Performance Enhanced Optimization based Image Retrieval System

Tessy Annie Varghese
Karunya University
Coimbatore, India

ABSTRACT

Image retrieval is a system for browsing, searching the query image and retrieving similar images from large databases. A wide variety of features can be used for image retrieval. This process selects a subset of relevant features from a group of features of the image. It also helps to acquire better understanding about the image by describing which the important features are. The accuracy can be improved by increasing the number of features selected. But this increases the complexity of the retrieval system. The performance can be improved by removing the irrelevant and redundant features from taking into consideration. This is known as optimization. Many optimization techniques can be used. Ant Colony Optimization (ACO) is the technique proposed in this paper. With ACO, the image features are selected and images are retrieved from databases with high accuracy.

Keywords

Optimization, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO).

1. INTRODUCTION

With the rapid developments in the areas of information technology and the use of the Web services there has been an increase in the growth of the digital media and image collections. To gain the full potential of the existing technologies, there should be effective mechanisms to search large image collections. The applications of image retrieval system include areas such as medical imaging, security and criminal investigation, copyright protection, and computer aided design. For many decades, optimization has been an active area of interest research. Many optimization algorithms have been developed. Before the optimization is done, the features from the image have to be extracted. Several textural features are used for image analysis and these features are available in [1].These features are basically second order statistical features. The mathematical formulas for these features are available in [2].

In this paper, the texture features are used for image retrieval. From the co-occurrence matrix (CCM) we are extracting a number of features. CCM calculates the probability of occurrence of the same pixel value between each pixel and its neighbors [3]. From the extracted features a subset of the features has to be selected.

Many algorithms have been used for the selection of the features in different researches. Genetic Algorithm (GA) is used as the optimization technique for the classification of abnormal retinal images in [4]. GA was implemented along with neural network for pattern recognition and to determine the object orientation in [5]. The result showed good optimization by reducing the number of hidden nodes and the time required to train the neural network. Particle Swarm Optimization (PSO) is used develop a modified counter propagation neural network for abnormal MR brain image classification in [6]. An enhancement of the PSO known as chaotic PSO is used for image classification in [7]. The results showed a considerable increase in the image classification accuracy with the proposed algorithm. Apart from the image processing applications these algorithms can also be used in other areas such as neural networks. A comparison of the various algorithms is done in [8]. A feature selection strategy based on rough sets and PSO was proposed in [9]. Experiments carried out using UCI data compared the proposed algorithm with GA-based approach and other deterministic rough set reduction algorithms. The results showed that PSO is efficient for rough set-based feature selection.

In this paper we use ant colony optimization (ACO) as the technique for feature selection which is inspired by the behavior of ants and has numerous applications in various fields such as image processing, networking and neural networks. ACO was used for image edge detection in [10]. The proposed edge detection method takes advantage of the improvements introduced in ant colony system. With suitable parameter values, the algorithm was able to successfully identify edges in the canonical test images. In [11] parallel ACO is used for the segmentation of MR brain tumor. The proposed method has advantage that it can effectively segment the fine details of the image. It has higher accuracy as compared to the existing algorithms. ACO is used for remote sensing image classification in [12]. ACO is implemented in this work as an optimization technique using the Davies-Bouldin (DB) index and Euclidean Distance (ED) as functions to be minimized. This yielded good results with an increase in the overall accuracy of 4.312% and an improvement in Kappa coefficient of 0.0496.

2. BLOCK DIAGRAM

From the query image the features are extracted. The features are autocorrelation, contrast, cross correlation, cluster prominence, cluster shade, energy, entropy and homogeneity. From these features we select only a subset of the features that are useful for the image recognition. This is done using ant colony optimization. Then using the Euclidean Distance measure images are retrieved from the database.
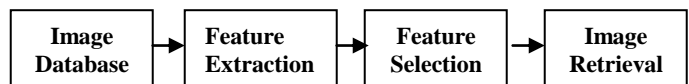
| Image Database | → | Feature Extraction | → | Feature Selection | → | Image Retrieval |

Figure 1. Block Diagram of the image retrieval system

3. FEATURE EXTRACTION

3.1 Co-occurrence Matrix (CCM)

This paper proposes a co-occurrence matrix which is a tabulation of how often different combinations of pixel values occur in an image. The co-occurrence matrices are very large; therefore they can be used to get more useful features of the image. These features generated are known as Haralick features. The features include contrast, correlation, energy and homogeneity. A CCM $P_d[i,j]$ is defined by first specifying a displacement vector **d**=(dx,dy) and counting all pairs of pixels separated by **d** having gray levels i and j.

