International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 3619-3623

Dimensionality Reduction of Biometric Features based on Entropy Properties

Yarob A. M. Istitieh[1], Mohamed A. El-Sayed[1,2]

[1]Department of Computer Science, College of Computers and IT, Taif University, Hawia, Taif, KSA.
[2]Department of Mathematics, Faculty of Science, Fayoum University, Fayoum, Egypt.

Abstract:

In this article, we will study method depend on biometric features, entropy properties and dimensionality reduction. It consists of two main phases, first phase, we will construct a biometric identification technique with multi-model features extraction. In second phase, the various properties of entropy applied on multi-model features extraction to increase the accuracy factor of security recognition in this system. Therefore, predictable that the presented way can grant new and useful style for biometric verification. It may be raise the precision factor of safety recognition. To illustrate the quality and flexibility of the approach, We will provide experimental outputs and compared the results against previous methods.

Keywords: Biometric features, entropy, verification, authentication, features reduction.

INTRODUCTION

Biometric recognition indicates the automated discrimination of users depends on their physiological and behavioral merits of persons such as face, retina, fingerprint and vein. The curse of dimensionality can affect the achievement of the biometric authentication system when the classes number raises that dimension volume of the biometric form becomes very large. Feature Selection (FS), likewise know dimensionality reduction, is generally implemented before classification processes. The main goal from classes/clusters identification accomplishment that produced of pertinent features among huge datasets. FS also allows getting less complex representations that can have higher quality when compared to the representation with the authentic features in high dimensional datasets with large precision[1,2].

Yuhang Dinget al [3], have shown the theoretical foundation and difficulties of hand vein recognition, at first. secondly, it deeply studied the improved conditional thinning and new threshold segmentation technique that applied on hand vein images. The feature extraction is studied using the crossing points and end points in hand vein image. The matching process depend on intervals is applied to correspond hand vein samples.

For traditional algorithm, it is needful to employ images of high-quality, which need complex set equipment. Shi Zhao et al, proposed [4] a biometric technique using hand-dorsa, extracting vein structures. The used approaches may be work utilizing simple equipment conceivable.

The method of analyses of subspace is an efficient approach [5] for features discrimination. In subspace analysis step appropriate low-dimensional subspace is found, it effect directly on recognition accomplishment. Xie et al [6] presented way of random projection to decrease high dimensional space of features into low dimension space. Kaur and Ada [7] presented the usefulness of biometric identification systems, their types, principle of operation, application area, instruments required, advantages and disadvantages.

Feature selection is a technique that allows selecting a subset of pertinent features that correspond to the original features in problem domain with large precision. Feature selection is the famous method for decreasing the dimensions space. An enormous set of systems need powerful recognition platforms to either assert or locate the identity of a person seeking their duties and services. A easy mechanism of authentication using SVM is given for identification of retinal features [8].

The paper [9] offered a technique of authentication based on identification of retinal features. The approach can be used to other images in which it is of benefit to find the vascular consisting of a blood-vessel. This way covers style for blood-vessel classification, feature selection creation and the comparing of these features. The standard available databases STARE, VARIA and DRIVE are utilized to compute the implementation analysis.

The reminder of this paper is organized as follows: Section 2 briefly explains entropy and its properties. Section 3 outlines the K-means via Rényi entropy for clustering. Section 4 presents how Rényi entropy used to reduce the dimensionality. The experiment, charts and numerical results are given in Section 5. Finally, the conclusion is covered in Section 6.

SOME PROPERTIES OF ENTROPY

Rudolf Clausius,1865, introduced the concept of entropy, its properties and how it used as a measurement tool of energy. Claude Shannon,1948, add entropy concept into information theory as measurement of uncertainty [10]. Many fields such as classification, data analysis and dimension reduction utilizes entropy minimization to discover patterns or similarities in a big data volume. The randomness region in image is characterized by large entropy values, while the region of data structure are characterized by small values of entropy [11]. The common well-known use of entropy in the data mining field is decision trees for gaining information, since the values of a numeric features are discrete in hierarchy discretization form. The clusters number of the distribution data can be determine using entropy as an information measure, the data belonging to a cluster represent as one group. These groups have a histogram appears cluster distribution of features, a classification with high confidence has low value of entropy [11,12].

