Natural Language Querying for Content Based Image Retrieval System

Sreena P. H.[1], David Solomon George[2]

M.Tech Student, Department of ECE, Rajiv Gandhi Institute of Technology, Kottayam, India[1,]

Asst. Prof, Department of ECE, Rajiv Gandhi Institute of Technology, Kottayam, India[2]

Abstract: Content based image retrieval (CBIR) system is a database management system for retrieval of images based on the similarity of image content with the query image. The issue of semantic gap causes retrieval of irrelevant images from database. In the proposed CBIR system, fuzzy clustering of the Tamura texture features extracted from the database is used to overcome this problem. Natural language querying is implemented by use of fuzzy data space. The fuzzy membership values derived using fuzzy c-means clustering is used as similarity measure. The proposed technique is implemented in Matlab and its effectiveness is verified using the standard Brodatz texture database.

Keywords: Content Based Image Retrieval System, Fuzzy c-means clustering, Natural language query, Semantic retrieval

I. Introduction

With increased accessibility to internet and low cost digital storage devices, large amount of digital information is available to common man. Taking the example of images, this large volume of data is used in fields of education, entertainment, commerce, biomedicine, and crime prevention. The highly evolved, and much sophisticated visual system help human beings to select the required image from a large image database within a few minutes or less. But it is hard to teach a machine the interpretation of what we see. Yet, tremendous efforts have been made over the past few decades to make the machine understand, index and annotate the pictures over wide range with much progress. Image retrieval systems are such image database management systems that try to understand the image before presenting it to user.

Different image retrieval techniques are in use [1]. They are categorized on the basis of nature of query like keyword based, free text based, image content based, or composite of these. Newer techniques, which use user interaction, help to reduce the semantic gap between the retrieved images and user query [2]. Early image retrieval techniques based on textual queries became inadequate as a result of advances in internet and digital image storage technologies [3], [4]. The systems that use the properties inherent to images would be able to handle large image databases. The CBIR systems were developed on this idea. A few commercial and experiment prototypes CBIR systems such as QBIC [5], Photobook [6], Virage [7], Netra [8], VisualSEEK [9], SIMPLIcity [10] have been developed.

Fig. 1 gives general overview of a CBIR system. The first step in developing a CBIR system is feature database creation. The feature database should contain feature vector corresponding to each image in the database. A feature vector is selected depending on the nature of database and application of the system. Commonly used content based features or the low level features are color, texture and shape. Features can be classified as general features- color, texture and shape and domain specific feature- face, fingerprint, etc, as they depend on application. As image perception differs for different person, there is no single best feature representation.

On presenting a query image to the system, its feature vector is extracted and similarity/distance with all feature vectors in database is measured. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly. Various similarity measures like Minkowski-Form distance, Quadratic Form distance, Mahalanobis distance etc.are found in literature [11].Images with largest similarity or shortest distance are made available to the user after sorting.
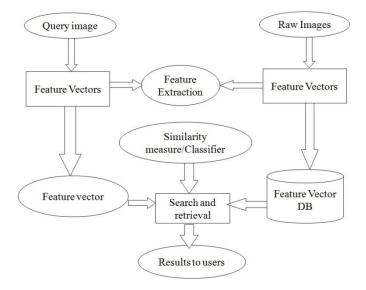
Fig. 1 Flow Chart of CBIR System

The main problem faced by CBIR systems is twofold. One is the semantic gap problem which is the lack of coincidence between the image representation and the human interpretation for an image. The low level feature bears no information about the semantic content of the object being retrieved. Other is the perception subjectivity. Different users, or even the same user under different circumstances, may interpret an image differently. Moreover, the way users define the similarity between two images may be quite different. The perception subjectivity problem exists at each semantic level and depends on individual experience.

Recent works in CBIR mainly aims at reducing the semantic gap between the query and retrieved images. Common methods used to address the problem of semantic gap are: using object ontology to define high-level concepts, using machine learning tools to associate low-level features with query concepts, introducing relevance feedback (RF) into retrieval loop for continuous learning of user's intention, generating semantic template (ST) to support high-level image retrieval, and making use of both the visual content of images and the textual information obtained from the Web for WWW (the Web) image retrieval [12].

