A Rule-based Approach to Image Retrieval

Dhaval Mehta, E.S.V.N.L.S.Diwakar, C.V.Jawahar
Centre for Visual Information Technology
International Institute of Information Technology
Gachibowli, Hyderabad - 500019, INDIA
jawahar@iiit.net

Abstract—In this paper, a rule-based approach is introduced for retrieving images from an image database. Compared to image-based queries, this approach allows the user to query in a more natural language. Performance is demonstrated with simple queries on a generic image database and sophisticated queries on a database of face images.

1. Introduction

With rapid growth of storage capacity and processing power, retrieval of relevant information from large multimedia databases has become an important problem in Information Technology. Conventional methods for processing numerical or textual data are found to be grossly insufficient for retrieval from image, video and audio databases. Manual annotation of images for content representation is a primitive solution to this problem. This has failed to meet the diverse requirements arising in a generic image (or multimedia) database system for retrieval of content specific data. Also in large databases, manual annotation is prohibitively expensive. Content Based Image retrieval (CBIR) systems aim at efficient retrieval of relevant images from large image databases based on characteristic features [5]. These features are typically extracted from shape, texture, or color properties of the images. The appropriateness of images in the database is ranked for the query according to similarity measures computed from the feature representation.

The notion of "content" in an image is often abstract. In many situations, it is difficult for the system to explicitly represent this content. User often finds it difficult to express the content. In such situations, query-by-image is preferred. The user gives a sample query-image and most similar images are retrieved from the database. However, higher level interactions are not possible with such systems. When a user can characterize sun-set image as *an image with top middle yellow and bottom middle black*, an image-based approach is inappropriate. This paper puts forth a rule-based approach for retrieval of images. During indexing, a set of low-level features are extracted for each individual regions [2]. The features can be based on color, texture, geometric and spatial properties of the region. This can provide a powerful interface for generic databases. If the database is domain specific, like medical image database or satellite image database, one can think of defining human understandable concepts in terms of the image features computed. Such systems can in tern support higher level queries which are human friendly.

0-7803-7651-X/03/$17.00

In this paper, we consider a domain specific database of face images and demonstrate the utility of a rule-based approach for retrieval of queries like *show laughing people* and *old or beautiful ladies* In face database, we extract a set of secondary features from the low-level features. These include features like face location, face size, cloth properties, orientation of face, side view or front view of face, properties of various facial features like eye, nose, mouth, hair etc [6].

Section 2 provides the background material on rule based systems and fuzzy rule based approaches. Feature extraction and representation is discussed in Section 3. Inferencing schemes employed in this paper are described in Section 4. Section 5 provides the results for many sample queries in a generic and a face database.

2. Preliminaries

Rule-based Systems

A rule-based system consists of a set of IF-THEN rules, a set of facts, and an interpreter controlling the application of the rules, given the facts. The rules and facts are used to convert the high-level query given by the user to a low-level query that can directly use the extracted features.

The components of the proposed rule based system are [9]:

1. The Rule-base, where the relevant rules are stored. For example in the face database system, a rule can be:
If mouth is open and mouth is wide then person is laughing.
2. The fact-base, which represents the current state of the world. This is where the facts or declarative knowledge is stored. For example facts can be:
White color has RGB values around $< 255, 255, 255 >$.
Round face has height to width ratio around $< 1.0 >$
3. An inference engine, which operates on the rules and facts. The interpreter parses the query, expands the rules and reasons with the help of computed features.

Fuzzy Sets and Fuzzy Rule-based Systems

Classically, an element may or may not belong to a set. This concept of crisp sets may be extended to fuzzy sets with the introduction of the idea of partial truth. Thus, in fuzzy theory, an object (element) may be a member of a set to a certain extent depending on the membership function characterizing

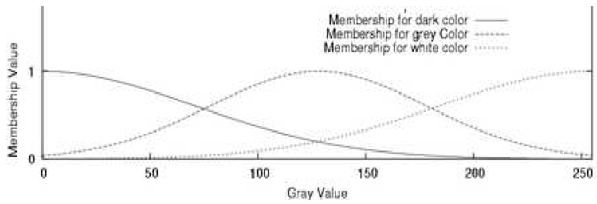

Figure 1. Membership Functions for "dark", "gray" and "white"

the set. For example, in classical sets, if we want to classify pixel by their grey values into three groups - Dark, Gray and White we define these sets as

$$f(x) = \begin{cases} dark & \text{if } 0 \leq x \leq 60 \\ gray & \text{if } 60 < x \leq 200 \\ white & \text{if } 200 < x \leq 255 \end{cases}$$

But a pixel with grey value 62 is too dark to be classified as a gray pixel. If by incorporating fuzziness, we can rewrite the membership functions as shown in Figure 1. Now, if we want to classify a pixel with gray value 62, it can be classified as a pixel with 0.7 dark, 0.5 gray and 0.0001 white in color.

