

# Face recognition based on manifold learning and Rényi entropy

Wen-Ming Cao, Ning Li

Intelligent Information Processing Key Laboratory, Shenzhen University, Shenzhen, China; [wmcao@szu.edu.cn](mailto:wmcao@szu.edu.cn)

Received 10 September 2009; revised 29 October 2009; accepted 9 November 2009.

## ABSTRACT

Though manifold learning has been successfully applied in wide areas, such as data visualization, dimension reduction and speech recognition; few researches have been done with the combination of the information theory and the geometrical learning. In this paper, we carry out a bold exploration in this field, raise a new approach on face recognition, the intrinsic  $\alpha$ -Rényi entropy of the face image attained from manifold learning is used as the characteristic measure during recognition. The new algorithm is tested on ORL face database, and the experiments obtain the satisfying results.

**Keywords:** Manifold Learning; Rényi Entropy; Face Recognition

## 1. INTRODUCTION

Face recognition has becoming a research hotspot in the fields of image processing, pattern recognition and artificial intelligence in recent years. Numerous research papers appear on the famous international publications, a great deal of capital and manpower is invested to this research and its relevant application system developments. However, the performance of the face recognition could be influenced by many factors, such as illumination, gesture, age, facial expressions, image resolution, and noise, etc; which cause difficulties for the computer processing of face recognition, and turn it into a challenging task at the same time.

The existing face recognition methods can be roughly classified into two categories [1]: local feature based, and global feature based. A local feature based method symbolizes a human face with the extracted feature vectors (eyes, nose, mouth, hair, and face contours), designs certain kinds of classifiers to make recognition. On the other hand, a global feature based method employs the whole images as the input feature vectors, and then low-dimensional features are extracted by some learning

algorithms. The main difference between the two categories lies in the way how the features are extracted. In a local feature based method, features are designed completely by the algorithm designers; while in global feature based method, features are automatically extracted or learned by some self-learning algorithms.

Local feature based or learning based methods can be further divided into two classes: 1) statistical learning, such as artificial neural networks (ANN) [2-4], support vector machine (SVM) [5,6], and Boosting [7,8]; 2) manifold learning(or dimensionality reduction), such as linear methods like PCA [9,10], LDA [11,12], and nonlinear methods like Isomap [13], LLE [14], Laplacian Eigenmaps [15,16].

In recent years, nonlinear dimensionality reduction (NLDR) methods have attracted great attentions due to their capability to deal with nonlinear and curved data [1]. All these algorithms rely on an assumption that the image data lie on or close to a smooth low-dimensional manifold in a high-dimensional input image space. A big limitation of NLDR algorithms is the way how to estimate the intrinsic dimension of the manifold. LLE and Laplacian Eigenmaps do not give method to estimate the intrinsic dimension; Isomap shows a simple way to estimate the intrinsic dimension by searching for the “elbow point” where the residual error decreases significantly. However, for some real data, it is difficult to find an obvious “elbow point” to indicate the intrinsic dimension.

The intrinsic dimensionality estimation of a data set is a classical problem of pattern recognition. From the math point, the intrinsic dimension of a manifold is the dimension of the vector space that is homeomorphic to local neighborhoods of the manifold; in other words, intrinsic dimension describes how many “degrees of freedom” are necessary to generate the observed data. When the samples are drawn from a large population of signals one can interpret them as realizations from a multivariate distribution supported on the manifold. The intrinsic entropy of random samples obtained from a manifold is an information theoretic measure of the complexity of this distribution. The entropy is a finite value when the distribution satisfies the restriction of

Lebesgue integral on the lower dimensional manifold.

In 2003, Costa and Hero proposed an algorithm that can jointly estimate both the intrinsic dimension and intrinsic entropy of a dataset randomly distributed on a smooth manifold [17,18]. The algorithm constructs the Euclidean K-NN graph over all the sample points and use its growth rate to estimate the intrinsic dimension and entropy by simple linear least squares and method of moments procedure. This approach allows for the estimation of the desired quantities using algorithms with low computational complexity that avoid reconstructing the manifold or estimating multivariate distributions.