A co-occurrence matrix **C** is defined over an **n x m** image **I** as:

$$C_{\Delta x, \Delta y}(i,j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, if\ I(p,q) = i\ and\ I(p+\Delta x, q+\Delta y) = j \\ 0, \quad otherwise \end{cases} \quad (1)$$

where $\Delta x, \Delta y$ are offset values.

From the CCM we are extracting the image features. These include autocorrelation, contrast, cross correlation, cluster prominence, cluster shade, energy, entropy and homogeneity.

a. Auto Correlation- For an M x N image, the autocorrelation function, S(x,y), takes the form:

$$\sum_{i=1}^{M-x} \sum_{j=1}^{N-y} \frac{I(i,j) \times I(i+x, j+y)}{(M-x) \times (N-y)} \quad (2)$$

where I(x,y) has a value of 1 if the pixel located at (x,y) meets the user criteria (e.g., contains the phase of interest) and 0 otherwise.

b. RMS Contrast-

$$\sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \bar{I})^2} \quad (3)$$

where intensities I_{ij} are the *i*-th *j*-th element of the two dimensional image of size *M* by *N*. $\bar{I}$ is the average intensity of all pixel values in the image. The image *I* is assumed to have its pixel intensities normalized in the range [0,1].

c. Cross Correlation-

$$1 - \frac{\sum_i \sum_j [F(x_i, y_j) - \bar{F}][G(x_i^*, y_j^*) - \bar{G}]}{\sqrt{\sum_i \sum_j [F(x_i, y_j) - \bar{F}]^2 \sum_i \sum_j [G(x_i^*, y_j^*) - \bar{G}]^2}} \quad (4)$$

where F(x$_i$,y$_j$) is the pixel intensity or the gray scale value at a point (x$_i$,y$_j$) in the undeformed image. G(x$_{i*}$,y$_{j*}$) is the gray scale value at a point (x$_{i*}$,y$_{j*}$) in the deformed image. $\bar{F}$ and $\bar{G}$ are mean values of the intensity matrices F and G, respectively.

d. Contrast-

$$\sum_i \sum_j (i-j)^2 C(i,j) \quad (5)$$

e. Cluster Shade-

$$\sum_i \sum_j (i - M_x + j - M_y)^3 - C(i,j) \quad (6)$$

f. Cluster Prominence-

$$\sum_i \sum_j (i - M_x + j - M_y)^4 - C(i,j) \quad (7)$$

g. Homogeneity-

$$\sum_i \sum_j \frac{1}{1 - (i-j)^2} C(i,j) \quad (8)$$

h. Energy-

$$\sum_i \sum_j C^2(i,j) \quad (9)$$

where $M_x = \sum_i \sum_j iC(i,j)$, $M_y = \sum_i \sum_j jC(i,j)$ and

$C(i,j)$ is the (i,j)th element of the co-occurrence matrix.

i. Entropy-

$$\sum_{i=0}^{M-1} P_i \log_2 \left(\frac{1}{P_i}\right) \quad (10)$$

4. FEATURE SELECTION

Feature selection is important because all the features are not helpful in image retrieval systems. Some features may act as interfering elements and will decrease the success rate of the retrieval system. The main aim of feature selection is to select a set of best features from a large number of features. It maximizes the accuracy and also simplifies the procedure of retrieval. This paper uses ant colony optimization (ACO) as the technique for feature selection. ACO is a population based metaheuristic which is inspired by the behavior of ants. It is used to find solutions for various optimization problems. In ACO, a colony of ants cooperates to look for solutions for the problem. Artificial ants incrementally build a solution by adding components to a partial solution under construction.

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. The feature selection problem representation exploited by the artificial ants includes the following:

* *n* features that constitute the original set, $F = \{f_1, \ldots, f_n\}$.
* A number of artificial ants to search through the feature space (*na* ants).
* T_i, the intensity of pheromone trail associated with feature f_i.
* For each ant *j*, a list that contains the selected feature subset, $S_j = \{s_1, \ldots, s_m\}$.

4.1 ACO Algorithm

The steps of the ACO algorithm are as follows:

1. Initialization:
- Set $T_i = cc$ and $\Delta T_i = 0$, $(i = 1, …, n)$, where cc is a constant and ΔT_i is the amount of change of pheromone trial quantity for feature f_i.
- Define the maximum number of iterations.
- Define k, where the k-best subsets will influence the subsets of the next iteration.
- Define p, where $m − p$ is the number of features each ant will start with in the second and following iterations.

2. If in the first iteration,
- For $j = 1$ to na,
  - Randomly assign a subset of m features to S_j.
- Goto step 4.

3. Select the remaining p features for each ant:
- For $mm = m − p + 1$ to m,
  - For $j = 1$ to na,
    - Given subset Sj, Choose feature fi that maximizes $USM_i^{Sj.}$
    - $S_j = S_j \cup \{f_i\}$.
- Replace the duplicated subsets, if any, with randomly chosen subsets.

