A Probabilistic Self-Organizing Map For Facial Recognition

Grégoire Lefebvre and Christophe Garcia
Orange Labs - 4 rue du clos Courtel, 35512 Cesson Sévigné, France
{firstname.lastname}@orange-ftgroup.com

Abstract

This article presents a method aiming at quantifying the visual similarity between an image and a class model. This kind of problem is recurrent in many applications such as object recognition, image classification, etc. In this paper, we propose to label a Self-Organizing Map (SOM) to measure image similarity. To manage this goal, we feed local signatures associated to the regions of interest into the neural network. At the end of the learning step, each neural unit is tuned to a particular local signature prototype. During the labeling process, each image signature presented to the network generates an activity vote for its referent neuron. Facial recognition is then performed by a probabilistic decision rule. This scheme offers very promising results for face identification dealing with illumination variation and facial poses and expressions.

1. Introduction

Over the past two decades, face recognition has been an important research subject in the pattern recognition field that has been extensively investigated. Due to its potential commercial applications, such as surveillance, human-computer interactions, vision systems and video indexing, identifying human faces remains a challenging problem. The main difficulties are due to illumination constraints, facial expressions and orientations.
Whereas holistic matching methods use the whole face region and face feature-based methods consider local regions as the eyes, nose and mouth, we investigate the "bag of features" representation [3] which models an object by a set of local signatures. Based on interest point detection, we assume that the relevant salient biometric information is sufficiently redundant whatever view is considered. For each salient point, we focus on its near influence area to describe the signal singularity. The edge descriptor should then compute a stable signature, regarding geometric transformation. This large amount of training information is then organized thanks to a Self Organizing Map [8]. A decision rule based on conditional probability is then defined, using a learning by example strategy from facial feature stimulation on SOM neurons.
This paper is organized as follows. In section 2, we first present our face recognition scheme based on SOM learning from local descriptions. Section 3 is presents some experimental results that illustrate the performances of the proposed method. And finally, conclusions are drawn.

2. Supervised face classification scheme

2.1. System architecture

As underlined in [4], a classification scheme is generally composed of three main steps: pre-processing, feature extraction and classification. In our approach, the pre-processing step consists in detecting some salient points in the image to be compared, reducing thus the zones of interest to a limit number of regions considered as perceptually relevant. From each detected salient point, a salient patch is extracted and a local feature vector is calculated. Each local feature vector is then fed into a SOM network resulting in a neural activity map composed of all winning cells, as shown in our previous study [10]. Finally, to complete the feature extraction step of our classification scheme, the obtained activity map is used in order to label each SOM unit by measuring the frequency of SOM prototype appearance, individual by individual. Assessing the similarity between a test image and an individual model is then reduced to compute a new decision rule from a maximum *a posteriori* probability (MAP) using previous labeled neuron stimulations. The different computational steps used in this method are detailed in the next sections.

2.2. Regions of interest description

According to the active vision mechanisms [7], the

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

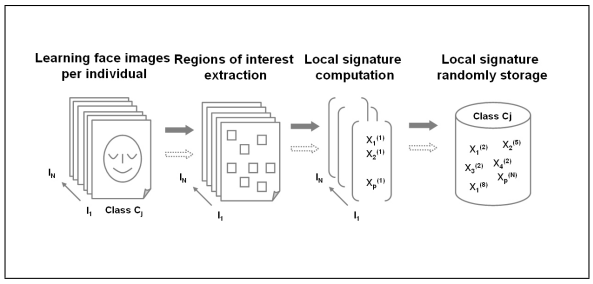

Figure 1. Facial feature description.

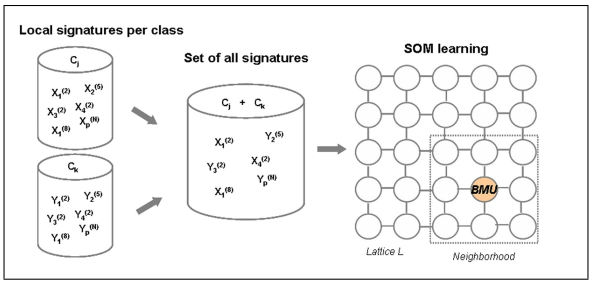

Figure 2. SOM learning process.

goal of salient point detectors is to find perceptually relevant image locations. Many detectors have been proposed in the literature [6, 2]. The salient locations selected by human visual system contain generally high contrast, lines and edges. Following this observation, we focus our interest on the salient point detector [9] that uses a Haar wavelet analysis to find pixels on sharp region boundaries. The facial features are then described in some regions of interest around each salient point (cf. Figure 1). Many local descriptors have been proposed [14, 11, 1], but we chose the Regularity Foveal Descriptor [13] using foveal wavelet to describe the 2D signal singularity. Using again wavelets is justified by the consideration of the human visual system for which multi-resolution, orientation and frequency analysis are of prime importance.

