Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance

PAUL B. HIBBARD * and SAMIRA BOUZIT

School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JP, Scotland, UK

Received 8 August 2004; accepted 24 September 2004

Abstract—Binocular correspondence must be determined if disparity is to be used to provide information about three-dimensional shape. The current study investigated whether knowledge of the statistical distribution of disparities in the natural environment is employed in this process. A simple model, which produces distributions of distances similar to those found in the natural environment, was used to predict the distribution of disparities in natural images. This model predicts that crossed disparities will be more likely as (i) stimulus elevation decreases below fixation and (ii) fixation distance increases. To determine whether these factors influence binocular correspondence for human observers, ambiguous stereograms were presented to observers, as stimulus elevation and fixation distance were manipulated. Clear biases were observed in the depth perceived in these stereograms, which were more likely to be seen as closer than fixation (i) for stimuli presented below fixation and (ii) as fixation distance increased. These results suggest that binocular correspondence is determined in a manner consistent with the distributions of disparities expected in natural scenes.

Keywords: Binocular correspondence; natural image statistics; distance; elevation.

INTRODUCTION

Differences between the retinal images formed in the left and right eyes provide important depth information. Binocular disparity, the difference in the locations of corresponding points in the left and right images, provides a particularly important cue to depth (Julesz, 1960, 1971). In order to make use of this information, it is first necessary to determine how each point in the left eye’s image should be matched with a point in the right eye’s image. This is known as the binocular correspondence problem, and finding a solution to this problem is far from trivial (Scharstein and Szeliski, 2002). For every point in the left image, there are many potential matches in the right image. Despite this difficulty, human observers can readily solve the correspondence problem, and accurately perceive depth relations,
even in stimuli such as the random dot stereograms introduced by Julesz that consist of many identical elements, and therefore huge numbers of potential matches.

Physiological evidence demonstrates that the first stage of binocular processing in the visual cortex does not provide a solution to this problem. The most widely accepted mechanism by which disparity selectivity arises in area V1 is the binocular energy model (Fleet et al., 1996; Ohzawa et al., 1990; Qian, 1994). In this model, responses of complex Gabor filters to the left and right retinal images are combined to form a binocular energy response. Due to the quasi-periodic nature of such responses, an individual binocular energy mechanism cannot be said to have a unique disparity preference. Rather, such a mechanism would produce comparable responses to many stimuli with disparities separated by whole-number multiples of the periodicity of the underlying filter functions (Fleet et al., 1996). Consistent with this limitation, the responses of neurons in V1 do not correlate with the perception of depth from disparity. These responses are disambiguated at a later stage of processing (Cumming and Parker, 2000a, b; Thomas et al., 2002).

One possible mechanism by which this disambiguation may be achieved is to pool information across spatial scales (Fleet et al., 1996; Marr and Poggio, 1979; Smallman, 1995; Tsai and Victor, 2003). As the locations of the false peaks in the responses of energy neurons occur at a rate determined by the spatial scale of the underlying filters, only real peaks (corresponding to the actual disparity) would be expected to coincide across scales. Combining information across scales should therefore localise true peaks (Fleet et al., 1996; Tsai and Victor, 2003).

Alternatively, true matches may be distinguished from mismatches by the application of matching constraints. A number of constraints have been proposed in disparity matching algorithms. These include the uniqueness constraint (Marr and Poggio, 1979); the disparity gradient limit (Burt and Julesz, 1980); coherence and smoothness constraints (Marr and Poggio, 1976, 1979) and the nearest-disparity constraint (Mitchison and McKee, 1985, 1987; McKee and Mitchison, 1988; Prince and Eagle, 2000; Read, 2002a, b). Each of these can be seen as embodying some knowledge of, or assumptions about, the structure of the environment. The uniqueness constraint for example embodies the fact that each point in the world will project to at most one point in each image. Assumptions such as the disparity gradient limit, or smoothness or nearest disparity preferences, embody assumptions about the likely distribution of disparity within a local spatial region of the image.

Such assumptions will tend to bias the choice of correspondence when faced with ambiguous information. If these assumptions are to be useful, then the biases they introduce ought to improve the likelihood of obtaining correct binocular matches. To achieve this, they must represent a true reflection of the distributions of disparities that will be encountered in our natural environment. The current study made use of simple stimuli in which binocular matching is ambiguous in order to determine whether observers demonstrate consistent biases in how correspondence is achieved, and whether such biases are related to the expected distribution of disparity in the natural environment.