**Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2007)**

# Face Recognition System Using Ant Colony Optimization-Based Selected Features

Hamidreza Rashidy Kanan[1], Karim Faez[1] and Mehdi Hosseinzadeh[2]
[1]Image Processing and Pattern Recognition Lab., Electrical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Hafez Avenue, Tehran, Iran, 15914
[2]Department of Computer Engineering, Faculty of Engineering, University of Isfahan, Hezar Jerib Ave., Isfahan, Iran, 81744
{rashidykanan, kfaez}@aut.ac.ir
hosseinzadeh@comp.ui.ac.ir

*Abstract*-**Feature Selection (FS) is a most important step which can affect the performance of pattern recognition system. This paper presents a novel feature selection method that is based on Ant Colony Optimization (ACO). ACO algorithm is inspired of ant's social behavior in their search for the shortest paths to food sources. In the proposed algorithm, classifier performance and the length of selected feature vector are adopted as heuristic information for ACO. So, we can select the optimal feature subset without the priori knowledge of features. Simulation results on face recognition system and ORL database show the superiority of the proposed algorithm.**

## 1. INTRODUCTION

Feature Selection (FS) is extensive and spread across many fields, including document classification, data mining, object recognition, biometrics, remote sensing and computer vision [1]. Given a feature set of size ***n***, the FS problem is to find a minimal feature subset of size ***m*** (m<n) while retaining a suitably high accuracy in representing the original features. In real word problems FS is a must due to the abundance of noisy, irrelevant or misleading features [2].

Usually FS algorithms involve heuristic or random search strategies in an attempt to avoid this prohibitive complexity. However, the degree of optimally of the final feature subset is often reduced [2].

Among too many methods which are proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method and Ant Colony Optimization (ACO)-based method have attracted a lot of attention. These methods attempt to achieve better solutions by using knowledge from previous iterations.

Metaheuristic optimization algorithm based on ant's behavior (ACO) was represented in the early 1990s by M. Dorigo and colleagues [3]. ACO is a branch of newly developed form of artificial intelligence called Swarm Intelligence. Swarm intelligence is a field which studies "the emergent collective intelligence of groups of simple agents" [4]. In groups of insects which live in colonies, such as ants and bees, an individual can only do simple task on its own, while the colony's cooperative work is the main reason determining the intelligent behavior it shows [5].

ACO algorithm is inspired of ant's social behavior. Ants have no sight and are capable of finding the shortest route between a food source and their nest by chemical materials called pheromone that they leave when moving.

ACO algorithm was firstly used in solving Traveling Salesman Problem (TSP) [6]. Then has been successfully applied to a large number of difficult problems like the Quadratic Assignment Problem (QAP) [7], routing in telecommunication networks, graph coloring problems, scheduling and etc. This method is particularly attractive for feature selection as there seems to be no heuristic that can guide search to the optimal minimal subset every time [2]. In the other hand, if features are represented as a graph, ant will discover best feature combinations as they traverse the graph.

In this paper a new modified ACO-Based feature selection algorithm has been introduced. The classifier performance and the length of selected feature vector are adopted as heuristic information for ACO. So proposed algorithm needs no priori knowledge of features. Proposed algorithm is applied to Discrete Wavelet Transform (DWT) Coefficients in the application of face recognition system and finally the classifier performance and the length of selected feature vector are considered for performance evaluation.

The rest of this paper is organized as follows. Section 2 presents a brief overview of feature selection methods. Ant Colony Optimization (ACO) is described in section 3. Section 4 explains the proposed feature selection algorithm and finally, section 5 and 6 attain the experimental results and conclusion.

## 2. AN OVERVIEW OF FEATURE SELECTION (FS) APPROACHES

Feature selection algorithms can be classified into two categories based on their evaluation procedure [2, 8]. If an algorithm performs FS independently of any learning algorithm (i.e. it is a completely separate preprocessor), then it is a filter approach (open-loop approach). This approach is based mostly on selecting features using between-class separability criterion [8]. If the evaluation procedure is tied to the task (e.g. classification) of the learning algorithm, the FS algorithm employs the wrapper approach (closed-loop approach). This method searches through the feature subset

**1-4244-0700-1/07/$20.00 ©2007 IEEE**

space using the estimated accuracy from an induction algorithm as a measure of subset suitability.

The two mentioned approaches are also classified into five main methods which they are Forward Selection, Backward elimination Forward/Backward Combination, Random Choice and Instance based method.

