Facial biometry by stimulating salient singularity masks

Grégoire Lefebvre, Christophe Garcia

Facial Biometry by Stimulating Salient Singularity Masks

Grégoire Lefebvre and Christophe Garcia

France Telecom R&D
4, Rue du Clos Courtel
35512 Cesson Sévigné - France
{gregoire.lefebvre,christophe.garcia}@orange-ftgroup.com

Abstract

We present a novel approach for face recognition based on salient singularity descriptors. The automatic feature extraction is performed thanks to a salient point detector, and the singularity information selection is performed by a SOM region-based structuring. The spatial singularity distribution is preserved in order to activate specific neuron maps and the local salient signature stimuli reveals the individual identity. This proposed method appears to be particularly robust to facial expressions and facial poses, as demonstrated in various experiments on well-known databases.

1. Introduction

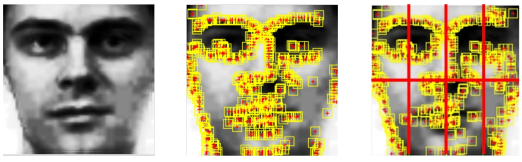
In many computer vision applications, evaluating image content is fundamental. Over the past two decades, face recognition has been an important research subject in the pattern recognition field that has been extensively investigated. Due to its potential commercial applications, such as surveillance, human-computer interactions, vision systems and video indexing, identifying human faces remains a challenging problem. The well-known difficulties continue to be illumination constraints, facial expressions, and facial orientations.

In this paper, we consider these three problems in recognizing human faces. Overcoming the illumination changes, various studies propose to use thermal imageries [1] or near-infrared images [9]. Dealing with facial expressions [10], some algorithms achieve high recognition rates for frontal face images when the size and the position of the face is normalized. Accounting for face variations, the Blanz *et al.* study [2] simulates the process of image formation in 3D space, and they estimates 3D shape and texture of faces from single images. Whereas these systems are competitive, they are still not adequate for many applications in a real life environment.

Consequently, it is still needed to develop a viewpoint-independent face recognition algorithm with illumination change robustness.

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 from natural image categorization [4] 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 compute a stable signature, regarding geometric transformation. This large amount of training information is then organized thanks to a topological map structuring.

In order to build our "bag of facial features", we improve the Tan *et al.* works [14]. In their implementation, the original image is divided into non-overlapping sub-blocks with equal size. Each sub-block is described by the concatenation of each composing pixel value. Our approach differs by only considering the facial salient points and by describing the related patches with their signal singularities. Then, we build an activation mask per individual, using a multi-SOM (*Self Organizing Map* [7]) structure preserving salient region distribution. Consequently, our system uses less information and our region-based multi-SOM scheme creates a more discriminative face recognition model well evaluated by a cumulative minimal quantization error function.



(a) (b) (c)

Figure 1: Interest point locations on a face image (a) with the wavelet-based detector (b) and the wavelet-based detector on a D2x4 grid (c).

This paper is organized as follows. In Section 2, we present our face recognition scheme based on the stimulation of our salient singularity mask. Then, in Section 3, we demonstrate our system's performances with some experimental results. Finally, we put forward several conclusions.

2. Singularity Mask Stimulation

2.1 Salient Point Detection

Building our salient singularity mask, we focus on specific facial information. The goal of salient point detectors is to find perceptually relevant image locations. Many detectors have been proposed in the literature [6, 3, 8]. In this paper, we investigate the wavelet-based point detector proposed in [8], focusing on edges and singularities, by observing that salient locations selected by the human visual system generally contain singularities. Consequently, our system can be trained on small image patches centered on these salient points. The salient point detector in [8] uses wavelet analysis to find pixels on sharp region boundaries. Working with wavelets is justified by the consideration of the human visual system for which multi-resolution, orientation and frequency analysis are of prime importance.

In this paper, we will combine the wavelet-based salient point detectors [8] with a face subdivision to preserve the spatial singularity distribution (see Figure 1). The face subdivision proposes to split the face into 2×2, 2×4 or 4×4 equal regions and the wavelet-based point detectors extract the locations inside each subdivision. Indeed, it seems to be very important to find relevant point in each subdivision, to keep the global face structure because sometimes the lack of points in a particular region is as informative as a strong compactness. For example, the number of interest points should be greater in an eye area than in a cheek region. However, here we want the same number of salient points by subdivision to be compared during face orientation or illumination changes.

