Research Article

Chaotic Neural Network for Biometric Pattern Recognition

Kushan Ahmadian and Marina Gavrilova

Department of Computer Science, University of Calgary, Calgary, AB, Canada T2N 1N4

Correspondence should be addressed to Kushan Ahmadian, kahmadia@ucalgary.ca

Received 8 December 2011; Revised 24 February 2012; Accepted 25 February 2012

Academic Editor: Sheryl Brahnam

Copyright © 2012 K. Ahmadian and M. Gavrilova. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Biometric pattern recognition emerged as one of the predominant research directions in modern security systems. It plays a crucial role in authentication of both real-world and virtual reality entities to allow system to make an informed decision on granting access privileges or providing specialized services. The major issues tackled by the researchers are arising from the ever-growing demands on precision and performance of security systems and at the same time increasing complexity of data and/or behavioral patterns to be recognized. In this paper, we propose to deal with both issues by introducing the new approach to biometric pattern recognition, based on chaotic neural network (CNN). The proposed method allows learning the complex data patterns easily while concentrating on the most important for correct authentication features and employs a unique method to train different classifiers based on each feature set. The aggregation result depicts the final decision over the recognized identity. In order to train accurate set of classifiers, the subspace clustering method has been used to overcome the problem of high dimensionality of the feature space. The experimental results show the superior performance of the proposed method.

1. Introduction

Growing efforts are devoted to develop and implement new security systems based on biometric features. System subjects could be either human or virtual reality entities, and they are accepted or rejected by the system based on biological or behavioral biometric features. Biometric pattern recognition includes recognition of fingerprint, face, gait, signature, voice, ear, iris, or other physiological or behavioral features.

As one of the key biometric features, facial biometrics plays a key role in user authentication. It consists of a set of high-dimensional vectors representing topological, color, or texture information. The feature set is very complex and may contain hundreds of features. This makes it a difficult biometric pattern recognition problem to deal with [1]. Many of the earlier face recognition algorithms are based on feature-based methods. These methods identify a set of geometrical features on the face such as eyes, eyebrows, mouth, and nose [2].

Properties and relations between the feature points, such as areas, distances, and angles are used as descriptors for face recognition. Statistical methods are usually used to lower the number of dimensions; however, there are no universal answers to the problem of how many points give the best performance. In addition, there is no clear answer on what the important features are or how to extract those features (Figure 1).

Alternative face recognition techniques based on elastic graph matching [3] and support vector machines (SVMs) [4] have been investigated as well.

On the other hand, appearance-based face recognition algorithms are used as a tool to project an image into the subspace and finding the closest point set [5]. One of the very well-known linear transformation methods that are used vastly for dimensionality reduction and feature extraction is the principal component analysis (PCA) [6]. In this method, object classes that are closer together in the output space are often weighted in the input space to reduce potential misclassification. PCA can be used over raw face image to extract the Fisherface or by similar manner on the eigenface to obtain the discriminant eigenfeatures [6]. Many variations of PCA combining feature representation methods that utilize the strengths of different realizations have been developed recently. LDA (Linear discriminant
2 Advances in Artificial Intelligence

Figure 1: Data samples for FERET facial database.

analysis) [7], Kernel PCA [8], and generalized discriminant analysis (GDA) using a kernel approach [3] are among such methods.

However, a simple modification of PCA features or attempts to find a better fitting of those features onto linear or nonlinear subspaces alone is not sufficient to deal with high dimensionality of data. Thus, hybrid approaches were recently introduced to overcome shortcomings of individual methods. For example, a combination of support vector machines with skin color regression model was successfully used for face recognition from video sequences. Line edge map approach [2] to extract lines from a face edge map as features is based on a combination of template matching and geometrical feature matching. The nearest feature line classifier method [9] attempts to extend the capacity covering variations of pose, illumination, and expression for a face class by finding the candidate person owning the minimum distance between the feature point of query face and the feature lines connecting any two prototype feature points.

