Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 5, Number 1 (2013), pp. 1-13

http://www.irphouse.com

An Adaptive Recognition Technique named SOMPF based on Palmprint and Face Using Neural Network Based Self Organizing Maps

A.S. Raja[1] and V. Joseph Raj[2]

[1]Research Scholar, Sathyabama University,
Jeppiar Nagar, Chennai, Tamil Nadu, India
e-mail: csehod@capeitech.org
[2]Professor, Kamaraj College, Thoothukudi, Tamil Nadu, India
e-mail: v.jose08@gmail.com

Abstract

Biometrics has been gaining attraction due to the ever-growing demand of this field of research on access control, public security, forensics and e-banking. However, there are still many challenging problems in improving the accuracy, robustness, efficiency, and user-friendliness of these biometric systems. In this manuscript we propose a new adaptive multi-modal biometric framework based on self organizing maps for the recognition of individuals using palm print and face. We show that the proposed framework helps to improve the performance and robustness of recognition when compared to some standard methods in literature. The major focus of our approach is to keep the framework adaptive and robust, thereby, capable of being used in a wide variety of environments. Moreover we also discuss some new directions on which SOM shall be effectively used in biometrics community. We show all our findings with experimental results.

Keywords - Biometrics, Palm print, Face, Self Organizing Maps (SOM).

Introduction

The increase of terrorism and other kinds of criminal actions, such as fraud in e-commerce, increased the interest for more powerful and reliable ways to recognize the identity of a person [1, 2]. To this end, the use of behavioral or physiological characteristics, called biometrics, is proposed. Biometrics is best defined as measurable physiological and or behavioral characteristics that can be utilized to verify the identity of an individual [1]. Many physiological characteristics of humans, i.e., biometrics, are typically invariant over time, easy to acquire, and unique to each individual. Therefore the biometrics traits are increasingly adopted for civilian applications and no longer confined for forensic identification.

The recognition of individuals without their full co-operation is in high demand by security and intelligence agencies requiring a robust person identification system. Many biometric recognition algorithms have been proposed so far [3, 4, 5, 6, 7, 8]. Some of these include algorithms related to recognition of face, face, palm print, iris and voice (See Handbook of Biometrics [9]). A multimodal system is a combination of face and palm print (for instance) or any other combination of biometrics. Multimodal biometrics can be used to overcome some of the limitations of a single biometric. For instance, it is estimated that 5% of the population does not have legible fingerprints [1], a voice could be altered by a cold and stand alone palm print recognition systems are susceptible to changes in ambient light and the pose of the subject.

A typical biometric system usually consists of that specific biometric detection scheme followed by an extraction methodology (which shrinks the dimensionality of useful information) and then a classifier to make the appropriate decision. The dominant approach towards using palmprint for recognition is based on the statistical features. For statistical based palmprint recognition approach, the works that appear in the literature include eigenpalm (where the original palm print images were projected to a relatively lower dimensional space called eigen palms) [10], fisherpalms (which uses fisher linear discriminant to reduce the dimension) [11], Gabor filters [12], Fourier Transform [13], and local texture energy [14] (Out of these approaches, eigen palm and fisher palm are used for comparing with our approach). Chang et al. [9] used PCA on face and face, with a manual land marking method. With a larger dataset of 111 subjects, they achieved a combined recognition rate of 90%. Rahman and Ishikawa [15] also used PCA for combining face and face, they used profile images and manually extracted features. On a dataset of 18 subjects of profile face and face, the recognition rate was 94.44%. Middendorff and Bowyer [16] used PCA/ICP for face/face, manually annotating feature landmarks. On a 411 subject dataset they were able to achieve a best fusion rate of 97.8%. Yuan eral. [17] used FSLDA (full-space linear discriminant analysis) algorithm on 75 subject database with 4 images each (USTB) and on the ORL database of 75 subjects, achieving a best recognition rate of 98.7%. Despite of all these advancements, when it comes to practical usage in real life environment, there have been issues. Therefore, an adaptive framework which shall work in all scenarios is very much required. Here we provide an adaptive framework for appearance-based multi-modal recognition based on self organizing maps. This framework shall be easily extended to address very interesting questions faced by the biometrics community.

