 ISSN: 2321-8169
2508 - 2512

Acquisition of Images using Neural Network

N Lohitesh Kumar
Dept of CSE
CIT
Tumkur, India
e-mail: lohiteshkn@gmail.com

Vinay D R
Dept of CSE
CIT
Tumkur, India
e-mail:vinu.ise@gmail.com

***Abstract*—** The application of computer vision to the image retrieval problem is Content-based image retrieval (CBIR). The interest in digital images is growing day by day. Users in professional fields are make use of the opportunities offered by the ability to access and manipulate remotely-stored images in different ways. The problems in image retrieval are becoming widely accepted, and the finding solution is an active area for research and development.

This dissertation work aims at developing a hybrid scheme for intelligent image retrieval system using neural networks. Each image in the database is indexed by a visual feature vector, which is extracted using color moments and discrete cosine transform coefficients. The query is characterized by a set of predefined semantic labels. A novel method of similarity measure using dot product is used for ranking and retrieval for improved performance of the system

***Keywords**-content Based Image Retrieval, Discrete Cosine Transform, Image Retrieval*

*****

I. INTRODUCTION

In the recent years, the growth of the internet has enormously increased the number of image collections. The accumulation of these images is attracting more and more users in various professional fields such as fashion, geography, medicine, satellite, architecture, advertising and design. Meanwhile, the study of image retrieval which is concerned with effectively and efficiently accessing desired images from large database has become more challenging. Retrieval of images is concerned with techniques for storing and retrieving images both efficiently and effectively [1].

Over the past few years, many advanced techniques have been evolved in Content- Based Image Retrieval (CBIR) systems. Applications like art, medicine, entertainment, education, manufacturing, etc, make use of a vast amount of visual data in the form of images. In recent years, CBIR systems will be based on features like color, shape, texture, spatial layout, object motion, etc [2]. Among all the visual features, color is the most dominant and distinguishing one in almost all applications.

CBIR system involves a number of critical areas where research is needed which includes representation of data, extractions of features and indexing, query matching and user interface. The motivation for developing CBIR systems is to release the workload of manually annotating image data using text based keywords. The existing CBIR systems can be categorized into two classes [3, 4]. The first scheme extracts low-level visual features from images, and then uses a similarity measure to calculate the distance between a query and images from the database for ranking the images. The second scheme is a semantic content-based method, where semantics are automatically extracted from images, and then a construction key is made from these semantic. The query is characterized using combinations of the semantics extracted from the images, and the retrieval is achieved by counting the semantic items occurred. Currently, most of all CBIR systems fall into the first category, where semantic information of the image is not utilized during retrieval.

To overcome the gap between the low-level visual features and the high-level semantics, pattern recognition techniques can be used to extract semantics. To obtain high-level semantics, which is desirable in image retrieval, region information is not sufficient. For some applications, object extraction can be ignored in design of CBIR systems. The reason for this is the objective of the CBIR system is to retrieve some semantic relevant images from databases rather than to recognize objects from images.

Neural network is a modeling tool which is best for image retrieval tasks. It has been successfully applied in intelligent image retrieval systems, especially for semantics recognition and learning similarity measure. The proposed image retrieval (IR) system takes low-level visual features as the system input and the semantic labels as its output. Off-line learning takes place before performing retrieval tasks [5].

The rest of the dissertation is organized as follows. Section2 describes the image retrieval system in details. Section3 presents the implementation and experimental results. Section4 represents the conclusion and future scope.

II. SYSTEM DESIGN

System design aims to identify the modules that should be in the system and also the specifications of these modules to interact with each other to produce the desired results.

A. System Architecture

An image retrieval system is a computing method which is used for browsing, searching and retrieving images from a database. The Fig 2.1 shows the Block diagram of the proposed Image Retrieval system using neural network. The proposed system consists of modules which is shown below

- Feature Extraction Module: This module is used to extract the features such as color moments, DCT of image which is given as input and outputs the visual features. This module is responsible for extraction of color features from the image.
- Database Module: The feature which is extracted from the image database is stored in this module.