2.2 Singularity Description

In our study, we are interested in a local facial edge descriptor. In [11], Lowe proposes the SIFT method (*Scale Invariant Features Transform*). The author has chosen to describe a region by the spatial distribution of the gradient magnitude. It allows us to compute a locally stable representation of an image regarding, affine, scale and illumination changes. In a recent study [13], it has been shown that an edge or more generally a singularity can also be efficiently characterized by considering its Hölder exponents that estimate the edge regularity.

Then, for each singularity point, the Hölder exponent is estimated with foveal wavelets. Orientations θ and Hölder exponent α are then jointly used and we approach their distribution with 3D histograms. To build such histograms, we consider a 32×32 patch around each interest point that we split into 16 sub-regions and we quantify the number of times each pair (α, θ) appears in each sub-region (see Figure 2). We use three Hölder exponent bins into the range $[-1.5, 1.5]$ and eight orientation bins into $[-\frac{\pi}{2}, \frac{\pi}{2}]$. All 3D histograms are concatenated to form the final signature : the Regularity Foveal Descriptor (RFD). The dimension is : 8 orientations $\times$ 3 Hölder bins $\times$ 16 sub-regions = 384. Thus, in the experiment section 3.3.1, the RFD proves its superiority to the SIFT descriptor and to various MPEG7 descriptors [12]. Consequently, the RFD descriptor appears to evaluate better the facial singularity smoothness for our biometric system.

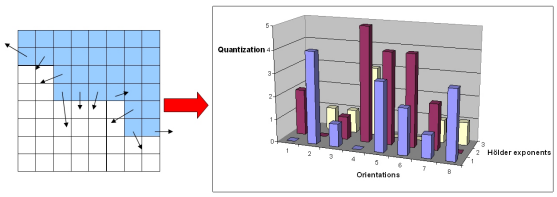
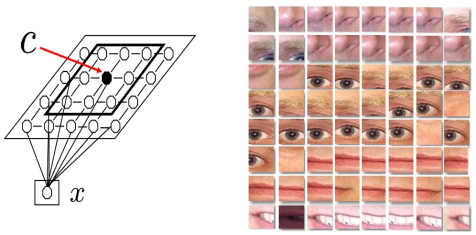


Figure 2: Orientations and Hölder exponents for a sub-region, resulting in a 3D histogram.

2.3 Spatial Information Structuring

In order to structure these local facial vectors into a "bag of facial features", we propose a system based on Kohonen topological maps. The Kohonen model [7] is based on the construction of a neuron layer in which neural units are arranged in a lattice L as shown in Figure 3(a). 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 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 during the learning process to capture the input information and the topology existing in the input space. These two properties can be considered as a competitive learning and a topological ordering. At the end of the learning process, the face patches are clustered in terms of common visual similarity and each SOM unit synthesizes the most recurrent local signature for each visual concept, composing here our "bag of facial features" (See Figure 3(b)).

Let us now describe the SOM algorithm by assuming a SOM lattice structure composed of U neural units. Let $X = x(t)$ be a set of observable samples with $x(t) \in \Re^d$, $t \in \{1, 2, \dots\}$ being the time index. Supposing $M = m_i(t)$ is a set of reference vectors with $m_i(t) \in \Re^d$, $i \in \{1, 2, \dots, U\}$.



(a) (b)

Figure 3: (a) BMU c on the SOM lattice. (b) Face patch projection on the SOM lattice.

If $x(t)$ can be compared simultaneously to all $m_i(t)$ by using a distance measure $d(x(t), m_i(t))$ in the input space, then the best matching unit (BMU) c is defined by :

$$c = \arg\min_i d(x(t), m_i(t)), \forall i = 1, 2, \ldots, U. \quad (1)$$

A kernel-based rule is used to reflect the topological ordering. 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 :

$$m_i(t+1) = m_i(t) + \lambda(t)\phi_{ci}(t)[x(t) - m_i(t)], \quad (2)$$

where $\lambda(t)$ designates the learning rate i.e. a monotonically decreasing sequence of scalar values with $0 < \lambda(t) < 1$. $\phi_{ci}(t)$ represents the neighborhood function. Classically, a Gaussian function is used, leading to :

$$\phi_{ci} = \exp -\frac{||r_c - r_i||^2}{2\delta(t)^2}. \quad (3)$$

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

2.4 Proposed Scheme Overview

A face recognition scheme is generally composed of three main steps: pre-processing, feature extraction and feature classification.