Let we have n outcomes denoted by $Z = \{\zeta_1, \zeta_2, ..., \zeta_n\}$, and associated probabilities ω_1, ω_2, …, ω_n, Shannon Entropy is defined as $\ell(Z) = -\sum_{i=1}^{n} \omega_i \ ln(\omega_i)$. Since $-\omega_i \ ln(\omega_i) \geq 0$ for $0 \leq \omega_i \leq 1$, that $\ell(Z) \geq 0$. If one of the ω_i =1; all others are equal to zero then $\ell(Z) = 0$ and $0\ln(0) = 0$. The entropy measure is a sign of any probability distribution that is more beneficial than the other. The minimum value of entropy provides minimal uncertainty, thus limiting knowledge about the system and its structure. For example, the minimum value of entropy is required in data classification field. In probability distribution and subject to some the specific restrictions is a problem of evaluating a minimal entropy by the global minimization of Shannon entropy measurement. All the probability mass can be moved from one state to another by grouping, that is, decrease the number of states and decrease entropy. Frenken [13] referred to some sober properties of Shannon entropy:

1. Shannon entropy measure is concave and nonnegative value in associated probabilities ω_1, ω_2, …, ω_n, this does not change by including the probability of zero.
2. It has a value equal to zero with a completely certain outcome and equal to any probability positive with uncertain outcome.
3. In the uniform distribution, gives a fixed number of outcomes.
4. In the joint distribution of two independent distributions, the sum of the individual entropies gives the final entropy value.
5. For any two values $\eta \neq 0$ and γ, the entropy $\ell(Z) = \ell(\eta Z + \gamma)$ that depend on the unordered probabilities and not on Z.

Assume that the outcomes $Z = \{\zeta_1, \zeta_2, ..., \zeta_n\}$ can be aggregated into a smaller number of sets C_1, C_2, …, C_K in such a way that each outcome is in only one set C_k, where k = 1, ..., K. The probability that outcomes are $\omega_k = \sum_{i \in C_k} \omega_i$ of set C_k Let $\ell_0(Z)$ is the between group entropy, the relationship between the entropy $\ell_0(Z)$ at the level of sets and the entropy $\ell(Z)$ at level of the outcomes is given by the entropy decomposition theorem, $\ell_0(Z) = -\sum_{k=1}^{K} \omega_k \ ln(\omega_k)$. Shannon entropy can then be written as $\ell(Z) = \ell_0(Z) + \sum_{k=1}^{K} \omega_k \ \ell_k(Z)$, and $\ell_k(Z) = -\sum_{i \in C_k} \frac{\omega_i}{\omega_k} \ ln(\frac{\omega_i}{\omega_k})$.

A property of this relationship is that $\ell(Z) \geq \ell_0(Z)$ because ω_k and $\ell_k(Z)$ are nonnegative. This means that after data grouping, there cannot be more uncertainty (entropy) than there was before grouping.

Rényi [11,14] has more ability to expand Shannon entropy to a continued set of entropy measurements. The Rényi's entropy measure $\ell_\alpha(Z)$ of order α of associated probabilities ω_1, ω_2, …, ω_n, is defined as:

$$\ell_\alpha(Z) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{n} \omega_i^{\alpha} \quad (2)$$

where $\alpha \neq 1$ is a positive real parameter. The Rényi entropy of order α for each distribution is defined as:

$$\tau^*(0.5) = 2Arg \max_{\tau \in C} [ln \sum_{i=0}^{\tau} \sqrt{\omega_i / P_A} + ln \sum_{i=\tau+1}^{K} \sqrt{\omega_i / P_B}]$$

[14, 15] . The Rényi entropy and information content converge to the Shannon entropy for α→1. [15]

DIMENSION REDUCTION USING RÉNYI ENTROPY

In this section we outline the K-means via entropy minimization. The method of this section enables us to perform learning on the data set, in order to obtain the similarity matrix and to estimate a value for the expected number of clusters based on the clustering requirements or some threshold. The Bayesian inference is therefore very suitable in the development of the entropy criterion. Suppose that after clustering the data set Z, we obtain the clusters { C_j, j = 1, ..., K } by Bayes rule, the posterior probability $P(C_j \mid Z)$ is given as

$$P(C_j \mid Z) = P(Z \mid C_j) P(C_j) / P(Z) \propto P(Z \mid C_j) P(C_j)$$

since $P(C_j \mid Z)$ given in the following is the likelihood and measures the accuracy in clustering the data. The prior $P(C_j)$ measures consistency with our background knowledge.