4. Evaluate the selected subset of each ant using a chosen classification algorithm:
- For $j = 1$ to na,
  - Estimate the Mean Square Error (MSE_j) of the classification results obtained by classifying the features of S_j.
- Sort the subsets according to their MSE. Update the minimum MSE (if achieved by any ant), and store the corresponding subset of features.

5. Using the feature subsets of the best k ant:
- For $j = 1$ to k, /* update the pheromone trails */

$$\Delta T_i = \begin{cases} \dfrac{\max\limits_{g=1:k}(MSE_g) - MSE_j}{\max\limits_{h=1:k}\left[\max\limits_{g=1:k}(MSE_g - MSE_h)\right]} & if\ f_i \in S_j \\ 0 & otherwise \end{cases} \quad (11)$$

$$T_i = \rho . T_i + \Delta T_i \quad (12)$$

where ρ is a constant such that $(1 - \rho)$ represents the evaporation of pheromone trails.
- $\square$ For $j = 1$ to na,
  - Randomly produce $m − p$ feature subset for ant j, to be used in the next iteration, and store it in S_j.

6. If the number of iterations is less than the maximum number of iterations, goto step 3.

5. EXPERIMENTAL RESULTS

This paper adopts the texture features of the image for the image retrieval. To analyze the effect of the proposed method, experiments were conducted on image databases having more than 100 images. The performance of the proposed method was found to be superior when compared to other methods. Table 1 shows the list of features selected using ACO.

Table.1 Features selected using ACO

Features extracted	Features selected
Autocorrelation, Contrast, Cross correlation, Cluster prominence, Cluster shade, Energy, Entropy and Homogeneity	Autocorrelation, Contrast, Cluster prominence, Energy, Homogeneity

A large number of features are extracted from the image. This increases the complexity of the system. So using ACO we are extracting only the features that are useful in retrieving images from the database. This simplifies the system and increases its accuracy.

Table.2 Performance of the system

No. of images in the database	No. of images retrieved	Accuracy
120	110	91.6%

The experiments were conducted on a database of 120 images. An accuracy of 91.6% is obtained which is sufficiently higher for an image retrieval system.

6. REFERENCES

[1] Haarlick.R.M,1979. "Statistical and structural approaches to texture", IEEE Transactions on system, man and cybematics 67, 786-804

[2] D J Hemanth, D Selvathi & J Anitha ,2010 "Artificial Intelligence Techniques for Medical Image Analysis: Basics, Methods, Applications", VDM –Verlag, ISBN-9783639248258

[3] Chuen-Horng Lin , Rong-Tai Chen , Yung-Kuan Chan, 2009. "A smart content-based image retrieval system based on color and texture feature", Image and Vision Computing 27pp658–665

[4] J.Anitha, C. Kezi Selva Vijila & D.Jude Hemanth, 2010. "A hybrid genetic algorithm based fuzzy approach for abnormal retinal image classification", International Journal of Cognitive Informatics and Natural Intelligence", Vol.4, No.3, pp 29-43,

[5] Yas Abbas Alsultanny, Musbah M. Aqel,2003. "Pattern Recognition using Multilayer Neural-Genetic Algorithm", Neurocomputing 51, pp237-247

[6] D.Jude Hemanth, C. Kezi Selva Vijila & J.Anitha, 2010. "Performance improved PSO based modified counter propagation neural network for abnormal MR brain image classification", Int. J. Advance. Soft Comput. Appl., Vol.2, No.1, March.

[7] Krishna Chandramouli, Ebroul Izquierdo, 2006. "Image Classification using Chaotic Particle Swarm Optimization ", IEEE Transactions on Image Processing pp3001-3004

[8] Auralia I. Edwards, Andries P. Engelbrecht, 2005. "Comparing Different Optimization Strategies for Nonlinear Mapping and Mapping Large Datasets using Neural Networks", World Congress on Evolutionary Computation, pp306-313

[9] 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, pp 459-471

[10] Anna Veronica Baterina, Carlos Oppus, 2010. "Image Edge Detection Using Ant Colony Optimization", International Journal of Circuits, Systems and Signal Processing, Issue 2, Vol.4,

[11] J.Jaya, K. Thanushkodi,2010. "Segmentation of MR Brain Tumor Using Parallel ACO", International Journal of Computer and Network Security, Vol.2, No.6, pp150-153

[12] Fizazi Hadria, Hannane Amir Mokhtar, "Remote Sensing Image Classification using Ant Colony Optimization"