Gaze-Centric Image Analysis for Efficient Visual Search

Umesh Rajashekar, Thomas Arnow, Alan C. Bovik and Lawrence K. Cormack

Despite recent advances in computer vision, pattern recognition, and image processing, many visual tasks remain unsolved. Consider, for example, *visual search* – the problem of finding a target in a background of distracters. Whenever we look for a familiar face in an audience, or search for a misplaced item, we engage in visual search. Given the infinite variations of a target's features (size, orientation, color), and background conditions (lighting, occlusion), it is a marvel that humans excel at searching and distinguishing objects.

One aspect of the human visual system (HVS) critical to its success as an efficient searcher is its *active* nature of looking. To avoid the data glut accompanying its large field of view, the HVS processes only a tiny central region at high resolution, while the resolution of processing falls rapidly towards the periphery. To assimilate visual information from this *foveated* input (Fig. 1), the HVS uses a combination of steady eye *fixations* linked by rapid ballistic eye movements called *saccades* (Fig. 2).

Fig 1: Foveated image when fixating the climber.

Fig 2: Typical eye scan pattern. Fixations are red circles and saccades are black lines.

Understanding how the HVS selects and sequences fixations should help the design of active artificial vision systems with great potential utility in applications such as automated pictorial database query, autonomous vehicle navigation, and semi-automated inspection of medical radiographs.

Theories for automatic gaze selection broadly fall into top-down and bottom-up categories. Top-down approaches emphasize a high-level, cognitive or semantic understanding of the scene. Bottom-up approaches assume that eye movements are strongly influenced by low-level image features such as contrast and edge density. Given the rapidity and sheer volume of saccades during search tasks, it is also reasonable to suppose that there is a significant random component to fixation locations.

Inexpensive accurate eye tracking devices make it possible to compute image statistics at observers' points of gaze or *fixation points*. In our recent work, we have been using precision eye trackers to record eye movements of subjects performing visual tasks. We apply methods of visual psychophysics and image processing to extract image statistics at fixation points. By determining statistical features that differ significantly at human fixations and ones randomly placed by computer on the same image, we are able to use the statistical features as predictors to automatically compute likely fixations in natural images. Our experiments[1] have followed two scenarios: *target search* and *visual surveillance*.

In our *target search* experiments, we use a noise-based reverse correlation technique to find whether observers use structural cues to direct their fixations when searching for simple targets embedded in naturalistic *1/f* noise stimuli at low SNRs. We have been able to demonstrate idiosyncratic, target-dependent features used by observers in these tasks. Even in noisy displays, observers do not search randomly, but deploy fixations to regions that resemble aspects of the target[2].

In our *visual surveillance* experiments, we record the fixations of observers as they freely viewed calibrated natural images. We have attempted to quantify the differences in feature statistics in image patches centered on human and randomly selected fixations. As expected, humans tend to fixate on regions with higher contrast. More interestingly, regions that *differed* from their surrounding (in their luminance and contrast) more strongly attract visual fixations.

Based on this work we are developing automatic fixation selection algorithms that use either the classification images (for target search) or a linear combination of luminance and contrast image features (for visual surveillance) as cues to decide fixations. Thus far, the distributions of computed fixations have been found to correlate quite well with observers' fixations. We quantify the similarity between fixation patterns using the information-theoretic Kullback-Leibler distance.

One interesting application we are studying is *visual search for corners*[3]. We use principles of foveated visual search and automated fixation selection to accomplish the corner search, in an attempt to demonstrate a case study of both foveated search and feature detection. The result is a new algorithm for finding corners which will be used to drive object recognition and as a corner-based algorithm for directing fixations for machine vision.

Automatic fixation selection algorithms are fundamental for the design of active vision systems. By studying the interplay between eye movements and gaze-centric image statistics, we have been gaining insight into how humans deploy fixations. The strong influence of low-level image attributes on fixation locations in our findings suggests a significant bottom-up component in gaze selection. We are currently investigating the influence of motion, color, and stereo primitives in drawing fixations. Using an information-theoretic framework and recent models of natural scene statistics, we are also developing optimal fixation algorithms to extract the maximum amount of structural information from a scene using the minimum number of fixations[4].

References

1. http://live.ece.utexas.edu/research/eyefix
2. U. Rajashekar, A. C. Bovik, and L. K. Cormack, *Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis*, **Journal of Vision 6** (4), pp. 379-386, 2006. doi:10.1167/6.4.7.
3. T. L. Arnow and Alan C. Bovik, *Finding corners in images by foveated search*, in J. G. Apostolopoulos and A. Said (eds.), **Proc. SPIE Int. Soc. Opt. Eng. 6077** (60770Y), 2006. doi:10.1117/12.644430
4. R. Raj, W. S. Geisler, R. A. Frazor, and A. C. Bovik, *Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy*, **J. Opt. Soc. Am. A 22** (10)**,** pp. 2039-2049, 2005. doi: 10.1364/JOSAA.22.002039