In this paper the semantic gap and perception subjectivity problems are addressed by using fuzzy logic. Semantic clusters are defined by fuzzy c-means clustering. Fuzzy linguistic terms are defined which helps in natural language querying. A statistical texture measurement called Tamura texture of the standard Brodatz database [13] is extracted to create the feature space. Each Tamura feature space is fuzzy clustered. Multiple queries are combined using fuzzy min-max aggregation technique.

The proposed system is discussed in detail in the following section. The Tamura feature extraction and fuzzy clustering and aggregation techniques used in the system is explained in section II. Section III gives the results for some example query and the precision results for the retrieval system.

II. PROPOSED CBIR SYSTEM

CBIR system proposed in this paper uses Tamura texture features for retrieval and fuzzy clustering for semantically grouping the images. Brodatz digital album is used as image database. The proposed system is shown in figure 2 below.
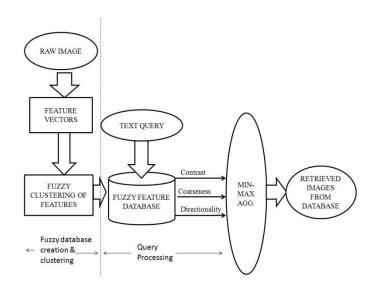
Fig. 2 Flow Chart of the proposed CBIR System

The steps for implementation of the proposed system are
- Feature extraction
- Fuzzy clustering
- Query processing

After the first two steps fuzzyfied feature space is obtained, which is stored in the system database and is used for the query process, where actual retrieval takes place.

A. *Feature extraction*

As mentioned earlier Tamura texture features are used as texture features. A set of six texture features taking on the basis of human psychological experiments [14]. Tamura gave a set of images from Brodatz album to a group of people and asked to describe the texture. From the result mathematical expression for six features were developed that effectively describe all input patterns and were well distributed within these patterns. The features include coarseness, directionality, contrast, line-likeness, regularity and roughness. Many CBIR have effectively used Tamura features for texture description.

Out of six three important Tamura features of all the database images are extracted. Modified versions of coarseness, contrast and directionality are used. The other three features, line-likeness, regularity and roughness are derived from the former three features, hence not used.

Coarseness describes the size of the texture particles and its measurement is given by size operator. Around each pixel, a neighbourhood of size 2k x 2k is defined (k = 0, 1 … 5). The k value that maximizes the gradient of pixel values over non-overlapping neighbourhoods in both horizontal and vertical orientation is taken as S_{best}. A histogram of S_{best} is made, which is the coarseness feature vector of the image.

Contrast modification of image is usually done by changing its gray level distribution. Tamura et al. also considered other factors like dynamic range of gray levels, polarization of the distribution of black and white on gray-level histogram. The first is measured using the standard deviation of grey levels and the second the kurtosis α_4. The contrast measure is therefore defined as ratio of standard deviation and n^{th} power of α_4.

Directionality is measured as a global property, without considering orientation. i.e., same figures with different orientation should have same directionality. Histogram of local edges probabilities against directional angles is used which is shown to sufficiently represent global features of the input picture as long lines and simple curves.

The features are stored in feature database. Instead taking texture feature vector as element of 6-D space (corresponding to 6 Tamura features), each of the 6 feature is taken as a vector. A histogram of size measures S_{best} is made with respect to different values of 2^k for k=0, 1...5. This forms vector for coarseness. The directionality is represented by 16-D direction histogram HD. Direct histogram measure for contrast doesn't exist. So it is represented as a vector with each component the contrast measure of one of the 128 x 128 image sub blocks. Thus a small amount of spatial information is added to the measurement.

B. Fuzzy clustering

After feature extraction fuzzy clustering of feature space takes place. The extracted features are crisp values in the range [0, 1]. Fuzzyfication of these features is done by grouping the feature space to multiple classes and assigning membership values to them. Membership value indicates the degree to which a feature belongs to the class. In fuzzy set theory each class is semantically indexed using linguistic labels. This helps to bring natural language interpretation of the features.

