

# **Visual Search using Principal Component Analysis**

**Project Report**

**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 the human visual system gathers data about the real world using fixations and saccades. In this report, we investigate the results of modeling this active scanning process using principal component analysis in a visual search task.

## 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 sequences 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 known as the fovea and the sampling rate drops almost exponentially away from the fovea. As a result, humans see with very high resolution at the fixation point and the resolution falls away from the fixation point. 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 [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 has 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. In addition to 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. 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 a 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.

## **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 *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.

### *2.1. Image features and fixations*

Privitera 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). Their results 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.

### *2.2. 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 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.

### *2.3. 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.

## **3. Visual Search using Principal Component Analysis**

The goal of the following visual search experiment was to investigate the presence of features in images that attract a subject's eye in a "target-search" task. The idea behind this experiment was to extract features in fixation regions that resemble the target and hence forcing the eye to fixate on these regions. While several "attractors"

have been discussed in [1], the use of principal component analysis (PCA) seemed attractive due to its success in face recognition systems [5].

### *3.1 SVD based face recognition*

The following is a brief overview of a Singular Value Decomposition [SVD] based algorithm for face recognition. The SVD [6] decomposes a matrix  $A$  into its left singular vectors  $L$ , right singular vectors  $R$  and the corresponding singular values  $V$ . It can be shown that the singular vectors represent the “shape” information while the singular values are representative of the “gain” in the image. In face recognition systems, the face database is normalized with the singular values of the input (search) image and an SVD index is computed as the sum of absolute difference of the face database image and the target image. The database image with the lowest SVD index is chosen as the recognition result. The recognition rate can be as high as 85% [5]. A similar matching approach is used in the visual search experiment described below.

### *3.2 Experiment details*

The experiment was conducted to investigate if the eye uses a technique similar to PCA for finding targets in images. The following is a description of the experimental setup.

Subjects were asked to search for targets such as Fig. 1 in a larger image such as Fig. 2 as fast as possible. Their eye movement was recorded using the Model 504 remote eye tracker from Applied Science Laboratories (ASL). The target image was chosen to be 100\*100 pixels and the image to be searched in was 1024\*768 pixels. Subjects were seated at 32” from a 21” flat screen monitor on which the images were displayed. To avoid the complications of color in the search process, all images were gray scale.

Further, since cognition of the image makes the analysis of the search results very complex, all images selected were abstract computer generated images taken from <http://www.visualparadox.com>. Also, to avoid quick head movements, subjects were instructed to use a chin rest during the data recording process. The experiment was conducted on 4 subjects and each subject was shown 7 images. The EYENAL data analysis software from ASL was used to analyze the eye scan paths into fixations and saccades. MATLAB was used to perform all other computations.

### *3.3 Data analysis*

Once the scanpaths were analyzed into fixations and saccades, a region (the same size as the target) centered at each fixation points was extracted from the image and an SVD index with respect to the target image was computed as described earlier. Fixations that lie at image boundaries and outside image boundaries are not amenable to processing and hence are ignored. The SVD index is set to  $-1$  for these fixations. Fig. 4 shows a plot of SVD index for the fixation pattern shown in Fig. 3.

### *3.4 Interpreting the results*

A lower SVD index corresponds to a good match to the target while a negative index corresponds to invalid data. It is interesting to note in the plot that many of the eye's fixations have a small SVD index, which might indicate the use of PCA for target search by the eye. However since the experiment has been conducted only for a limited number of subjects, this generalization cannot be made with certainty. Another interesting point to note is the sudden jump the eye makes from points to high SVD index to one of low values.

#### 4. Conclusions and future work

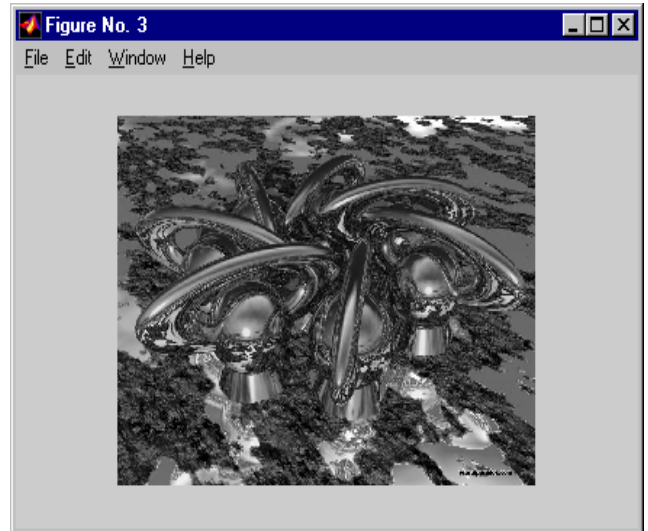
The SVD seems to be a promising tool for visual search. However conclusions about the efficacy of the SVD algorithm cannot be made with certainty unless more experiments are performed. If these results seem consistent, the SVD index can be used to generate a probability map of fixations. However, this experiment was a good opportunity to familiarize with the eye tracker's operation and set up a test bed for more experiments. Future experiments will involve analyzing the image to investigate dependency of the fixation point and regions of the image with a 3 cycles per degree of visual angle spatial frequency content. This is due to the fact that the eye is sensitive to the above spatial frequency. Another interesting experiment is to foveate the image at a given fixation point and predict the next fixation point based on analysis of the foveated image, the fixation point being the previous fixation point.

