Recognition Methodology: Algorithms & Architecture

Robert M. Haralick

Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195

1. Overview

Computer recognition and inspection of objects is, in general , a complex procedure requiring a variety of kinds of steps which successively transform the iconic data to recognition information. We hypothesize that the difficulty of today's computer vision and recognition technology to be able to handle unconstrained environments is due to the fact that the existing algorithms are specialized and do not develop one or more of the necessary steps to a high enough degree. Our thesis is that there are no shortcuts. A recognition methodology must pay substantial attention to each of the following five steps: conditioning, labeling, grouping, extracting, and matching.

These five steps of conditioning, labeling, grouping, extracting, and matching constitute a canonical decomposition of the recognition problem, each step preparing and transforming the data in just the right way for the next step.

1.1 Conditioning

Conditioning is based upon a model which suggests that the observed image is composed of an informative pattern modified by uninteresting variations which typically add to or multiply the informative pattern. Conditioning estimates the informative pattern on the basis of the observed image. Thus conditioning surpresses noise which can be thought of as random unpatterned variations affecting all measurements. Conditioning can also perform background normalization by surpressing uninteresting systematic or patterned variations. Conditioning is typically applied uniformly and is context independent.

1.2 Labeling

Labeling is based upon a model which suggests that the informative pattern has structure as a spatial arrangement of events, each spatial event being a set of connected pixels. Labeling determines in what kinds of spatial events each pixel participates. For example, if the interesting spatial events of the informative pattern are only events of high valued pixels and events of low valued pixels, then the thresholding operation can be considered as a labeling operation. Other kinds of labeling operations include edge detection, corner finding, and identifying pixels which participate in varieties of shape primitives.

1.3 Grouping

The labeling operation labels pixels with the kinds of primitive spatial events the pixel participates in. The grouping operation identifies the events by collecting together or identifying maximal connected sets of pixels participating in the same kind of event. If the labels are symbolic then the grouping is really a connected components operation. If the labels are the gray levels, then the grouping operation is in fact a segmentation. If the labels are step edges, then the grouping operation constitutes edge linking, etc.

The grouping operation is the operation in which there is a change of logical data structure. The observed image, the conditional image, the labeled image are all digital image data structures. Depending on the implementation, the grouping operation can produce an image data structure in which each pixel is given an index which is associated with the spatial event to which it belongs or the grouping operation can produce a data structure which is a collection of sets. Each set corresponds to a spatial event and contains the pairs of (row, column) positions which participate in the event. In either case, there is a change in the logical data structure. The entities of interest before the grouping step are pixels. The entities of interest after the grouping step are sets of pixels.

1.4 Extracting

The grouping operation determines the new set of entities. But after the grouping step the new entities are naked. The only thing they posses is their identity. The extracting operation computes for each group of pixels a list of its properties. Example properties might include its centroid, its area, its orientation, its spatial moments, its gray tone moments, its spatial-gray tone moments, its circumscribing circle, its inscribing circle, etc. Other properties might depend on whether the group is considered as a region or as an acr. If the group is a region, then number of holes might be a useful property. If the group is an arc, then the average curvature might be a useful property.

Extracting also can measure topological or spatial relationships between two or more groupings. For example, an extracting operation may make explicit that two groupings touch or are spatially close or that one grouping is above another.

1.5 Matching

After the completion of the extracting operation, the events occurring on the image have been identified and measured. But the events in and of themselves have no meaning. The meaning of the observed spatial events occurs when a perceptual organization has occurred in which it is recognized that a specific set of spatial events in the observed spatial organization constitutes an imaged instance of some previously known object such as a "chair" or the letter "A."

It is the matching operation which determines the interpretation of some related set of image events associating these events with some given 3D object or 2D shape. The association determined by matching establishes a correspondence between each spatial event on the image in the related set of events with some spatial event on the 3D object or 2D shape. The association is one which in some sense best matches both the character of the spatial events and their spatial relationships. Thus, after matching, two primitive image events which stand in some spatial relationship will have associated with them two object events which stand in a similar relationship.

There is a wide variety of image operations which are matching operations. The classic one is template matching which is effective only if the variety of instances expected to be encountered is limited. For example, rotation and size variations must be very small. The background must be near uniform. Random shape deformations must be minimal..

Simple shapes will correspond to a primitive spatial event and the property measurement from the primitive spatial event will often be adequate to permit recognition of the shape. In this case, the matching operation amounts to matching the vector of properties measured from the image spatial event with the vector of properties of a prototype representative. Such matching is what constitutes statistical pattern recognition.

Complex shapes will correspond to a set of primitive spatial events. Here, recognition must proceed by using the property vector of each observed spatial event as well as the spatial relationships between the events. In this case, the matching amounts to determining a relational homomorphism with unary constraints established by the required matching of the property vectors of the observed image events with the property vectors of the prototype primitives. Such a matching is what constitutes structural pattern recognition.

2. Algorithms and Computer Hardware Architecture

There are many varieties of algorithms employed in each of the five phases of machine vision recognition. We make no attempt in this section to make a scholarly documentation of them. However, we do make some brief remarks relative to some of the more obvious relationships we have noticed.

The most useful and perhaps most used image operators are those of convolution and morphology. The convolution of an image f with a kernel k is defined by

$$(f * k)(i,j) = \sum_{m,n} f(i-m,\ j-n)\ k\ (m,n)$$

There are two fundamental morphologic operations. They are dilation and erosion. The dilation of an image f with a structuring element k is defined by

$$(f \oplus k)(i,j) = \max_{m,n}\{f(i-m, j-n) + k(m,n)\}$$

The erosion of an image f with a structuring element k is defined by

$$(f \ominus k)(i,j) = \min_{m,n}\{f(i+m, j+n) - k(m,n)\}$$

The convolution and morphology operators are structurally similar. Convolution is a shift, multiply, and sum. Dilation is a shift, add, and max. Erosion is a shift, subtract, and min. The pipeline implementations are obviously similar.

The three opeartors each have an associative or associative-like property that permits operations with large domains to be done as a concatenation of operations with smaller domains. That is

$$\begin{aligned} f * (k_1 * k_2) &= (f * k_1) * k_2 \\ f \oplus (k_1 \oplus k_2) &= (F \oplus k_1) \oplus k_2 \\ f \ominus (k_1 \oplus K_2) &= (f \ominus k_1) \ominus k_2 \end{aligned}$$

The pipeline stages of the basic operation can be joined together in a slightly different way to permit the pipeline implementation of the associative relation.

All this suggests that hardware capable of long pipeline is useful. However, the algorithms which use convolution and morphlogy also make use of other operations. There are, for example, all the point operators. They consist of the arithmetic operations of add, subtract, multiply, and divide; the boolean operations of AND, OR, XOR, NAND, NOR, and negate; the table look up operators and thresholding. Then there are algorithms which can make use of two parallel pipelines which are joined together by a point opeartor and whose result is then pipelined with another operation. Our image of the simple pipeline which we associate with simple convolution and morphology must now get replaced by a network of pipelines, each pipeline being considered as a branch in the network. Pipelines join and split at nodes which perform some kind of point operator.

Of course, the pipeline is not the only parallel implementation structure. There is the SIMD parallel paradigm. The transformation of the pipeline into a network of pipelines generalizes to the SIMD implementation. Instead of a unified SIMD we can visualize a partitionable SIMD machine, with the processes in each partition block all executing the same instruction and with a capability of memory sharing between partition blocks as a means of communication.

In both the pipeline and the SIMD implementation some degree of reconfiguration will need to be present.