ESTIMATING MOVEMENT DIRECTION WITH A NEURAL NETWORK

William C. Treurniet

Communications Research Centre
Communications Canada

ABSTRACT

Various schemes have been developed to identify the direction of motion in a scene, often to assist in compressing the information required to broadcast a sequence of images. These schemes typically make sequential comparisons of blocks of pixels in order to arrive at a most likely direction of motion. This paper investigates how an artificial neural network may be used to perform the same task. The approach is interesting because these networks perform computations in parallel, thus allowing a form of top-down as well as bottom-up processing. Also, because computations are conceptually performed in parallel, it would be possible to consider performing the task in real time with an appropriate hardware implementation. Preliminary results show that a properly trained network has interesting properties similar to those of real neurons and can, indeed, report direction of movement based on binary pixel values.

KEYWORDS: movement estimation, neural network

1. Introduction

Improvement of television image quality can readily be achieved, but usually at the cost of a considerable increase in the bandwidth required to broadcast the images. Thus, current work on defining high-definition television systems includes development of methods for image data compression. Movement-adaptive encoding of television images is one method under consideration. This scheme encodes an area of the television frame more compactly if it is determined to be very similar to an adjacent area in the previous frame. In that case, the image content is presumed to have moved from one location to another. Transmitting only the direction and extent of movement of display areas can result in a considerable saving in the amount of information that needs to be transmitted. A number of techniques to measure the similarity of two blocks of pixels have been developed (e.g., Huang, 1981; Sabri, 1984; Storey, 1986; Tsuda and Hiraoka, 1986; Puri, Hang, and Schilling, 1987; Lee and Griffiths, 1987). Motivation for this work was the need for an efficient algorithm which could perform the necessary comparisons within the time required.

The central issue for movement-adaptive encoding of television signals is recognition of the direction and extent of movement in an image. Tracking movement in a visual field conceptualized as an array of points is not an easy task for image processing systems. In computational vision research, following individual points through space in successive time intervals has been termed the correspondence problem (e.g., Marr, 1982). That is, a given point at an instant in time must be identified with one of many possible points at a later instant. When only local information is available, even the motion of edges is ambiguous. Under certain conditions, however, the velocities of non-parallel edge segments belonging to the same pattern offer constraints which unambiguously identify the exact direction of motion (Glazer, 1981; Adelson and Movshon, 1982). The measurement of the velocity gradient may be further improved by computing a vector that best fits each set of neighbouring vectors. This procedure helps to compensate for discontinuities, intensity fluctuations, and noise in the sampled image (Glazer, 1981).

In human vision, Ramachandran and Anstis (1983) showed that perceived motion in different parts of a display are not always independent. Bistable patterns presented simultaneously in random locations all showed the same direction of motion. Apparently, the perceived direction of movement in the visual field may be affected by global as well as local relationships in an image. Therefore, methods that process an image in both a top-down and a bottom-up manner should be useful for ascertaining the direction and extent of movement of image content.

The need for both local and global