ISSN: 2321-8169
2508 - 2512

- Matching and Ranking Module – To rank and retrieve the relevant images, the similarity measure is used. Dot product is used as similarity measure for ranking.
- User Module - An interactive user interface will be provided for user to give the query image and to display the result.

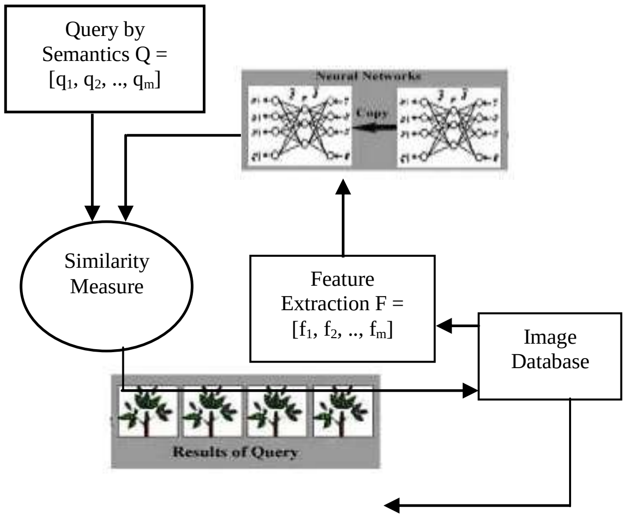

Fig 2.1: Block diagram of the proposed IR system using neural nets

B. Feature Extraction

The method of extracting visual features of images for indexing, ranking & retrieval is known as feature extraction. Feature extraction is a used to reduce dimensions. When the input data is too large, then the input data will be converted into a reduced set of features which is named as feature vector. Transforming the input data into the set of features is called feature extraction [6]. The visual features used in image retrieval which affects the system performance directly is denoted by F = [f1, f2. . . . fm]. The commonly used method to extract feature is color..

In the proposed IR system, the RGB color model is used to calculate the color moments and discrete cosine transform (DCT) coefficients of the images which is called as feature vectors. The DCT coefficients corresponding to lower frequencies provide more information to an image than higher frequencies.

1) Color Moments

In the proposed system, the feature extraction method used is color moments. Color moments can be defined as measures which are used to differentiate the images based on their features of color. These moments will be used for color similarity between images.

The three color moments used are mean, standard deviation and skewness. Moments are calculated for each of color channels in an image. Therefore, an image consists of 9 moments i.e., 3 moments for each 3 color channels.

Mean can be defined as the average pixel values in the image.

$$E_i = \sum_{N}^{j=1} \frac{1}{N} P_{ij} \qquad \text{--------- (2.1)}$$

Where P_{ij} is the color value of the i^{th} color component of the j^{th} image pixel and N is the total number of pixels in the image.

Standard Deviation: The standard deviation is calculated using the equation 2.2

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{N}^{j=1}\left(P_{ij} - E_i\right)^2} \qquad \text{----------- (2.2)}$$

Where E_i is the mean

Skewness can be calculated using the equation 2.3.

$$s_i = \sqrt{\frac{1}{N}\sum_{N}^{j=1}\left(P_{ij} - E_i\right)^3} \qquad \text{----------- (2.3)}$$

2) Discrete Cosine Transform

A discrete cosine transform (DCT) is a sequence of finite data points in terms of a sum of cosine functions oscillating at different frequencies [7]. DCT helps separate the image into spectral sub-bands of differing importance with respect to the image's visual quality. A signal or an image in spatial domain can be re-represented by a set of coefficients in frequency domain. Lower frequencies provide additional information to an image than higher frequencies. Therefore, the DCT coefficients corresponding to higher frequencies are rejected. The DCT coefficients can be classified into 7 groups which are shown in Fig.2.2. The DCT features are calculated by mean and variances of the DCT coefficients for three channels.

