

# Face recognition based on fuzzy probabilistic SOM

Laura Lanzarini, Franco Ronchetti, César Estrebou, Luciana Lens  
III-LIDI (Institute of Research in Computer Science LIDI)  
Faculty of Computer Science, National University of La Plata  
La Plata, Buenos Aires, Argentina  
{laural, fronchetti, cesarest@lidi.info.unlp.edu.ar}; lucianasollens@gmail.com

Aurelio Fernández Bariviera  
Department of Business  
Universitat Rovira i Virgili  
Reus, Spain  
aurelio.fernandez@urv.net

**Abstract**—Face recognition is a topic of great interest in different areas, especially those related to security. The identification of a person by the image of her face is a difficult task because of changes experienced by the face due to various factors, such as facial expression, aging and even the lighting. This paper presents a new face recognition technique based on the combination of a competitive fuzzy neural network and a probabilistic decision criterion. The results of this technique on two images database offer satisfactory results. We also discuss the limitations of the proposed technique and future research lines.

## I. INTRODUCTION

Face recognition is a biometric technique widely used in various areas such as security and access control, forensics and police check points.

The identification of a person by the image of her face is a process of comparing an image of a given person with a set of previously stored images. These previous images are in a database. In order to allow more flexibility to the recognition process, the database usually includes several images of the same person in the different situations that can occur when a new image is captured. These situations includes different facial expressions, changes of the head position (i.e. not only frontal images), scale changes, etc.

The information to search in the database of images is not a trivial search for the original image but a characterization (distinctive features) of it. This paper uses descriptors generators based on SIFT (Scale Invariant Feature Transform) method, defined by [7]. The advantage of these descriptors is that they are invariant to scale, rotation, perspective, partial occlusion and lighting. These descriptors transform each image into a set of numerical vectors. The images in the database are stored with this representation.

The technique we propose uses vector based images to train a fuzzy Self Organizing Maps (SOM) neural network, which identifies the most relevant characteristics of each person. Once the network is trained, its implementation is straightforward. The network, based on SIFT image of the person to be recognized, it is able to indicate who is the closest image in the database.

This paper is organized as follows: Section 2 briefly mentions some related literature, section 3 describes the main characteristics of SOM and fuzzy SOM, section 4 highlights the most relevant features of the method described by Lowe in order to obtain SIFT descriptors, section 5 details the

information inputs of the neural network, section 6 presents a pseudocode of the recognition process, section 7 shows the results. Finally section 8 presents the main conclusions of this study and explores lines of future research.

## II. RELATED LITERATURE

The existing literature provides several solutions to the problem of face recognition using SIFT descriptors.

Aly [3] shows that the use of SIFT features for face recognition gives better results than Eigenfaces and Fisherfaces algorithms. Different sizes were used in the training data set, where it was found that performance declines when this set is smaller. Regarding to the significant number of SIFT features required for reliable comparison, it was found that using a fewer characteristics the performance is better than with Fisherfaces and Eigenfaces.

One of the main drawbacks using SIFT vector recognition are false positives. This situation leads sometimes to incorrect recognition. To solve this issue, [4] proposes the use of a variant of Particle Swarm Optimization (PSO) defined in [5]. The optimization technique is used in order to select the most representative SIFT vectors. As a result, there is not only a reduction in false positives but also in the required computation time for processing and storing the images' database.

Regarding competitive SOM networks, there are techniques applied to face recognition that focus only on certain parts of the image, e.g. [6]. This approach does not favor the invariance of results and increases the training and recognition times. Therefore in this paper we chose to transform the image through vectors with their most representative features.

## III. FUZZY SOM

SOM neural networks were developed by Teuvo Kohonen in 1982 [1]. Its main application is clustering of available information, preserving its topology. Unlike other clustering methods of the type *winner-take-all* such as K-means, SOM incorporates the concept of neighborhood between groups so that groups that are close at the output architecture are close in input space. This can be used for different purposes: to operate on either with clusters or to reduce the dimension of the input space. The neighborhood criterion also helps in the adaptive initialization process. The feature of SOM is the reason why we select this architecture.

SOM is a neural network of two layers: the input layer and the hidden or competitive layer. The grouping is done by means of centroids, i.e., every element of the competitive layer is associated to a vector of the same dimension as the input space and is considered the prototype or the representative of the group.

In other words, each input vector is considered represented by (or associated with) the competitive neuron having the closest weight vector (or centroid) according to a given similarity measure. Since it is an unsupervised adaptation process, the expected response is replaced by a distance measure.