The remainder of this paper is organized as follows: In section 2 we discuss an object identification technique suitable for face and palm print. In section 3 we discuss how Self Organizing Maps (SOM) shall be used for dimensionality reduction. In section 4 we explain how SOM is used with Face and Palm print for recognition. In section 5 we discuss the results obtained using SOMPF method. Paper concludes with conclusion and future direction.

Dataset

We used the dataset related to face and palm print retrieved from [18] for our experiments[1]. There are 107 subjects in this dataset. We use only the right face and right hand palm print for our experiments. The images were scaled to a size of 60x60 for palm print and to a size of 60x60 to reflect the rectangular dimensions of the face. Some of the sample palm print and face images detected from the database is mentioned in Fig 1.

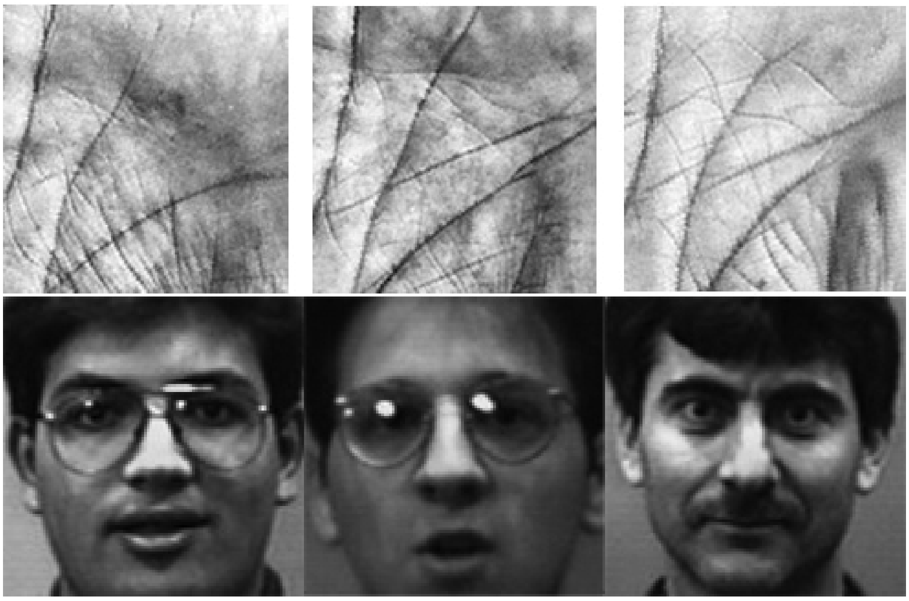

Fig. 1 Example images from the datasets [18]

[1]Multimodal datasets are not readily available as a standard in literature. The dataset that we have used is a virtual database for our multimodal study. The reason for referring this as virtual is because of the fact that the original dataset does not contain the combined palm print and face of every individuals. Instead, one dataset[20] had the face images and the other dataset[21] had the palm print images of individuals separately. We had combined them by taking 3 palm print images of an individual from the palm print dataset with 3 face images of an individual from the face dataset. There were 107 face subjects and 165 palm print subjects. We had considered all the 107 face subjects and we had chosen randomly 107 subjects from the palm print subjects.

SOM & Methodology Used in This Paper

Self Organizing Map (SOM) is a special kind of unsupervised computational neural network [22] that combines both data projection (reduction of the number of attributes or dimensions of the data vectors) and quantization or clustering (reduction of the number of input vectors) of the input space without loss of useful information and the preservation of topological relationships in the output space.

