CORRESPONDENCE IN APPARENT MOTION: DEFINING THE HEURISTICS

MARC GREEN

PSYCHOLOGY DEPARTMENT
YORK UNIVERSITY
NORTH YORK, ONTARIO M3J 1P3

ABSTRACT

Correspondence matching in apparent motion is based on two heuristics: match images if they 1) have a similar form and 2) are in close proximity. Psychophysical experiments are used to define these heuristics. Observers judged motion path between images in a competition paradigm. Results showed that the tokens used in form matching are spatial frequency and orientation. Further, proximity is defined in a 3-D spatial reconstruction rather than 2-D retinal coordinates. A possible representation for the computation of correspondence is a multidimensional detector space, with dimensions including spatial frequency, orientation, X, Y and Z (or disparity) coordinates.

INTRODUCTION

A remarkable property of biological vision systems is the ability to deduce that two images, seen at different places and/or times, represent the same physical object. The advantages of this property are nicely exemplified in the phenomenon of apparent motion: when viewing a series of static pictures, or "frames", each object in one frame moves to the location of the corresponding object in the subsequent frame. Coherent motion is perceived only if the visual system, after considering successive frames, can properly match images corresponding to the same object. This "correspondence problem" is presumably solved by application of heuristics to provide a "preference metric" (13) which evaluates the affinity between potential matches. Preference metrics can be derived by two general classes of heuristic: 1) match images of similar form and 2) match images with the greatest spatial proximity. At first glance, this recipe for matching seems simple enough, but real difficulties arise when implementation is attempted. These heuristics need to be more precisely defined by answering the following questions. First, what form primitives are used as tokens in correspondence matching? Second, is proximity defined in two-dimensional retinal coordinates or in an internal, 3-D reconstruction of space. This question has important implications since use of a 3-D metric requires that a depth must be assigned each form token before matching can proceed. The studies described below are addressed to answer each of these questions.

THE FORM HEURISTIC

The notion that correspondence matching is based partly on form similarity has been around for a long time. However, it has proved suprisingly difficult to identify correspondence tokens since apparent motion tends to be independent of form similarity. Early studies (8, 14) found that when there was only one image in each frame, the apparent motion seen with two identical images was readily perceived with two different images. The first would deform gradually into the second with no loss of motion continuity. More recently experimenters (3,11) have used competition methods and have likewise concluded that form similarity plays no role in token matching.

Why has it proved so difficult to identify correspondence tokens? Two possible explanations come to mind. First, stimuli were either geometric forms, circles, squares, letters, etc. or alphabetic characters, which differ in high spatial frequency content, but are similar in low spatial frequencies. Both human psychophysical (1) and computer (9) experiments have resulted in the view that one representational stage in early visual processing is the activity in arrays of detectors which are sensitive to edges at different resolution. At each resolution level, the detectors are activated only by a narrow band of spatial frequencies. Geometric shapes and alphabetic characters would all stimulate similar populations of coarse, low resolutions detectors. If activity in detectors at different resolution were tokens, then there would be strong affinity between all such images. Second, previous investigators have used a flash technique which produced a luminance transient (i.e., a D. C. offset) accompanying the presentation of form. The luminance flux per se might be used as a token for matching. This seems plausible because it has been suggested (4) that the visual system contains two parallel sets of detectors for analyzing spatio-temporal luminance change. The detectors are modeled as difference of Gaussians (DOG's) with different temporal properties. If the inhibitory Gaussian is developed simultaneously with the excitatory, then the detector is "sustained" and is highly tuned to aspects of form such as spatial