2.3. SOM learning

For face recognition, we learn a global SOM from all face signatures, without considering each identity (cf. Figure 2). The Kohonen model [8] is based on the construction of a neuron layer in which neural units are arranged in a lattice L. Usually, the lattice is two dimensional (rectangular or hexagonal). The neural layer is innervated by d input fibers, called *axons*, which carry the input signals and excite or inhibit the cells via synaptic connections. The Kohonen network aims at preserving the topology of the input space and at tuning each cell to a particular set of stimuli. To reach these goals, the excitation of neurons has to be restricted to a spatially localized region in L and the location of this region has to be determined by those neurons that respond most intensively to a given stimulus. Moreover, L acts as a topographic feature map if the location of the most strongly excited neurons is correlated in a regular and continuous fashion with a restricted number of signal features of interest [12]. Neighboring locations in L correspond thus to stimuli with similar features. To satisfy these properties, a neighboring function between cells must be added in the network model. For this purpose, each cell $i \in L$ is connected to a set $N_L(i)$ of neighboring cells, defining thus a topological ordering. The goal of the Kohonen learning algorithm is then to adapt the shape of L to the distribution of the input vectors. The 2D lattice shape changes to capture the input information and the topology existing in the input space. Those two properties can be considered as a competitive learning and a topological ordering. Let M be the input space and $X = x(t)$, $t \in \{1, 2, \ldots, T\}$ be a set of observable samples with $x(t) \in M \subset \Re^d$, t being the time index. Supposing $M = m_i(t)$, $i \in \{1, 2, \ldots, N\}$, is a set of reference vectors with $m_i(t) \in \Re^d$, randomly initialized. The best matching unit (BMU) $m_c(t)$ is then defined by the index c :

$$c = \arg\min_i ||x(t) - m_i(t)||, \forall i = 1, \ldots, N \quad (1)$$

A kernel-based rule is used to reflect the topological ordering observed in the human visual cortex. The updating scheme aims at performing a stronger weight adaptation at the BMU location than in its neighborhood. This kernel-based rule is defined by :

$$m_i(t+1) = m_i(t) + \lambda(t)\phi_c^{(i)}(t)\,[x(t) - m_i(t)] \quad (2)$$

where $\lambda(t)$ designates the learning rate, i.e. a monotonically decreasing sequence of scalar values with $0 < \lambda(t) < 1$ and $\phi_c^{(i)}(t)$ designates the neighborhood function that governs the strength of weight adaptation as well as the number of reference vectors to be updated. Classically, a Gaussian function is used, leading to :

$$\phi_c^{(i)}(t) = \exp\left(-\frac{||r_c - r_i||^2}{2\delta(t)^2}\right) \quad (3)$$

Here, the Euclidian norm is chosen and r_i is the 2D location for the i^{th} neuron in the network. $\delta(t)$ specifies the neighborhood width decreasing during the time.

2.4. SOM labeling

In the following, let us use the following notations to build a SOM labeling process : N the number of SOM

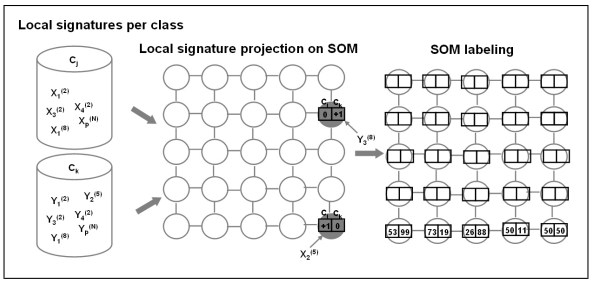

Figure 3. SOM labeling.

