The problem of feature selection can be stated as follows: given the original set, *F*, of *n* features, find subset *S*, which consists of *m* features ($m < n$, $S \in F$), such that the classification accuracy is maximized.

In the first iteration of ACO, each ant will randomly choose a feature subset of *m* features. Only the best *k* subsets, $k < na$, will be used to update the pheromone trial and influence the feature subsets of the next iteration. In the second and following iterations, each ant will start with *m - p* features that are randomly chosen from the previously selected k-best subsets, where *p* is an integer that ranges between 1 and *m* - 1. In this way, the features that constitute the best *k* subsets will have more chance to be present in the subsets of the next iteration. However, it will still be possible for each ant to consider other features as well. For a given ant *j*, those features are the ones that achieve the best compromise between pheromone trails and local importance with respect to *Sj*, where *Sj* is the subset that consists of the features that have already been selected by ant *j*. The Updated Selection Measure *(USM)* is used for this purpose is defined as:

$$USM_i^{s_j} = \begin{cases} \dfrac{(\tau_i)^{\eta}\left(LI_i^{s_j}\right)^{\alpha}}{\sum_{g \notin s_j}(\tau_g)^{\eta}\left(LI_g^{s_j}\right)^{\alpha}} & \text{If } i \notin s_j \\ 0 \end{cases}$$

where, $LI_i^{s_j}$ is the local importance of feature f_i given the subset S_j. The parameters η and α control the effect of pheromone trail intensity and local feature importance respectively. $LI_i^{s_j}$ is measured using the MIEF measure and defined as:

$$LI_i^{s_j} = I(C, f_i)\times\left[\frac{2}{1+\exp\left(-\alpha D_i^{s_j}\right)}-1\right]$$

Where

$$D_i^{s_j} = \min\left[\frac{H(f_i)-I(f_i,f_s)}{H(f_i)}\right]\times\frac{1}{|s_j|}\sum_{f_s\in s_j}\left[\beta\left(\frac{I(C,\{f_i,f_s\})}{I(C,f_i)+I(C,f_s)}\right)^{\gamma}\right]$$

The parameters α, β, γ are constants, $H(f)$ is the entropy of f_i, $I(f_i, f_s)$ is the mutual information between f_i and f_s, $I(C,f_i)$ is the mutual information between the class labels and f_i *and* $|s_j|$ is the cardinal of s_j

5. IMAGE RETRIEVAL

After the optimized features are calculated the distance between query image features and data set images is calculated using manhattan distance. Manhattan distance is given by

$$D = \sum_{i=1}^{n} |X_i - Y_i|$$

The minimum distance value signifies an exact match with the query. The fact that the distances in each dimension are modulated before summation, places great emphasis on those features for which the dissimilarity is large. The images for which distance is less are retrieved and shown at output.

6. RELEVANCE FEEDBACK

Human perception of image similarity is subjective, semantic, and task-dependent. Although content-based methods provide promising directions for image retrieval, generally, the retrieval results based on the similarities of pure visual features are not necessarily perceptually and semantically meaningful. In addition, each type of visual feature tends to capture only one aspect of image property and it is usually hard for a user to specify clearly how different aspects are combined. To address these problems, interactive *relevance feedback*, a technique in traditional text-based information retrieval systems, was introduced [9]. With relevance feedback, it is possible to establish the link between high-level concepts and low-level features. Relevance feedback is a supervised active learning technique used to improve the effectiveness of information systems. The main idea is to use positive and negative examples from the user to improve system performance [10] [11] [12]. For a given query, the system first retrieves a list of ranked images according to a predefined similarity metrics. Then, the user marks the retrieved images as relevant (positive examples) to the query or not relevant (negative examples). The system will refine the retrieval results based on the feedback and present a new list of images to the user. Hence, the key issue in relevance feedback is how to incorporate positive and negative examples to refine the query and/or to adjust the similarity measure. By the positive and negative examples given by the user, using support vector machine the images of data base are classified into positive and negative set. Then the images are retrieved from the positive set by calculating similarity measures and the images with high similarity are retrieved.

7. EXPERIMENTAL RESULTS

The proposed method is tested on COREL dataset, which contain 1,000 images. The visual features extracted form images include color (color

ISSN: 2348-4098
VOLUME 02 ISSUE 06 JULY 2014

coherence vector), texture (discrete cosine transform coefficients) and shape (edge histogram descriptor). An image is represented using 16 features (8 colors are considered) using color coherence vector. Coefficients of DCT of the image which are calculated as mentioned above give 48 features. Edge histogram descriptor are calculated by dividing the image into 4 sub images and for each plane (R, G, B) of sub image, edge densities are calculated along vertical, horizontal, 45 degree, 135 degree, non directional. So 80 features are obtained using edge density histogram. So totally 144(16+48+80) features are used for representing an image. After calculating the 144 features for all images in the database, only 15 features from 144 features are selected using ant colony optimization technique.

