

Fuzzy Shannon Entropy for Face Recognition

Abdullah Gubbi, Mohammad Fazole Azeem, Mohammed Rafeeq, Shoab Kamal

Abstract— This paper presents a new local feature called Fuzzy Shannon entropy based features for face recognition. These features are computationally simple and capture the variation in the face image. The new Fuzzy membership functions are defined which are not like standard Gaussian or trapezoidal membership functions, but they are computed with normalized Image intensity values. This paper gives the flexibility to design of new membership functions to cater the need of the problem. The Information set is the combination of fuzzy membership and Information source (attributes). By applying this concept, we have extracted features from ORL and Faces94 databases with Support Vector Machine and K Nearest Neighbour classifier the results obtained with the proposed method are better than other published contemporary techniques.

Keywords— Face recognition, Fuzzy logic, Shannon Entropy, K- Nearest Neighbour, Support Vector Machine and Information set

I. INTRODUCTION

Face is one of the most frequently used biometric trait for recognition. Face Recognition (FR) is a predominant field in image processing. It can be utilized in many applications such as biometrics, surveillance, law enforcement, Airport, Robotics, Human-computer interaction (HCI) systems and computer vision. Face recognition system basically utilizes two major modules i.e., training and testing. The features extracted from the face image is utilized for training and testing set for a classifier. Face recognition is effortlessly done by human beings, perhaps it is a complex task for the machines since it involves lots of variations of image appearance, such as (i) Occlusion, (ii) Pose variation (front, non-front), (iii) illumination condition, (iv) Image orientation and (v) Facial expression.

Many novel face recognition methods have been proposed in the last several decades owing to the increase demands for applications related to it. The methods involved in face recognition are predominantly classified in detail [1]. Some of the popular algorithms for FR are based on Holistic methods (Principle Component Analysis, Linear Discriminant Analysis Successive PCA and Prioritized PCA) and Template based methods (Features extracted from eyes, nose, ears, chin, lips etc.)

Authors [2] proposed a context-aware local binary feature learning (CA-LBFL) method for face recognition which exploits the contextual information of adjacent bits by constraining the number of shifts from different binary bits,

Revised Manuscript Received on April 12, 2019.

Abdullah Gubbi, Dept. of Electronics and Communication Engineering, Bearys Institute of Technology, Mangalore, India. (E-mail: abdullahgubbi@yahoo.com)

Mohammad Fazole Azeem, Dept. of Electrical Engineering, King Khalid University, Aseer Province, KSA. (E-mail: mf.azeem@gmail.com)

Mohammed Rafeeq, Dept. of Electronics and Communication Engineering, Bearys Institute of Technology, Mangalore, India. (E-mail: rafeeqatpace@gmail.com)

Shoab Kamal, Dept. of Electronics and Communication Engineering, Bearys Institute of Technology, Mangalore, India. (E-mail: shoabkamal87@gmail.com)

in order to exploit robust information for face representation. In this process, they first extract the pixel difference vectors in local patches, and create a discriminative mapping in an unsupervised manner. They then created a codebook made up of histogram for classification.

In this paper [3], authors have proposed an approach in order to overcome the representation and matching problems for an age invariant face recognition. A new maximum entropy feature descriptor (MEFD) was developed that encodes the facial images into a set of discrete codes using maximum entropy. They also proposed a new matching method, known as identity factor analysis (IFA) to estimate the probability that two faces have the same underlying identity.

Authors [4] here proposed a Local Quantized Gradient Direction (LQGD) descriptor which utilizes eight levels of quantized gradient magnitude and direction information from the horizontal and vertical gradients from 3x3 local neighbourhood. Contourlet transform is used for image pre-processing. From LQGD they construct the histogram for extracting the facial features. These features are fused with the contourlet transform features and they have applied Linear Discriminant Analysis (LDA) for obtaining the final features.

The authors [5] have used Principle Component Analysis for dimensionality reduction on full faces and half faces (both left side and right side). They have used half faces based on the fact that the face is vertically symmetrical. For classification they have used Euclidean classifier on ORL database.