A few concepts are useful to understand the workings of the technique. The input space (also called signal) is the set of input data we employ to feed the algorithm; the set of input data in our case refers to the set of images that we use for training; typically, the observations are multidimensional and are thus expressed by using a vector for each of them. In our case the observations refer to the pixels present in each image (in our case the dimension/vector size of each palmprint image is 60x60=3600 and face image is 60x60=3600). On the contrary, the output space (trained network, network or SOM) refers to the low-dimensional universe in which the algorithm represents the input data. It usually has two-dimensions, and is composed of a set of elements called neurons (or nodes) which are interconnected, hence the network. What the algorithm does is to represent the input space onto the output space, keeping all the relevant information and ordering observations in a way such that topological closeness in the output space implies statistical similarity in the input space.

The input space is composed of n-dimension vectors we want to visualize/cluster in a low-dimensional environment. We can express the input vector t as:

$$x=[\xi 1(t),\ \xi 2(t),\ \ldots,\ \xi n(t)]T\in R^n,$$

where$\xi i(t)$ represents the value for each dimension.

The output space is an array of p by q neurons (nodes) topologically connected following a kind of geometrical rule (the most common topologies being circles, squares and hexagons). In our case p =11 and q = 11. Each of the nodes is assigned a parametric real vector of initially random values that we call model, and express as:

$$mi=[\mu in,\ \mu in,\ \ldots,\ \mu in]T\in R^n$$

Last, we may also define as d(x, mi) any distance metric between two vectors x and mi. The most widely used is the Euclidean distance, although other specifications are also valid.

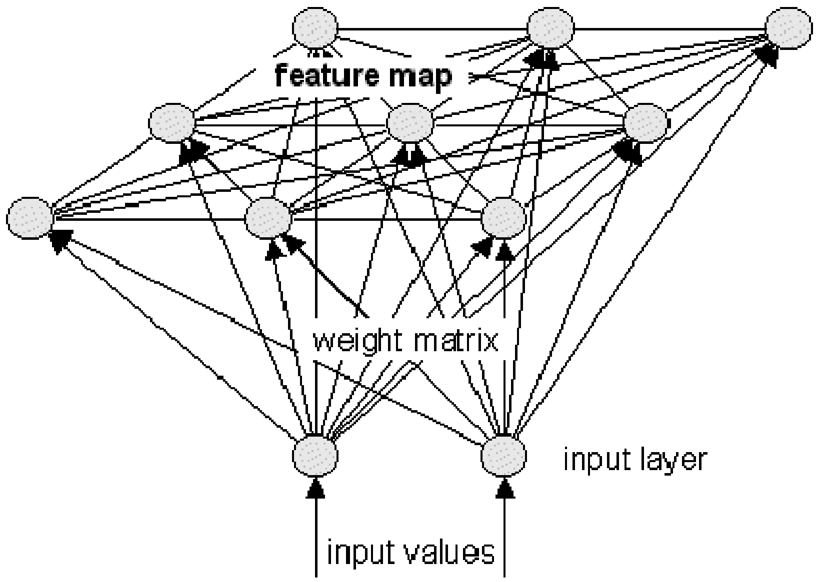

Fig. 2 Self Organizing Map (SOM)

What we are looking for is a topologically-ordered representation of the signal space into the network. That is done by the SOM in an iterative process called training, in which each signal vector is sequentially presented to the output space. The best matching unit (b.m.u.) for x is defined as the neuron minimizing the distance to x. When this is found, the b.m.u. is activated and an adaptive process starts by which such neuron and its topological neighbours are modified by the following scheme:

$$mi(t+1)=mi(t)+hci(t)[x(t)-mi(t)],$$

where t and t + 1 represent, respectively, the initial and the final state after the signal has activated the neuron; hci(t) is called neighbourhood function and expresses how the b.m.u. and its neighbours are modified when activated by a signal; usually, the linear or Gaussian versions are used. This process is repeated over many cycles before the training is finished. The neighbourhood function depends on several parameters relevant for this stage: the distance between the b.m.u. and the modified neuron (so the further away the neuron is, the smaller the adjustment); a learning rate $\alpha(t)$ that defines the magnitude of the adjustment, and gradually decreases as the training cycles advance; and the neighbourhood radius, which decides which of the surrounding neurons of the b.m.u. are also modified, and also decreases over the training stage and the self arranging (organization) of the input observations.