$$P(Z \mid C_j) = \prod_{\zeta_i \in C_j} P(\zeta_i \mid C_j) = \exp(\sum_{x_i \in C_j} \ln p(\zeta_i \mid C_j))$$

By the Bayes approach, a classified data set is obtained by maximizing the posterior probability $P(C_j \mid Z)$. In the case a data set that has spherical clusters noted that the performance of K-means algorithm is well and very quiet [13]. Since the model is based on the K-means, each cluster has Gaussian distribution with mean values C_j, j = 1, ..., k and constant cluster variance. In this way , for any specified cluster C_j ,

$P(\zeta_i \mid C_j) = \exp(-[(\zeta_i - c_j)^2/(2\sigma^2)])/\sqrt{2\pi\sigma^2}$.

When take the natural log in two sides of previous equation and skipping constant, we get, $\ln P(\zeta_i \mid C_j) = -(\zeta_i - c_j)^2/(2\sigma^2)$.

Suppose that k^* be the desirable number of clusters in the K-means algorithm. After clustering, the entropy will be minimum based on the clustering requirement. The value of entropy is decreased when the clusters are combined. Therefore if we start with some large number of clusters $K > k^*$, The clustering algorithm will reduce K to k^* because clusters with probability zero will disappear. Note that convergence to k^* is guaranteed because the entropy of the partitions is bounded below by 0. M. Figueiredo and Jain [16] found that the initial number of clusters K equal to the square root of two. The posterior probability becomes:

$$P(C_j \mid Z) \propto \exp(-\sum_{\zeta_i \in C_j}[(\zeta_i - c_j)^2/(2\sigma^2)] - \lambda\ell(Z))$$

Notice that since the entropy term is nonnegative, the equation is minimized if entropy is minimized. Therefore the equation is the required clustering criterion. We note that when $\lambda = 0$, The term $\sum_{\zeta_i \in C_j}[(\zeta_i - c_j)^2/(2\sigma^2)] + \lambda\ell(Z)$ is identical to the cost function of the K-Means clustering algorithm. The Rényi Entropy K-means (REK) algorithm is given in the following. Multiple runs of REK are used to generate the similarity matrix. Once this matrix is generated, the learning phase is complete.

REK Algorithm

1. Let k is the initial number of clusters.
2. Let ε is a value for the stopping criteria.
3. Initialize the cluster centers $\theta_i(\upsilon)$, a priori probabilities ω_i, $i = \overline{1,k}$, randomly an λ.
4. Set the counter $\upsilon = 0$.
5. Repeat
6. Set $\upsilon = \upsilon + 1$
7. For each input vector $\zeta_j \in C_r$, $j = 1,2 \ldots, n$, $r = 1, \ldots, k$

   Classify each ζ_j to get the partition C_i such that

   $[\zeta_j - \theta_r(\upsilon)]^2 - \lambda\ln(\omega_r)/n \le [\zeta_j - \theta_i(\upsilon)]^2 - \lambda\ln(\omega_i)/n$
8. Update the cluster centers $\theta_i(\upsilon+1) = \frac{1}{C_i}\sum_{j \in C_j}\zeta_j$
9. The priori probabilities of clusters $\omega_i(\upsilon+1) = C_i/n$.
10. Until satisfy the convergence; $\max_i |\theta_i(\upsilon+1) - \theta_i(\upsilon)| < \varepsilon$

CLUSTERING BASED ON RÉNYI ENTROPY

The entropy is a good measure of the quality of clustering. Therefore propose an entropy method to handle the problem of dimension reduction. The algorithm for dimension reduction consist of two main steps: First, find out reduced dimension with good clustering by entropy method. Second, identify clusters in the dimensions found. To identify good cluster, we set a threshold τ . A reduced dimension has good clustering if its entropy is below the threshold. The real number α is a entropic index that characterizes the degree of non-extensivity. The term belongs to BGS entropy when the limit $\alpha \to 1$.The approach starts by finding a large one-dimensional space with good clustering, this is the used to generate candidate 2-dimensional spaces which are checked against the data set to determine if they have good clustering. The process is repeated with increasing dimensionality until no more spaces with good clustering are found. The Rényi Entropy Dimension Reduction (REDR) algorithm is given as following:

REDR Algorithm

1. Let $\upsilon = 0$ and C_k be 1- dim space.

2. $\upsilon = \upsilon + 1$

3. For each space $c \in C_\upsilon$ do

3.1 $\ell(c) = renyi - ntropy(density(c))$.