Authors [6] have proposed a technique in which they use discrete cosine transform and logarithmic transform for robust facial recognition system. The entropy is calculated from the DCT coefficients extracted from the image. Feature vector is created based on the pixels that preserve maximum entropy. Then, Kernel Entropy Component Analysis (KECA) is applied on the feature dataset. They have used the nearest neighbour classifier for different similarity measures.

Fuzzy Logic (FL) theory deals with uncertainties [7], where the boundaries are not crisp or deals with partial truth. and it has got wide applications in the field of control and signal processing. FL has been able to touch almost all the fields. Fuzzy logic generally refers to fuzzy sets, as a tool to model the ambiguity and vagueness in complex systems [8]. Fuzzy Logic (FL) is effectively utilized for the complicated and vague processes in order to obtain the solutions for the scenario where there is no mathematical model or the mathematical model developed is highly non-linear. Since its origin, FL has obtained numerous revenues.

The motivation for the present work comes from the two evident drawbacks of fuzzy sets such as treating the information source (attributes) and their membership functions separately in all the problems dealing with the fuzzy logic theory and not accounting for probability (the occurrence) of information sources. The membership function that delivers the degree of association of a particular information source value in the fuzzy sets is always considered ignoring the information source value. In the real world, the information is a combination of information source and an agent's opinion about it (membership function). An attempt is made to overcome these problems in the present study by introducing the concept of information set that represents the amount of information or uncertainty in an information source. The theory of Information sets and Information processing developed by authors [9] has great potential to expand the scope of fuzzy sets and fuzzy modelling in biometrics.

The organization of the paper is as follows: Section II describes the concept of entropy. The details about information sets and information processing is explained in section III. Section IV discusses about the implementation of fuzzy Shannon feature along with various classifiers. In Section V the results are discussed and compared with the results of the contemporary method. The conclusions are drawn in Section VI.

II. CONCEPT OF ENTROPY

In information theory, The entropy also known as average information of an image is a measure of the degree of randomness in the image. The entropy usually refers to the Shannon entropy (SE), which quantifies the expected value of the information contained in a message. The probabilistic behaviour of a source of information can be estimated by entropy. If A is a random event with probability P(A), then the information contained in the event A is given by

$$I(A) = \log \left[\frac{1}{p(A)} \right] = -\log[p(A)] \tag{1}$$

In case of random variable, if events are given e_1, e_2, \dots, e_m occurring with probabilities, p_1, p_2, \dots, p_m the Shannon entropy is defined as [10]:

$$H = \sum p(-\log p) = \sum p \log p^{-1} \tag{2}$$

Where p is probability and log p is logarithmic gain.

The Shannon entropy can also be applied to a face image. The probability distribution of a gray scale image is estimated based on the number of gray values and their occurrences. The entropy will be lower if the image contains no variation in the intensities. On the contrary, the entropy will be higher, if there is large variation in the intensities of an image. The entropy of an image is single valued; this single value if used as a feature leads to an increased classification error. In order to overcome the above problem in this work the image is partitioned into smaller patches and we compute the entropy of the window using the information set concept proposed by Hanmandlu et al [9]. We have adopted the SE computation differently which is explained in the implementation.

The present work is concerned with the formulation of the Information set based on Shannon entropy. The motivation for the present work comes from the three drawbacks one is from Shannon entropy and two from the fuzzy set. The problem with Shannon entropy for an image is that it gives a single value, which cannot be treated as a feature. The first drawback of fuzzy set is that it treats information source (attributes) and their membership functions separately. The second drawback deals with the probability that is not taken into the account, for (the occurrence) of information sources. In the real world, the information is a combination of information source and an agent's opinion about it (membership function).

The support vector machine (SVM) classifiers [11] and K nearest neighbour (KNN) [12] have been used for the purpose of classification.

III. INFORMATION SETS AND INFORMATION PROCESSING

For holistic approaches like PCA which computes the covariance matrix consisting of eigen vectors at the end we may not be able to know which eigen vector belongs to which face. Instead of considering global representation of image, we consider the local features of the image by partition image into small patches. While we partition we don't consider it to be part of eye, forehead, lips and ears of face image. These patches are non-overlapping type referred as granulization, from which we collect local information. The membership function defines the uncertainty associated with Fuzzy set, the membership function also gives the degree of association of each element to the fuzzy set. The total uncertainty is computed by entropy function, the intensity values are termed as information source values (attributes) for an image.