The set of centroids is obtained through an iterative process that initially assigns random values. The process is repeated until centroids exhibit no significant changes. In other words, the iteration process ends when each input vector is represented by the same competitive neuron from the previous iteration.

Let  $W = \{w_1, w_2, \dots, w_n\}$  be the set of centroids. Each of them is associated with a competitive neuron. In each iteration, for each input vector  $X = (x_1, x_2, \dots, x_m)$  the neuron that will represent it, is determined as follows:

$$c = \arg \min_j (dist(W_j, X_i)) \quad j = 1 : N \quad (1)$$

where  $X_i$  is the input vector and  $dist$  is a previously established measure of similarity or distance. The goal is to identify each input vector with the closest centroid. Usually the Euclidean distance is used as a measure of distance, but it could be changed by another norm, based on the characteristics of the problem.

The neuron that is closest to an input vector is called winning neuron, since it is the one that wins the competition for the representation of a vector (the closest so far). The process continues by updating the weight vector of that neuron and its vicinity according to equation (2)

$$W_j = W_j + \alpha * h * (X_i - W_j) \quad (2)$$

where  $X_i$  is the input vector,  $j$  the competitive neuron whose vector is to be updated,  $h$  a neighborhood function that controls the scope of the change and  $\alpha$  a value between 0 and 1 representing the learning factor. Equation (2) have some variants that can be found in [2].

The concept of neighborhood is used to allow the network to adapt properly. This implies that competitive neighboring neurons represent similar input patterns. Therefore, during the training process (obtainment of centroids) neighborhood definition is wide and then groups boundaries are fine-tuned along the iterations.

In this work we use a fuzzy SOM network, because it is considered more stable to changes in the input data. Its operation is very similar to the SOM defined in [1], but when calculating the winning neuron our model incorporates the concept of membership degree by which a single input vector can belong to several groups at once. Equation (3) specifies

how to calculate it:

$$G_{ij} = \begin{cases} 1 & \text{if } X_i = W_j \\ 0 & \text{if } X_i = W_l, l \neq j \\ \left( \sum_{l=1}^m \left( \frac{dist(X_i, W_j)}{dist(X_i, m_l)} \right)^{\frac{1}{\alpha-1}} \right)^{-1} & \text{otherwise} \end{cases} \quad (3)$$

Then, equation (2) is modified to take into account this measure fuzzy proximity between vectors as follows:

$$W_j = W_j + \alpha * G_{ij} * h * (X_i - W_j) \quad (4)$$

For variations of equation 4 we refer to [9].

#### IV. SIFT DESCRIPTORS

In [7], Lowe defined a method to extract features from an image and use them to find matches between two different views of the same object. These features, called SIFT (Scale Invariant Feature Transform) features, are invariant to image scale and rotation, and quite invariant to affine distortion, as well as changes in point of view and lighting. They are also highly distinctive.

The process to determine SIFT features for an image consists in four steps:

- First, the location of potential points of interest within the image is determined. These points of interest correspond to the extreme points calculated from plane subsets of Difference of Gaussian (DoG) filters applied to the image at different scales.
- Then, the points of interest whose contrast is low are discarded. This is an improvement from the definition in [8].
- After this, the orientation of relevant points of interest is calculated.
- Using the previous orientations, the environment is analyzed for each point and the corresponding feature vector is determined. As a result of this process, a set of 128-length feature vectors that can be compared with those from another image of the same object with a different scale, orientation, and/or point of view, is obtained. This comparison can be done directly by measuring the distance and establishing a similarity threshold. More detailed information about this method is available in [7].

#### V. MODEL BASED ON FUZZY SOM

Images in the database, represented by their respective sets of SIFT descriptor are the data that will be used to train a fuzzy SOM neural network. This means that the training set consists of numerical vectors of dimension 128, corresponding to the various images of all subjects in that database. This implies that for a given SIFT descriptor, a person is assigned. Each picture is associated with a person and therefore so is each SIFT descriptor.

The method used to train the neural network was described in Section III.

After the training, each neuron of the network can represent more than one person. To determine this, you enter the

training data and calculated for each of them, the degrees of membership of each neuron as indicated in (ref eq: Grade). Then each training vector is assigned in proportion to their degree of belonging to the  $k$  most representative competitive neurons.

Since each vector is associated with a person, each with a related competitive neuron list proportion possessing SIFT descriptor for each subject.