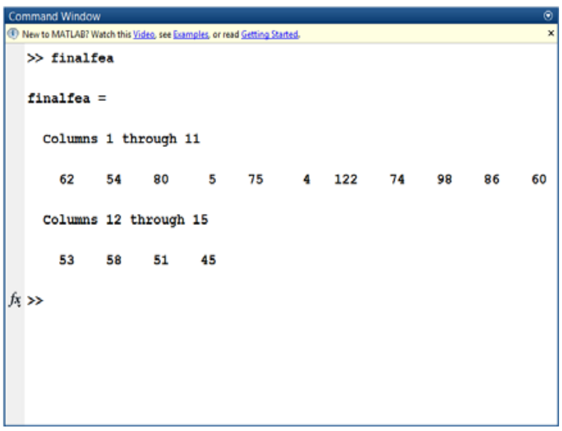

Figure 3: Selected feature set selected using ACO

Therefore by using ACO optimization technique the features numbered [62,54,80,5,75,4,122,74,98,86,60,53,58,51,45] are selected from 144 features. For the input query image the features are calculated and the features numbers obtained from optimization technique are selected. Distances between query image and data base images are calculated using manhattan distance and the images which have less distance are retrieved.

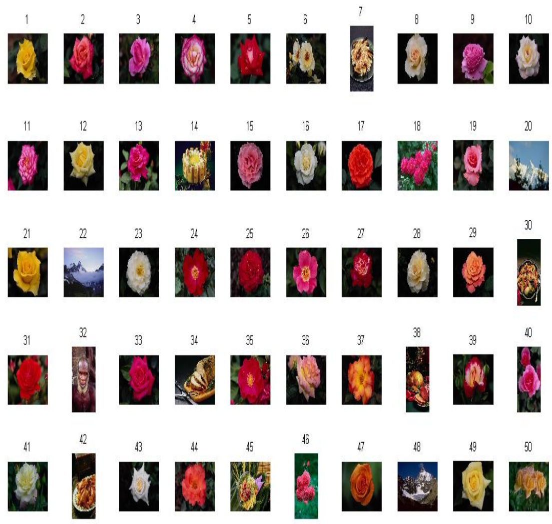

Figure 4: Retrieved images using similarity measurement

Wrongly retrieved = [20 22 30 32 34 38 42 46 48]

Wrongly retrieved image numbers are given as input by the user as a feedback to the system. Then by changing the weights of the features using SVM, retrieval is done by calculating the distances again and the retrieved are displayed.

ISSN: 2348-4098
VOLUME 02 ISSUE 06 JULY 2014

Figure 5: Retrieved images after feedback from the user

From figures 4 and 5, it is clear that the number of relevant images retrieved has been increased and also the average time taken for retrieval system has decreased from 4.08 seconds to 3.4 seconds. So the precision and recall time has improved by the proposed system.

8. CONCLUSION

Feature selection technique for extracting the most relevant features and dropping the irrelevant was done to improve system efficiency in terms of both Retrieval Efficiency, Retrieval Time and by using relevance feedback technique, semantically more meaningful retrieval results were obtained, so that retrieval efficiency is improved.

ACKNOWLEDGEMENT

I would like to express my deep sense of indebtedness and sincere gratitude to G. Prathibha, Assistant professor, ECE Department , Acharya Nagarjuna University college of Engineering & Technology, AP, India for her valuable guidance and appreciation throughout the work.

REFERENCES

[1] “Fundamentals of Content-Based Image Retrieval” by Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng.

[2] J. Huang, et al., "Image indexing using color correlogram," IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 762-768, Puerto Rico, June 1997.

[3] "CBIR Using DCT for Feature Vector Generation" by Vibha Bhandari, Sandeep B.Patil in "International Journal of Application or Innovation in Engineering & Management (IJAIEM)".

[4] "Comparison of Color Features for Image Retrieval" by S.R. Kodituwakku in "Indian Journal of Computer Science and Engineering".

[5] “Content-Based Image Retrieval (CBIR) System Based on the Clustering

ISSN: 2348-4098
VOLUME 02 ISSUE 06 JULY 2014

and Genetic Algorithm" by Eng. Ahmed K. Mikhraq.

[6] D. Park, Y. Jeon and C. Won. "Efficient Use of Local Edge Histogram Descriptor," Int. J. of Computer Vision, vol.7-1, pp. 11-32, 2003.

[7] P. Zhang and H. Zhu, "Medical Image Retrieval Based on Co-Occurrence Matrix and Edge Histogram," Multimedia Technology (ICMT), International Conference, pp.5434-5437, 26-28, July 2011.

[8] "Feature Subset Selection Using Ant Colony Optimization" by Ahmed AI-Ani in International journal of computational intelligence volume 2 number i 20051ssn:1304-2386.

[9] "Relevance Feedback in Content Based Image Retrieval" a Review by Pushpa B. Patil, Manesh B. Kokare.

[10] "Incorporate Support Vector Machines To Content-based Image Retrieval with Relevant Feedback" by Pengyu Hong, Qi Tian, Thomas S. Huang.

[11] "Performance Enhanced Optimization based Image Retrieval System" by Tessy Annie Varghese in *IJCA Special Issue on "Evolutionary Computation for Optimization Techniques" ECOT, 2010.*

[12] Xiangyang Wang, Jie Yang, Xiaolong Teng, Weijun Xia, Richard Jensen,2007. "Feature Selection based on Rough sets and Particle Swarm Optimization", Pattern Recognition Letters 28.

BIOGRAPHY

	C.V. Rashmi received B. Tech degree in electronics and communication engineering from vignan's lara institute of technology and science, Guntur in 2012. She is currently pursuing M. tech in communication and signal processing in university college of engineering and technology, Acharya Nagarjuna university.
	Er. G. Prathibha is working as an Assistant professor in university college of engineering and technology, Acharya Nagarjuna university. She received her B.Tech (ECE) degree from R.V.R & J.C college of engineering, Guntur, M. Tech (system and signal processing) from JNTU, Hyderabad. Her areas of interest are medical image processing, video processing and pattern recognition.