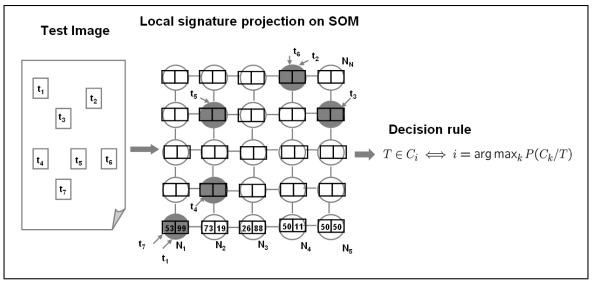

Figure 4. Decision rule.

neurons; K the number of individuals; T the number of local signatures for an image; P the number of learning images for the individual $k \in K$; I_l a learning image; I_t a test image; $A_k(n)$ the activation for the neuron n from the class k learning images ; $L_k(n)$ the label for the neuron n from the class k learning images; $Z_l(n)$ the local signature set from the learning image I_l where the neuron n is the BMU.

As shown on Figure 3, we label the previous learning SOM with the activation of each local signatures from the learning image database. That is to say that we focus on each learning signature projection on SOM lattice and we store each best matching unit stimulation, class by class. Consequently, at the end of the labeling process, each SOM unit is labeled with the number of times that it has been activated as a winner neuron for each individual. This label value is defined by the equation 4 :

$$L_k(n) = \frac{A_k(n)}{\sum_{q=1}^{K} A_q(n)} \tag{4}$$

$$A_k(n) = \sum_{l=1}^{P} card\{Z_l(n)\} \tag{5}$$

$$Z_l(n) = \{x(t) \in I_l, ||x(t) - m_n|| < ||x(t) - m_j||, j \neq n\} \tag{6}$$

2.5. Decision rule

The decision rule consists of maximizing the *a posteriori* probability $P(C_i/I_t)$ to deduce the identity of the individual I_t (cf. Figure 4 and Equation 7). Following the Bayes's theorem and the equiprobability of each individual belonging, we obtain the equation 8. Using the local signature independence for a test image, the probability $P(I_t/C_i)$ can be deduce from its local signature projections on SOM lattice with the associated learning label element $L_k(n)$ as in Equation 9.

$$I_t \in C_i \iff i = \arg\max_k P(C_k/I_t) \tag{7}$$

$$P(C_k/I_t) = \frac{P(I_t/C_k)}{\sum_{q=1}^{K} P(I_t/C_q)}, P(C_k) = \frac{1}{K} \tag{8}$$

$$P(I_t/C_k) = \sum_{n=1}^{N} card\{Z_t(n)\} \times L_k(n) \tag{9}$$

$$Z_t(n) = \{x(t) \in I_t, ||x(t) - m_n|| < ||x(t) - m_j||, j \neq n\} \tag{10}$$

3. Experimental Results

For all the experiments, we configure our SOM with the following rules to reach good learning results from 500 RFD individual signatures in term of accurate input data representation [8] : $T = 500 \times N$; $\lambda(t) = T/(T + 99t)$; $\delta(t)$ decreases linearly from $\sqrt{2}N/2$ to $1/2$.

In these experiments, we focus on two face databases (cf. Figure 5) : The first database named ORL[1] is collected by AT&T and Cambrigde University Laboratories. 40 distinct subjects are available with 10 image samples. For some subjects, 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 second database, YALE[2], contains 165 views of 15 persons. The 11 face images per person present illumination variations and facial expressions as happy, sad, sleepy or surprised. In our experiments, all faces are extracted using the face detector proposed in [5] and resized to 200×200 pixels. In order to evaluate the system performances, we use a 3-fold cross validation method in the following experiments.

Figure 6 presents our method result with different SOM size and Table 1 proposes a comparison with well-known face recognition algorithms published by Yang et al. [15] who use a leaving-one-out cross validation. Then, it is very interesting to observe that our proposal challenges the statistical methods with respectively 98.67% and 94.44% of good identification on the

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[2] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