This procedure may be used as a visualization tool for multidimensional datasets as well as a clustering method. In the first case, we would want to see how the different observations are mapped into the SOM to discover (dis)similarities, making use of the topological preservation of the statistical characteristics, and study how the different dimensions are distributed; in the second one, the network would have a relatively small number of neurons (as many as clusters we would want to obtain) and we would focus on analyzing which observations are grouped with which. In our case, images which have similar face characteristics gets grouped together within the respective nodes/maps.

The description of SOM given above (also referred as unsupervised SOM in some literature) focuses on unsupervised exploratory analysis. However, SOMs can be used as supervised pattern recognizers, too. This means that additional information, e.g., class information, is available that can be modeled as a dependent variable for which predictions can be obtained. The original data are often indicated with X; the additional information with Y. An approach suggested by Kohonen [27] for supervised SOM is to perform SOM training on the concatenation of the X and Y matrices.

Although this works in the more simple cases, it can be hard to find a suitable scaling so that X and Y both contribute to the similarities that are calculated. Melssen et al. [28] proposed a more flexible approach where distances in X and Y -space are calculated separately. Both are scaled so that the maximal distance equals 1, and the overall distance is a weighted sum of both:

$$D(o, u)=Dx(o, u)+(1-\alpha)Dy(o, u)$$

where D(o, u) indicates the combined distance of an object o to unit u, and Dx and Dy indicate the distances in the individual spaces. Choosing $\alpha = 0.5$ leads to equal weights for both X and Y spaces. Scaling so that the maximum distances in X and Y spaces equal one takes care of possible differences in units between X and Y . Training the map is done as usual; the winning unit and its neighborhood are updated, and during training the learning rate and the size of the neighborhood are decreased.

One shall extend the principle used for supervised SOM to more than one layer as well, the result of which is being referred in literature as super-organized SOM. This is the idea which is used in this framework. For every layer a similarity value is calculated, and all individual similarities then are combined into one value that is used to determine the winning unit.

$$D(o, u) = \sum_i \alpha_i D_i(o, u)$$

where the weights i are scaled to unit sum. These weights are the only extra parameters (compared to classical SOMs) that need to be provided by us. The super-organized map accounts for individual types by using a separate layer for every type. When compared to other neural network based approaches, it shall be noted that in SOM - the neurons are arranged on a flat grid not as a multilayer perceptron (input, hidden, output).

4. Proposed Approach (SOMPF)

We use SOM for face and palm print recognition and hence we call the approach as SOMPF. The set of input data in our case refers to the set of images that is used; the observations refer to the pixels present in each image. First we apply SOM to palm print and face separately. In this case for face, the dimensionality of the input vector is 3600 (this is because of the normalized size of the face image that is used –60x60 size) and in the case of palmprint the dimensionality of the input vector if 1170 (this is because of the normalized size of the face image that is used –60x60 size). The output space is an array (separate for both face and palm print) of p by q neurons (nodes) topologically connected following a kind of geometrical rule (a rectangular topology has been used). In our case p=11 and q = 11 for palm print and p=11 and q = 11 for face. With the same setup, we do a supervised mode SOM analysis (where we use some images for training and some images for testing). In the end (SOMPF approach), we combine both the palm print and face using super organized SOM. In other words, in SOMPF we get multiple layers (as opposed to supervised SOM where there are only two layers X and Y). We play with the weight between palm print and face layers and determine the optimum weightage for the recognition experiment under consideration. All these interesting experimental results obtained using SOM in unsupervised mode, supervised mode, super organized mode are explained in the next section.

5. Experiments & Results

As mentioned earlier, this paper uses the Palm print & Face dataset obtained from [18]. There are 107 subjects. Each subject has 3 images each for face and 3 images for palm print.