FS methods may start with no features, all features, a selected feature set or some random feature subset. Those methods that start with an initial subset usually select these features heuristically beforehand. Features are added (Forward Selection) or removed (Backward Elimination) iteratively and in the Forward/Backward Combination method features are either iteratively added or removed or produced randomly thereafter.

The disadvantage of Forward Selection and Backward Elimination methods is that the features that were once selected/eliminated cannot be later discarded/re-selected. To overcome this problem, Pudil et al. [9] proposed a method to flexibly add and remove features. This method has been called floating search method.

In the wrapper approach the evaluation function calculates the suitability of a feature subset produced by the generation procedure and compares this with the previous best candidate, replacing it if found to be better. A Stopping criterion is tested every iteration to determine whether the FS process should continue or not.

Other famous FS approaches are based on the Genetic Algorithm (GA) [10], Simulated Annealing [2] and Ant Colony Optimization (ACO) [2, 5].

## 3. ANT COLONY OPTIMIZATION (ACO)

In the early 1990s, ant colony optimization (ACO) was introduced by M. Dorigo and colleagues as a novel nature-inspired metaheuristic for the solution of hard combinatorial optimization (CO) problems [11].

The ability of real ants to find shortest routes is mainly due to their depositing of pheromone as they travel; each ant probabilistically prefers to follow a direction rich in this chemical. The pheromone decays over time, resulting in much less pheromone on less popular paths. Given that over time the shortest route will have the higher rate of ant traversal, this path will be reinforced and the others diminished until all ants follow the same, shortest path (the "system" has converged to a single solution) [2].

In general, an ACO algorithm can be applied to any combinatorial problem as far as it is possible to define:

- Appropriate problem representation. The problem must be described as a graph with a set of nodes and edges between nodes.
- Heuristic desirability ($\eta$) of edges. A suitable heuristic measure of the "goodness" of paths from one node to every other connected node in the graph.
- Construction of feasible solutions. A mechanism must be in place whereby possible solutions are efficiently created.
- Pheromone updating rule. A suitable method of updating the pheromone levels on edges is required with a corresponding evaporation rule. Typical methods involve selecting the n best ants and updating the paths they chose.
- Probabilistic transition rule. The rule that determines the probability of an ant traversing from one node in the graph to the next.

### 3.1 ACO for Feature Selection

The feature selection task may be reformulated into an ACO-suitable problem. ACO requires a problem to be represented as a graph. Here nodes represent features, with the edges between them denoting the choice of the next feature. The search for the optimal feature subset is then an ant traversal through the graph where a minimum number of nodes are visited that satisfies the traversal stopping criterion. Figure 1 illustrates this setup. The ant is currently at node ***a*** and has a choice of which feature to add next to its path (dotted lines). It chooses feature ***b*** next based on the transition rule, then ***c*** and then ***d***. Upon arrival at ***d***, the current subset *{a; b; c; d}* is determined to satisfy the traversal stopping criterion (e.g. a suitably high classification accuracy has been achieved with this subset). The ant terminates its traversal and outputs this feature subset as a candidate for data reduction.

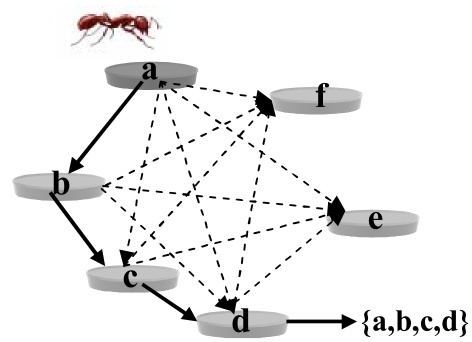


Fig. 1 ACO problem representation for FS

A suitable heuristic desirability of traversing between features could be any subset evaluation function - for example, an entropy-based measure [12], rough set dependency measure [13] or the Fisher Discrimination Rate (FDR)[14]. The heuristic desirability of traversal and edge pheromone levels are combined to form the so-called probabilistic transition rule, denoting the probability of an ant at feature **i** choosing to travel to feature **j** at time **t**:

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}.[\eta_{ij}]^{\beta}}{\sum_{l \in J_i^k} [\tau_{il}(t)]^{\alpha}.[\eta_{il}]^{\beta}} & if\ \ j \in J_i^k \\ 0 & otherwise \end{cases} \quad (1)$$

Where ***k*** is the number of ants, $\eta_{ij}$ is the heuristic desirability of choosing feature **j** when at feature **i** ($\eta_{ij}$ is optional but often needed for achieving a high algorithm performance [11]), $J_i^k$ is the set of neighbor nodes of node **i** which have not yet been visited by the ant **k.** $\alpha \succ 0, \beta \succ 0$ are two parameters that determine the relative importance of the pheromone value and heuristic information (the choice of $\alpha, \beta$ is determined experimentally) and $\tau_{ij}(t)$ is the amount of virtual pheromone on edge (i,j).