The calculation is as follows:

- Let  $L$  be the number of people in the database.
- Let  $T$  the total number of SIFT descriptors corresponding to all the images in the database. Each descriptor corresponds to one image of one person.
- Let  $N$  be the number of neurons of the network. For each descriptor,  $d_j$  with  $j = 1..T$  the degree of membership to each neuron is computed. Neurons are ordered (descending) according to their degree of membership. Let  $n_1, n_2, \dots, n_k$  the  $k$  competitive neurons with the greatest membership degrees, such that

$$G(d_j, n_1) > G(d_j, n_2) > \dots > G(d_j, n_k)$$

and, in addition,

$$G(d_j, n_k) > G(d_j, n_r) \quad \text{with } r = k + 1 : N$$

- The  $k$  competitive neuron with the greatest membership degrees will share the representation of SIFT descriptor  $d_j$  in proportion to:

$$Prop(n_i, d_j) = \frac{\exp^{-i^2/2}}{\sum_{x=1}^k \exp^{-x^2/2}} \quad i = 1 : k \quad (5)$$

It should be highlighted that the membership degree is not a variable in 5. It is only taken into account the established order among neurons. The denominator in equation (5) normalize proportions in the interval [0,1] in such way that, if all values of  $Prop(n_i, d_j)$  for all SIFT descriptors  $d_j$  corresponding to the same individual  $s_l$  are added, would reach a equivalent value to the total amount of these SIFT vectors that are represented by neuron  $n_i$ . This is not an integer value, since it corresponds to the sum of proportions and will be considered the degree of recognition of neuron  $n_i$  for individual  $s_l$ .

- The correspondence between the SIFT descriptor and the person to whom it corresponds is indicated as follows

$$Q(s_k, d_j) = \begin{cases} 1 & \text{if } d_j \text{ corresponds to person } s_k \\ 0 & \text{if not} \end{cases} \quad (6)$$

Then, the relative frequency of persons recognized by the  $i$ -th neuron can be expressed as follows

$$P(s_k | n_i) = \frac{\sum_{t=1}^T Prop(n_i, d_t) \cdot Q(s_k, d_t)}{\sum_{t=1}^T Prop(n_i, d_t)} \quad \forall s_k \in L \quad (7)$$

It should be noted that, after calculating the conditional probabilities given in (7), the network is able to estimate the similarity between a new image and those stored in the database. The following section describes this process.

## VI. IDENTIFICATION MECHANISM

The identification mechanism is probabilistic. It consists in applying to the trained network all SIFT descriptors for the image of the subject to be recognized. For each descriptor, a winning neuron is calculated. We should bear in mind that each neuron usually represents several people.

The person who is identified by the network is one that abides by the expression (8)

$$s_k \in L \iff s_k = \arg \max_r (P(s_r)) \quad \forall r/s_r \in L \quad (8)$$

where, applying the total probability theorem,  $P(s_k)$  is computed in the following way:

$$P(s_k) = \sum_{i=1}^N PN(n_i) \cdot P(s_k | n_i) \quad (9)$$

donde

$$PN(n_i) = \frac{\sum_{t'=1}^{T'} Prop(n_i, d_{t'})}{T'} \quad (10)$$

being  $T'$  the number of SIFT descriptors that represent the input image.

Equation(10) computes the winning probability (proportion) of the  $i$ -th neuron after entering the  $T'$  SIFT descriptors of the image.

Since each neuron can determine the probability with which recognizes the different candidates, equation (9) allows the to obtain the network response for each person to be recognized. As indicated in the expression (8), it will selected the subject with the highest probability.

Figure 1 shows the pseudo-code of the proposed method.

## VII. RESULTS

In order to measure the effectiveness of the proposed method we use two databases obtained from [10]. The first one is YALE facebase, containing 165 face images of 15 subjects with 11 different images per person. Each image has a resolution of 320x243 pixels. The second base is AT&T facebase, which contains 400 images of 40 people with 10 images per individual. The size of each image is 112x92 pixels.