When the fuzzy set is considered in the context of the information theoretic entropy, the pair becomes a product leading to the concept of the Information set. Some aspects of the theory of Information sets have been developed recently by Prof. M. Hanmandlu [9]. The objective is to compute the entropy of the information source values from the patches.

The Shannon Entropy is specially designed to handle the probabilities. If, the Intensities of the image are treated as probabilities in the Shannon entropy then their total should be equal to one. The drawback related to information sources (intensities of the image) values taken as probabilities results in a very small value as we are dividing the complete image with a global maximum to normalize the information source values. Due to this, the sum of probabilities may not be equal to one. By assigning the free parameters and the membership function computed for the window may be considered as the fuzzy set. The set formed by the membership function and the gray scale values may be treated as the information set.

With reference to the above recollect the Shannon entropy function (Shannon 1948) which has no free parameters to represent the statistics of a fuzzy set. Shannon entropy



purely deals with probabilities. If the information source values are utilized in these entropies then they are treated as probabilities and their sum must be equal to 1. The problems encountered regarding the information sources as probabilities are that they are too small which results in the entropy to be small. The smaller values of entropy results in lesser discriminating power in solving the problem. In order to avoid this, the information source values are normalized by a global maximum and the probabilities are replaced by these values. By this, the sum of information source values may not be unity as is required for the sum of the probabilities in (3). In this context, let us recall the Shannon entropy function (Shannon 1948) expressed as this local information set constitute a fuzzy set. Fuzzy set is a combination of gray levels $I = \{I_{ij}\}$ in a window and membership value denoted by $\{\mu_{ij}\}$. $\{\mu_{ij}\}$ is calculated for each window in the image. The membership function values represent the degree of association of gray levels to the set. Thus, the pair formed by the gray level and the membership function is considered as fuzzy set. The nature of uncertainty in a fuzzy set can be statistical and non-statistical.

The normalized information source values are computed using the global maximum as

$$I_n(i, j) = \frac{I(i, j)}{I_{g_{max}}} \quad (3)$$

Where $I(i, j)$ represents the intensity value of a pixel and $I_{g_{max}}$ represents the global maximum intensity value. Now $I_n(i, j)$ is called normalized information source value. The Membership function in the form of exponential form is given in equation 4.

One can use the local statistics I_{ref} as I_{wmax} or I_{wmin} or I_{wavg} for each window. Where I_{wmax} , I_{wmin} , and I_{wavg} represent maximum, minimum and average of window under consideration.

The fuzzifier f_h^2 in (3) is devised by Hanmandlu et al. in [8] and it gives the spread of attribute values with respect to the chosen reference Where f_h is given by

$$f_h^2 = \frac{\sum_{i=1}^W \sum_{j=1}^W (I_{ref} - I(i, j))^4}{\sum_{i=1}^W \sum_{j=1}^W (I_{ref} - I(i, j))^2} \quad (4)$$

$$\mu_{ij}^e = e^{-\left\{ \frac{|I(i,j) - I_{ref}|}{f_h^2} \right\}} \quad (5)$$

Membership function in the form of exponential form is given in equation 5

$$\mu_{ij}^g = e^{-\left[\frac{I(i,j) - I_{ref}}{\sqrt{2} f_h} \right]^2} \quad (6)$$

Equation 6 represents the membership function in the Gaussian form.

IV. IMPLEMENTATION

A. Database used and implementation

Through the information set, new facial features are created. To extract the features, the face image is partitioned into small patches i.e., 3x3, 5x5 or 7x7. The borders can be padded with zeros to have all the information captured wherein each patch delivers a feature. The membership function represents the distribution of the source which can be Gaussian, trapezoidal etc. In our implementation we have considered the membership function given by equation 6

$$\mu_{ij} = e^{-\left[\frac{I(i,j) - I_{ref}}{f_h^2} \right]^2} \quad (7)$$

In fuzzy logic, the degree of association of the information source (attributes) to that of the fuzzy set is represented by the value of the membership function. The product of normalized information source value and the membership function value leads to information set. Once the μ_{ij} is computed the information set is computed for each value in the window using equation 8

$$H_1 = \sum_{i=1}^3 \sum_{j=1}^3 I_n(i, j) \log \mu_{ij} * I_n(i, j) \quad (8)$$