Considering the fact that on the various conditions of distance, direction, illumination or gestures, one object can form different face images; the corresponding high-dimensional dataset of those images can be viewed as a nonlinear low-dimensional embedded manifold determined by factors of illumination, location, scale, gesture, etc. Based on Costa’s algorithm, we take Rényi entropy as the characteristic vector; present a novel face recognition method—Face Recognition Based on Manifold Learning and Rényi Entropy.

## 2. MANIFOLD LEARNING USING K-NEAREST NEIGHBOR GRAPHS

Costa’s algorithm is based on minimal Euclidean graph methods. Let  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  be  $n$  independent identically distributed (i.i.d.) random sample points in a compact subset of  $\mathbb{R}^d$ , with multivariate Lebesgue density  $f$ .  $x_n$  is also called the set of random vertices on a minimal Euclidean graph. First, a Euclidean k-nearest neighbors (k-NN) graph is constructed over all the sample points. Let  $N_{k,i}(\mathcal{X}_n)$  be the set of  $k$ -nearest neighbors of  $x_n$  in  $X_i$ . Then the total edge length functional of the Euclidean k-NN graph is:

$$L_\gamma^{K-NN}(\mathcal{X}_n) = \sum_{i=1}^n \sum_{X \in N_{k,i}(\mathcal{X}_n)} d^\gamma(X, X_i)$$

where  $\gamma$  is a power weighting constant,  $d(X, X_i)$  is the Euclidean distance between  $X$  and  $X_i$ . The k-NN graph exhibits an asymptotic behavior of their total edge functional:

$$\lim_{n \rightarrow \infty} L_\gamma(\mathcal{X}_n) / n^\alpha = \beta_{d,L_\gamma} \exp\{(1-\alpha)H_\alpha(f)\}$$

where  $\beta_{d,L_\gamma}$  is a constant independent of  $f$ , and

$$H_\alpha^{\mathbb{R}^d}(f) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^\alpha(x) dx$$

is the extrinsic Rényi  $\alpha$ -entropy of the multivariate Lebesgue density  $f$ . So the asymptotic unbiased and strongly consistent estimator of the  $\alpha$ -entropy is:

$$\hat{H}_\alpha(\mathcal{X}_n) = \frac{1}{1-\alpha} \left[ \log L_\gamma(\mathcal{X}_n) / n^\alpha - \log \beta_{d,L_\gamma} \right]$$

The convergence results for k-NN graphs are extended from Euclidean spaces to general Riemannian manifolds. Let  $(M, g)$  be a compact smooth Riemann  $m$ -dimensional manifold. Suppose  $y_n = \{Y_1, \dots, Y_n\}$  are i.i.d. random elements of  $M$  with bounded density  $f$  relative to the volume element  $\mu_g$ . Let  $\hat{L}_\gamma$  be the total edge length of the k-NN graph. Assume  $m \geq 2, 1 \leq \gamma < m$  and let  $\alpha = (m - \gamma) / m$ . Then

$$\lim_{n \rightarrow \infty} \frac{\hat{L}_\gamma(y_n)}{n^\alpha} = \beta_{m,L_\gamma} \int_M f^\alpha(y) \mu_g(dy)$$

where  $\beta_{m,L_\gamma}$  is a constant independent of  $f$  and  $M$ . Accordingly, the asymptotic unbiased and strongly consistent estimator of the  $\alpha$ -entropy  $H_\alpha^{(M,g)}(f)$  is:

$$\hat{H}_\alpha^{(M,g)}(y_n) = \frac{m}{\gamma} \left[ \log(L_\gamma^M(y_n) / n^{(m-\gamma/m)}) - \log \beta_{m,L_\gamma} \right]$$

Define  $l_n = \log \hat{L}_\gamma(y_n)$ , the convergence theorem (5) suggests the log-linear model below:

$$l_n = a \log n + b + \varepsilon_n$$

where

$$a = (m - \gamma) / m,$$

$$b = \log \beta_{m,L_\gamma} + \gamma / m H_\alpha^{(M,g)}(f)$$

and  $\varepsilon_n$  is an error residual that goes to zero as  $n \rightarrow \infty$ .