However, 6 features can be defined for each region i.e. 2 feature for each color channel. Totally there are 42 features for each image is used. The DCT co-efficient will be calculated using the equation 2.4

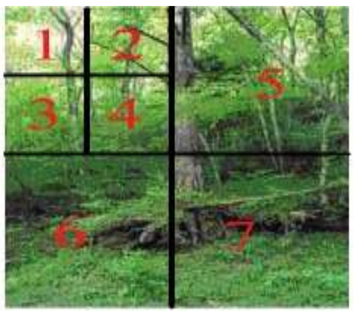

Fig 2.2: DCT Transformation

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \text{ - (2.4)}$$

C. Off-line Semantics

Semantics can be extracted manually from image, and named as semantic vector, S = [s_1, s_2, . . . s_m], to represent the presence or absence of semantics within the image, where *m* is the number of semantics predefined by domain experts, and s_i takes binary values [0 or 1]. Example, $S = [1; 0; 0]$ can be read as that the image contains semantics 1, but it does not contain semantics 2 and 3.

The semantic vector and feature vector can be defined as pair of both feature vectors F and semantic feature vector S i.e., $G = \{(F, S)\}$ which is stored in database. In the proposed system, the three layer architecture is used for neural network that is 51-30-4 is included in the IR system. The output layer classifies 4 classes corresponding to 4 semantics which is used in the system.

The visual feature of the image database is given as input to the neural network and has semantic labels as its output. The network is trained with 10,000 epochs without the momentum term and with a learning rate as 0.1 for the off-line learning using the back propagation learning algorithm [8, 9, and 10].

D. *Similarity Measure and Ranking*

The query is represented by an vector $Q = [q_1, q_2, \ldots q_m]$, where q_k takes binary values with 1(0) representing the presence (absence) of a semantics in the target images. During the retrieval process, the low-level features are given as input for neural network and the output will be [0 or 1] which is represented by $O = [o_1, o_2, \ldots o_m]$. The weighted dot product (DP) is used for similarity measure which is calculated using equation 2.5

$$D(Q, O) = \sum_{k=1}^{m} \alpha_x q_x O_x \qquad \text{----- (2.5)}$$

The top most images which are relevant to query image retrieved from the database.

III. RESULT AND ANALYSIS

In this chapter, the details of the experiment conducted and results obtained are presented.

A. *Experimental Setup*

The proposed IR system using neural network is implemented using MatLab. An artificial image database with nature scene images containing the semantic concepts Cloud, Tree, Yellow Flower and Red Flower is used. Table 3.1 shows some basic statistics of the database.

Table 3.1 Database Statistics I

Semantic	Number of Images
Cloud	62
Tree	62
Yellow Flower	62
Red Flower	62

In the database, there are 80 images containing one semantic, 36 images containing two semantic concept each (ex: (cloud & tree)), 24 images containing 3 semantics combination 6 images containing all four semantic concepts, which are shown in Table 3.2.

Table 3.2 Database Statistics II

Semantic Number	Number of Images
1 Semantic	80
2 Semantic	36
3 Semantic	24
4 Semantic	6

Table 3.3 Transformation Methods

Sl. No	Method	Parameters
1.	Resize	0.8
2.	Resize	1.2
3.	Rotation (Clockwise)	90°
4.	Rotation (Clockwise)	180°
5.	Rotation (Clockwise)	270°
6.	Salt-Pepper Noise	0.03
7.	Gaussian Noise	$\mu=0$, $\sigma = 0.06$
8.	Vertical Mirror	
9.	Horizontal Mirror	

For evaluation purpose, each image was transformed into 9 different images. The transformation detail is shown in Table 3.3. So totally there are 1,460 images in the database. A 5-fold cross validation scheme is used to evaluate the performance of the retrieval system. In each fold, the database is divided into 730 images in training dataset and 730 images in test dataset randomly.

Fig 3.1 shows the system performance for a specific query utilizing 1 semantic concept "Red Flower".

