VISUAL ROUTINES: INGREDIENTS, DESIGN, AND CONTROL

Marc H.J. Romanycia

Department of Computer Science
The University of British Columbia
Vancouver, B.C., Canada, V6T 1W5

ABSTRACT

The concept of Visual Routine is introduced. A description is given of an implemented computer system which can correctly compute in images of simple 2-D geometric shapes eleven common properties and relations. A visual routine programming language is outlined. Issues relevant to the control of visual-routine-based search are discussed. The results of testing the system are reported.

KEYWORDS: visual routines, visual attention, non-local image properties and relations, computational vision.

1 Introduction

Visual routines are proposed by Ullman (1984) as a means of accounting for how abstract shape properties and spatial relations can be derived from early visual representations. Abstract shape properties like ***convex, closed,*** and ***symmetric***, and abstract spatial relations like ***inside, outside***, and ***parallel*** cannot be efficiently computed with simple local computations. To compute them efficiently we must attend to specific points or regions, and apply operations selectively at these places. Visual routines are sequences of such operations appropriately applied.

Visual routines are ideally suited to deducing the computational restrictions surrounding the task of visual perception. They form a natural language for dissecting visual tasks into their component tasks, and at the same time, they translate naturally into standard computer terminology. This latter feature makes them easy for us to build and work with on conventional computers. It allows us to draw on our large body of programming experience to suggest possible algorithms and approaches, and it allows us to make direct use of existing theoretical results when we analyze the complexity of the visual routine algorithms.

In this paper we describe an implemented computer system which makes concrete Ullman's visual routine proposal. This system computes eleven shape properties and relations: ***centred-in, closed, connected, convex, horizontal, inside, outside, parallel, part-of, touching,*** and ***vertical.*** Examples of these routines are given, and the visual routine language in which the routines are written is outlined.

The domain in which the system searches for properties and relations is the 2-D world of simple geometric straight-edged shapes, such as squares, triangles, and line segments.

The visual routines and their control logic are embedded in a Question-and-Answer system which can correctly reply to such queries of an image as "Find all the triangles inside squares" and "Find any three instances of a vertical bar outside a simple closed curve." The Question-and-Answer system serves as a visual routines development testbed. With it we are able to conveniently test the correctness and efficiency of the visual routines and the control logic.

The work described in this paper presents a fourfold contribution to intermediate level Computational Vision:

- It provides a concrete instantiation of Ullman's visual routine proposal.
- It argues for the importance of certain basic operations out of which visual routines can be composed.
- It provides a set of working visual routines for eleven common properties and relations.
- It identifies eight key issues for the control of visual routines.

This work follows the tradition of Marr (1982) in attempting to give a computational account of vision. Thus, our system is not built as an engineering project to suit some practical needs, and, although an effort is made to maintain a correspondence with what is currently known about the human visual system, the system is not presented as a model of the human system.

2 The Visual Routine Framework

Ullman (1984) presented visual routines as a framework for undertaking the study of intermediate level vision. This framework has several components, as illustrated in Figure 1.

Two components of the visual routine framework are the base maps and the parallel methods which create them. The base maps (also called "base representations") are uniform retinotopic maps describing the most basic properties of points or small patches in the image: orientation, colour, intensity, stereo disparity, and motion. These representations are typically derived directly from the image without taking into account any high level information about what may be in the scene. Hence they are created by "bottom-up" processes. This notion of base map feature is closely related to the idea of ***preattentive*** or "pop-out" feature that perceptual psychologists have developed (Julesz, 1983, 1984; Treisman, 1985, 1986).