analyzing their spatial/structural binding (location similarity) and temporal/sequential binding (order of fixations). Based on their experiments, the aROI generated by wavelet decomposition of the image (which is inherently multi-resolution) seemed to match the hROIs well. Symmetry and contrast also seemed to be strongly correlated with fixation. Their results also indicate that the fixation point prediction can be no better 50% i.e. only half the predictions made are accurate. While the results of this paper are definitely promising, the techniques to determine fixation points do not seem to account for the fact that the next fixation point selection is dependent on the current fixation point. Further, a weighted result of using multiple image processing algorithms might produce better prediction of aROIs.

2. *Probability models*

Klarquist and Bovik [3] propose an alternative technique for fixation point selection in 3D space. The fixation point selection was developed for FOVEA - "an active vision system platform with capabilities similar to sophisticated biological vision systems"[3]. FOVEA uses a probabilistic approach to fixation point selection and hence makes the selection of the fixation point less rigid and also contingent on the features around the current fixation point. The probability model is developed using a number of criteria. The fixation point selection process is independent of the criteria and hence creates a clear dichotomy between the selection criterion and the selection process. The selection criterion is based on local information content (gradient information), proximity of the candidate fixation point to the current fixation point and the surface map in the vicinity of the current fixation point. However no indication of the performance of their system with human scanpaths is provided.

3. *Saliency models for image understanding*

Henderson [4] proposes a more robust method towards fixation point selection in images. The model incorporates the cognition factor involved in fixation point selection. The initial fixation map is derived by analyzing low-level features (contrast, edges) in the image. Based on the task at hand (search for a target), the model is trained to "understand" the image. Incorporating cognition into a model is a difficult task since cognition is task specific. However the proposed model facilitates both the prediction of the fixation points and the duration of fixations.

3. Probability models for Visual Search

For my project, I propose to investigate probability models in a constrained visual search environment. Human subjects will be presented with a target that he/she will need to search in a scene. The eye movements will be recorded using the Model 504 remote eye tracker from Applied Science Laboratories (ASL). The scanpaths will be analyzed for extracting fixation points and saccades. The region around the fixation points will then be analyzed to understand the similarity between the target and region the eye fixated on. The analysis of the similarity will be based on correlation, edge co-occurance and eigen decomposition techniques to name a few. This data will be analyzed to determine "attractors" - features that force the eye to fixate around a point as opposed to other points. This result will then be extended to determine a probability fixation model based on density of "attractors" in the image. These results will then be tested with human scan path data. The experiments will be conducted using the MATLAB psychophysics toolbox and ASL provided software for scanpath analysis.

4.Conclusion

Determination of fixation points is an important step in applications that use foveation strategies. Development of a probability model in a constrained "visual search" environment is proposed. This work will be extended in future to the "visual surveillance" which differs from visual search in that there is not target to be found and hence generalizes to the case of finding fixation points in an arbitrary scene.

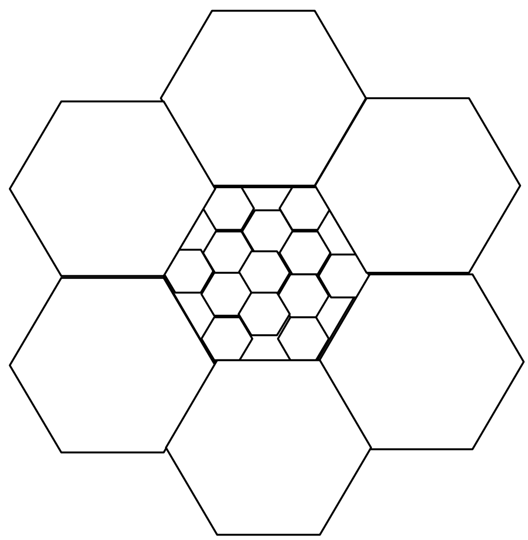

Figure 1:Sampling structure in the retina

Figure 2A: Original uniform resolution image of Lena

Figure 2B:Foveated image of Lena

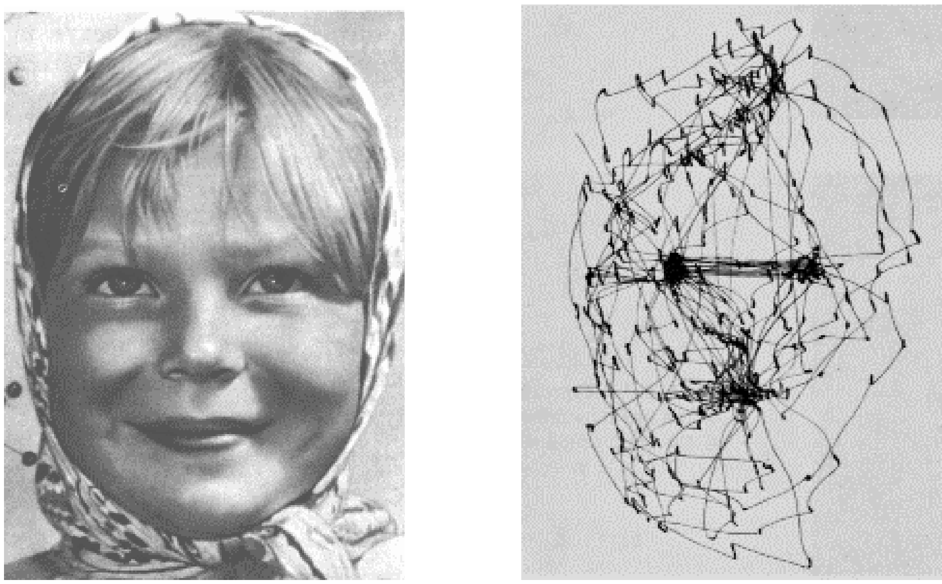

Figure 3: Fixation and saccades in viewing an image [1]

4. References

1. A. L. Yarbus, *Eye movements and Vision* New York:Plenum Press, 1967
2. C. M. Privitera and L. W. Stark, "Algorithms for Defining Visual Regions of Interest: Comparison with Eye Fixations," *IEEE Tran. On Pattern Analysis and Machine Intelligence,* September 2000, Vol 22, No 9, Pg. 970-982
3. W. N. Klarquist and A. C. Bovik, "FOVEA: A Foveated Vergent Active Stereo Vision System for Dynamic Three-Dimensional Scene Recovery," *IEEE Trans on Robotics and Automation,* October 1998,Vol 14,No 5, Pg 755-770
4. J. M. Henderson, "Eye movement control during visual object processing: Effects of intial fixation position and semantic costraint" ,*Journal of Experimental Psychology*, 1993, Vol 47, Pg 79-98