Fig 3.1 Image retrieval

B. *Performance Analysis*

The performance of image retrieval system is analyzed by two statistical measures: recall and precision. For the given query, the recall is defined as the fraction of relevant images returned by the system with respect to the total number of relevant images:

$$\mathrm{Recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|} \quad (3.1)$$

 ISSN: 2321-8169
2508 - 2512

A higher value of recall therefore indicates that the top-ranked hits more target images.

The precision is defined as the fraction of the relevant images returned by the system with respect to the size of the selection returned by the system.

$$Precision = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\{\text{retrieved documents}\}\}|} \quad (3.2)$$

Because of the database used in this study, the standard *recall* and *precision* calculation formulas cannot be directly applied to characterize the system performance. Therefore, modified *recall* calculation formula is used. The measure adopted in the performance evaluation is the so called mean rank μ, which is defined by

$$\mu = \frac{N(N+1)}{2\sum_{i=1}^{N} rank(i)} \quad (3.3)$$

Where rank (i) is the rank of the relevant image I (i.e., position of retrieved relevant image I in the retrieved images), which is an integer between 1 and 730 and N is the number of total relevant images in the test database.

1) Precision/Recall for training datasets

The results shown in this section for a certain number of semantic concepts are the average rates for all possible combination, for example, 2 S (2 semantic concepts) is the average performance for all 2 semantic concepts combination including <Cloud, tree>, <Cloud, Yellow flower>, <Cloud, Red Flower> etc.

Table 3.4: Recall performance for the training datasets (%)

Recall	1S	2S	3S	4S
Fold 1	99	96.25	89.5	80.6
Fold 2	98.5	94.66	77.98	72.55
Fold 3	95.2	93.05	83.4	52.33
Fold 4	97.76	96.61	87.66	85
Fold 5	97.7	95.5	84.1	68.3
Avg.	98.23	95.81	84.52	71.756

It is observed that the average recall and μ rates for queries with different semantics monotonically decrease with the increased number of semantics in the images which is shown in Fig 3.2 and 3.3.

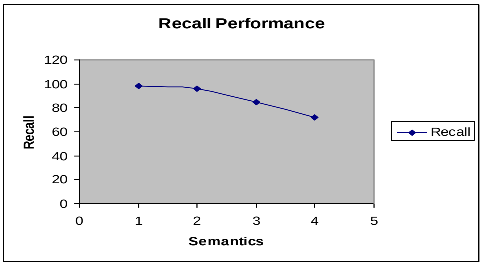

Fig 3.2 Recall performance

Table 3.5: μ performance for the training datasets

μ	1S	2S	3S	4S
Fold 1	0.99	0.967	0.9	0.844
Fold 2	0.97	0.966	0.860	0.899
Fold 3	0.98	0.954	0.875	0.631
Fold 4	0.97	0.95	0.90	0.94
Fold 5	0.98	0.95	0.897	0.84
Avg.	0.978	0.954	0.88	0.83

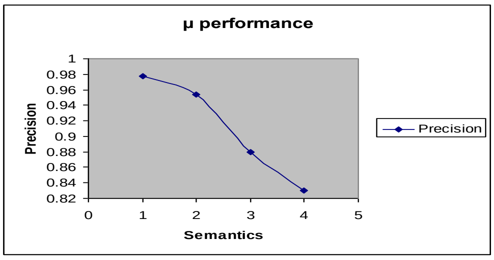

Fig 3.3 μ performance

2) Precision/Recall with different features

Three different feature sets i.e., the color moment (CM) features, DCT features and the mixed features are examined under the image retrieval system framework respectively. It is observed that the average recall and μ performances for the system using the mixed features is much better than that obtained by the separated ones which are shown in Fig 3.4 and 3.5.