We apply the following methods to both image bases:

- *SIFT*: This technique follows the criterion established in [7] where, for each image, the total number of coincidences is computed *vis-à-vis* each individual within the image base. The image with the greatest value is selected. This process implies to compare each descriptor of the input image with the set of descriptors for each of the images in the base. It will be considered that a descriptor corresponds (there is coincidence) to the image if the

$Descr$  = Descriptors used for identifying (there are  $T'$ )  
 $N$  = number of network neurons after training  
 $Prop(i, j) = 0, i = 1..N ; j = 1 : T'$

{ Distribution of SIFT descriptors of the input image }

**for all** element  $d$  from  $Descr$  **do**  
    Determine  $n_1, n_2, \dots, n_k$   
    **for**  $x = 1..k$  **do**  
        Calculate  $Prop(n_x, d)$  according to (5)  
    **end for**  
**end for**

{ Identification }

**for all**  $s_k \in L$  **do**  
     $P(s_k) = 0$   
    **for**  $i = 1..N$  **do**  
        Calculate  $PN(n_i)$  as indicated in (10)  
         $P(s_k) = P(s_k) + PN(n_i) \cdot PR(s_k | n_i)$   
    **end for**  
**end for**

$k = max_r(P(s_r)), \forall s_r \in L$   
 $s_k$  is the person identified by the model

Fig. 1. Pseudo-code for the probabilistic system

distance to the closest descriptor is less than 50% of the distance with the second closest. This percentage is a parameter of the algorithm and was selected following [7].

- *SIFT+PSO*: this technique is an improvement of the previous method, defined in [4]. It uses an optimization technique to select the most representative SIFT descriptors. The recognition process is the same as in method *SIFT*, but with the advantage of operating on a smaller number of descriptors (see figure 2)
- *ProbSOM*: is a crisp version of our proposal, and defined in [11]. In this case, both during the training of the SOM network and the construction of the model, descriptors are associated to a single centroid.
- *Fuzzy ProbSOM*: Our proposed method, described in section V and section VI.

Since the hit rate varies depending on the number of images used to construct the model, we performed in the 4 cases and for both bases, 9 cuts to form the set of training images, these are: 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 % and 90 %. For each of them, there were 30 independent runs of each method. In each case we used the same selection of images for training.

Figure 3 shows, for each of the methods, the average hit rate (of 30 runs) for each cut using YALE base. As shown, the proposed method is the one that achieves the best results. Since in all cases the deviations were equivalent and that it is a large sample, we performed ANOVA test along with Tukey's test. Our results show significant differences (at 0.05 level) in all cases except at the lowest percentage where SIFT and

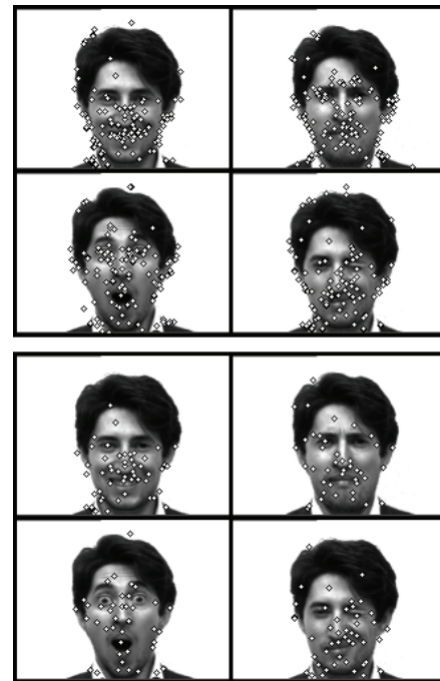


Fig. 2. SIFT descriptors of a person of the YALE database. The top row shows all descriptors found while the bottom row shows only the descriptors selected by *SIFT+PSO*

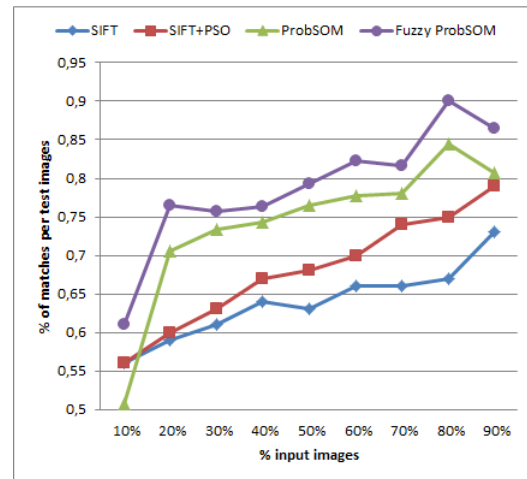


Fig. 3. Hit rate of each method for YALE image base. Each value corresponds to the average percentage rate of 30 independent runs.

*SIFT+PSO* methods were equivalent.