Fuzzyfication of texture features involves dividing the texture data to multiple classes and assigning membership values to features in each class. The algorithm for classification and membership assignment is done by modifying Lin's algorithm [15]. Three classes are specified for each texture feature. A linguistic term is associated with a class. The linguistic term bridges the gap between the low-level feature and high-level semantics of the texture. Tamura texture features and linguistic terms associated with each feature are shown in table I. The class of the image is represented by the cluster centers and its membership value in each cluster. The cluster centres and membership value of a feature is obtained using fuzzy c-means clustering algorithm. These linguistic terms and membership values are stored for query processing.

TABLE I
LINGUISTIC TERMS FOR TAMURA FEATURES USED FOR RETRIEVAL

Tamura Texture	Linguistic terms		
Coarseness	Fine	Medium	Coarse
Contrast	Low	Medium	High
Directionality	Isotropic	Medium	Directional

Fuzzy c-means algorithm is the fuzzy extension of the k-means algorithm [16]. The algorithm is described as described below:

Given the data set $\mathbf{Z}$, choose the number of cluster $1 < c < N$ the weighting exponent $m > 1$, the termination tolerance $\varepsilon > 0$ and the norm-inducing matrix $\mathbf{A}$.

Initialize the partition matrix $\mathbf{U}$, randomly, such that $\mathbf{U}^{(0)} \in M_{fc}$.

Repeat for $l = 1, 2, \ldots$

Step 1: Compute the cluster prototypes (centre):

$$v_t^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{tk}^{(l-1)})^m z_k}{\sum_{k=1}^{N} (\mu_{tk}^{(l-1)})^m}, \quad 1 \le i < c. \tag{1}$$

Step 2: Compute the distance

$$D_{tk\mathbf{A}}^2 = (z_k - v_t^{(l)})^T \mathbf{A}(z_k - v_t^{(l)}), 1 \le i < c, 1 \le k \le N \tag{2}$$

Step 3: Update the partition matrix:

For $1 \le k \le N$, If $D_{tk\mathbf{A}} > 0$ for all $i = 1, 2, \ldots, c$

$$\mu_{tk}^{(l)} = \frac{1}{\sum_{j=1}^{c} \left(D_{tk\mathbf{A}} / D_{jk\mathbf{A}} \right)^{2/(m-1)}} \qquad (3)$$

Otherwise $\mu_{tk}^{(l)} = 0$ if $D_{tk\mathbf{A}} > 0$, and $\mu_{tk}^{(l)} \in [0,1]$ with $\sum_{t=1}^{c} \mu_{tk}^{(l)} = 1$.

Until $\left\| \mathbf{U}^{(l)} - \mathbf{U}^{(l-1)} \right\| < \varepsilon$

Here, v_t is the tth cluster centre, μ_{tk} is the membership value for the kth image in cluster t, and elements of the partition matrix $\mathbf{U}$ are μ_{tk}.

The result of clustering is N classes to which the data vectors are assigned. Each vector can be a part of more than one cluster, depending on the membership value. Fuzzy c-means clustering scheme is used in this thesis to identify the cluster to which the database images belong. The membership value gives the degree of inclusion of an image to a cluster thus the indecisiveness in classification is also brought to consideration.

C. Query Processing

In this step a textual query is given to the system. Predefined query terms are used. They are the fuzzy linguistic terms as given in table 5.1. Multiple queries are connected using logic operators like OR, AND and NOT. Examples for fuzzy querying are as follows.

- Coarse and medium contrast and isotropic.
- Fine and high contrast or directional.

The connectives AND, OR, and NOT are used in the system to aggregate multiple features. The results of these operations are also fuzzy in nature, with membership value as defined in table II given below.

TABLE II
FUZZY AGGREGATION RULES

Fuzzy Connector * X=A*B	Membership value $\mu_X(x)$
AND	$\min(\mu_A(x), \mu_B(x))$
OR	$\max(\mu_A(x), \mu_B(x))$
NOT	$1 - \mu_A(x)$

The aggregation combines the three features and assigns membership value to images as given in table II. For example, for the queries given above the resulting membership values are $\min(\min(\mu_1(x), \mu_2(x), \ldots \mu_n(x)))$ and $\max(\min(\mu_1(x), \mu_2(x), \ldots \mu_n(x)))$ respectively, where $\mu_i(x)$ is the membership value of the ith image in the database.

Once the membership value is decided the images are to be ordered for retrieval. The overall membership value is taken as the similarity measure. Here ordering is done on the basis of the new membership value. Greater the membership, more will the image satisfy our query.