Among new approaches, learning techniques have emerged very recently. Some of the promising new approaches incorporate neural-network-based learning methods for biometric feature processing or fuzzy-logic-based multimodal biometric pattern classification systems [10, 11]. However, direct application of neural-network method is not sufficient to overcome the main obstacle, high complexity and high dimensionality of biometric data, especially if various features of a single or multiple biometrics must be taken into account. Dimensionality reduction methods are ideal for this task. The goal is to transform data from a high-dimensional space into a lower-dimensional one without loss of important information. Normally, the lower dimension maximizes the variance of data. High dimensionality of data is a common problem in recognition systems where a set of features from the training samples are used to create a learner. The complexity of designing algorithms for the recognition purpose grows significantly as the number of dimensions grows. A common set of methods to reduce the dimensionality of space is the clustering approach. In clustering, the objects are grouped according to their similarity by some similarity measures. Usually clustering is used to design a set of boundaries to better understand the data (based on structured data). Other usage of clustering involves indexing and data compression. We aim to target the second goal in order to create a meaningful subspace of the original space.

The high dimensionality problem arises when the feature set comes either from a single complex resource (such as combination of geometric, appearance-based, and color-based facial features) or from different resources (face/ear). In this paper, we utilize the concepts discussed above in order to create a secure and precise system for face recognition. The unique characteristic of such system is that it is based on chaotic neural network for feature selection and training.

In the next section we describe the proposed neural-network methodology with focus on the dimensionality-reducing methods. In Section 3, experimental results are presented. Concluding summary is found in Section 4.

2. Proposed Methodology

The proposed system consists of training different chaotic learners based on unimodal biometric data coming from a single source (i.e., a single subject). We start by proposing to use the chaotic Hopfield neural network for storing the biometric patterns. When a new pattern is introduced to the network, the network tries to convert the introduced pattern to the closest pattern saved in the memory and thus arrives to the user whose biometric is the closest to one given as an input. In order to train the network, we first obtain a set of vectors from user input (i.e., facial biometric) and then feed them as features of the new pattern to the CNN-based learning engine. Clustering is performed to group feature vectors with similar features and to further reduce complexity of feature vector space. Having the vector of weights from the candidate clusters, the next step in the
proposed methodology is defining an energy model for the associative memory to learn the data patterns. The benefit of this approach is that it is a learner system that converges the given set of vectors to the stored pattern. In the classical formulation of Hopfield network for optimization problem, the method usually traps in one of many local minima of the energy function since no simulated annealing or noise injection policies have been applied. To remedy this problem, we take benefit from chaotic noise injection policy.

2.1. Overall System Architecture. One of the objectives of this paper is to investigate if neural networks combined with dimension analysis provide benefits in improving biometric system performance and circumvention, that is, resistance to low-quality data or absence of one of biometric traits all together.

First, we aim to find the best subset of biometric features derived from the original dataset. The new subset could be either a portion of the original dataset or a reconstructed set of samples with a reduced dimensionality. As an example, when using the principal component analysis method (PCA), the main goal is to find the principal components of the distribution of features (such as faces, ears, and other biometrics utilized in the system), which are eigenvectors of the covariance matrix of the set of biometric images. These eigenvectors can be considered as a set of features that together characterizes the variation between biometric samples.

Here, we propose an alternative system. As the primary biometric sample, we chose face images due to significant variability of image quality and also popularity of the biometric for testing data analyzing methods.

Next, we propose a new principle for subspace analysis and dimensionality reduction based on a generalized description of spherical coordinates. Each face image in the training set can be represented in terms of a linear combination of the original images. The number of possible classes is equal to the number of face image groups in the training set. However, the faces can also be approximated using only the “best” eigenfaces—those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images. The primary reason for using fewer eigenfaces is computational efficiency. This approach allows not only move compact space representation, but also a convenient tool for subsequent clustering and learning of common patterns. Next, we propose to use neural networks as a fast and reliable way for a biometric system to learn the pattern from the previously extracted subspaces. The neural network approach is based on my chaotic noise injection strategy which is the leading strategy for neural network training. In experimental section, I show that this approach has a high potential for success in biometric research studies. The advantages are the ability to learn and later recognize new biometric samples in an unsupervised manner and that method is easy to implement using the proposed neural network architecture. The traditional architecture is shown in Figure 2.
The new system architecture that we propose is based on neural network representation (see Figure 3).

The novelty of the proposed system relies on the representation of the feature space, which is not limited to single biometrics nor to the number of dimensions. While the system is capable of handling large number of feature vectors, it makes it also capable of learning complex biometric patterns faster using neural network learner.

