Image Classification Decreasing of Principal Component Analysis

Afshin Shaabany[1] and Fatemeh Jamshidi[2]
Science and Research Branch, Islamic Azad University, Fars, Iran

***Abstract*— This paper presents the application of Image Classification Decreasing of Principal Component Analysis. Recently, research in image procedure has excited much interest in the security systems community. In this paper, we take advantage of the simplified features and classifier to categorize images object with the wish to detect weapons effectively. In order to legalize the efficiency of the classifier, several classifiers are used to compare the overall accuracy of the system with the express admiration from the features decreasing. These classifiers include most remote First, Density-based Clustering and k-Means methods. The final result of this research clearly shows that model has the ability in improving the classification accuracy using the extracted features from the multi-dimensional feature decreasing. Besides, it is also shown that model is able to quickness the computational time with the reduced dimensionality of the features compromising the frail decrease of accuracy.**

***Keywords*—**Accuracy, Classification Decreasing, Principal Component Analysis and Imaging Procedure

I. INTRODUCTION

Image classification is an essential procedure in image procedure and its major issue lies in categorizing images with huge number of input features using traditional classification algorithm. These algorithms tend to produce unstable prediction models with low generalization performance [1]. To subdue high dimensionality, image classification usually depends on a pre-procedure step, specifically to extract a reduced set of meaningful features from the initial set of huge number of input features. Recent advances in classification algorithm have produced new methods that are able to handle more complex problems. Security systems are becoming essential in situations where personal safety could be imperilled due to criminal activities [2]. Formal security systems require the constant attention of security personnel, who monitor several locations simultaneously [3]-[4], therefore, the forward movement in image procedure techniques has become an advantage to the security systems to improve on the operational activity for monitoring purpose.

We used on a set of data which was available freely in the internet [5] to carry out some experimental research on the classification. We appraise the selected algorithms using the drilling dataset which contains 15 features with their associate class labels. Besides, 4 test dataset that contain the same features value of the image objects for each class have been identified. Feature extraction procedure was carried out to extract all useful features from 256 binary black and white images to represent the characteristics of the image object. From the image analysis and feature extraction, 15 important and useful features of the image object as the characteristics of the dataset were, the value of the feature will not change. We took the invariance of each feature into consideration and the features include of compactness, ratio of major axis length and minor axis length, hull ratio, moment, area ellipse ratio, axis ratio, ratio between area of the enclosing box minus area of the spot and area of the enclosing box, ratio between the height and the width of the enclosing box, ratio between the squared perimeter and the area of the spot, roughness, ratio of the area of the spot and the area of the enclosing box and compactness circularity of the spot.

II. PROPOSED METHOD

Characteristic subset estimation is done to look for combinations of characteristics whose values divide the data into subsets containing a strong single class majority [6]. The search is in favor of small feature subsets with high class persistence. This persistence subset evaluator uses the persistence metric presented by H. Liu et al.:

$$Consistency_s = 1 - \frac{\sum_{i=0}^{J} |D_i| - |M_i|}{N} \quad (1)$$

where s is an characteristic subset, J is the number is the number of distinct combinations of characteristic values for s, $|D_i|$ is the number of occurrences of the i characteristic value combination, $|M_i|$ is the cardinality of the majority class for the i characteristic value combination and N is the total number of instances in the data set [6].

To use the Persistence Subset Evaluator, the dataset needs to be discredited with numeric characteristics using any suitable method such as the method of U. M. Fayyad et al. [7]. The search method that can be used is the forward selection search which is to produce a list of characteristics [8]. The characteristics are then ranked according to their overall contribution to the persistence of the characteristic set.

Principal component analysis is one of the most popular multi dimensional features decreasing products derived from the applied linear algebra. PRINCIPAL COMPONENT ANALYSIS is used copiously because it is a simple and non-parametric technique of extracting relevant information from complex data sets. The goal of PRINCIPAL COMPONENT ANALYSIS is to reduce the dimensionality of the data while retaining as much as possible of the variation in the original dataset.