Thus, by introducing fuzzy logic, we can use natural language terms like high temperature, young man, big size. etc. By incorporating fuzziness in the features extracted from the images the queries can be given in a more natural language.

The building blocks of the rule-based system can be crisp or fuzzy. We use a fuzzy-rule based system, which is a generalization of the conventional one. In a fuzzy system, memberships makes the same ontological commitment as logic. Fuzzy logic takes the features as fuzzy predicates and returns the truth value of the feature as a number between 0 and 1 rather than just true or false. Fuzzy logic takes a complex sentence like 'A and B' (where A and B are two features) and determines its truth value as a function of the truth values of its components.

3. Region Extraction and Representation

Knowledge acquisition refers to ways in which a knowledge-based system evolves its knowledge representation [1]. Systems acquire their knowledge either by being told by experts or by learning from experience. In either case knowledge acquisition is an integral part of indexing the image database. An image may be retrieved based on the local or global 'content' in the image. A user may be interested in a specific region/object with some properties or may be interested in their relative spatial arrangement. We need a mechanism to extract this information from the image and represent it in an appropriate form.

Developing a segmentation algorithm which will meaningfully segment all images is yet an open problem in image

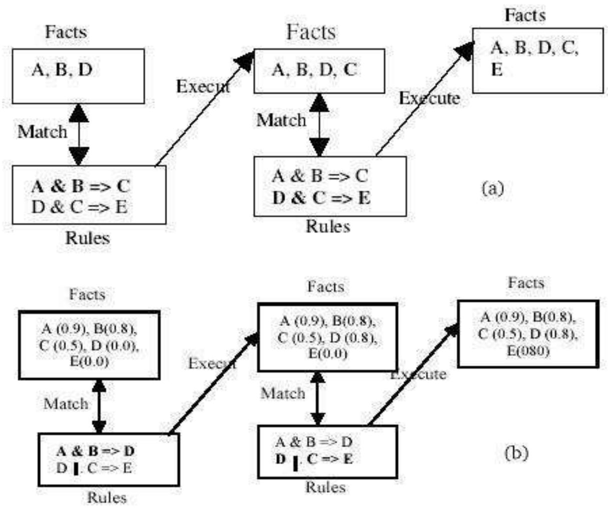

Figure 2. Inferencing Chain for Rule-Based System

analysis. Therefore, for a generic database, we use a simple spatial partitioning as segmentation, while for a domain specific case (face database), we use domain information for feature extraction [4]. For some specific cases, a semi-automatic region extraction is also employed.

Numeric values corresponding to low or high-level features are computed from the segmented regions. In situations related to human perception, it is not necessary to process the exact numerical values and what one is usually interested in a fuzzy representation of the feature. A query to extract "brown" images never really insists on any precise RGB value. Instead, a RGB value around $< 165, 42, 42 >$ is sufficient. Therefore, it is preferable to define the notion of color [8] (or as a matter of fact many other features) as fuzzy sets for supporting systems where a user prefers to query more intuitively and in natural language.

Mathematical basis of fuzzy sets and fuzzy logic [3] allows us to represent these sets with partial attachments for individual images. Membership is characterized by a real $[0, 1]$ valued function. Conventional boolean binary logic can be extended to this multivalued framework easily.

4. Inferencing

The Inferencing engine is the heart of a rule-based system. This module processes the raw features extracted from the images with the help of available knowledge in the form of rules to identify and thereby retrieve the relevant images. Given a rule `If A AND B THEN C`, and a query to retrieve images with `C`, the system searches for images where `A` and `B` are true. Thus, the knowledge based system employs the small fragments of human know-how (rules) along with the features to retrieve the images. The retrieval process may not be this simple always. A continuous chain of successive inferences may be necessary for the appropriate retrieval.