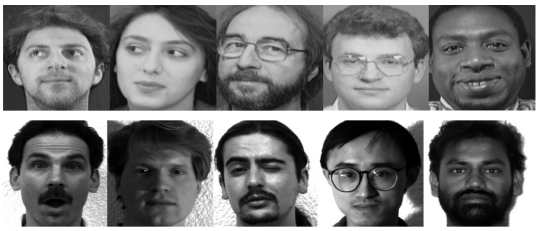

Figure 5. ORL and YALE image samples

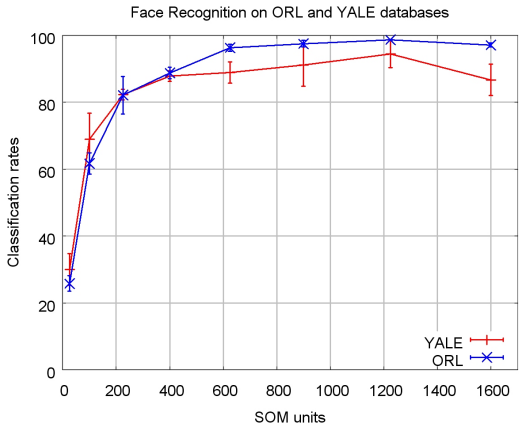

Figure 6. Face recognition results.

Approach	ORL	YALE
ICA[3]	93.75%	71.5%
Eigenfaces	97.5%	71.5%
Kernel eigenfaces	98%	75.8%
Fisherfaces	98.5%	91.5%
Kernel fisherfaces	98.75%	93.9%
Our method	98.67%($\pm$0.12)	94.44%($\pm$4.15)

Table 1. Classification rate comparison.

ORL and YALE databases. The respective standard deviation are 0.12 and 4.15. These performances appear when the SOM size achieves 1225 units. With this configuration, the learning SOM process clusters more precisely the different face signatures and the individual frequency labeling proposes a better classification result. From a size of 225 neurons, we perform more than 80% of good recognitions but when this number is upper than 1225, we are confronted to the classical overlearning issue. Increasing the number of neurons means growing the SOM learning time but the decision result is still immediate, independently to the number of persons and the learning examples.

[3]Independent Component Analysis

4. Conclusion

In this paper, we propose an original face recognition system using directly local signature information. Based on the two main properties of SOM, which are dimension reduction and topology preservation, this architecture features all facial identities by neural activity counts. In order to quantify the visual similarity between a test image and the global neural model, we build a probabilistic decision rule. This solution implemented for facial recognition gives us very promising results. However, a growing and pruning strategy or a hierarchical SOM could be useful to determine automatically the SOM size from learning data.

References

[1] Bay H., Tuytelaars T., and Van Gool L.J. Surf: Speeded up robust features. In *ECCV*, pages 404–417, 2006.

[2] Bres S. and Jolion J.-M. Detection of Interest Points for Image Indexation. In *VISUAL*, pages 427–434. Springer-Verlag, 1999.

[3] Csurka G., Bray C., Dance C., and Fan L. Visual Categorization with Bags of Keypoints. In *ECCV*, pages 327–334, Prague, Czech Republic, May 2004.

[4] Duda R.O., Hart P.E., and Stork D.G. *Pattern Classification*. John Wiley & Sons, 2^{nd} edition edition, 2001.

[5] Garcia C. and Delakis M. Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE TPAMI*, 26(11):1408–1423, 2004.

[6] Harris C. and Stephens M. A Combined Corner and Edge Detector. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[7] Hoffman J.E. and Subramaniam B. The Role of Visual Attention in Saccadic Eye Movements. *Perception & Psychophysics*, pages 787–795, 1995.

[8] Kohonen T. *Self-Organizing Maps*. Springer, 2001.

[9] Laurent C., Laurent N., Maurizot M., and Dorval T. In Depth Analysis and Evaluation of Saliency-based Color Image Indexing Methods using Wavelet Salient Features. *Multimedia Tools and Application*, 31:73–94, 2006.

[10] Lefebvre G., Laurent C., Ros J., and Garcia C. Supervised image classification by SOM activity map comparison. In *ICPR*, volume 2, pages 728– 731, 2006.

[11] Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.

[12] Ritter H., Martinetz T., and Schulten K. *Neural Computation and self-Organizing Maps - an introduction*. Addison-Wesley, New York, 1992.

[13] Ros J. and Laurent C. Description of local singularities for image registration. In *ICPR*, volume 4, pages 61–64, 2006.

[14] Schmid C. and Mohr R. Local Grayvalue Invariants for Image Retrieval. *IEEE TPAMI*, 19(5):530–535, 1997.

[15] Yang M-H. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. *IEEE ICAFGR*, pages 215–220, 2002.