With respect to AT&T image base, we perform the same tests and we conclude that with the exception of the *SIFT* method, the remainder provides equivalent results, i.e. no significant differences. Figure 4 illustrates the results. In this case, it is important to note that, having a model based on a neural network it allows classification with lower computational time, despite the reduction in the number of descriptors obtained with *PSO* in [4].

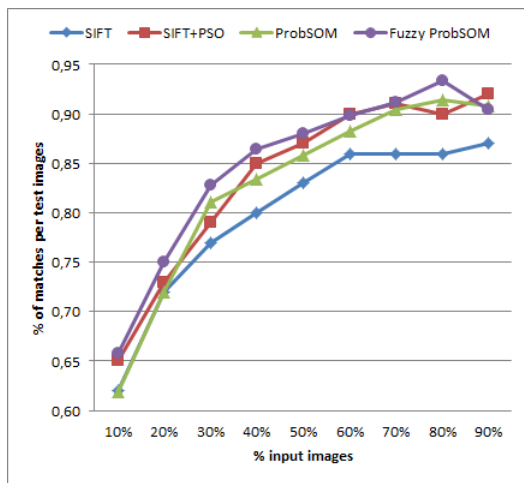


Fig. 4. Hit rate of each method for AT&T image base. Each value corresponds to the average percentage rate of 30 independent runs.

### VIII. CONCLUSIONS AND FUTURE WORK

We presented a new face recognition technique based on the combination of a competitive neural network with fuzzy probabilistic decision criterion. Its application on two test databases has yielded higher hit rates than conventional SIFT method defined in [7].

The selected bases are very different, while AT&T has similar faces, YALE includes images of the same individual with significant changes (e.g. different expressions or use of glasses) as shown in Figure 2. These factors makes the recognition process more difficult. Consequently, the hit rates using YALE are inferior than the hit rates using AT&T. As shown in figure 3, the proposed method is the most robust, offering the best results to large changes in SIFT vectors.

A weakness of the proposed method is its inability to operate on a kind of rejection. This situation is repeated in the other methods. Note that all four methods look within the database for the most likely subject that correspond to the input image. The problem arises when the subject to identify is not in the base. In this case, the use of a threshold to make a decision does not guarantee to obtain a correct answer. For that reason we think that incorporating a second neural network to the model could assist in the recognition process.

### REFERENCES

- [1] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. Biological Cybernetics. Springer Berlin/ Heidelberg. Vol43, nro. 1. Pp.59-69. Ed. 1982. ISSN 0340-1200
- [2] Teuvo Kohonen. Self-organizing Maps. 2nd Edition. Springer. 1997. ISBN 3-540-62017-6.
- [3] Aly Mohamed. Face recognition using sift features. CNS186 Term Project, 2006.
- [4] Maulini J., Lanzarini L. Face Recognition using SIFT descriptors and Binary PSO with velocity control. Published in Computer Science & Technology Series. XVII Argentine Congress of Computer Science Selected Papers. ISBN 978-950-34-0885-8. Pags. 43-53. EDULP, 2012.
- [5] Lanzarini L., López J., Maulini J., and De Giusti A. A new binary pso with velocity control. In Advances in Swarm Intelligence, Part I, volume 6728, pages 111-119. Lecture Notes in Computer Science. Springer, 2011.

- [6] Qiu Chen, Koji Kotani, Feifei Lee and Tadaihiro Ohmi (2010). Face Recognition Using Self-Organizing Maps, Self-Organizing Maps, George K Matsopoulos (Ed.), ISBN: 978-953-307-074-2, InTech, DOI: 10.5772/9173.
- [7] David G. Lowe. Distinctive image features from scale-invariant keypoints. International. journal of computer vision, 60, 2004.
- [8] D.G. Lowe. Object recognition from local scale-invariant features. In International Conference on Computer Vision, pages 1150-1157, 1999.
- [9] Chen, Ning. Fuzzy Classification Using Self-Organizing Map and Learning Vector Quantization. Lecture Notes in Computer Science. Data Mining and Knowledge Management. Springer Berlin Heidelberg. pp 41-50. ISBN=978-3-540-23987-1. 2005
- [10] Face recognition homepage. URL=[www.face-rec.org/databases](http://www.face-rec.org/databases)
- [11] Estrebow C., Lanzarini L., Hasperué W. Voice recognition based on probabilistic SOM. Conferencia Latinoamericana de Informática. CLEI 2010. Paraguay. October 2010. ISBN 978-99967-612-0-1. (CD Proceedings)