The overall process of ACO feature selection can be seen in figure 2. The process begins by generating a number of ants, k, which are then placed randomly on the graph (i.e. each ant starts with one random feature). Alternatively, the number of ants to place on the graph may be set equal to the number of features within the data; each ant starts path construction at a different feature. From these initial positions, they traverse edges probabilistically until a traversal stopping criterion is satisfied. The resulting subsets are gathered and then evaluated. If an optimal subset has been found or the algorithm has executed a certain number of times, then the process halts and outputs the best feature subset encountered. If neither condition holds, then the pheromone is updated, a new set of ants are created and the process iterates once more.

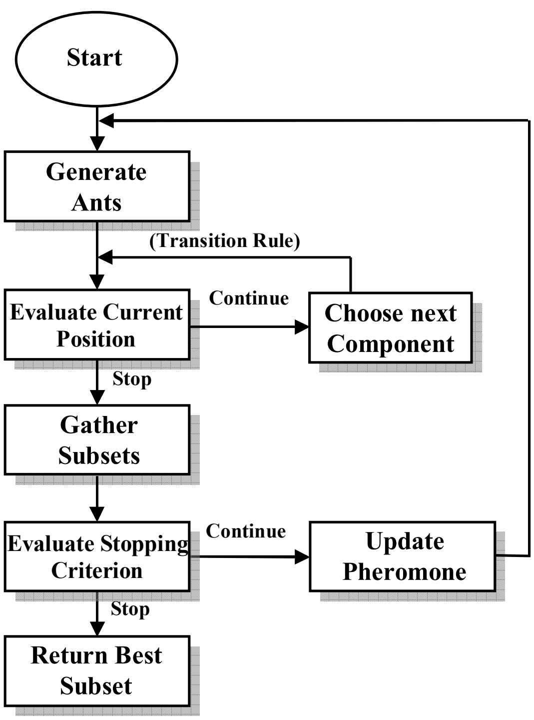


Fig. 2 ACO-based feature selection overview

The pheromone on each edge is updated according to the following formula:

$$\tau_{ij}(t+1) = (1-\rho).\tau_{ij}(t) + \rho.\Delta\tau_{ij}(t) \qquad (2)$$

Where:

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{n} (\gamma'(S^k) / |S^k|) \qquad (3)$$

This is the case if the edge (i;j) has been traversed; $\Delta\tau_{ij}(t)$ is 0 otherwise. The value $0 \le \rho \le 1$ is decay constant used to simulate the evaporation of the pheromone, $S^k$ is the feature subset found by ant **k**. The pheromone is updated according to both the measure of the "goodness" of the ant's feature subset $\gamma'$ and the size of the subset itself. By this definition, all ants update the pheromone.

As a mentioned earlier, the first ACO algorithm, Ant System (AS), was developed by professor Dorigo in 1992 [11]. This algorithm was introduced using the TSP as an example application. Then gradually AS is extended for improving the performance of optimization algorithm.

These extensions include Elitist AS (EAS) [6], Rank-Based AS ( $AS_{rank}$ ) [15] and Max-Min AS (MMAS) [16]. The main differences between AS and these extensions are the way the pheromone update is performed, as well as some additional details in the management of the pheromone trials.

## 4. PROPOSED FEATURE SELECTION ALGORITHM

The main steps of proposed algorithm are as follows:

1) Initialization
    - Determine the population of ants (p).
    - Set the intensity of pheromone trial associated with any feature.
    - Determine the maximum of allowed iterations (k)
2) Generation ants and evaluation of each ants
    - Any ant ( $A_i$ , i=1:p) randomly is assigned to one feature and it should visit all features and build solutions completely. In this step, the evaluation criterion is Mean Square Error (MSE) of the classifier. If any ant could not decrease the MSE of the classifier in three successive steps, it finished its work and exit.
3) Evaluation of the selected subset of each ant
    - In this step the importance of the selected subset of each ant is evaluated through classifier performance. Then the subsets according to their MSE are sorted and some of them are selected according to ACS and $AS_{rank}$ algorithms.
4) Check the stop criterion
    - If the number of iterations is more than the maximum allowed iteration exit, otherwise continue.
5) Pheromone updating
    - For features which are selected in the step 3 pheromone intensity are updated.
6) Go to 2 and continue

## 5. EXPERIMENTAL RESULTS

For experimental studies we have considered ORL gray scale face image database. This database contains 400 facial images from 40 individuals in different states. So, the number of classes in our experiments is 40. The total number of images in each class is 10.