Table 3.6: Recall performance with different features for training datasets (%)

Recall	1S	2S	3S	4S
Mixed	98.23	95.81	84.52	71.756
CM	74.25	48.50	39.74	12.46
DCT	84.25	70.194	51.68	13.99

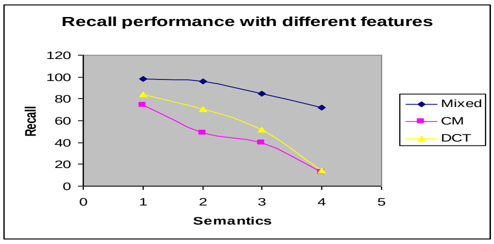

Fig 3.4 Recall performance with different features

Table 3.7: μ performance with different features for training datasets

μ	1S	2S	3S	4S
Mixed	0.978	0.954	0.88	0.83
CM	0.83	0.72	0.58	0.25
DCT	0.88	0.74	0.58	0.18

 ISSN: 2321-8169
2508 - 2512

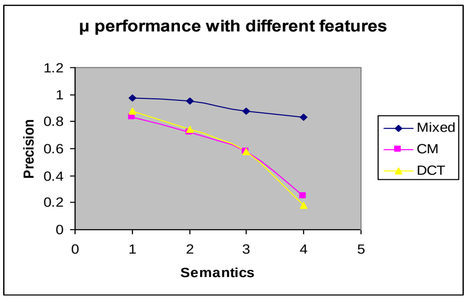

Fig 3.5 Recall performance with different features

IV. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This thesis explores the successful design and development of a CBIR system. To overcome the gap between image semantic content and its corresponding representation using low-level visual features, a hybrid scheme is adopted. Following are the important outcome of the developed IR system:

- The query is and characterized by a set of visual feature vector (color moments and discrete cosine transform coefficients) and user predefined semantic labels which has improved the performance of the system.
- A novel method of similarity measure called dot product is used to rank the images which increased the precession and recall rate of the system nearly 3 times.
- A combination of color moments and discrete cosine transform coefficients are used for extraction visual feature vector which improved the performance of the system.

B. Future Scope

In the present system when the number of semantics in the image increases more than two the performance of the system monotonically decreases. This is can be improved by adding the system with user's relevance feedback. The interactive relevance feedback with on-learning strategy could enhance the recall performance in the HIR system.

REFERENCES

[1] S. Flank, P. Martin, A. Balogh, and J. Rothey. PhotoFile. "A digital library for image retrieval". In Proc. of International Conference on Multimedia Computing and Systems, Washington, DC, IEEE Computer Society, vol. 1908, pp. 292-295. May 1995.

[2] V. N. Gudivada and V. V. Raghavan. "Content-based image retrieval systems. Computer", IEEE, vol. 28, pp. 18-22, September 1995.

[3] F. Crestani and G. Pasi. "Soft Computing in Information Retrieval: techniques and applications", Physica Verlag (Springer Verlag), vol.14, pp 393, 2000.

[4] A. Yoshitaka and T. Ichikawa. "A survey on content-based retrieval for multimedia databases", IEEE Trans. On Knowledge and Data Engineering, vol. 11, pp. 81-93, 1999.

[5] Nenad S. Kojic, Slobodan K. Cabarkapa. "Implementation of Neural Network in CBIR Systems with Relevance Feedback", Journal of Automatic Control, vol. 16, pp. 41-45, 2006

[6] Daniela Stan Raicu. "Image Feature Extraction", Visual Computing Workshop on Image Processing, pp 1-13, may 2004

[7] N. Ahmed. "Discrete Cosine Transform", Digital Signal Processing, vol. 1, pp. 4-5, 1991.

[8] White, "Artificial Neural networks: Approximation and Learning Theory", Oxford, UK: Blackwell, pp. 224-258, 1992.

[9] Hornik k, "Approximation capabilities of Multilayer feed forward networks", In Proceedings of the Neural Networks, Vol. 4, pp. 251-257, 1991.

[10] D. E. Rumelhart, G. E. Hinton and R. J. Willianms Learning representations of back-propagation errors, *Nature*, vol. 323, pp. 533–536, 1986