MULTIMEDIA DATA MINING FOR BUILDING RULE-BASED IMAGE RETRIEVAL SYSTEMS

Dianhui Wang   *Xiaohang Ma*

Department of Computer Science and Computer Engineering
La Trobe University, Melbourne, VIC 3086, Australia

ABSTRACT

This paper presents a framework for building rule-based image retrieval (RBIR) systems. Soft computing based multimedia data mining techniques are employed to extract and optimize fuzzy rules that infer semantic concepts for image retrieval. Using fuzzy inference system (FIS), a novel RBIR system with semantic query is developed. Experimental results on an artificial image databse are reported.

1. INTRODUCTION

While retrieving images, one looks at the similarity between a query and the images from a database in terms of some sense such as semantic relevance or statistically characteristic closeness. However, computers retrieve images based on a distance metric defined over a high dimensional feature space [1]. Due to the limitations of representing high-level semantic concepts using low-level visual features, such as color histogram, texture or shape information, it is quite challenging and difficult to exploit advanced image retrieval systems, where semantic concepts within images may be deduce by a single or a set of semantic query input(s).

Multimedia data mining is a process of extracting useful patterns or rules from large volumes of multimedia data [2]. For complicated data sources such as images or videos, interesting patterns or associative relations are usually vague and favorably described by linguistic variables. Therefore, mining fuzzy recognition rules to infer image contents from high dimensional feature vectors is significant to design the RBIR systems. Although content-based image retrieval (CBIR) techniques have been paid considerable attentions in the past, up to date, only few works about the RBIR systems can be found in references.

This paper aims at developing a framework for building the RBIR systems using soft computing techniques. The reminder of the paper is organized as follows: Section 2 describes our technical contributions. Section 3 evaluates the proposed RBIR system and reports simulation results. A conclusion is given in the last section.

2. MINING RULES FOR RBIR SYSTEMS

2.1. System Architecture

Our proposed RBIR system consists of the following main modules:

1. Feature Extraction Module — extracts low-level visual features from images, and segments the images based on the features.
2. Rule Initialization (Extraction)Module — provides a program with friendly user-interface to domain workers, who can manually extract patterns for semantic concepts representation.
3. Rule Optimization Module — optimizes the rule base in terms of the numbers of the antecedant and rules, and the parameters involved in the membership functions.
4. Fuzzy Inference Module — encodes the images using the degrees of confidence on semantic concepts through a fuzzy inference method.

The framework of our proposed RBIR system is depicted in Figure 1. The following subsections will give some details on these modules.

2.2. Feature Extraction

Firstly, all images in the database (DB) are transformed from the RGB colour space to the YUV colour space, where the Y encodes luminance, the U and V together encode chromaticity, respectively. We partition the images into small blocks. The selection of block size b is experimental. A smaller b can preserve more information but increase computational workload, whilst a larger b can decrease the quality of the local feature representation power. A 12-dimensional feature vector, namely $F = [f_1, f_2, \ldots, f_{12}]$, is extracted for each small blocks. Of these features, six of them are the first and the second order of color moments for the three colour channels, and the remains are from the mean and the square root of the second order moment of the coefficients of three