2.2. Chaotic Neural Networks. Having the vector of weights from the candidate clusters, the next step in our methodology is defining an energy model for the associative memory to learn the data patterns. As it will be shown in Section 2.4, the inputs of the neural network are the obtained candidate vectors from the dimensionality reduction phase. The candidate vectors are generated from the original vectors of images each of which consists of a set of grayscale pixels.

The nature of associative network is based on minimizing the energy function assigned to the neural network. The energy function simply shows the distance of the stored pattern to the introduced pattern. The lower the energy function gets the more stable the network is. On the other side, patterns are stored as the weights between pairs of neurons. By other means when two values are changing with a high correlation, the weights between the neurons corresponding to those values get higher.

The benefit is the learner system that converges the given set of vectors to the stored pattern. In order to learn the patterns, the chaotic dynamics of single neuron of noise chaotic neural is described as follows:

\[ x_{ij}(t) = \frac{1}{1 + e^{-y_{jk}(t)/\tau}}, \]

\[ y_{jk}(t + 1) = ky_{jk}(t) + \alpha \left\{ \sum_{i=j, i \neq j}^{N} \sum_{l=1, l \neq k}^{M} w_{jkl}x_{jk}(t) + I_{ij} \right\} - z(t) \left( x_{jk}(t) - I_{0} \right) + n(t), \]

where \( z(t + 1) = (1 - \beta_{1})z(t) \) and \( n(t + 1) = (1 - \beta_{2})n(t). \)

Here, \( x_{jk} \) and \( y_{jk} \) are outputs and inputs of neuron \( jk \), respectively, and \( w_{jkl} \) is the weight of the connection between neurons \( jk \) and \( il \). \( K \) is the damping factor, \( I_{ij} \) is the input bias of neuron \( ij \), and \( \beta_{1} \) and \( \beta_{2} \) are damping factors of neural self-coupling and stochastic noise. \( \alpha \) is a positive scaling factor, \( \varepsilon \) is steepness parameter of the output function, \( n(t) \) is the random injected noise, and \( I_{0} \) is the initial value which controls the chaotic behavior. The single-neuron dynamics of the noisy chaotic neural network correspond and are controlled by different values of \( I_{0} \). In our experiments, the chaotic phase can be reached at \( I_{0} = 0.3 \), which has been used to inject chaotic noise into the network.

In the next sections, we consider several dimensionality reduction methods for facial biometric and will compare the resulting biometric verification system performance against other popular approaches.

2.3. Knowledge-Based Methods. Knowledge-based facial recognition methods are rule-based methods that utilize the knowledge of geometry of human face structure and expected position and symmetry of facial elements (i.e., eyes). To finely adjust the position of each face image of the database, we use the knowledge-based method augmented with the skin tone detection to first identify the face. The knowledge-based methods are relying on the simple set of rules. For example, every face has symmetric features (eyes), and the color of the eyes area is normally darker compared to the other facial areas (i.e., forehead, cheeks). By using these rules, we identify facial areas. The main challenge for this category of methods is to find a proper set of rules. If the rules set are too general, then the false acceptance rate increases dramatically while specific types of rules may result in high false rejection rate due to facial gestures and artifacts.

By using the color tone, we overcame many problems in our application. The color tone is very simple to use while very effective for normalized pictures. There are different approaches to use the color tone. Researchers have proposed to use both RGB and HSV color models to recognize the skin area of the image. We have used the same parameters to detect the face area while a postdetection face was required in some of the images to detect the face area more accurately. Figure 4 shows an example of detection, normalization, and selection of the area of interest. The parameters are set according to the following criteria:

\[ 0.4 \leq r \leq 0.6, \quad 0.22 \leq g \leq 0.33, \quad r > g > (1 - r)/2, \quad (2) \]

\[ 0 \leq H \leq 0.2, \quad 0.3 \leq S \leq 0.7, \quad 0.22 \leq V \leq 0.8. \quad (3) \]

Since each new image contains a large number of pixels (256 * 256 in case of FERET database), we need to remove the redundant or correlated pixels which are our features but will not be important for the final learning pattern. For example, the background of the images in the database should not be considered in the final learning machine. This is illustrated in Figure 5 where the background is a smooth pattern and face recognition phase should ignore that area.
We propose to use a specialized subspace clustering method to remove the redundant features and gain a lower-dimensional subspace where the new features show less correlation and are more meaningful and easier to use in the verification process.

