Top-Down Vision in Humans and Robots

Lawrence W. Stark

University of California at Berkeley
Berkeley, California, USA

The scanpath theory suggests that a top-down internal cognitve model of what we "see" controls not only our vision, but also drive the sequences of rapid eye movements and fixations, or glances, that so efficiently travel over a scene or picture of interest. The contrary belief is that features of the external world control eye fixations and vision in a bottom-up mode by impinging on the retina and sending signals to the brain.

Philosophers have speculated that we "see in our mind's eye", but until the scanpath theory, little evidence supported this conjecture. Eye movements are an essential part of vision because of the dual nature of the visual system -- (i) the fovea, a narrow field, about 1/2 to 2 degrees, of high resolution vision; and (ii) the periphery, a very wide field, about 180 degrees, of low resolution vision, sensitive to motion and flicker. Eye movements must carry the fovea to each part of a scene or picture or page of reading matter to be processed with high resolution. An illusion of clarity exists, that we 'see' the entire visual field with high resolution, but this cannot be true.

Prof Stark 特別講演会, 1991年12月3日.

The cognitive model of what we expect to see is what we actually 'see'. This internal model drives our eye movements in a "scanpath", a repetitive, sequential set of saccades and fixations over subfeatures of the picture or scene, so as to check out and confirm the model. These scanpath sequences are idiosyncratic to the subject and to the picture. Experiments have shown that when we look at ambiguous pictures patterns of eye movement change with the mental image we have of the ambiguous figure. When we engage in visual imagery, looking at a blank screen and visualizing a previously seen figure, our scanpath eye movements are similar whether viewing the figure or the blank screen. This provides strong evidence that the internal cognitive model and not the external world (since this is absent in visual imagery) drives the scanpath. Recent evidence uses string editing distances to quantitate the similarity and dissimilarity between scanpaths. Also, studies of visual search indicates that a primitive form of pre-cognitive spatial model controls a 'searchpath' sequence of eye movements.

Buttressed by these new views of top-down human vision, we have applied the scanpath theory to robotic vision. Here we use our knowledge of the spatial layout of the robotic working environment, including position and orientation of the video cameras, and the nature of the robots and the work-pieces to develop a "cognitive model". This computer model then controls the image processing. Regions of interest, ROIs, are generated so that image processing, such as local thresholding and centroid calculations, can be carried out efficiently and robustly. Only those subfeatures essential for identification and control are processed, reducing the computational task greatly. The model not only controls image processing, as in human vision in the scanpath mode, but can also control the robots, the cameras, and displays for the supervisory human teleoperators. The model also serves to reduce communication bandwidth requirements since only commanded and correction model parameters are transmitted. Thus a top-down visual scheme satisfies a visual feedback control system for robots.