0-7803-9332-5/05/$20.00 ©2005 IEEE

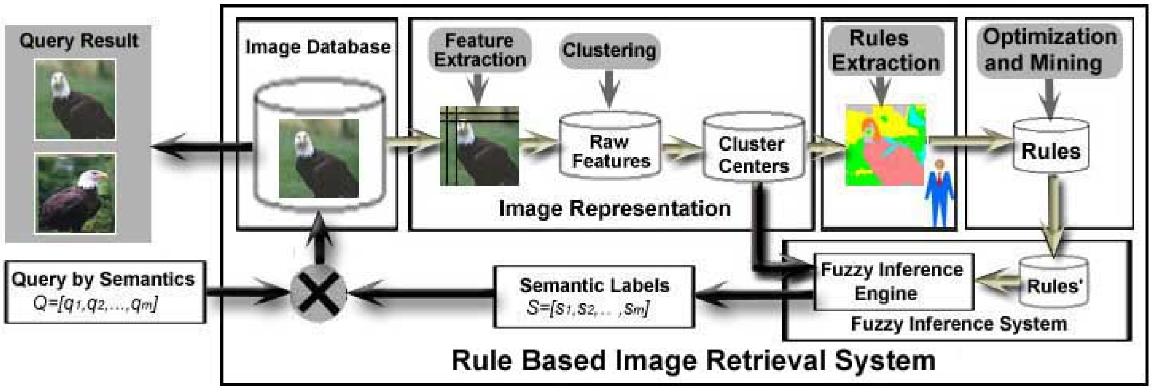

Fig. 1. Framework of the proposed rule-based image retrieval system

high-frequency bands in a wavelet transform [3]. The motivation of using the features extracted from high-frequency bands is that they reflect some texture information of the images. We only apply this wavelet transfom to the luminance component because human's eyes are more sensitive to luminance than colours. The four frequency bands are denoted by LL, LH, HL and HH, respectively. Each band contains $b/2 \times b/2$ coefficients. Without loss of generality, we denote the coefficients in the HL band by $c_{i,j}^{HL}$, where $1 \leq i \leq b/2$ and $1 \leq j \leq b/2$. The features extracted from this band can be calculated by

$$f_m^{HL} = \frac{4}{b^2} \sum_{i=1}^{\frac{b}{2}} \sum_{j=1}^{\frac{b}{2}} c_{i,j}^{HL}, \tag{1}$$

$$f_s^{HL} = \left(\frac{4}{b^2} \sum_{i=1}^{\frac{b}{2}} \sum_{j=1}^{\frac{b}{2}} c_{i,j}^{HL^2} \right)^{\frac{1}{2}}. \tag{2}$$

We take the features only for LH, HL and HH bands. The image contents can be characterized by a set of feature vectors $V = \{F_1, F_2, \ldots, F_{(M\times N)/b^2}\}$, where M and N are the image sizes and $F_i = [f_1^i, f_2^i, \ldots, f_{12}^i]$. A standard k-means algorithm is applied to cluster the feature vectors in V into several groups [4], namely, $CF = \{CF_1, CF_2, \ldots, CF_k\}$, $\bigcup_{i=1}^{k} CF_i = V$, where CF_i is represented by its cluster centre, denoted by $C_i = [c_1^i, c_2^i, \ldots, c_{12}^i]$, and k is the number of clusters (usually predefined by the designer). Feature vectors belonging to the same group charaterize similar visual appearance. Therefore, original image can be segmented by mapping each CF_i onto a region in the image (as shown in Figure 2). In addition, as an important part of the patterns associated with semantic concepts in images, a percentage value is recorded for each cluster center, which is given by

$$p_i = \frac{b^2 \times CN_i}{M \times N}, i = 1, 2, \cdots, k, \tag{3}$$

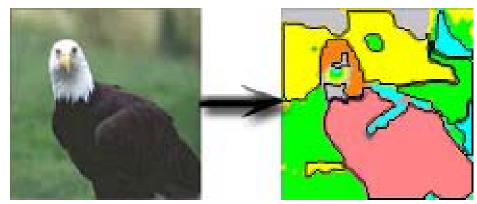

Fig. 2. Image clustering example

where CN_i is the number of feature vectors in CF_i.

In such a way, an image $I \in DB$ can be encoded by a set of center-percentage pattern pairs (PP), denoted by $P_I = \{< C_1, p_1 >, < C_2, p_2 >, \ldots, < C_k, p_k >\}$.

Remarks: (1) the number k is subjectively determined, it can vary for different images and may cause empty clusters; (2) the regions associated with the cluster centers may not necessarily form an entire object, but a link to some relevant concepts.

2.3. Rule Initialization

A program with user friendly interface, as shown in Figure 3, is developped to assist domain workers to extract rules from some typical images (also called training dataset and denoted by TDS). During the course of extracting the initial prototype patterns to form fuzzy recognition rules, the following principles are applied:

1. Select the obvious regions to represent the semantic concepts contained in the images.

2. If there are several regions with similar visual appearance, they can be merged together by taking the average of the mean values as the new cluster center, and the sum of their percentage values as the new percentage value.