#### 4. References

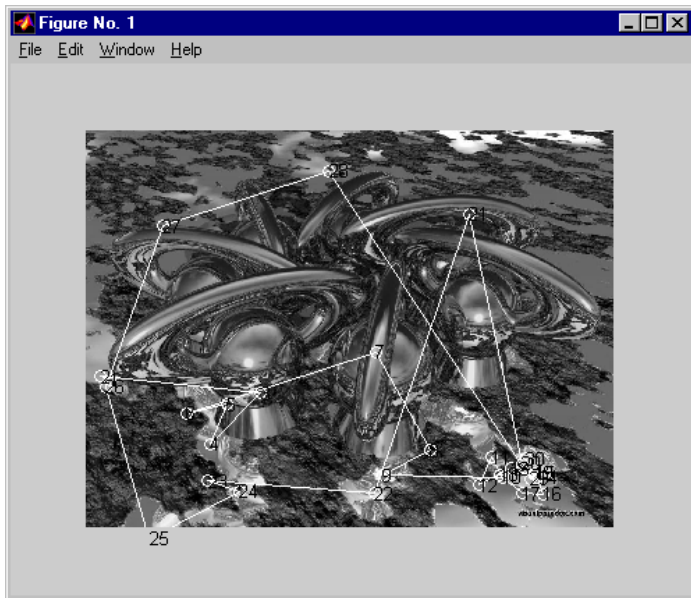
1. A. L. Yarbus, *Eye Movements and Vision*, New York:Plenum Press, 1967.
2. C. M. Privitera and L. W. Stark, "Algorithms for Defining Visual Regions of Interest: Comparison with Eye Fixations," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Sep 2000, Vol 22, No 9, pp. 970-982.
3. W. N. 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*, Oct 1998, vol 14, No 5, pp 755-770.
4. J. M. Henderson, "Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint", *Journal of Experimental Psychology*, 1993, Vol 47, Pg 79-98.
5. M. A. Turk and A. P. Pentland, "Face recognition using Eigenfaces," *Computer Vision and Pattern Recognition*, 1991 pp 586 –591
6. G. Strang, *Linear Algebra and its applications*, Third edition, Harcourt College Publishers; 1988.



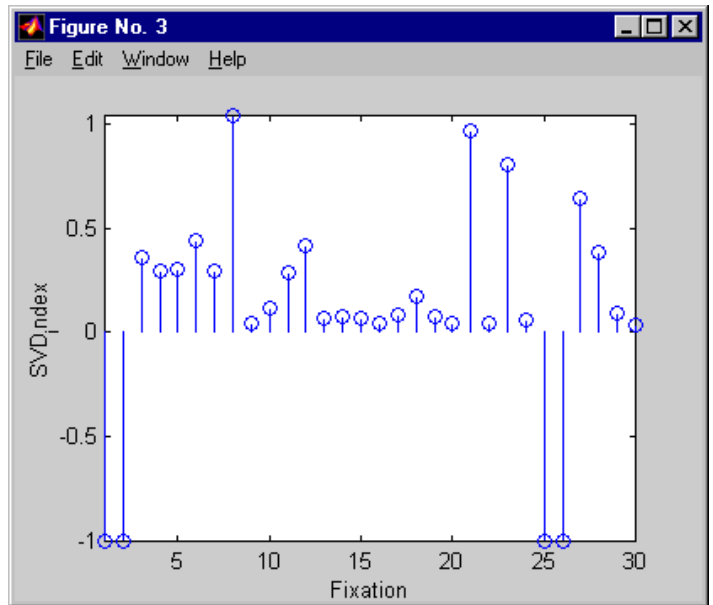
**Fig. 1 :Target**



**Fig. 2: Image containing target**



**Fig. 3: Fixation pattern**



**Fig. 4: SVD index vs Fixation**