III. RESULTS

The proposed system was applied to 111 images of Brodatz database in GIF format. The system is implemented in MATLAB. Measurements for directionality and coarseness are made without partitioning the image into sub images. Local contrast measurements are taken by dividing images to 128 x 128 pixel blocks. The queries used to verify the effectiveness of the system are given below:

1) Coarse, High contrast and Isotropic.
2) Fine, Low contrast and Directional

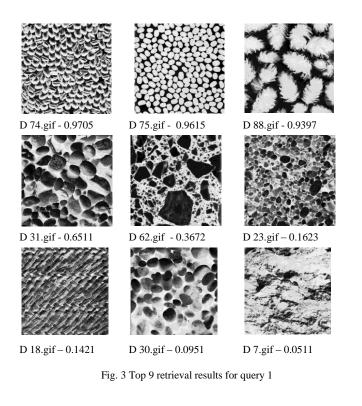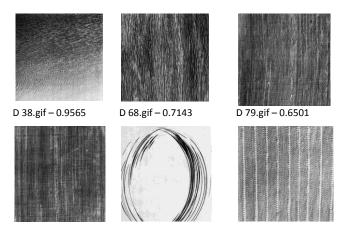# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

3)   Coarse, Low contrast and Directional

The top 9 retrieval results for the above three query are shown in figure 3, figure 4 and figure 5, respectively. The retrieved image name and corresponding membership value is given along with the image. The similarity is measured basis of membership value obtained after aggregation. The images are arranged in decreasing order of the membership value. It can be seen that as membership value decreases the images is less likely to satisfy all the three requirements.

In figure 3 the first 6 retrieved images fully satisfy the query, being high contrast, coarse and isotropic. The seventh image is slightly directional last one has poor contrast. But one may decide not to choose these images as its membership to these groups is very low.

D 74.gif - 0.9705   D 75.gif - 0.9615   D 88.gif - 0.9397

D 31.gif - 0.6511   D 62.gif - 0.3672   D 23.gif – 0.1623

D 18.gif – 0.1421   D 30.gif – 0.0951   D 7.gif – 0.0511

Fig. 3 Top 9 retrieval results for query 1

D 38.gif – 0.9565   D 68.gif – 0.7143   D 79.gif – 0.6501

D 78.gif – 0.6375   D 44.gif – 0.5094   D 85.gif – 0.2401

D 1.gif – 0.1950   D 94.gif – 0.1179   D 19.gif – 0.0875

Fig. 4 Top 9 retrieval results for query

D 47.gif – 0.8541   D 50.gif – 0.5486   D 96.gif – 0.1706

D 45.gif – 0.0748   D 13.gif – 0.0701   D 71.gif – 0.0646

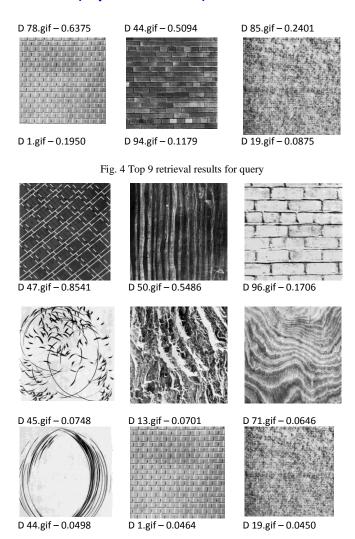D 44.gif – 0.0498   D 1.gif – 0.0464   D 19.gif – 0.0450

Fig. 6 Top 9 retrieval results for query 3

The observations made for query 1 can also be applied to query 2 and query 3. It can also be observed that images D44 and D1 are retrieved for both query 2 and query 3. But from the membership values it can be concluded that they are of fine, low contrast and directional (query 2) rather than coarse, low contrast and directional. i.e. the images are part of more than one cluster. Thus the issue of perception subjectivity is taken in to by the proposed system.