Figure 3 shows some samples images of this database.

Fig. 3 Some images of ORL database

Discrete Wavelet Transform (DWT) Coefficient features have been extracted from each face image. Then proposed ACO-based feature selection methods is applied to feature set and finally, the length of selected feature vector and classifier performance are considered for evaluating the proposed algorithm.

The details of experiments are as follows:

### 5.1 Feature Extraction

After preprocessing (histogram equalization) of facial images, we extract the DWT coefficients as a feature vector.

In this step, Discrete Wavelet Transform is applied to any face images. Since the face images are not continuous, we used Haar wavelet which is also discrete. We applied pyramid algorithm to each preprocessed image for decomposing it into 3 resolution levels. Then we used the approximation of images at level 3 and converted them into vectors by concatenating the columns. Dimensions of ORL database images is $92 \times 112$, so after decomposing them, the length of wavelet feature for each image is 168 [17, 18, 19, 20].

For scale invariancy of extracted features, we normalized them.

### 5.2 Feature Selection

After the extraction of DWT Coefficients, ACO is used to select the optimal feature sets.

We consider our system as a block diagram that is shown in Figure 4.

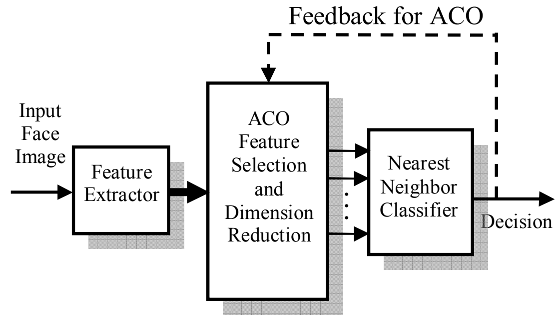


Fig. 4 Block Diagram of proposed feature selection scheme

In this step, we have applied proposed algorithm to the extracted features in the formats of ACS and $AS_{rank}$ with the same parameters.

Various parameters for leading to better convergence are tested and the best parameters that are obtained by simulations are as follows:

$\alpha=1$, $\beta=0.1$, $\rho=0.2$, the initial pheromone intensity of each feature is equal to 1, the number of Ant in every iteration p=50 and the maximum number of iterations k=25.

Selected features of each method are classified using nearest neighbor classifier and the obtained MSE is considered for performance evaluation. The results of this step are summarized in Tables 1 and 2.

Table 1. MSE and execution time of Two Different Methods

| Method | MSE | Time (s) |
|---|---|---|
| **ACO (ACS)** | 1% | 1320 |
| **ACO ($AS_{rank}$)** | 0.25% | 960 |

ACS and $AS_{rank}$ algorithms have comparable performance. The ACS method is faster than $AS_{rank}$ method however it has lower performance.

Also, Table 2 gives the optimal selected features for each method.

Table 2. Selected Features of Two Different Methods

| Method | Selected Features (DWT Component) | Number of Selected Features |
|---|---|---|
| ACO (ACS) | 4, 5, 20, 25, 29, 30, 37, 42, 44, 49, 58, 62, 68, 70, 73, 93, 94, 95, 96, 100, 102, 109, 112, 113, 114, 118, 120, 121, 125, 132, 138, 139, 141, 147, 149, 152, 156, 157, 158, 159, 163, 168 | 42 |
| ACO ( $AS_{rank}$ ) | 2, 6, 18, 21, 22, 42, 49, 57, 58, 73, 75, 83, 93, 95, 96, 100, 116, 118, 122, 125, 136, 138, 144, 147, 149, 153, 157, 158, 160, 167 | 30 |

Tables 1 and 2 show that using proposed method, we can achieve 99.75% and 98.5% recognition rate only with 30 selected features.

Finally, selected features of each method are classified and the obtained recognition rates are shown in Figures 5.