3.2 if $\ell(c) < \tau$ then $S_\upsilon = S_\upsilon \bigcup c$ else $FS_\upsilon = c \bigcup FS_\upsilon$.

4. $C_{\upsilon+1} = orderdlist(FS_\upsilon)$

5. if no empty($C_{\upsilon+1}$) then goto 2.

6. Output of $\bigcup_\upsilon S_\upsilon$.

RESULTS

The REDR algorithm and the classical matching algorithms were tested on some synthetic image and data from the UCI data repository such as the Iris data. The algorithm was run 200 times on the synthetic images and iris data in order to obtain the similarity matrix and the average number of clusters.

(a) Iris Versicolor (b) Iris Virginica (c) Iris Setosa

Figure 1. different types of Iris plants used for classification

we tested the algorithm on the different data obtained from the UCI repository and got satisfactory results. The results presented in this section are on the Iris data. In 1935, a study was made of the differences in measurements, 4 dimensional and consists of a plants' morphology (width and length of petals and sepals) of 3 closely related iris flowers by Anderson. The Iris consists of three types of Iris plants: Iris Versicolor, Iris Virginica, and Iris Setosa with 50 instances per class. These flowers are difficult to recognize, since the genes number are in doubles forms : 6, 4 and 2 respectively, also the genetic structure locates their physical characteristics (one was a hybrid) was at the time, and now unknown. One class Iris Versicolor is well separated from the other two. The algorithm was able to obtain the 3-cluster solution when using the entropy constant $\alpha \to 1$ and λ's of 10.5 and 11. Two cluster solutions were also obtained using entropy constants of 14.5, 15, 15.5 and 16, Figure 1 shows the results of clusters number of different values of λ for the iris data and Percentage of correct classification.

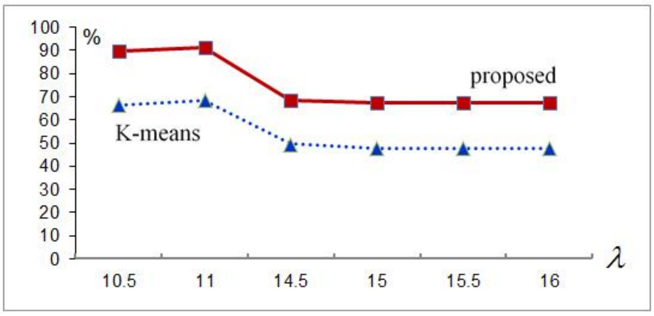

Figure 2 different values of λ with and Percentage of correct classification for the iris data.

To evaluate the performance of the algorithm, the percentage of data that were correctly classified for three cluster solution are determined. The results of it is compared direct K-means. The algorithm had a 95% correct classification while the direct K-means achieved only 70% percent correct classification, see Figure 2. Entropy is good measurement of correct classification, since each cluster can be computed using the entropy as follow

$$\ell(C_j) = -\sum_{j=1}^{k} (n_j^i / n_j) ln(n_j^i / n_j)$$

where n_j is the cluster size of C_j and n_j^i is sets size from cluster C_i that were allocated to cluster j. The overall entropy of the clustering is the sum of the weighted entropy of each cluster and is given by

$$\ell(C) = -\sum_{j=1}^{k} n_j \ell(C_j) / n$$

where n is the number of input patterns. The entropy is given in Figure 3. The lower the entropy the higher the cluster quality.

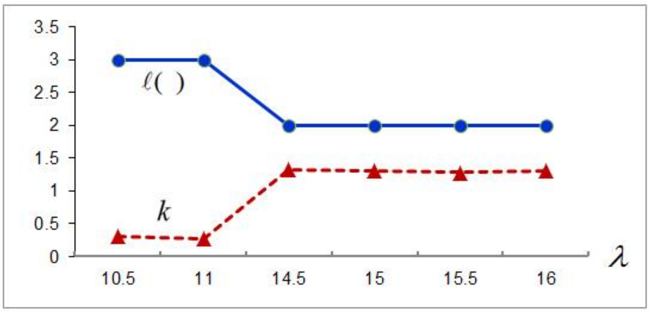

Figure 3 different values of λ with clusters number and entropy values for the iris data

The effect of λ and the different cluster sizes on the average value of k obtained. The results are given in Figure 4, show that for a given λ and different k value the average number of clusters converge.