The first experiment which was performed was to find the total number of output nodes which are required. Unsupervised SOM was ran over the given 107 subjects related to palm print and face dataset. In the plot shown in Figure. 3, the background color of a unit corresponds to the number of samples mapped to that particular unit; one shall observe that they are reasonably spread out over the map (one unit is empty for face; two units are empty for palm print; no samples have been mapped to them). The plot in Figure. 4 shows the mean distance of objects, mapped to a particular unit, to the codebook vector of that unit. A good mapping should show small distances everywhere in the map. These show that the number of output nodes which are chosen (11x11) are good enough for our purpose.

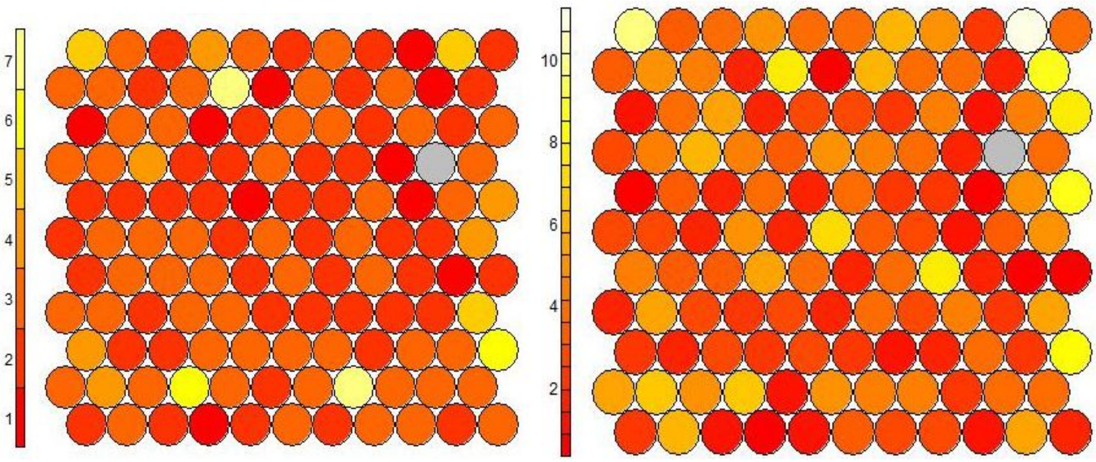

Fig. 3 Counts plot of the map obtained from the face and the palm print dataset. Empty units are depicted in gray. The color in each cell represents the number of (a) face and (b) palm print images which went into that went into that cell.

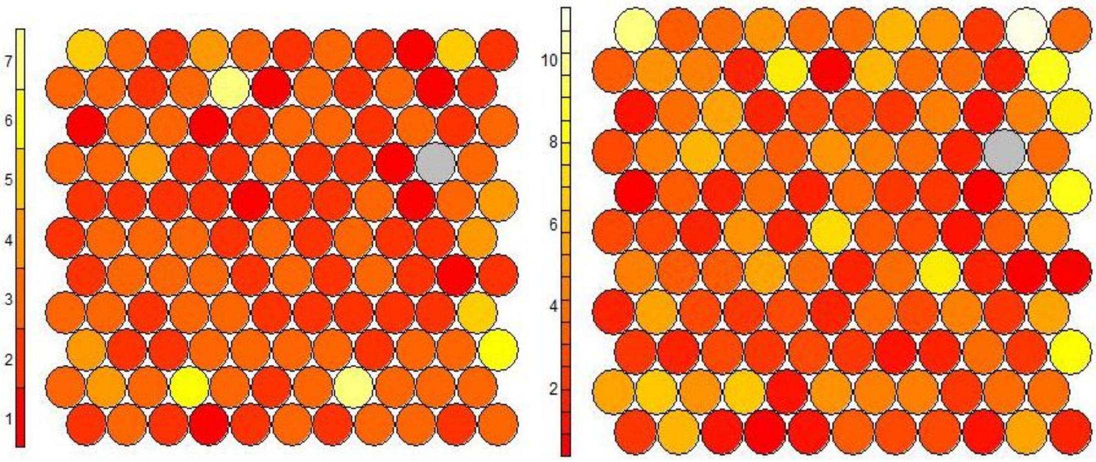

Fig. 4 Shows the quality of the mapping; the biggest distances between x and mi vectors are found in the bottom left of the map for (a) face (b) palm print