Precision measures the retrieval accuracy of a CBIR system [17]. It is the ratio of number of relevant images retrieved to total number of images retried. The relevant images are those which satisfy the query to the system.

$$Precision = \frac{No.\,of\ relevant\ images\ retrieved}{Total\ No.\,of\ images\ retrieved}$$

The precision measures for the three queries are given in table III. Precision when the total number of images retrieved is 5, 9, 15, 20 and 25 are given. The number of relevant images is given in brackets.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

TABLE III

PRECISION AT DIFFERENT NO. OF RETRIEVED IMAGES

Total No. of images retrieved	Query 1	Query 2	Query 3
5	1 (5)	0.8 (4)	0.6(3)
9	0.7778 (7)	0.7778 (7)	0.5556(5)
15	0.6000 (9)	0.6000 (9)	0.4667 (7)
20	0.5000 (10)	0.5500 (11)	0.4000 (8)
25	0.4400 (11)	0.5200(13)	0.3200 (8)

From the precision analysis it can be seen that when all the database images are retrieved precision approaches zero. That is the later retrieved images are irrelevant.

IV. CONCLUSION

A CBIR system that uses a natural language query is proposed in the paper. Tamura texture measures of the image is extracted and fuzzyfied using fuzzy c-means algorithm. It is stored in database as the low level fuzzy feature vector of the images. Multiple queries are aggregated using fuzzy min-max algorithm and the resulting membership value is used as similarity measure. Greater the membership the image better satisfy the query. The effectiveness of system is verified using Brodatz texture database. The precision results show that the images with larger membership are more relevant to the query.

REFERENCES

[1] R. Dutta, D. Joshi, J. Li, and J. Z. Wang, "Images Retrieval: Ideas, Influences, and Trends of New Age", *ACM Computing Surveys*, Vol. 40, No. 2, 2008.
[2] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE *Transaction on Circuits Video Technology*, vol 8, no. 5, pp. 644–655, 1998.
[3] N. S. Chang and K. S. Fu, "Query-by pictorial-example", *IEEE Transactions on Software Engineering*, vol.6., 1980.
[4] S. F. Chang, J. R. Smith, M. Beigi, and A. Benitez, "Visual information retrieval from large distributed online repositories", *Communication ACM (Special Issue on Visual Information Retrieval)*, pp. 12–20, Dec. 1997.
[5] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker, "Query By Image and Video Content: The QBIC System", *IEEE Computer*, Vol. 28, No: 9, pp. 23-32, September 1995.
[6] A. Pentland, R. Picard and S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases", *International Journal of Computer Vision*, Vol. 3, pp. 233-254, 1996.
[7] A. Gupta, "Visual Information Retrieval: A Virage Perspective", Virage Inc, San Diego, Technical Report Revision 4, CA 92121, 1996.
[8] W. Y. Ma, B. Manjunath, "Netra: a toolbox for navigating large image databases", *Proceedings of the IEEE International Conference on Image Processing*, 1997, pp. 568–571.
[9] J. R. Smith, S. F. Chang, "VisualSeek: a fully automatic content based query system", *Proceedings of the Fourth ACM International Conference on Multimedia*, 1996, pp. 87–98.
[10] J.Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries", *IEEE Transactions Pattern Analysis and machine intelligence*, vol. 23, no. 9, pp. 947–963, 2001.
[11] F. Long, H.J. Zhang, D.D. Feng, *Fundamentals of content-based image retrieval*, Multimedia Information Retrieval and Management, Springer, Berlin, 2003.
[12] Y. Liu, D. Zhang, G. Lu, Wei-Ying Ma, "Asurvey of content-based image retrieval with high-level semantics", *Pattern Recognition* ,vol 40, pp. 262 – 282
[13] P. Brodatz, Textures, A Photographic Album for Artists & Designers, Dover, New York, NY, 1966.
[14] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. Smc-8, No. 6, June 1978.
[15] H. C. Lin, C. Y. Chiu, S. N. Yang, "A Fuzzy Logic CBIR System",*Proceedings of IEEE International Confereene on Fuzzy Systems*, pp. 1171-1176, 2003,
[16] J. C. Bezdek, R. Ehrlich, W. Full, "The Fuzzy c - means clustering algorithm", *Computers and Geosciences*, Vol-10, pp. 191-203, 1984.
[17] H. Muller, W. Muller, D. M. Squire, S. M. Maillet, T. Pun, "Performance Evaluation of Content Based Image Retrieval: Overview and Proposals", *Pattern recognition Letters -21*, pp. 593-601, 2001.