Bootstrap resampling is used here to estimate mean graph length  $E[L_\gamma^M(y_n)]$ , and linear least squares (LLS) is applied to jointly estimate slope  $\hat{a}$  and intercept  $\hat{b}$  from sequence  $\{\log E[L_\gamma^M(y_n)], \log n\}_n$ . After that, the following estimates of dimension  $\hat{m}$  and entropy  $\hat{H}$  are obtained by inversion of the relations:

$$\hat{m} = \text{round}\{\gamma / (1 - \hat{a})\}$$

$$\hat{H}_\alpha^{(M,g)} = \frac{m}{\gamma} (\hat{b} - \log \beta_{\hat{m},L_\gamma})$$

The specific steps of the algorithm are described as:

- 1) Using entire database of signals  $y_n = \{Y_1, \dots, Y_n\}$  to construct K-NN graph.
- 2) Estimate the dimension and  $\alpha$ -Rényi entropy of the manifold that the sample sets lie in.
  - a) Set parameters  $M, N, p_1, \dots, p_Q, (p_1 < \dots < p_Q \leq n)$ ;
  - b) Initialize  $\bar{m} = 0, \bar{H} = 0, M' = 1$ ;
  - c) Choose  $P$  from  $p_1, \dots, p_Q$  in turn,
    - I.  $\bar{L} = 0; N' = 1$ ;

II. Randomly select a subset of  $P$  signals  $y_p$  from  $y_n$ ; Compute graph total edge length  $L_p$ ;

$$\bar{L} = \bar{L} + L_p;$$

III.  $N' = N' + 1$ ; if  $N' = N$ , goto step IV, else goto II;

IV. Compute sample average graph length

$$\hat{E}[\hat{L}(y_p)] = \bar{L} / N;$$

d) Use **Eq.9** to estimate dimension  $\hat{m}$  and entropy

$$\hat{H} \text{ from } \left\{ \hat{E}[\hat{L}(y_p)] \right\}_{p=p_1}^{p_0};$$

$$\bar{m} = \bar{m} + \hat{m}_{M'}, \bar{H} = \bar{H} + \hat{H}_{M'};$$

e)  $M' = M' + 1$ , if  $M' = M$ , goto step (f), else goto (c).

f) The estimate dimension  $\hat{m} = \bar{m} / M$ ;  $\alpha$ -Rényi entropy  $\hat{H} = \bar{H} / M$ .

g). End.

The parameters  $M$  and  $N$  are used to provide a trade-off between the bias and variance performance of the estimators for finite  $n$ .  $\beta_{m,L}$  is the limit of the normalized length functional of the corresponding Euclidean entropic graph for a uniform distribution on the unit cube  $[0,1]^m$ . It can be determined by performing Monte Carlo simulations of the entropic graph length on the unit cube  $[0,1]^m$  for uniform random samples.

Unlike previous solutions, Costa's algorithm can prove statistical consistency of the obtained estimators under weak assumptions of compactness of the manifold and boundedness of the (Lebesgue) sampling density supported on the manifold. This approach allows for the estimation of the desired quantities using algorithms with low computational complexity ( $O(n \log n)$ ) that avoid reconstructing the manifold or estimating multivariate distributions.

### 3. FACE RECOGNITION ALGORITHM BASED ON MANIFOLD'S RÉNYI ENTROPY

We applied the method to a real high-dimensional dataset with unknown manifold structure and intrinsic dimension and entropy---face images. We chose "ORL Face Database" [19] of AT&T Laboratories Cambridge. There are 10 different images of each of 40 distinct people (40 sample sets). The images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses).

The original image is in PGM format, 92\*112 pixels, with 256 grey levels per pixel. We first processed each image into BMP format, 64\*64 pixels, normalized the pixel values between  $[0, 1]$ . Then we arranged each im-

age in a 4096\*1 matrix using the common lexicographic order.