Here, the first step consists of salient point detection thanks to the point detector exposed in the section 2.1.

The second step is the salient region description using the RFD descriptor and signature structuring thanks to a face subdivision based multi-SOM approach, resulting in individual singularity masks. The face activation mask represents the stimulation of our subdivision multi-SOM scheme by the salient singularity signatures. Each mask synthesizes the minimal quantization error of each best matching unit during the learning process. Each model is unique, and corresponds to one person.

The last step of our facial feature classification is done by comparing the different masks using two approaches evaluated on Section 3.3.1 :

- a KNN[1] algorithm determines the nearest learning face mask from the test activation mask,
- the test image class is defined by the smallest value of the cumulative minimal quantization error (CMQE). This value corresponds to the minimal reconstruction error from a test mask to a learning mask, revealing thus its category.

In Figure 4, we present our face recognition algorithm 1. Our system's architecture consists of three steps to build the individual face model.

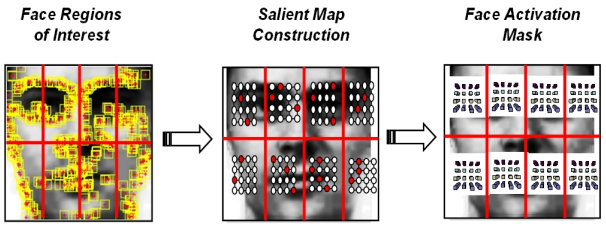


Figure 4: The Proposed System Architecture

First, for each individual, we detect salient points in each face subdivision. We keep the same number of points in each subdivision and the locations are listed in order of saliency.

Secondly, singularity descriptions are computed for each region of interest and project to specific SOMs, one per face subdivision. Then, an activation matrix, corresponding to our multi-SOM face mask, is updated with the best matching unit stimulation inside each SOM. This stimulation is represented by the minimal quantization error between the BMU weight and the local singularity signatures projected on the current SOM.

Finally, the CMQE is calculated for each individual model to evaluate the test image category. The CMQE is the sum of the activation matrix elements, and this value can be regarded as a reconstruction error from a test model to a learning model.

Consequently, the interest point distribution is preserved allowing to identify people even if an occlusion appears. Furthermore, the subdivision based information structuring allows us to select the more recurrent singularities and this property is very useful to deal with facial orientations and expression variations.

[1] K-Nearest Neighbors

Algorithm 1 Salient Face Model Creation

for each individual $j = \{1, \ldots, J\}$ **do**
  for each learning image **do**
    for each face subdivision $d = \{1, \ldots, D\}$ **do**
      Detect the salient points.
      for each salient point **do**
        Compute the local signature $x(t)$ with the RFD.
        The SOM_d corresponding to the current subdivision receives the signature $x(t)$ and and we estimate the minimal quantization error (MQE) for this BMU c by :

$$mqe_c(t) = ||x(t) - m_c(t)||. \quad (4)$$

        The activation matrix M_j corresponding to the full face mask is then updated, with t the time :

$$M_j[c_x, c_y](t+1) = M_j[c_x, c_y](t) + mqe_c(t), \quad (5)$$

        with c_x and c_y respectively the x and y position of c in the multi-SOM architecture.
      end for
    end for
  end for
  The cumulative minimal quantization error $\mathbf{CMQE}_j$ is calculated for each individual map :

$$\mathbf{CMQE}_j = \sum_{x,y} M_j[x, y], \forall (x, y) \in (U \times D)^2. \quad (6)$$

end for

3. Experiments

3.1 Face Databases

In this experimentation section, we focus on three face databases :

- the first database is extracted from the FERET dataset that has been built for the Facial Recognition Technology program[2]. We test our system on 46 individuals with the *fa* and *fb* expressions, corresponding to the regular and alternative facial expressions,
- the second database named ORL[3] 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 YALE[4] database 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 the interior-face area is cropped as shown on Figure 5 with the red rectangle, and resized to 200×200 pixels. In order to evaluate the system performances, we use a leaving-one-out cross validation method in the following experiments.