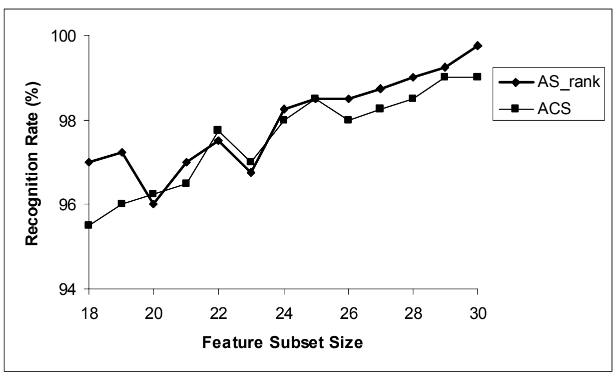


Fig. 5 Recognition Rate of Subsets Obtained Using $AS_{rank}$ , ACS

## 6. CONCLUSION

In this paper a novel ACO-Based feature selection algorithm is presented. In the proposed algorithm, the classifier performance and the length of selected feature vector are adopted as heuristic information for ACO. So, we can select the optimal feature subset without the priori knowledge of features. Proposed approach is simulated in the ACS and $AS_{rank}$ algorithm formats. Simulation results on face recognition system and ORL database show the superiority of the proposed algorithm.

## ACKNOWLEDGMENTS

This research was supported by the Iran Telecommunication Research Center (ITRC).

## REFERENCES


[1] A. Al-Ani, "An Ant Colony Optimization Based Approach for Feature Selection", Proceeding of AIML Conference, 2005.

[2] R. Jensen, "Combining rough and fuzzy sets for feature selection", Ph.D. Thesis, University of Edinburgh, 2005.

[3] M. Dorigo and G. Di Caro, "Ant Colony Optimization: A New Meta-heuristic"Proceeding of the Congress on Evolutionary Computing, 1999.

[4] E. Bonabeau, M. Dorigo and G. Theraulaz, "Swarm Intelligence: From Natural to Artificial Systems", New York: Oxford University Press, 1999.

[5] B. Liu, H. A. Abbass and B. McKay, "Classification Rule Discovery with Ant Colony Optimization", IEEE Computational Intelligence, Vol.3, No.1, 2004.

[6] M. Dorigo, V. Maniezzo and A. Colorni, "The Ant System: Optimization by a Colony of Cooperating Agents" IEEE Transactions on Systems, Man, and Cybernetics, Part B, 26(1), 29-41, 1996.

[7] V. Maniezzo and A. Colorni, "The Ant System Applied to the Quadratic Assignment Problem" Knowledge and Data Engineering, 11(5), pp 769-778, 1999.

[8] R. O. Duda and P. E. Hart, "Pattern Recognition and Scene Analysis Wiley", New York, 1973.

[9] P. Pudil, J. Novovicova and J. Kittler, "Floating search methods in feature selection" Pattern Recognition Letters, 15, 1119-1125, 1994.

[10] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection", Pattern Recognition Letters, Vol. 10, No. 5, pp. 335–347, 1989.

[11] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey", Theoretical Computer Science 344, 243–278, 2005.

[12] R. Jensen and Q. Shen, "A Rough Set-Aided System for Sorting WWW Bookmarks", Web Intelligence: Research and Development, pp. 95–105, 2001.

[13] Z. Pawlak, "Rough Sets: Theoretical Aspects of Reasoning About Data", Kluwer Academic Publishing, Dordrecht, 1991.

[14] H. H. GAO, H. H. YANG and X. Y. WANG, "ANT COLONY OPTIMIZATION BASED NETWORK INTRUSION FEATURE SELECTION AND DETECTION", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 2005

[15] B. Bullnheimer, R. Hartl and C. Strauss, "A new rank-based version of the Ant System: A computational study", Central European Journal for Operations Research and Economics, 7 (1), 25-38, 1999.

[16] T. Stützle and H. H. Hoos, "MAX-MIN Ant System", Future Generation Computer Systems, 16 (8), 889-914, 2000.

[17] Hamidreza Rashidy Kanan, Karim Faez and Mehdi Ezoji, "Face Recognition: An Optimized Localization Approach and Selected

PZMI Feature Vector Using SVM Classifier", Proceeding of the International Conference on Intelligent Computing (ICIC 2006), LNCS 4113, PP. 690-696, 2006.

[18] Hamidreza Rashidy Kanan, Karim Faez and Mehdi Ezoji, "An Efficient Face Recognition System Using a New Optimized Localization Method", Proceeding of the 18th International Conference on Pattern Recognition (ICPR'2006), 2006.

[19] Hamidreza Rashidy Kanan and Karim Faez, "ZMI and Wavelet Transform Features and SVM Classifier in the Optimized Face Recognition system", Proceeding of the 5th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2005), PP. 295-300, 2005.

[20] Hamidreza Rashidy Kanan and Karim Faez, "PZMI And Wavelet Transform Features In Face Recognition System Using A New Localization Method", Proceeding of the 31st Annual Conference of the IEEE Industrial Electronics Society (IECON 2005), PP. 2690-2694, 2005.