M

Mining Images for Structure

Terry Caelli
Australian National University, Australia

INTRODUCTION

Most data warehousing and mining involves storing and retrieving data either in numerical or symbolic form, varying from tables of numbers to text. However, when it comes to everyday images, sounds, and music, the problem turns out to be far more complex. The major problem with image data mining is not so much image storage, per se, but rather how to automatically index, extract, and retrieve image content (content-based retrieval [CBR]). Most current image data-mining technologies encode image content by means of image feature statistics such as color histograms, edge, texture, or shape densities. Two well- known examples of CBR are IBM's QBIC system used in the State Heritage Museum and PICASSO (Corridoni, Del Bimbo & Pala, 1999) used for the retrieval of paintings. More recently, there have been some developments in indexing and retrieving images based on the semantics, particularly in the context of multimedia, where, typically, there is a need to index voice and video (semantic-based retrieval [SBR]). Recent examples include the study by Lay and Guan (2004) on artistry-based retrieval of artworks and that of Benitez and Chang (2002) on combining semantic and perceptual information in multimedia retrieval for sporting events.

However, this type of concept or semantics-based image indexing and retrieval requires new methods for encoding and matching images, based on how content is structured, and here we briefly review two approaches to this.

BACKGROUND

Generally speaking, image structure is defined in terms of image features and their relations. For SBR, such features and relations reference scene information. These features typically are multi-scaled, varying from pixel attributes derived from localized image windows to edges, regions, and even larger image area properties.

MAIN THRUST

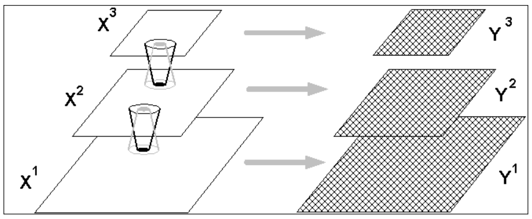
In recent years there has been an increasing interest in SBR. However, this requires the development of methods for binding image content with semantics. In turn, this reduces to the need for models and algorithms that are capable of efficiently encoding and matching relational properties of images and associating these relational properties with semantic descriptions of what is being sensed. To illustrate this approach, we briefly discuss two representative examples of such methods: (1) Bayesian Networks (Bayesian Nets) for SBR, based first on multi-scaled image and then on image feature models; (2) principal components analysis (also termed latent semantic indexing or spectral methods).

Bayesian Network Approaches

Bayesian Nets have recently proved to be a powerful method for SBR, since semantics are defined in terms of the dependencies between image features (nodes), their labels, and known states of what is being sensed. Inference is performed by propagation probabilities through the network. For example, Benitez et al. (2003) have developed MediaNet, a knowledge representation network and inference model for the retrieval of conceptually defined scene properties integrated with natural language processing. In a similar way, Hidden Markov Random Fields (HMRFs) have become a common class of image models for binding images with symbolic descriptions. In particular, Hierarchical Hidden Markov Random Fields (HHMRF) provide a powerful SBR representation. HHMRFs are defined over multi-scaled image pixel or features defined by Gaussian or Laplacian pyramids (Bouman & Shapiro, 1994). Each feature, or pixel, x, at a given scale is measured (observed) to evidence scene properties, states, s, corresponding to semantic entities such as ground, buildings, and so forth, as schematically illustrated in Figure 1. The relationships between states serves to define the grammar. The link between observations and states defines, in this approach, the image semantics. Accordingly, at each scale, l, we have a set of observations and states, where $p(o_l(x)/s_l(x))$ defines the dependency of the observation at scale, l, on the state of the world (scene).

Specifically, the HHMRF assumes that the state at a pixel, x, is dependent on the states of its neighboring pixels at the same or neighboring levels of the pyramid. A simple example of the expressiveness of this model is a forestry scene. This could be an image region (label:

Figure 1. The hierarchical hidden Markov random field (HHMRF) model for image understanding. Here, the hidden state variables, X, at each scale are evidenced by observations, Y, at the same scale and the state dependencies within and between levels of the hierarchy. The HHMRF is defined over pixels and/or feature graphs.



forest) dependent on a set of regions labeled *tree* at the next scale, which, in turn, are dependent on *trunk*, *branches*, and *leaves* labels at the next scale, and so forth. These labels have specific positional relations over scales, and the observations for each label are supported by their compatibilities and observations.

Consequently, the SBR query for finding forestry images is translated into finding the posterior maximum likelihood (MAP) of labeling the image regions, given the forestry model. Using Bayes' rule, this reduces to the following optimization problem:

$$s_l^*(x)) \propto \arg\max_S \{p(s_l(x)/o_l(x))\prod_{u,v} p(s_{l\pm v}(x \pm u))\}$$

where $l \pm v$ corresponds to the states above and below level l of the hierarchy. In other words, the derived labeling MAP probability is equivalent to a probabilistic answer to the query if this image is a forestry scene. There are many approaches to approximate solutions to this problem, including relaxation labeling, expectation maximization (EM), loopy belief propagation and the junction tree algorithm (see definitions in Terms and Definitions). All of these methods are concerned with optimal propagation of evidence over different layers of the graphical representation of the image, given the model and the observations, and all have their limitations. When the HHMRF model is approximated by a triangulated image state model, the junction tree algorithm is optimal. However, triangulating such hierarchical meshes is computationally expensive. On the other hand, the other approaches mentioned previously are not optimal, converging to local minima (Caetano & Caelli, 2004).

When the structural information in the query and image is defined in terms of features (i.e., regions or edge segments) and their relational attributes, again, HMRFs can be applied to image feature matching. In this case, the HMRF is defined over graphs that depict features and their relations. That is, consider two attributed graphs, G_s and G_x, representing the image and the query, respectively. We want to determine just how, if at all, the query (graph) structure is embedded somewhere in the image (graph). We define HMRF over the query graph, G_x. A single node in G_x is defined by x_i, and in the graph G_s, by s_α. Each node in each graph has vertex and edge attributes, and the query corresponds to solving a subgraph isomorphism problem that involves the assignment of each x_i a unique s_α, assuming that there is only one instance of the query structure embedded in the image, although this can be generalized. In this formulation, the HMRF model considers each node x_i in G_x as a random variable that can assume any of *S* possible values corresponding to the nodes of G_s.

- **The Observation Component:** Using HMRF formalities, the similarity (distance: *dist*) between vertex attributes of both graphs is consequently defined as the observation matrix model

$$B_{i\alpha} = p(y_x^i / x_i = s_\alpha) = dist(y_x^i, y_s^\alpha).$$

- **The Markov Component:** Here, we use the binary (relational) attributes to construct the compatibility functions between states of neighboring nodes. Assume that x_i and x_j are neighbors in the HMRF (being connected in G_x). Similar to the previously described unary attributes, we have

$$A_{ji;\beta\alpha} = p(x_j = s_\beta / x_i = s_\alpha) = dist(y_x^{ij}, y_s^{\alpha\beta}).$$

- **Optimization Problem and Solutions:** Given this general HMRF formulation for graph matching, the optimal solution reduces to that of deriving a state vector $s^* = (s^1, .., s^T)$ where $s^i \in G_s$ for each vertex $x^i \in G_x$ such that the MAP criterion is satisfied, given the model $\lambda = (A, B)$ and data

$$s^* = \arg\max_{s_\alpha .. s_\varsigma} \{p(x_1 = s_\alpha, ..., x_T = s_\varsigma / \lambda)\}.$$

Specifically, we introduce two sets of HMRF models for solving this problem.

Related Content