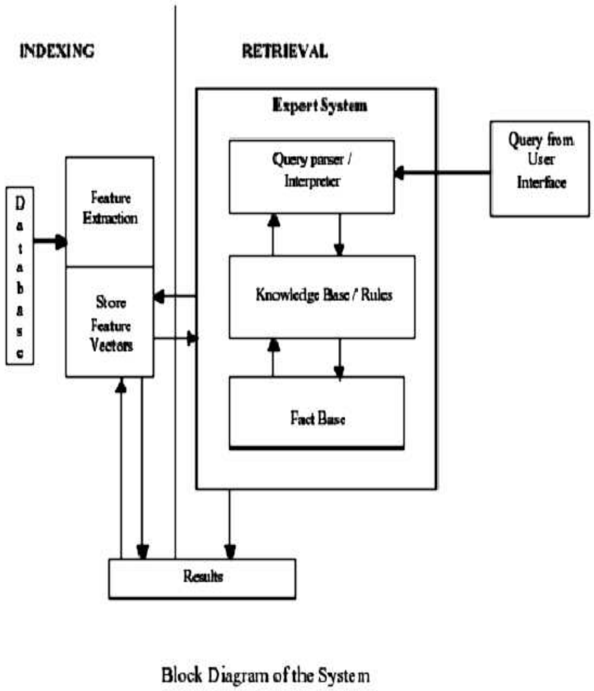

Figure 3. Block Diagram of the System

The query is parsed and expanded using the rules present in the rule base [9]. Then, we check antecedents as long as the inference chain can be expanded. Once we reach the lowest level, where explicit feature representations are available, truthness of the antecedents is verified. Then images whose features comply with the facts are retrieved. This inferencing process is explained in Figure 2 (a)

For example, consider that we have following rules in our rule base:

(A)Open mouth and (B) wide mouth => (C) laughing

(D)Fair skin and (C) laughing => (E) beautiful person

Subjectivity of the image perception is incorporated in the rules, which the system uses. Given a query to retrieve beautiful faces, the inferencing process shown in Figure 2(a) is employed. Here, the fact that the image has properties A, B and D. i.e., a specific sample face image has Open mouth, Wide mouth and Fair skin. For the first rule, it matches both the conditions of the antecedents. So after execution, fact C is also added to fact base. In the same way the second rule is also executed. At the end we conclude that E is true for this face image. i.e., we conclude that this is the face image of a beautiful person.

In image databases, we also need to rank images based on the appropriateness of the image to the given query. This can be achieved using a fuzzy rule based system where the image can satisfy a property even partly. The definition of the open mouth is often fuzzy. As the width to height ratio of the mouth increases, truthness $T(\cdot)$ corresponding to open mouth increases and beyond a specific threshold, all mouths are considered open. These partial truthness is expressed as fuzzy memberships. If the feature is a considered fuzzy, then it will have membership in all sets defined over the feature space.

The mathematical basis [3] of fuzzy logic allows us to process rules containing multivalued logical variables. If A and B are fuzzy sets and $T(A), T(B)$ are the memberships of an image in A and B, truthness of logical expressions like $AandB$ or $AorB$ can be evaluated as

$$T(AandB) = \min(T(A), T(B))$$
$$T(AorB) = \max(T(A), T(B))$$
$$T(not(A)) = 1 - T(A)$$

There are alternate definitions for many of these logical operations.

Therefore, `If A AND B THEN C` gives $T(C)$ as $\min(T(A), T(B))$. Thus, it is possible to have images with different $T(C)$, compared to $[0, 1]$.

A query which consists of descriptions of features bounded by binary operators is evaluated. The images are ranked according to this value and the best results are displayed.

The user can give modifiers like high, low , not etc., to characterize the features. These modifiers are additional constraints on the queries that can be given by the user to get better results. They consist of adding adjectives to the features. These are also called as linguistic hedges. To incorporate these linguistic hedges, the fuzzy distribution function is to be modified accordingly. For example, to incorporate the 'high' modifier, the fuzzy distribution function has to become narrower.

$$T(lightly(A)) = sqrt(T(A))$$
$$T(highly(A)) = T(A) * T(A)$$

For a fuzzy-based system we have a query (A)top middle yellow and bottom middle black or (C) middle yellow.

A and B => D

C or D => E

We want to search for images with E i.e. A and B or C. If for an image, truth value for A is 0.9, for B is 0.8 and for C is 0.5, the image has truth value of max(min(0.9,0.8),0.5) = 0.8. This is shown in Figure 2 (b).

Figure 4. Images with top right red and middle black

Figure 5. images with top right red and middle highly black

The proposed system has two major components. The indexing phase consists of extraction of features from each image in the database and representing in the fact base. The retrieval phase consists of parsing the query given by user. This is done by expanding the query by using knowledge base and fact base. Finally relevant images are retrieved. The block diagram of the system is shown in Figure 3

5. Results

Performance of the proposed approach is demonstrated here on a generic image database of 40,000 images. Sophisticated queries are demonstrated on a domain specific database – a database of face images. This database consists of frontal views of faces, in uniform and cluttered background, studio photographs etc.

Color based Retrieval from a general database:— The proposed approach is verified on a generic image database. The system uses fuzzy sets to represent colors present in the image. The image is divided into nine regions (top left, top middle, top right, middle left, middle, middle right, bottom left, bottom middle and bottom right) and the color characteristics of the nine regions are expressed as memberships in color-fuzzy-sets. The whole color space is divided into nine color-fuzzy sets - Red, Green, Blue, Black, Brown, Yellow, Orange, Gray and White. For example, brown is considered as "around brown" such that the maximum membership is assigned at $< 165, 42, 42 >$ and membership monotonically decreases as color departs from this mean value. These mean color coordinates are assigned by a human being based on perceptual definitions. Exponential membership functions are assigned to each of these colors such that the crossover takes place somewhere in the middle of two peaks.