2.4. Subspace Clustering Method. Clustering aims at dividing datasets into subsets (clusters), where objects in the same subset are similar to each other with respect to a given similarity measure, whereas objects in different clusters are dissimilar [12]. There are many applications of cluster analysis in biometric, GIS, and oil and gas fields.

When clustering high-dimensional data of biometric features, we face a variety of problems [12]. The presence of irrelevant features or of correlations among subsets of features heavily influences the appearance of clusters in the full-dimensional space. The main challenge for clustering here is that different subsets of features are relevant to different clusters, that is, the objects cluster in subspaces of the data space but the subspaces of the clusters may vary. Additionally, different correlations among the attributes may be relevant for different clusters. This assumption implies that different biometric features or a different correlation of biometric features may be relevant for varying clusters the local feature relevance or local feature correlation.

A common way to overcome problems of high-dimensional data spaces where several features are correlated or only some features are relevant is to perform feature selection before performing any other data mining task. Due to the problem of local feature relevance and local feature correlation, usually no global feature selection can be applied to overcome the challenges of clustering high-dimensional biometric data.

Instead of a global approach to feature selection, we propose to use a local approach accounting for the local feature relevance and/or local feature correlation problems. Since traditional methods, like feature selection, dimensionality reduction, and conventional clustering, do not solve the previously sketched problems, novel methods need to integrate feature analysis into the clustering process more tightly.
The main idea of the method is to project $d$-dimensional vector of biometric feature points in the parameter space represented through a $(d-1)$-dimensional hyperplane. The algorithm for projecting biometric feature points is adapted from [13]. In order to detect those linear hyperplanes in the data space, the task is to search for points in the parameter space where many sinusoidal curves intersect. For that purpose, for each grid cell the number of intersecting sinusoidal curves is aggregated. Below, this interpretation is discussed in more detail.

Since higher dimensions cannot be depicted easily, we have chosen 3 pixels of the image of a single person to explain the process which forms a 3D vector (see Figure 4). Normally, the intensity value of a pixel in RGB space is a value between 0 and 255. We normalize this value to the range of 0 to 1. Later we incorporate other features and aim to normalize all the vectors to the scale. This preserves the intercorrelation of pixels within a single image. However, we can make a global correlation map by creating vector from the same pixels of different images. The new scheme is shown in Figure 6.

As it has been shown in Figure 5, the first step in processing the input image is to create the input vectors of the subspace clustering method. The vectors are created based on grouping same pixels from all the input images, for example, if we have 100 images, each of which featuring 256 pixels, the output of the grouping phase is 256 vectors of 100 dimensions each. By utilizing such a method, the relation between the points is preserved and they could be grouped as a single class if they show high correspondence (as correlation in PCA). Next, using the parameterization functions, the points are translated into sinusoidal curves. The intersection of the curves would be the passing plane from the points. The plain is considered as a class or cluster.

With the proposed concepts, one can transform the original subspace clustering problem (in data space) into a grid-based clustering problem (in parameter space) [12]. More details on how to apply this method to biometric research domain can be found in [11].

3. Exploring Dimensionality: Reduction

In this section we look at the dimensionality reduction methods to compare their performance against the proposed system.
3.1. Isomap. The Isomap algorithm is designed based on the Floyd-Warshall algorithm and classic multidimensional scaling [14]. The Isomap method performs the same while considering the neighbors only. The kernel matrix can be used to show the distance of Isomap. The common form of geodesic distance matrix is as follows:

\[ k = \frac{1}{2} hD^2H, \]  \hspace{1cm} (4)

where \( D^2 \) is the square of the geodesic distance matrix and \( H \) is matrix:

\[ H = I_n - \frac{1}{n} e_n e_n^T. \]  \hspace{1cm} (5)

In order to choose best \( k \) nearest neighbors (kNN) for the Isomap method we have tried different kNN values for the pixels of the original space. Based on these values we have reconstructed the images based on the first two eigenvectors of the highest eigenvalues as it can be seen in Figure 7.