Figure 5: FERET (a), ORL (b) and YALE (c)

3.2 System Configuration

In all experiments, the SOM networks are configured using the following rules to ensure optimal performance in terms of accurate data representation : the learning steps T are 500 times the cell number U; the learning rate forms a monotone decreasing sequence : $\lambda(t) = \frac{T}{T+99t}$; the neighborhood width $\delta(t)$ which decreases linearly with t from $\frac{\sqrt{2}}{2}U$ to 0.5.

3.3 Experimental Results

3.3.1 SOM Structuring Importance

The first part of the experiment assesses the discriminative power of the singularity descriptor based on the extraction of 3000 points of interest with the detector [8] on a 4×4 face subdivision. The RFD descriptor is computed in its traditional way to obtain a signature dimension of 384. The SIFT descriptor is used with the classical parameters to get a signature of size 128. The classical parameters are to describe a region by 4×4 histograms of 8 orientations. The comparison is made with MPEG-7 descriptors : HCD (*Histogram Color Descriptor*) and HTD (*Histogram Texture Descriptor*). The classification rate is obtained by a vote algorithm for each region of interest of the test image.

[2] http://www.itl.nist.gov/iad/humanid/feret/
[3] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[4] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

Thus, we can see on Table 1 that the RFD description is the most efficient. However, the global classification rate is only 68.48%, so it appears the system should synthesize, cluster and select the crucial information inside a "bag of facial features" to improve the classification performances. Therefore, we present different strategies to structure the salient RFD information to build a model on the FERET database.

A single model corresponds to one model for everyone, that is to say one SOM is used to structure all individuals (Single-SOM). A multi model corresponds to the concatenation of the 46 models representing each individual. Thus, a mask is composed of its personal activation model but also contains its activation strengths to other person models. A multi model can use one SOM per individual (Multi-SOM) or can be built with our face subdivision, proposing several SOMs per individual (Subdivision Multi-SOM).

In this second experiment, we apply five approaches :

1. a KNN using the Single-SOM model;
2. a KNN using the Multi-SOM model;
3. a KNN using the Subdivision Multi-SOM model;
4. the minimal value of CQME determines the test category using the Single-SOM model;
5. the minimal value of CQME determines the test category using the Subdivision Multi-SOM model.

Approach	Classifier	Class Rate
Local HCD	KNN+L2	48.08%
Local HTD	KNN+L2	55.76%
Local SIFT	KNN+L2	65.26%
Local RFD	KNN+L2	68.48%
1 - RFD+Single-SOM20x20	KNN+L2	92.39%
2 - RFD+Multi-SOM20x20	KNN+L2	93.48%
3 - RFD+Multi-SOM5x5+D4x4	KNN+L2	95.62%
4 - RFD+Multi-SOM20x20	Min_CQME+L2	98.91%
5 - RFD+Multi-SOM5x5+D4x4	Min_CQME+L2	$\mathbf{100\%}$

Table 1: Classification rates for the FERET database with different strategy.

Consequently, the experiment shows the SOM structuring importance from a bag of local singularity signatures (cf. Table 1). Indeed, the classification rate grows from 68.48% to 92.39% with the adjonction of a single SOM on the FERET database. Morevover, we can see that our subdivision strategy (i.e. one SOM 5x5 per 4x4 subdivision) in order to build an activation mask which preserves the

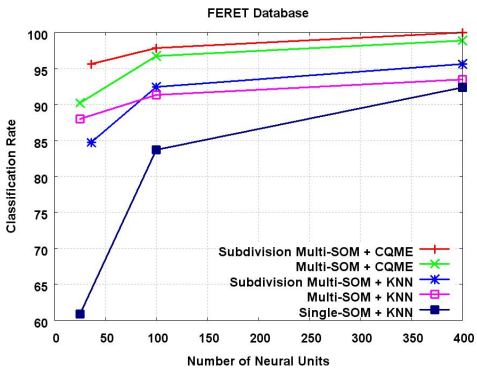


Figure 6: Classification rates for the FERET database.

spatial distribution performs better (95.62%) than a single SOM for each person (92.39%) or a single SOM per person (93.48%). We can remark that one single SOM for 46 persons is not efficient enough to cluster the full information, and that is why a multi scheme strategy is preferable.