Suppose $x_1, x_2, \cdots, x_N$ are $N \times 1$ vectors.

Step 1: Mean value is calculated as

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

Step 2: Each feature is used to subtract the mean value:

$$\Phi_i = x_i - \bar{x} \quad (3)$$

Step 3: Matrix $A = \begin{bmatrix}\Phi_1 & \Phi_2 & \dots & \Phi_N\end{bmatrix}$ is generated with $N \times N$ matrix and covariance matrix with the same dimension size is computed [9]:

$$C = \frac{1}{M}\sum_{i=1}^{N} \Phi_i \Phi_i^T = AA^T \quad (4)$$

The covariance matrix characterizes the distribution of the data.

Step 4: Eigen values are computed:

$$C = \lambda_1 > \lambda_2 > \ldots > \lambda_N \quad (5)$$

Step 5: Eigenvectors are computed:

$$C = \begin{bmatrix}u_1 & u_2 & \dots & u_N\end{bmatrix} \quad (6)$$

Since C is symmetric, $u_1, u_2, \ldots, u_N$ form a basis, and $(x - \bar{x})$ can be written as a linear combination of them:

$$x - \bar{x} = b_i u_1 + b_2 u_2 + \ldots + b_N u_N = \sum_{i=1}^{N} b_i u_i \quad (7)$$

Step 6: For dimensionality decreasing, it keeps only the terms corresponding to the K largest eigenvalues [10]

$$x - \bar{x} = \sum_{i=1}^{K} b_i u_i \quad \text{Where } K << N \quad (8)$$

The representation of x into the basis $u_1, u_2, \ldots, u_K$ is thus

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix} \quad (9)$$

The aim is to do comparison of supervised classification methods for classification of the image object to their known class from the reduced multi-dimensional features dataset. The issue in identifying the most promising classification method to do pattern classification is still in research. Therefore, we are interested in predicting the most promising classification method for pattern classification in terms of the classification accuracy achieved in detecting weapons. The algorithms considered in this study are UEM, Most remote First, Density-based Clustering and k-Means. The methodology for each classifier is presented with basic concept and background.

Most remote First is a unique clustering algorithm that combines hierarchical clustering and distance based clustering. It uses the basic idea of agglomerative hierarchical clustering in combination with a distance measurement criterion that is similar to the one used by K-Means. Most remote -First assigns a center to a random point, and then computes the k most distant points [7].

This algorithm works by first select an instance to be a cluster centroid randomly and it will then compute the distance between each remaining instance and its nearest centroid. The algorithm decides that the most remote instance away from its closed centroid as a cluster centroid. The procedure is repeated until the number of clusters is greater than a predetermined threshold value [8].

Density based algorithms typically regard clusters as dense regions of objects in the data space that are separated by regions of low density [9]. The main idea of density-based approach is to find regions of low and high density. A common way is to divide the high dimensional feature space into density-based grid units. Units containing relatively high densities are the cluster centers and the boundaries between clusters fall in the regions of low-density units [10].

This method of clustering also known as a set of density-connected objects that is maximal with respect to density-reach ability [12]. Regions with a high density of points depict the existence of clusters while regions with a low density of points indicate clusters of noise or clusters of outliers. For each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points, which is, the density in the neighborhood has to exceed some predefined threshold. This algorithm needs three input parameters, which include of the neighbor list size, the radius that delimitate the neighborhood area of a point, and the minimum number of points that must exist in the radius that delaminate the neighborhood area of a point [11].

K-Means is one of the simplest learning algorithms that solve clustering problem. K-Means algorithm takes the input parameter and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low [12]. Cluster similarity is measured in regard to the mean value of the object in a cluster which can be viewed as the centroid of the cluster. The k-Means algorithm randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar based on the distance between the object and cluster mean. Then, it computes the new mean for each cluster and this procedure iterates until the criterion function converges. The algorithm works well when the clusters are compact clouds that are rather well separate from one another. The method is relatively scalable and efficient in procedure large data sets because the computational complexity of the algorithm [3].

