Gaze-centric image analysis for efficient visual search

Umesh Rajashekar, Thomas Arnow, Alan Bovik, and Lawrence Cormack

We have determined some of the feature characteristics the human visual system employs in target searching and visual surveillance, and used them to automatically locate visually interesting regions.

Despite recent advances in computer vision, pattern recognition, and image processing, our understanding of many visual tasks remains incomplete. 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 possible in 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 critical to its success as an efficient searcher is its active way of looking. It uses a combination of steady eye fixations linked by rapid ballistic eye movements called saccades (Figure 1). This combined eye activity, executed over 15,000 times an hour, allows the visual system to assimilate information from a foveated scene, where the resolution is high only in a tiny central region, and falls off rapidly toward the periphery (Figure 2). This resolution gradient avoids the potential data glut associated with the visual system’s large field of view.

Understanding how the human visual system selects and sequences fixations should be useful for designing active artificial vision systems, and holds great potential for 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 two 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 an observer’s points of gaze, called fixation points. In our recent work, we have been using precise eye trackers to record the eye movements of subjects performing visual tasks, and applied the methods of visual psychophysics and image processing to extract image statistics at fixation points. Using predictors based on the statistical differences between features selected by human fixations and those randomly placed on the same image by computer, we have been able to automatically compute likely fixations in natural images. Our experiments have followed two scenarios: target search and visual surveillance.

In our target search experiments, we use a noise-based reverse correlation technique to determine if observers use structural cues to direct their fixations when searching for simple targets embedded in naturalistic 1/f noise at low signal-to-noise ratios. We have been able to demonstrate that observers do, indeed, choose features suggestive of the target in these tasks. Even in noisy displays, observers do not search randomly, but deploy fixations to regions that resemble aspects of the target.

Continued on next page
In our visual surveillance experiments, we recorded 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 classification images (for target search) or a linear combination of luminance and contrast image features (for visual surveillance) as cues. Thus far, the distributions of computed fixations have been found to correlate quite well with those of human observers. We quantify the similarity between fixation patterns using the information-theoretic Kullback-Leibler distance.

One interesting application we are studying is the visual search for corners. We accomplish the search by using the principles of foveated visual search combined with an automated fixation selection. Our effort is an attempt to demonstrate a case study of feature detection by means of foveated searching. The result has been a new algorithm for finding corners that will be used to drive object recognition, and to direct corner-based 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. Our findings show that low-level image attributes exert a strong influence on fixation locations, which suggests a significant bottom-up component in gaze selection. We are currently investigating what influences motion, color, and stereo primitives in attracting 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.

Author Information

Umesh Rajashekar, Thomas Arnow, Alan Bovik, and Lawrence Cormack
Department of Electrical and Computer Engineering
Center for Perceptual Systems
The University of Texas at Austin
Austin, TX
http://www.ece.utexas.edu/~umesh
http://live.ece.utexas.edu/people/bovik
http://live.ece.utexas.edu
http://www.psy.utexas.edu/psy/FACULTY/Cormack/Cormack.html
http://www.cps.utexas.edu

Umesh Rajashekar is currently a postdoctoral fellow in the Laboratory for Image & Video Engineering (LIVE) and the Center for Perceptual Systems at the University of Texas at Austin (UT-Austin). His research interests include image statistics at the point of gaze and didactic tools for education. He is the recipient of the TxTEC Graduate fellowship and the Lloyd A. Jeffress fellowship from UT-Austin. He was the Assistant Director of LIVE from Fall 2001 to Fall 2005.

Thomas L. Arnow has an M.S. in Computer Science and Systems Design from UT-San Antonio (1977), and an M.S. degree in Electrical Engineering from UT-Austin (1991). He is currently a Ph.D. candidate in ECE at UT-Austin, and a staff member at The UT Health Science Center at San Antonio.

Al Bovik is professor and the Curry/Cullen Trust Endowed Chair at UT-Austin, and director of LIVE. His research interests include digital video processing and computational visual perception. He has published 450 articles in these areas and has two U.S. patents. He is the author of the Handbook of Image and Video Processing (Academic Press) and Modern Image Quality Assessment (Morgan & Claypool).

Dr. Cormack is associate professor of psychology and neuroscience at UT-Austin, and adjunct associate professor of vision science at the University of Houston. His research interests include eye movements and visual search, computational modeling of early visual processes, and depth perception and disparity scaling.

Continued on next page
References

1. http://live.ece.utexas.edu/research/eyefix