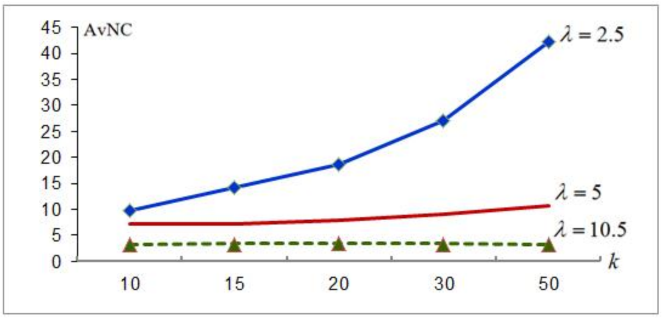

Figure 4 The average number of clusters (AvNC)for various λ and different k for the iris data

CONCLUSION

The paper presented a method based on biometric features, entropy properties and dimensionally reduction. The REK algorithm provided good estimates for the unknown number of clusters between biometric features. The approach worked on the data that we tested for dimensionally reduction, producing the required number of clusters. In second phase, the various properties of entropy applied on REDR algorithm to increase the accuracy factor of security recognition in this system. the paper provided REDR algorithm based on an entropy method that can be used for text dimension reduction of high dimensional data. It used coverage of data, density and correlation to determine the reduced dimension that have good clustering.

REFERENCES

[1] D. Lyon, “biometrics, identifications and surveillance”, Bioethics 22: 9,pp. 499-508, 2008.

[2] M. A. El-Sayed, proposed system of biometric authentication using palm print/veins with Tsallis entropy, Int. J. of Computer Science and Technology, vol. 6 (2), 9-14, 2015.

[3] Y. Ding, D. Zhuang , K. Wang, “A study of hand vein recognition methods”, Int. Conf. on mechatronics & automation, IEEE, Canada, 2005.

[4] S. Zhao, Y. Wang and Y. Wang, “Extracting hand vein patterns from low-quality images: a new biometric technique using low-cost devices”, 4 Int. Conf. on Image and Graphics, 2007.

[5] Cheng, J.; Liu, Q.; Lu, H. and Chen, Y.-W. "Supervised kernel locality preserving projections for face recognition". Neuro-computing. 67, 443-449,2005.

[6] H. Xie, J. Li, Q. Zhang, Y. Wang, "Comparison among dimensionality reduction technique based on random projection for cancer classification", Comput Biol Chem ,2016.

[7] A. G. Kaur, Ada, "A review on ear based biometric identification system", Int. J. of Advanced Research in Computer Science and Software Engineering. Vol 6 (3). Pp. 548-552, 2016.

[8] M. A. El-Sayed, M. Hassaballah, M. A. Abdel-Latif, "Identity verification of individuals based on retinal features using Gabor filters and SVM", J. of Signal and Information Processing, 7, 49-59 ,2016.

[9] M. A. El-Sayed and M. A. Abdel-Latif," Optimization methods for identity verification system using biometric features", Int. J. of Applied Engineering Research, 12(18), 7143-7153, 2017.

[10] S.-C. Fang, J. R. Rajasekera, and H.-S. J. Tsao, "Entropy optimization and mathematical programming", Kluwer Academic Pub., Norwell, 1997.

[11] M. A. El-Sayed, "Edges detection based on Renyi entropy with split/merge". Computer Engineering and Intelligent Systems, 3(9), 32-41 ,2012.

[12] M. A. El-Sayed, "Algorithm based on histogram and entropy for edge detection in gray level images", Int. J. of Computers & Technology, 11, 2207-2215, 2013.

[13] K. Frenken, "Entropy statistics and information theory", the elgar companion to neo-schumpeterian economics, Cheltenham,Edward Elgar Publishing, 544-555, 2007.

[14] A. Renyi, "On measures of entropy and information", in: Proceedings of the Fourth Berkeley Symposium on Math. Statist. Prob. 1, 547-561, 1961.

[15] M. A. El-Sayed, S. F. Bahgat, and S. Abdelkhalek, "New approach for identity verification system using the vital features based on entropy", Int. J. of Computer Science Issues, 10(6), 2, 11-17, 2013.

[16] M. Figueiredo and A. Jain, "Unsupervised learning of finite mixture models", Pattern Analysis and Machine Intelligence- IEEE Trans. , 24(3) , 381-396, 2002.