The second experiment which was performed was to do an exploratory analysis using unsupervised SOM. Figure.5 (a) and (b) shows the mapping of images related to unsupervised SOM. Each color/shape in the figure is used to represent a particular subject. From the dataset, one shall infer that each subject has 3 face images & 3 palm print images related to him which are more or less mapped into different unique cells. Figure 5 reveals this out clearly. For instance in Figure 5, if one looks at the first cell, approximately 3 similar units are mapped onto that cell for face and 3 similar units for palm print. The similar units indicate that they belong to the same subject. This explains that even without any training, unsupervised SOM was able to more or less grossly able to put the subjects into different cells. The error rate in grouping in this case was observed to be approximately 27% for face and 30% for palm print (out of the 321 images of 107 subjects, 225 went into the appropriate cells which belonged to similar subjects and 96 images did not gets mapped properly).

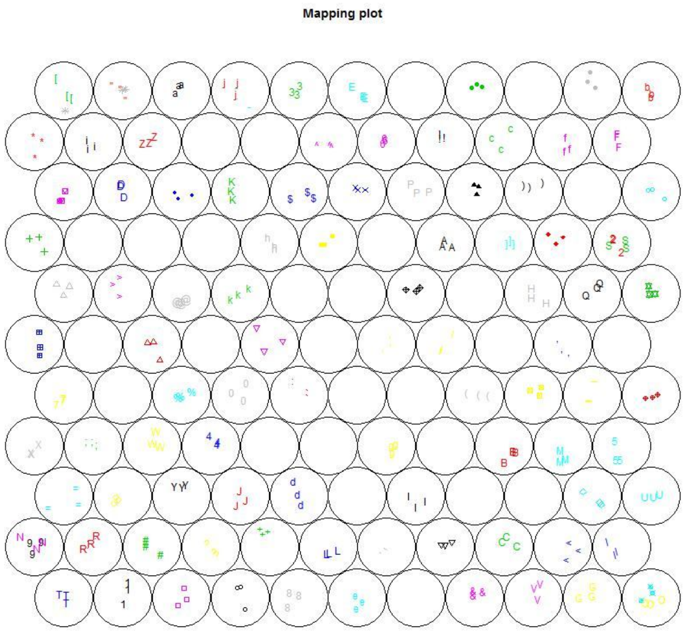

Figure.5 (a) Mapping of the 107 Palmprint subjects in a eleven-by- eleven SOM

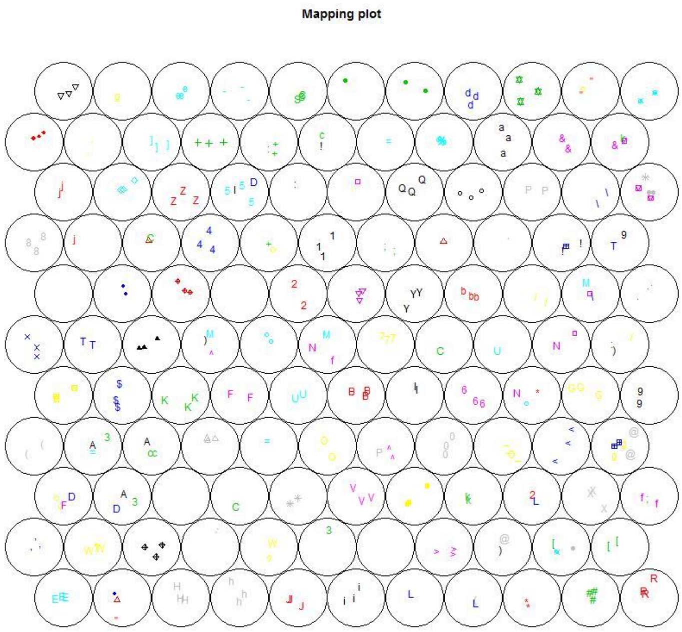

Figure.5 (b) Mapping of the 107 Face subjects in a eleven-by- eleven SOM

The third experiment that was used is to use the classifier information related to which image belonged to which subject using supervised SOM. In this experiment, the subject has been considered as the dependent variable (variable Y as explained in Section 3) and the pixel values of the image as the independent value (variable X as explained in Section 3). 1 random image from each subject has been chosen for training and the rest of the 2 images of each subject has been used for testing. The weights for X and Y has been varied with supervised SOM and the following characteristics as mentioned in Table 1 has been observed (the weights in a way indicate the relative strength between X and Y for recognizing a subject).

Table 1 (a). Error rate with supervised SOM by varying X and Y weights for face (b). Error rate with supervised SOM by varying X and Y weights for palmprint

X Weightage	Y Weightage	Error Rate
0.9	0.1	23.5
0.8	0.9	19.8
0.7	0.3	17.2
0.6	0.4	16.6
0.5	0.5	14.7
0.4	0.6	12.9
0.3	0.7	11.2
0.2	0.8	8.6
0.1	0.9	7.9

X Weightage	Y Weightage	Error Rate
0.9	0.1	26.6
0.8	0.9	24.3
0.7	0.3	21.4
0.6	0.4	19.9
0.5	0.5	16.7
0.4	0.6	14.9
0.3	0.7	12.1
0.2	0.8	10.7
0.1	0.9	9.2