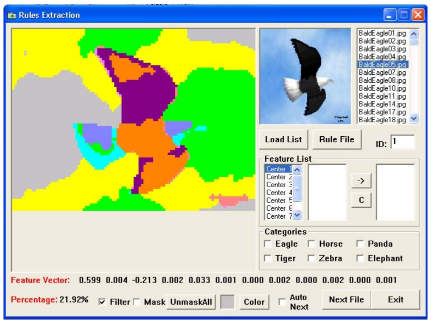

Fig. 3. Rule extraction program

By observing the PP for the TDS, we may obtain a collection of reduced PPs, denoted by $R_c = \{P_I : I \in TDS\}$. We denote a generic PP in the R_c with a corresponding semantic concept, namely, s, by $R_0 = \{< C'_1, p'_1 >, < C'_2, p'_2 >, \ldots, < C'_l, p'_l >, s\}$. This forms the base of fuzzy recognition rules, however the R_0 itself can not be directly used in a rule-based inference system for image recognition due to both conceptual uncertainties and descriptive incompleteness. Let $\tilde{C}'_i = [C'_i, p'_i], X_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,13}], i = 1, 2, \cdots, l$. Then, by assigning a membership function with each $\tilde{C}'_i$, we have a fuzzy recognition rule associated with the R_0 as follows:

$$Rule : X_1 \in \Xi(\tilde{C}'_1) \wedge \cdots \wedge X_l \in \Xi(\tilde{C}'_l) \Rightarrow s \in I, \quad (4)$$

where $X_i \in \Xi(\tilde{C}'_i)$ is read as X_i is around $\tilde{C}'_i$; the $\Xi(\cdot)$ represents a neighbourhood of $(\cdot)$, which is characterized by a membership function:

$$MF_{\tilde{C}'_i}(X) = e^{-[X-\tilde{C}'_i]W_i[X-\tilde{C}'_i]^T}, \quad (5)$$

where $W_i = diag[w_{i,1}, w_{i,2}, \cdots, w_{i,13}], w_{i,j} \geq 0$; s represents some semantic concept; the $\wedge$ is the logic AND operator; and the $\Rightarrow$ represents the implication operator.

For each $I \in TDS$, it will have a fuzzy rule attached. The whole rule base and the parameter set are denoted as Rs and Ws, respectively.

2.4. Fuzzy Inference Mechanism

Given an image I_q, and suppose $P_{I_q} = \{\tilde{C}^q_1, \tilde{C}^q_2, \cdots, \tilde{C}^q_m\}$. The following three steps of inference will give a degree of confidence on occurrence of the semantic concept s in the image I_q, denoted by $deg(s, I_q)$.

Step 1: Calculate the maximum firing degree between the antecedents of the fuzzy rule and $\tilde{C}^q_i, i = 1, 2, \cdots, m$

$$deg(\tilde{C}'_j) = max_i MF_{\tilde{C}'_j}(\tilde{C}^q_i), j = 1, 2, \cdots, l. \quad (6)$$

Step 2: Calculate the matching degree between the image I_q and the rule

$$deg(Rule, I_q) = min_j deg(\tilde{C}'_j). \quad (7)$$

There may be multiple fuzzy recognition rules for identifying a same semantic item, so we use the fuzzy algebraic sum operator $\oplus$, i.e., $a \oplus b = a + b - ab$, to determine the final confidence degree of the image to contain s. Let Se be the collection of fuzzy rules that identify the semantic item s. Then, we have

Step 3: Calculate $deg(s, I_q)$

$$deg(s, I_q) = \oplus_{Rule \in Se} deg(Rule, I_q). \quad (8)$$

After the image I_q goes through the whole fuzzy rule base, an indicator vector $E_{I_q} = [e_1, e_2, \cdots, e_h]$ can be generated, where h is the total number of the semantic concepts represented by the fuzy rule base, element e_i represents the confidence degree between the image I_q and the i-the semantic concept.

This completes the encoding process of the image I_q.

2.5. Rule Optimization

The initial fuzzy recognition rules are extracted from individual images according to the domain worker's perception. The size of the rule base is relatively large and there exist redundant and repetitious rules. Therefore, it is necessary and important to optimize the rule base so that a good tradeoff between efficiency in fuzzy inference and recognition rate may be achieved. Note that the objects to be optimized include two aspects, i.e., the antecedents of the fuzzy rules and the parameters involved in the membership functions. So, the present task contains both structure optimization and parameter optimization simultaneously. In this study, we adopt an alternate optimization scheme (AOS) to perform this task. The basic idea of our proposed AOS is to optimize the parameters first for a fixed structure, then to discrad an antecedent in the light of a fitness function step by step. The second optimization procedure takes place until the fitness function value drops down to below a predefined threshold in terms of the recognition rate for the DB. This step takes some time and multiple loops may be possible during the course. Finally, we refine the parameters involved in the updated fuzzy rule base.