Figure 6. Laughing People

Example Query 1: Show all images with top right red and middle black

Let the color characteristics of top right region for an image be $< 215, 2, 30 >$. This RGB value will have membership in all the fuzzy color sets that were defined. But, we consider the membership of that RGB value in the fuzzy set of red color. Similar is the case with black color in the middle region. Thus, in a fuzzy framework, all images have red color in the top right region and middle black. However, their truthness differs. Using the definition of fuzzy *AND (min)*, image with largest truth value together for red in top right and black in middle are retrieved. Retrieved images are shown in Figure 4.

Example Query 2: Show all images with top right red and middle highly black

This query is similar to the above query but it has a modifier *highly* for the black color of the middle region. Hence, the membership function for black color in the middle will become narrower. Using the definition of fuzzy *AND (min)*, image with largest truth value together for red on top right and highly black in the middle are retrieved. Retrieved images are shown in Figure 5.

Rule based Retrieval on face database:— Retrieval from face databases is important for many multimedia applications. It has attracted researchers due to its wide range of applications in commercial and law enforcement activities, image retrieval engines, video conferencing, ease in human computer interaction etc. There are many applications in which face databases are used. The knowledge of how to deal with the face database and how to extract information in an image is very important for these applications. Some of the appli-

Figure 7. Old or Laughing people wearing white shirt

cations include Face Recognition Systems, Automated Video Conferencing Systems and Intelligent Human Computer Interaction Systems.

Face images are initially segmented into two parts – face and background. The segmentation can done by using skin color model or by Neural Network based method or by using Edge based method [7][4]. After face detection, various features pertaining to the image are to be extracted like size of the image, background properties, etc. Domain specific features like face location, average skin color, eye location, eye color, hair color, size of face, size of eye, mouth location, cloth color etc. are then extracted [6].

Example Query 3: Laughing people

Here we have rule that "If mouth is open and mouth is wide then the person is laughing". We have taken as facts that open mouth means distance between lips is at least 10 pixels and wide mouth means distance between two corners of the mouth is at least 20 pixels. So we conclude that the images where mouth gap is > 10 and mouth length is > 20 satisfies user requirement. The system searches in the database and retrieves images which satisfy the query in the best possible manner. The results are shown in Figure 6.

Example Query 4: Old or Laughing people wearing white shirt

The results are shown in Figure 7. Here the *or* operator has least precedence. For all the laughing images, we check for images with white shirt. We have fact that white color is RGB values around $< 255, 255, 255 >$ and shirt means cloth and so we have to check for white color cloth. For old, we have rule that "If hair is white then person is old". So we search for images with white hair in the feature vector for all images and get the appropriate results. There is *or* operator between two parts of the query so we include images from both the parts in the final result.

6. Conclusions

A rule based system for retrieval of images based on textual queries is presented in this paper. The system uses fuzzy notions for feature representation and inferencing. Results are shown on a generic heterogeneous database and a face database. In the generic database, color-based queries are demonstrated. In the face database, face properties are indexed along with associated information like cloth color, background area, hair color etc. Many of the face image parameters are automatically extracted, while some of them are measured semi-automatically. This helps in handling higher level and human friendly queries. Natural language processing of the human queries can be the next step in providing better interface. We are presently working on other knowledge based mechanisms for image retrieval. Application systems for medical and satellite images are also envisaged.

References

[1] Hih-Cheng Hsu, Wesley W.Chu, Ricky K.Taira, "A Knowledge-Based Approach for Retrieving Images by Content", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 6, No. 4, August 1996, pp. 522 - 532.

[2] Yong Rui, Thomas S. Huang and Shih-Fu Chang, "Image Retrieval: Past, Present and Future", *Journal of Visual Communication and Image Representation*, Vol. 10, pp. 1 - 23, 1999.

[3] H. J. Zimmermann, "Fuzzy Set Theory and its Applications", *Allied Publishers Limited*, 2nd Edition

[4] Ming-Husan Yang, David J. Kriegman and Narendra Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, January 2002, pp. 34 - 58

[5] Arnold W. M. et al., "Content Base Image Retrieval at the end of early years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, December 2000, pp. 1349 - 1380

[6] Karin Sobattka and Ioannis Pitas, "Extraction of Facial Regions and Features using Color and Shape Information", Department of Informatics, University of Thessalonikim Greece

[7] P.Peer and F.Solina, "Automatic Detection of Human faces in images", COST 254: *International Workshop on Intelligent Communications and Multimedia Terminals*, pp.39-42,Liublijana, Solvenia, 1998

[8] Yixin Chen and James Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1252-1267, 2002.

[9] Stuart J. Russell, Peter Norvig, "Artificial Intelligence: Modern Approach", *Prentice Hall*, 2st edition