Another set of features can be computed by changing the position of membership

$$H_2 = \sum_{i=1}^3 \sum_{j=1}^3 \mu_{ij} * I_n(i, j) \log I_n(i, j) \quad (9)$$

Where H_1 and H_2 represents the fuzzy Shannon features computed from the non-overlapping windows or patches. The above equation is mathematically represented by information set $H = \{I_n(i, j) \mu_{ij}\}$. whose elements $I(i, j)$ is the information source or information values and μ_{ij} is the membership values. Thus, different information forms can be obtained by changing μ_{ij} .

In fuzzy applications, only the membership function is given the importance. The value of the information source is rarely used. In general, the membership function can be any arbitrary function to cater the need i.e., it needn't be always Gaussian, trapezoidal, exponential etc. The membership function can be treated as an agent's view for a given problem; hence one can design his own membership function to enhance the scope of fuzzy set. It can be the complement of membership function also or the membership can be derived from the other form of same information source. Thus, the information set may be prepared by different membership functions keeping the information source same that leads to many new features

Fig.1. shows the gray scale face images of some of the subjects in ORL database and Fig.2. shows the color face images of some of the subjects in Faces94 database.





Fig. 1. Gray scale face images of some of the subject in ORL database



Fig. 2. Color face images of some of the subject in Faces94 database

B. Classifiers

K nearest neighbour classifier (KNN) is best suited for supervised classification. It doesnot require underlying structure for classification. Support Vector Machine (SVM) proposed by Vapnik et al [11] is a statistical tool which uses linear and polynomial kernel as the surfaces for the classification. More details on SVM and KNN can be found in [11,12]

V. RESULTS AND DISCUSSIONS

A. Results

The ORL database has 40 subjects with each subject having 10 samples. These images are resized so that they are exactly divisible by 3x3 ,5x5,7x7. Some of the samples have glasses on their faces. The faces94 database has a total of 153 subjects wherein 20 are female and rest are male.

Table I and II exhibit the results of fuzzy Shannon entropy method on ORL face database having original image size of 112X92 which is resized to 99X99 to get total of 1089 features . The windows are non-overlapping type. Table III and IV exhibit the results of fuzzy Shannon entropy method on Faces94 database having original image size of 180X200 which is resized to 90X90 to get total of 900 features.

The features computed using equation (8) represent the Fuzzy Shannon features for a face image. Using SVM classifier gives a recognition rate of 97.5% on ORL database, with polynomial degree 1 for 5x5 and 7x7 window. The same results are replicated for polynomial degree 2 and 3 with window size of 3x3 as well as for 7x7 which represents that the obtained features are robust for the classifier. KNN classifier has given a recognition accuracy of 96.66 % on ORL for 3x3 as well as 5x5 window and recognition rate of 94.16% with 7x7 window. The number of features reduces as the window size increases. The experiments are conducted utilizing 70% training and 30 % testing images.

The features computed using equation (9) represent the Fuzzy Shannon features for a face image. Using SVM classifier gives a recognition rate 97.5% on ORL database, with polynomial degree 1 for 5x5 and 7x7 window. The same results are replicated for polynomial degree 2 and 3 with window size of 3x3 as well as for 7x7. In our proposed similarity measure with KNN has given 95.83% result on ORL for 3x3 as well as 5x5 window and recognition rate of 96.66% with 7x7 window. The experiments are conducted utilizing 70% training and 30 % testing images.

TABLE I. RECOGNITION RATES (IN PERCENTAGE) WITH FUZZY SHANNON FEATURES COMPUTED USING EQUATION (8) FOR ORL DATABASE

Window size	KNN	SVM(PR-Tools)		
		Poly1	Poly2	Poly3
3x3	96.66	96.66	97.5	97.5
5x5	96.66	97.5	96.66	96.66
7x7	94.16	97.5	97.5	97.5

TABLE II. RECOGNITION RATES (IN PERCENTAGE) WITH FUZZY SHANNON FEATURES COMPUTED USING EQUATION (9) FOR ORL DATABASE

Window size	KNN	SVM(PR-Tools)		
		Poly1	Poly2	Poly3
3x3	95.83	96.66	97.5	97.5
5x5	95.83	97.5	96.66	96.66
7x7	96.66	97.5	97.5	97.5