In this paper, a simplest form of the evolution strategies (ES) [4], termed as (1+1)-ES, is employed in the parameter optimization. The following fitness function fit is used in the (1+1)-ES algorithm:

$$fit(Rs, Ws) = \frac{N_{right}}{N_{all}}, \quad (9)$$

where the N_{right} and the N_{all} represent the number of the images whose semantic concepts are correctly inferred by the current FIS, and the total image number in the database, respectively.

For structure optimization, each step we discard one antecedent from a fuzzy rule and calculate the fitness function value by (9). If this removal results in a sharp drop,

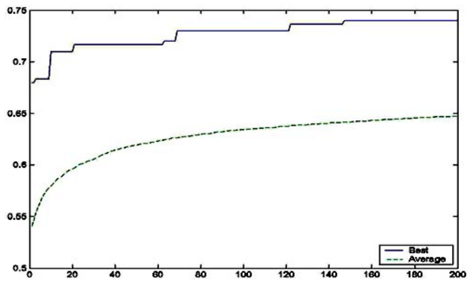

Fig. 4. Performance evolution using ES

Fig. 5. Retrieval result for "Bald Eagle"

say $fit(Rs, Ws) < threshold$, then the corresponding antecedent will be retained. Otherwise, this antecedent will be permanently discarded. The antecedent selection is random based, and many redundant and repetitious rules may be removed by this optimization strategy. It should be aware of that we optimize the rule base at the antecedent level rather than the rule level.

2.6. Query and Similarity Measure

Distinguishing from query by example (QBE) retrieval systems, the query in the proposed system is specified by an indicator vector $Q = [q_1, q_2, \ldots q_h]$, where q_i takes binary values with 1(0) representing the presence (*do not care*) of a semantic concept contained in the target images. During the retrieval process, a weighted dot product similarity measure $D(Q, S)$ is given by

$$D(Q, S) = \sum_{i=1}^{h} \alpha_i q_i e_i, \tag{10}$$

where $\alpha_k \geq 0$, subjected to $\sum \alpha_k = 1$, is a weighting factor which reflects the emphases on different semantics related to a specific query. In our simulations, we took these factors equally.

3. PERFORMANCE EVALUATION

The proposed RBIR system was implemented by MatLab 6.5 and Visual Basic 6.0, and evaluated by an image database with 300 animal images containing following semantic concepts: *Bald Eagle, Horse, Panda, Tiger* and *Zebra*. The sizes of the images are unified to be 300×400 or 400×300 and the image block size b is take as 4. A 2-D Daubechies-1 wavelet transform is applied to each small block. In this study, k is taken as 10 in the clustering algorithm. We selected 150 typical images as the training image set. 136 rules are extracted from those images, which include 242 antecedents in total.

The initial range for the parameters were set in $[0.5, 1.5]$ and the standard deviation was taken as 0.5. During the evolution process, the paramters are restricted to be equal and positive. The population was taken as 200 and evolved for 200 generations. Figure 4 shows the fit performance of the ES algorithm. Through rule optimization, the number of the rules is reduced from 136 to 65, and the total number of the antecedents is reduced from 242 to 91. As a result of rule optimization, the performance fit drops down from 74% to 60%. Figure 5 shows the system performance for a specific query with semantic concept "Bald Eagle".

4. CONCLUSION

Multimedia data mining techniques for developing RBIR systems has good potentials. From methodological viewpoints, *a priori* knowledge on semantic conepts plays a key role in rules mining. The ES algorithms demonstrated their feasibility and power to rule base optimization. This paper mainly contributes towards the development of a framework for building RBIR systems. Under this framework, rule base refinement and adaptation with user's relevance feedback will be interesting for further studies.

5. REFERENCES

[1] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 4, pp. 39–62, 1999.

[2] O. R. Zaane, J. Han, Z.-N. Li, S. H. Chee, and J. Y. Chiang, "Multimediaminer: A system prototype for multimedia data mining," *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 581 – 583, 1998.

[3] Y. Chen and J. Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1252–1267, 2002.

[4] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, Addison Wesley, England, 2002.