Our next task was to choose the proper number of eigenvectors for the recognition purpose. Experimentally, we found out that for the first 30 eigenvectors (sorted based on the corresponding eigenvalues) the correlation between the values is low, which makes these eigenvectors highly useful in the recognition phase. The recognition rate of Isomap for different \( k \)-values is shown in the following chart.

We can see that the best recognition rate is recorded for kNN values in the range from 12 to 14 and is below 40%.

Figure 11: (a) Reconstructed images for MVU based on the first two eigenvectors. (b) Recognition rates for different \( k \)-values of MVU.
3.2. Laplacian Eigenmaps. The Laplacian eigenmaps method is based on using spectral approach to reduce dimensionality [15]. The Laplacian eigenmaps method also assumes that important data is located on a lower-dimensional manifold in the high dimension space. The Laplacian eigenmaps method also utilizes the $k$ nearest neighbor method. However, instead of creating the weight matrix, in the Laplacian Eigenmaps method a graph is produced where each two vertices are connected if they are considered as neighbors. The graph is considered as a lower-dimensional manifold in the original high-dimensional space.

Similar to Isomap, in order to choose best $k$ nearest neighbors for the Laplacian method, we have tried different $k$-NN values for the pixels of the original space (Figure 9). Based on these values, we have reconstructed the images based on the first two eigenvectors of the highest eigenvalues as it can be seen in Figure 8.

The recognition rate of the Laplacian for different $k$-values is shown in the following chart in Figure 10. We can note recognition significantly higher than in the Isomap method and the range of values where it reaches its maximum potential of 80% is for $k$ values between 12 and 16.

3.3. Maximum Variance Unfolding. The final alternative to the subspace clustering method is maximum variance unfolding (MVU). MVU maps the data of high-dimensional space into a lower-dimensional space, preserving the distances between neighbor points [16]. MVU has to create the neighborhood graph $G$ first. In $G$ each data point is connected to its $k$ nearest neighbors. MVU tries to maximize the square Euclidean distances between all of the points or solves an optimization problem.

Using a singular value decomposition and by solving $K$ of the SDP, we can obtain the low-dimensional data representation $Y$.

The recognition rate of MVU for different $k$-values is shown in Figure 11. The method outperforms Isomap but has a higher volatility than Laplacian method, ranging from 38% to 78% recognition rate with the best rates achieved around $k = 11$ values. It is a close contender to Laplacian but cannot outperform it.

As it can be seen above, the results are reduced dimensions (the unnecessary details of each image have been removed). This makes the learning system able to process the data with less effort and the recognition rate gets higher due to the simpler nature of the system.
3.4. Subspace Clustering Method Performance. Here, we provide the comparison of our proposed subspace clustering (SC) method for dimensionality reduction against previously described Isomap, Laplacian, and LLE methods. The LLE method is not suitable for biometric security application due to highly inconsistent performance. The recognition rate of Isomap is not high enough to make the method useful in any commercial application. The results of the Laplacian method (the best method so far among all compared for the given database FERET) are thus the best so far. Now let us look at the performance of the proposed subspace clustering method. Figure 12 shows the resulted reconstructed image based on subspace clustering with kNN value of 14. Figure 13 presents recognition rates obtained for different $k$-values. The range of values with high recognition rate is from 10 to 16, and the recognition rates are higher than those of Laplacian or MVU method. It significantly outperforms the Isomap method as well. Figure 14 shows overall performance of resulting multi-modal system for both face and ear biometrics. As it can be seen, it outperforms single biometric system in FAR/FRR rates.

4. Conclusions

In this paper, we argued that the chaotic neural network approach for learning biometric patterns is a powerful tool to improve recognition rate of biometric system. We have discussed a biometric system based on a recently devised approach for feature dimension reduction on an example of face recognition system. We compared and contrasted the proposed methodology against other dimensionality reduction methods such as Isomap, Laplacian Eigenmap, and MVU. The experimental results show a better accuracy rate for the biometric system tested on FERET facial database among all tested methods. Further improvement of recognition can be obtained by combining facial image biometrics with other biometric traits such as ear biometric.

Acknowledgments

The authors would like to acknowledge support from National Sciences and Engineering Research Council of Canada, and University of Calgary, Canada, for partial support of this project. They are also grateful for valuable comments from reviewers and for assistance with paper revisions from Advances in Artificial Intelligence Journal Editorial members and staff.

References