For the approaches 1 and 2, we use a size of 20×20 for the SOMs to reach the best classification rates (cf. Figure 6). In the approach 3, the best configuration is a 4×4 subdivision and a dimension of 5×5 for each SOM.

Moreover, the direct CQME comparison between the different masks appears to perform better in order to reach 100% of good identification with the approach 5. Consequently, when we project the test image signatures on the different models, the minimal reconstruction error assigns its category. Moreover, with few SOM units (Multi-SOM3x3+D2x2), our multi-region based strategy achieves 95.65% which allows a very fast learning process for an immediate testing response (cf. Figure 6). The CQME criteria is very interesting because it allows to test an image with the individual models without comparing with all learning faces. Thus, when a new person is added to the system, only its activation mask is computed to take the final decision.

3.3.2 Comparing Performances

This experiment compares our strategy with state-of-the-art facial recognition methods. Here, we use the best configuration shown in the previous section and test the performances on several databases. We compare the proposed method to statistical projection approaches presented in Yang study [15]. We observe on Table 2 that on the ORL database and on the Yale database, our results are very competitive. Indeed, our system achieves respectively 100% and 93.33% of good face recognition on the two databases.

In order to investigate furthermore our salient region strategy, we artificially transform the subject 7 from the ORL database as shown in the Figure 7. It is very interesting

Approach	ORL CR	YALE CR
eigenfaces	97.5%	71.5%
fisherfaces	98.5%	91.5%
Independent Component Analysis	93.75%	71.5%
kernel eigenfaces	98%	75.8%
kernel fisherfaces	98.75%	93.9%
Our approach 4	**100%***	**91.5%**
Our approach 5	**100%***	**93.33%**

Table 2: Classification Rates (CR) on ORL and YALE.

to see how the proposed method our system can deal with partial occlusions and some transformations (contrast, blur, polarization), which is generally not the case with the statistical projection methods. Thus, the interest point detector find enough information even when an occlusion appears to recognize an individual, and the singularity descriptor avoid some color and illumination problems. Nevertheless, some experiments should be realized to show if these properties are repeatable in large databases.

Figure 7: ORL subject 7 recognized with transformations.

4. Conclusions

In this paper, we have presented a novel face recognition method, using the singularity information contained in regions of interest. Based on the main properties of SOM, which are dimension reduction, topology preservation and data accommodation, our scheme gives very promising results. Indeed, for three well-known face databases, our approach overcomes state-of-the-art statistical projection methods and our combination of interest point detection, salient region description and CQME information from a multi-SOM activation proves its robustness to facial orientations, facial expressions and illumination changes. Our multi-region-based architecture synthesizes well the singularity distribution between the different face subdivision and use this information selection to recognize individual faces. We will investigate if the method can be further enhanced by taking into account the spatial geometry between the salient regions.

References

[1] Arandjelovic, O., Hammoud, R., and Cipolla, R. On person authentication by fusing visual and thermal face biometrics. *IEEE AVSS*, pages 50–50, 2006.

[2] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *IEEE TPAMI*, 25(9), 2003.

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

[4] 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.

[5] Christophe Garcia and Manolis Delakis. 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] Kohonen T. *Self-Organizing Maps*. Springer, 2001.

[8] 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*, 2006.

[9] Stan Z. Li, RuFeng Chu, ShengCai Liao, and Lun Zhang. Illumination invariant face recognition using near-infrared images. *IEEE TPAMI*, 29(4), 2007.

[10] Ying li Tian, Takeo Kanade, and Jeffrey F. Cohn. Recognizing action units for facial expression analysis. *IEEE TPAMI*, 23(2):97–115, 2001.

[11] David G. Lowe. Object recognition from local scale-invariant features. In *IEEE ICCV*, volume 2, page 1150, 1999.

[12] Manjunath B. S., Ohm J-R, Vinod V. Vasudevan, and Akio Yamada. Color and texture descriptors. *IEEE TCSVT*, 11(6):703–715, 2001.

[13] Ros J., Laurent C., and Lefebvre G. A cascade of unsupervised and supervised neural networks for natural image classification. In *CIVR*, pages 92–101, 2006.

[14] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. *IEEE TNN*, 16(4):875– 886, 2005.

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