TABLE III. RECOGNITION RATES (IN PERCENTAGE) WITH FUZZY SHANNON FEATURES COMPUTED USING EQUATION (8) FOR FACES94 DATABASE

Window size	KNN	SVM(PR-Tools)		
		Poly1	Poly2	Poly3
3x3	100	100	80	95
5x5	97.66	98.5	85	98.5
7x7	96.16	96.66	90	100

TABLE IV. RECOGNITION RATES (IN PERCENTAGE) WITH FUZZY SHANNON FEATURES COMPUTED USING EQUATION (9) FOR FACES94 DATABASE

Window size	KNN	SVM(PR-Tools)		
		Poly1	Poly2	Poly3
3x3	99	99.05	97	96
5x5	98.16	98.75	97.5	97.5
7x7	96.63	96.87	95.87	99.05

B. Comparisons

The experimental results obtained for our method are compared with Face Recognition Using Facial Symmetry proposed by Singh et al [5] in which they have used ORL database for that they have extracted the Principle Component Analysis for full face, right face only and left face only with train to test ratio of 7:3. The results are tabulated as below

TABLE V. RECOGNITION ON ORL DATABASE USING PCA

Type of Face	Recognition Rate
Full face	95%
Right face only	91%
Left face only	93%

Our obtained results as shown in Tables 1-4 outperform the results presented by [5] in Table 5.

VI. CONCLUSION

This work proposes the new features for face recognition, which are simple to compute and captures the variations in the face image. The problem of PCA and LDA is that the sample size is small. We propose the features that are based on windows or patches which are non-overlapping type. When the features are computed from patches, every minute details are taken into account, which helps in improving the accuracy of recognition. It is also shown that as the window size is increased; the accuracy is approximately the same. But with increased window size the storage size can be reduced. Better recognition accuracy is obtained in both constrained and unconstrained environment and the results are compared with the contemporary literature works. Further work can be carried out by designing the new membership functions based on statistical parameters.

REFERENCES

1. Sayeed, Farrukh, and Madasu Hanmandlu. "Properties of information sets and information processing with an

application to face recognition." *Knowledge and Information Systems* 52.2 (2017): 485-507. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

2. Duan, Yueqi, et al. "Context-aware local binary feature learning for face recognition." *IEEE transactions on pattern analysis and machine intelligence* 40.5 (2018): 1139-1153.

3. Gong, Dihong, et al. "A maximum entropy feature descriptor for age invariant face recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

4. Parlewar, Manisha, Hemprasad Patil, and Kishor Bhurchandi. "A novel quantized gradient direction based face image representation and recognition technique." *2016 Twenty Second National Conference on Communication (NCC)*. IEEE, 2016.

5. Singh, Avinash Kumar, and Gora Chand Nandi. "Face recognition using facial symmetry." *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology*. ACM, 2012.

6. Kar, Arindam, et al. "A Face Recognition approach based on entropy estimate of the nonlinear DCT features in the Logarithm Domain together with Kernel Entropy Component Analysis." *arXiv preprint arXiv:1312.1520* (2013).

7. Liu, Yi-Hung, and Yen-Ting Chen. "Face recognition using total margin-based adaptive fuzzy support vector machines." *IEEE Transactions on Neural Networks* 18.1 (2007): 178-192.

8. Bansal, Mamta, and Madasu Hanmandlu. "A new entropy function for feature extraction with the refined scores as a classifier for the unconstrained ear verification." *Journal of Electrical Systems and Information Technology* 4.1 (2017): 135-158.

9. Hanmandlu, Madasu, and Anirban Das. "Content-based image retrieval by information theoretic measure." *Defence Science Journal* 61.5 (2011): 415-430.

10. Shannon, Claude E. "A note on the concept of entropy." *Bell System Tech. J* 27.3 (1948): 379-423.

11. Phillips, P. Jonathon. "Support vector machines applied to face recognition." *Advances in Neural Information Processing Systems*. 1999.

12. Parveen, Pallabi, and Bhavani Thuraisingham. "Face recognition using multiple classifiers." *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*. IEEE, 2006.