III. RESULTS

In this work, before any classification is applied on the dataset, PRINCIPAL COMPONENT ANALYSIS are used to explore the usefulness of each feature and reduce the multi dimensional features to simplified features with no underlying hidden structure. The distributions of each feature are attracted and analyzed Figure 1 shows the distributions for the features which are discarded after PRINCIPAL COMPONENT ANALYSIS implementation and these features include of hull ratio, axis ratio, ratio between area of the enclosing box minus area of the spot and area of the enclosing box, ratio of the area of the spot and the area of the enclosing box and compactness circularity of the spot

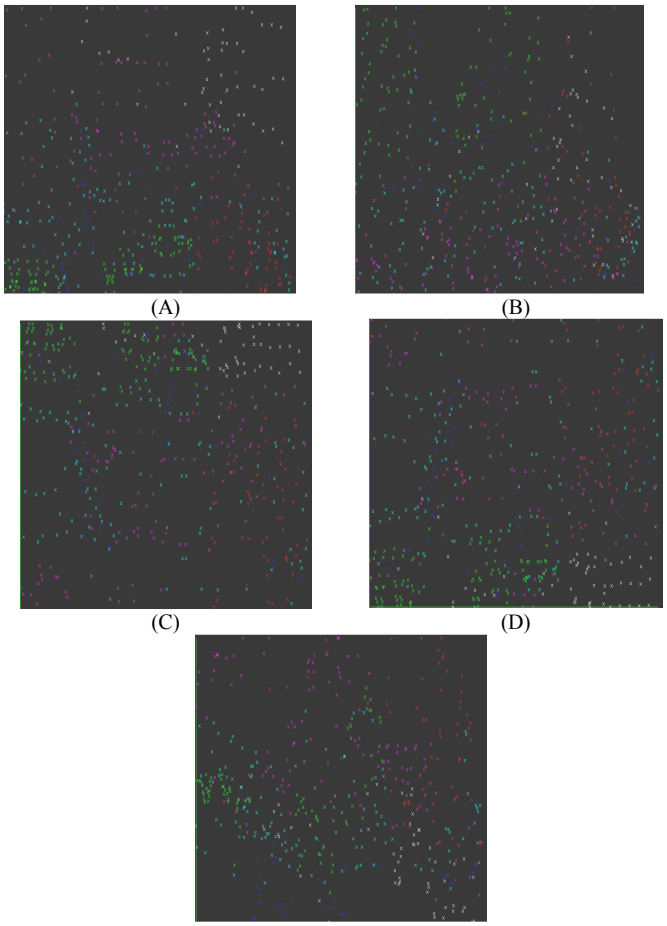

Fig. 1. The distributions of features which are being discarded after PRINCIPAL COMPONENT ANALYSIS implementation
(a) hull ratio, (b) axis ratio, (c) ratio between area of the enclosing box minus area of the spot and area of the enclosing box, (d) ratio of the area of the spot and the area of the enclosing box and (e) compactness circularity of the spot

In order to legalize the impact of multi dimensional feature decreasing methods of CSE and PRINCIPAL COMPONENT ANALYSIS, four types of dataset are used, namely the original data, data produced after CSE method, data produced after PRINCIPAL COMPONENT ANALYSIS method and data produced after CSE and PRINCIPAL COMPONENT ANALYSIS methods. The classifiers are analyzed and the accuracy appraisal is as shown in Table 1 with the computational speed. In this work, the model with the highest classification accuracy is considered as the best model for pattern classification of this dataset.