The above Table 1 shows that, if one uses the classification information also (using supervised SOM), then the recognition rate improves significantly (when compared to not using it - as earlier seen with unsupervised SOM). This is true both for palm print and face. The fourth experiment that was done was related to super-organized SOM. We modeled the palm print related pixel values as one layer, face related pixel values as the second layer and the class information as the third layer. A weight is associated to every layer to be able to define an overall distance of an object to a unit. We pose an optimization problem to optimize the weights in such a way that the recognition rate is the maximum. Interestingly, this also allows one to easily find out the dominant metric (palm print or face - based on the one which takes a higher weightage). To begin with, we seeded the initial weights to be of extremely high (1) for palm print and extremely low (0) for face. We noted down the results. We then optimized the weights for palm print and face as explained above and observed the results. The experimental results are presented in Table 3. It seems that face is a better metric when compared to palm print for the given standard dataset and a propositional weight of 7.:3 seems to give a better recognition rate.

Table. 3 Recognition rates for different palm print/face weights using super SOM

Face Weightage	Palm print Weightage	Combined Recognition Error Rate
0.9	0.1	6.6
0.8	0.9	5.4
0.7	0.3	4.6
0.6	0.4	5.1
0.5	0.5	5.6
0.4	0.6	6.2
0.3	0.7	6.7
0.2	0.8	7.2
0.1	0.9	9.2

The fifth experiment that was done was a comparative analysis of SOMPF with other methods related to multimodal biometrics involving palm print and face. Table 4 shows the comparative results between Palm print-Face-PCA (Principal Component Analysis), Palm print-Face-Sequential Float Feature Selection (SFFS) and Self Organizing Map for Face and Palm print (SOMPF).

The size of the training set varied from 1 to 2 images per person and the remaining of the images for each subject form the test set. For the PCA and the SFFS, the experiments that were conducted showed that all the training images during the training phase are classified correctly (Table 4). On the other face, the SOMPF could not classify correctly all the training images. Furthermore, Figure. Table 4 shows a greater improvement in the performed experiment with SOMPF than PCA or SFFS when using one number of training sample for each person. Using SOMPF with one image per person during training phase gives 3.4% error recognition rate against 11.9% error recognition rate using the PCA, and 9.6% error recognition rate using the SFFS method.