From each of the 40 sample sets  $s_i$ , ( $i = 1, \dots, 40$ ), we randomly chose 3,4,5,6,7,8 images as the training samples; hence there were 40\*3=120, 40\*4=160, 40\*5=200, 40\*6=240, 40\*7=280, 40\*8=320 samples in the training set respectively. All 400 images were treated as test samples. Each 5,6,7,8 images of the same sample set (4096\*5, 4069\*6, 4069\*7, 4069\*8 matrix) were trained at one time to get an estimate Rényi Entropy  $\hat{H}_{s_i}$  according to the algorithm introduced in section 2. The estimate entropy served as the characteristic vector of the recognition, each test image  $X$  was combined with every 3,4,5,6,7,8 sample images in one sample set (4096\*n,n=3,4,5,6,7,8 matrix respectively) to get an estimate Rényi entropy  $\hat{H}_X$ , then the classification criterion is :

$$\left\{ X \in s_k \mid \hat{H}_X - \hat{H}_{s_k} = \min_{i=1, \dots, 40} (\hat{H}_X - \hat{H}_{s_i}) \right\}$$

$X$  would be classified to the set where its combined entropy is nearest to the origin entropy of the training set.

### 4. EXPERIMENT RESULTS

The parameters used in the experiment were set as follows:  $k = 3, \gamma = 1, M = 1, N = 5, n = 6$ . **Figure 1** shows the 5 training images of face 1; and **Figure 2** displays the results of running 20 simulations of the Costa's algorithm



Figure 1. The 5 training images of face 1.

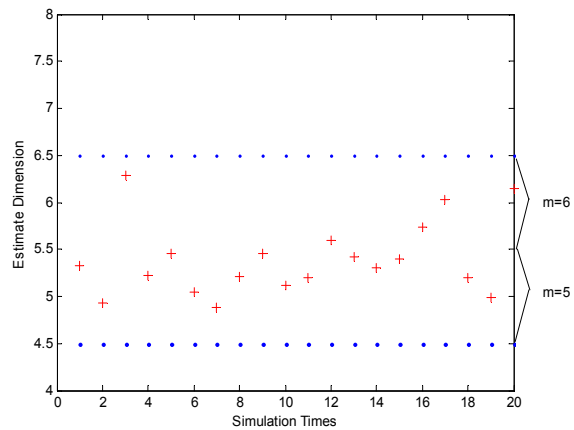


Figure 2. The estimate manifold dimension of face 1 is between 5~6.

using the 5 training images in **Figure 2**. In same conditions, as shown in Table 3, our Rényi Entropy method gives a weakly higher recognition accuracy compared with PCA and 2DPCA methods for ORL face database in **Figure 3**.

**Table 1** shows a part of estimate results of the training set from ORL face database.

Using the Face recognition algorithm in section3, the results of our method are showed in **Table 2**.

The results indicate that the face recognition algorithm based on Manifold's Rényi Entropy we present in this paper works effectively. The K-NN graph is employed so that we avoid the complex estimation of geodesic distances on the manifold.

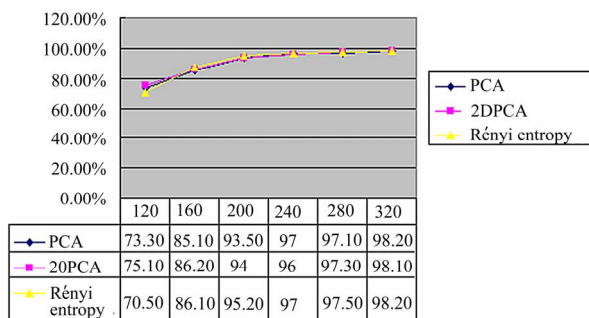
An important factor influenced the recognition efficiency is the sample number of the same set we used is a little small (only 10), which induce the lack of information needed for training samples. Besides, there must be certain inevitable information loss during the procedure of image pretreatment. Further work includes add bootstrap confidence intervals for the estimators and apply other classification method to minimize the error.