TABLE I
ACCURACY APPRAISAL AND COMPUTATIONAL SPEED OF EXPERIMENTAL METHOD

	Original data (15features)	PRINCIPAL COMPONENT ANALYSIS + Classifier
Most remote First	84.88 % (7.53ms)	83.38 % (5.66ms)
Density based Clusterer	86.20 % (8.35ms)	88.21 % (6.41ms)
K-Means	86.14 % (7.55ms)	89.58 % (5.79ms)

Based on TABLE I, As the dataset we used in this study is quite small and based on our research, the classifiers with features generated from PRINCIPAL COMPONENT ANALYSIS provide weakly less accuracy and computational speed compared to the classifiers using the predefined number of features. This is due to the reduced dimensional features offered by PRINCIPAL COMPONENT ANALYSIS which allow only the useful key features to participate in the classification procedure.

IV. CONCLUSION

In this article, the target is aimed to inquire into the performance and impact PRINCIPAL COMPONENT ANALYSIS on classification in the aspect of accuracy and computational speed. We stress on the analysis and usage of the multi-dimensional features decreasing on advanced classification method to classify weapons within an image. In order to legalize the efficiency of the feature decreasing method and classifier, several classifiers such as Most remote First, Density-based Clustering and k-Means methods are used to compare the overall accuracy of the classifiers. The potential of each classifier has been demonstrated and the hybrid method has shown a desirable result in detecting weapons compared to other classifiers.

REFERENCES

[1] M. A. Hall and G. Holmes, "Benchmarking Characteristic Selection Techniques for Discrete Class Data Mining," *IEEE Trans. Knowledge and Data Engineering*, vol. 15, no. 3, 2003.
[2] Kononenko, "Estimating characteristics: Analysis and extensions of relief," *in Proc. the 7th European Conf. Machine Learning,* 1994, pp. 171-182.
[3] M. A. Aizerman, E. M. Braverman, and L.I. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 826-837, 1964.
[4] Basilevsky, *Statistical Factor Analysis and Related Methods*, Wiley, New York, 1994.
[5] J. Han and M. Kamber, "*Data Mining: Concepts and Techniques*," Morgan Kaufmann, San Francisco, CA (2001).
[6] Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, 1997.

[7] F. Dellaert, "The Expectation Maximization Algorithm, College of Computing, Georgia Institute of Technology," *Technical Report,* 2002.
[8] S. D.Hochbaum and B. D. Shmoys, "A Best Possible Heuristic for the k-Center Problem," *Mathematics of Operational Research*, vol. 10, no. 2, pp. 180-184, 1985.
[9] S. Dasgupta and P. M. Long, "Performance guarantees for hierarchical clustering," *J. of Computer and System Sciences*, vol. 70, no. 4, pp. 555-569, 2005.
[10] X. Zheng, Z. Cai and Q. Li, "An experimental comparison of three kinds of clustering algorithms," *IEEE Int. Conf. Neural Networks and Brain*, 2005, pp. 767-771.
[11] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *2nd Int. Conf. Knowledge Discovery and Data Mining,* Portland, Oregon, USA, 1996.
[12] T. H. Cormen, C. E. Leiserson and R. L. Rivest, *Introduction to algorithms*, McGraw-Hill Book Company, 1990.

Email: afshinshy@yahoo.com
Afshin Shaabany was born in Marvdasht, Iran, in 1975. He received the Bs degree in Electerical Engineering from Tehran University, Tehran, Iran in 1999 and the Ms degree in Electrical Engineering from Polytechnic University, Tehran, Iran 2008. His research interests include IT, telecommunication, switching systems; Intelligent systems.

Email: fjamshidi59@yahoo.com
Fatemeh Jamshidi was born in Shiraz, Iran, in 1980. She received the Bs degree in Biomedical Engineering from the Jondi Shapour University, Ahvaz, Iran in 2002 and the Ms and PhD degree in Electrical Engineering from Shiraz University and Tarbiat Modares University, respectively. Her research interests include switching systems; Intelligent systems, Robust control, IT, telecommunication