Table 4. Test error recognition rate (%) with varying number of images per person

Number of training images per person	Number of testing images per person	Training phase			Testing phase		
		PCA	SFFS	SOMPF	PCA	SFFS	SOMPF
1	2	0	0	4.6	11.9	9.6	3.4
2	1	0	0	4.1	9.2	8.7	2.6

Table 4 shows that SOMPF can provide an improvement in error recognition rate when compared to the other approaches based on literature. Interestingly, self organizing maps shall also be used to address some interesting curiosities discussed in the biometrics community in a formal manner. For instance, there has been a curiosity/hypothesis which says that 'face as a biometric does not change over age when compared to other biometrics like palm print'. If one shall gather images of same subjects at different ages in a similar pose and background and do a supervised SOM across the different ages, one shall find out if face has been consistently performing when compared to palm print or some other biometric. Most of the results used in this paper are obtained using an opensource software framework named statistical R[19]. The archive of results and code used related to this paper is accessible at [20].

Conclusion

Neural Network based Self Organizing Maps has been used in this paper. The proposed approach SOMPF has been shown to perform well when compared to some standard methods from literature. This has been done by taking a standard dataset from literature.

References

[1] Jain AK, Bolle R, Pankanti S (eds) (1999) Biometrics: Personal identification in networked society. Kluwer Academic Publishers, Boston/Dordrecht/London

[2] Wechsler H, Phillips JP, Bruce V, Folgeman Soulie F, Huang TS (eds) (1997) Face recognition – From theory to applications. ASI NATO series, vol 163, Springer, Berlin Heidelberg New York.

[3] Zhao WY, Chellappa R, Rosenfeld A, Philips PJ (2000) Face recognition: A literature survey. UMD CfAR technical report CAR-TR-948

[4] Turk M, Pentland A (1991) Eigenfaces for face recognition. J Cognitive Neurosci 3(1):71–86

[5] Jain A, Flynn P, Ross AA, Handbook of Biometrics, Springer, Heidelberg (2008)

[6] Wiskott L, Fellous JM, Kru¨ ger N, von der Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE T Pattern Anal 19(7):775–779

[7] Kotropoulos CL, Tefas A, Pitas I (2000) Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. Pattern Recogn 33(12):1935–1947

[8] Penev PS, Atick J (1996) Local feature analysis: A general statistical theory for object representation. Network–Comp Neural 7(3):477–500

[9] de Vel O, Aeberhard S (1999) Line-based face recognition under varying pose. IEEE T Pattern Anal 21(10):1081–1088

[10] Z. Mu, L. Yuan, Z. Xu, D. Xi, S. Qi, Shape and structural feature based face recognition, in: Sinobiometrics, LNCS, vol. 3338, 2004, pp. 663–670, 2004.

[11] DeVel O, AeberhardS, Line-based face recognition under varying pose. IEEE T Pattern Anal, 21(10):10811088, 1999.

[12] G. Lu, D. Zhang, K. Wang, Palmprint recognition using eigenpalms features, Pattern Recognition Letters, 24 (9–10), 1473–1477, 2003.

[13] X. Wu, D. Zhang, K. Wang, Fisherpalms based palmprint recognition, Pattern Recognition Letters, 24, 2829–2838, 2003.

[14] W.K. Kong, D. Zhang, W. Li, Palmprint feature extraction using 2-D Gabor filters, Pattern Recognition, 36 (10), 2339–2347, 2003.

[15] Rahman M M, Ishikawa S (2005), Proposing a passive biometric system for robotic vision. In: Proc. 10th International Symp. on Artificial Life and Robotics, .

[16] Middendorf C, Bowyer K W (2008), Multibiometrics using face and ear. In: Handbook of Biometrics, pp. 315341. Springer, Heidelberg, .

[17] Yuan L, Mu Z -C, Xu X -N. (2007), Multimodal recognition based on face and ear. In:International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR2007, vol. 3(2-4), pp. 12031207.

[18] Face and Palmprint Database Source Link, http://www4.comp.polyu.edu.hk/ csajaykr/

[19] R Project http://www.r-project.org/

[20] Archive of software code, dataset and experimental results http://www.capeitech.org/research/asr/SOMPF/