**Table 1.** Dimension estimate  $\hat{m}$  and entropy  $\hat{H}$  for some training face images

	$\hat{m}$	$\hat{H}$ (bits)
Face 1	5	20.856
Face 4	6	19.890
Face 7	6	21.052
Face 14	5	20.175
Face 17	5	18.706
Face 39	5	17.804

**Table 2.** The results of Face recognition algorithm based on Manifold's Rényi Entropy.

Training samples	Computational complexity	Test samples	Error recognition	Correct Recognition rate
120	7	400	118	70.5%
160	8	400	51	87.2%
200	11	400	19	95.2%
240	13	400	12	97%
280	18	400	10	97.5%
320	21	400	7	98.2%



**Figure 3.** PCA, 2DPCA, Rényi entropy.

## 5. CONCLUSIONS

We have presented a new face recognition algorithm based on Manifold's Rényi Entropy. The algorithm is applied to ORL face database and the experiments obtain the satisfying results. At present we are studying the theoretic prove of this method, and trying to make further utilizations in other research fields such as pattern recognition and artificial intelligence.

## 6. ACKNOWLEDGMENTS

The work is supported by the National Science Foundation of China (No.60871093), and Pre-Research and Defense Fund of China: No.9140C80002080C80.

## REFERENCES

- [1] Lin, T. (2005) Machine perception of human faces: a grand challenge. The Korea Foundations for Advanced Studies (KFAS).
- [2] Rowley, H.A., Baluja, S. and Kanade, T. (1998) Neural network-based face detection [J]. *IEEE Trans Pattern Analysis and Machine Intelligence*, **20(1)**, 23-38.
- [3] Feraud, R., Bernier, O.J., Villet, J.E. and Collobert, M. (2001) A fast and accurate face detector based on neural networks [J]. *IEEE Trans Pattern Analysis and Machine Intelligence*, **22(1)**, 42-53.
- [4] Garcia, C. and Delakis, M. (2004) Convolutional face finder: a neural architecture for fast and robust face detection [J]. *IEEE Trans Pattern Analysis and Machine Intelligence*, **26(11)**, 1408-1423.
- [5] Osuna, E., Freund, R. and Giroso F. (1997) Training support vector machines: an application to face detection [J]. *Proc, IEEE Conf Computer Vision and Pattern Recognition*, 130-136.
- [6] Phillips, P.J. (1998) Support vector machines applied to face recognition [J]. *Adv. Neural Inform. Process. Syst.* **11**, 803-809.
- [7] Viola, P. and Jones, M. J. (2004) Robust real-time face detection [J]. *Int. J. Computer Vision*, **57(2)**.
- [8] Li, S.Z. and Zhang, Z. (2004) Float boost learning and statistical face detection [J]. *IEEE Trans Pattern Anal. Mach. Intel.*, **26(9)**, 1112-1123.
- [9] Kirby, M. and Sirovic, L. (1990) Application of the karhunen-loeve procedure for the characterization of human faces [J]. *IEEE Trans Pattern Anal. Mach. Intel.*, **12(1)**, 103-108.
- [10] Turk, M. and Pentland, A. (1991) Eigenfaces for Recognition [J]. *J. Cogn. Neurosci.*, **3**, 72-86.
- [11] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19(7)**, 711- 720.
- [12] Kim, T.K. and Kittler, J. (2005) Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27(3)**, 318-327.

- [13] Tenenbaum, J.B., de Silva, V. and Langford, J.C. (2000) A global geometric framework for nonlinear dimensionality reduction [J] *Science*, **290**, 2319-2323.
- [14] Roweis, S.T. and Saul L.K. (2000) Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, **290**, 2323-2326.
- [15] Belkin, M. and Niyogi, P. (2003) Laplacian eigenmaps for dimensionality reduction and data representation [J]. *Neural Computation*, **15(6)**, 1373-1396.
- [16] He, X., Yan, S., Hu, Y., Niyogi, P., and Zhang, H. (2005) Face recognition using laplacianfaces [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27(3)**, 328-340.
- [17] Costa, J.A. and Hero, A.O. (2003) Entropic graphs for manifold learning [J]. *In the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, 316- 320.
- [18] Costa, J.A. and Hero, A.O. (2004) Manifold learning using *k*-nearest neighbor graphs [J]. *Proc. of IEEE Int. Conf. on Acoust. Speech and Signal Processing*. Montreal, Canada.
- [19] ORL, *Face Database*, AT&T Laboratories Cambridge. [DB/OL].  
[http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att\\_faces.zip](http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.zip).