



HAL
open science

A New Approach to Face Image Coding using Gabor Wavelet Networks

Ahmed Chaari, François Cabestaing, Dorra Sellami-Masmoudi

► **To cite this version:**

Ahmed Chaari, François Cabestaing, Dorra Sellami-Masmoudi. A New Approach to Face Image Coding using Gabor Wavelet Networks. First Workshop on Image Processing Theory, Tools and Applications, IPTA'2008, Nov 2008, Sousse, Tunisia. pp.1-5, 10.1109/IPTA.2008.4743787. hal-00521102

HAL Id: hal-00521102

<https://hal.science/hal-00521102>

Submitted on 25 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Approach to Face Image Coding using Gabor Wavelet Networks

Ahmed Chaari*, François Cabestaing* and Dorra Sellami-Masmoudi†

* LAGIS Laboratory, UMR CNRS-8146, University of Lille
59655 Villeneuve d'Ascq, FRANCE

e-mail: ahmed.chaari@univ-lille1.fr, fcab@ieee.org

† National Engineering School of Sfax, ICOS Research unit,
BP W3038, Soukra St., University of Sfax, TUNISIA
e-mail: dorra.masmoudi@ieee.org

Abstract—In this paper, we describe a new approach to face image coding using Gabor Wavelet Networks (GWN). This type of network yields good results in many signal coding applications and was already used for face representation by Krüger in 2001. The main idea is to approximate the face image, considered as a two dimensional function, with a set of Gabor wavelets. We describe an improved network training technique, which uses several face images of the same person rather than a single one. This technique yields the model of the face image, but also statistical criteria defining the significance of each Gabor wavelet in this model.

I. INTRODUCTION

Recognizing a known individual using a single image of his/her face is a very challenging problem that has been widely addressed in the last years. Face recognition usually involves two separate steps. The first one, performed offline, aims at defining a *signature* of the individual. The second step, performed online, processes an image to recover a similar signature which is then searched for in a database.

In the scientific literature, one can distinguish between three categories of signature extraction methods, based on: a) a linear projection of statistical features (Principal Component Analysis — PCA, Independent Component Analysis — ICA, Linear Discriminant Analysis — LDA, etc...); b) a non-linear projection of these features (Neural Networks — NN, Support Vector Machines — SVM, etc...); and c) a direct coding of geometrical features extracted from the images.

Methods of categories a) and b) consider that the face is coded by a feature vector, i.e. a point in a feature space. Since the dimension of the original feature space is usually high, it must be reduced *a posteriori* using a linear or non-linear projection. The other approach consists in finding *a priori* an accurate feature vector defined in a rather small-dimensional space, therefore eliminating the need of a projection.

In this case, the face signature is more a *structured* description — or coding — of the face image than a blind description using a set of features extracted by a statistical method. In this paper, we describe a new approach to face coding, based on the approximation of the image function by a small number of 2D wavelets, organized as a wavelet network (WN).

In the literature, WNs were initially presented as special neural networks using wavelets as activation functions. They were first introduced by Zhang and Benveniste in 1992 in the context of non-parametric regression of parametric functions [1]. They have then been used in other fields, mainly in signal and image processing applications.

In the signal processing field, techniques based on wavelet networks have been used to solve many problems: detection of abnormal vibrations, echo elimination, segmentation and speech recognition [2]. In the image processing field, wavelet networks have been primarily used for human face detection and/or recognition and image compression.

In the first part of this paper, we present the wavelet network (WN) model with a special focus on a particular case, i.e. the Gabor Wavelet Network (GWN). In the second part, we describe the proposed face image coding technique, which adds the notion of *significance* to the GWN model previously proposed by Krüger and Sommer in [3]. Then we conclude the paper and present several prospects.

In the presented work, we consider that the image to be coded shows only the face of the individual, i.e. that it was previously extracted from a more complex image and cropped to a constant size. For the experiments, we have used the face images of the yalefaces database, which contains 165 labeled images of 15 individuals, each one with 11 different facial expressions.

II. WAVELET NETWORK

WNs are similar to Radial Basis Function (RBF) based networks except that their activation functions are wavelets rather than radial basis functions. More precisely, a WN coding a function defined on a compact subset of \mathbb{R}^n using N neurons is defined as [1]:

$$g(\mathbf{v}) = \sum_{i=1}^N w_i \psi(D_i R_i(\mathbf{v} - T_i)) + \bar{g}, \quad (1)$$

where $g(\mathbf{v})$, $\mathbf{v} \in \mathbb{R}^n$, is the resulting function approximation, w_i are the weights of the network output layer, $\psi(\cdot)$ is the mother wavelet, D_i is a diagonal matrix defined by n dilation

parameters, R_i is a rotation matrix defined by $n-1$ angles, T_i is a translation vector with n coordinates, and \bar{g} is the average (DC) value of the approximation.

In 1992, Zhang and Benveniste have described this original network approach as an alternative to other feed-forward neural networks for arbitrary approximation of nonlinear functions. They have shown that, thanks to wavelets properties, a WN is also a universal approximator of any real-valued function defined on a compact subset of \mathbb{R}^n .

The wavelets used in [1] were orthogonal, but other types of wavelets have then been used in several applications. For example, in [4], Feris *et al.* use non-orthogonal wavelets for face image processing, and Oussar [5] implements oblique wavelets structures for process modeling.

Training a wavelet network consists in adjusting wavelet parameters — dilation, orientation and translation — and network weights in order to achieve a correct coding of the input. Training is usually performed by minimizing an objective function, for example the squared difference between the initial function and its approximation.

A. Gabor Wavelet Network

Gabor wavelets, originally introduced in [6], are widely used in image processing applications. These wavelets achieve a good compromise between frequency and spatial resolution and are very efficient for coding image features [7], and detecting specific image components such as edges [4]. Gabor Wavelet Networks (GWNs) result from a combination of WNs and processing techniques based on Gabor wavelets. GWNs have been used in applications related to face image processing, particularly for feature extraction, face tracking [4] and face recognition [8].

Gabor wavelets have also been used in Elastic Bunch Graph Matching, yielding good results in the face recognition application [9]. The underlying idea is that a face is represented by a set of specific, meaningful feature points. Each of these feature points is described by the set of responses of 40 complex Gabor filters that were applied at that point. The filter set is fixed and usually contains Gabor filters with eight different orientations and five different central frequencies [10].

The Gabor mother wavelet is expressed as follows:

$$\psi_g(\mathbf{v}) = \exp\left(-\frac{1}{2}\mathbf{v}^T\mathbf{v}\right) \sin(\mathbf{v} \cdot (1 \ 0 \ \dots \ 0)^T), \quad (2)$$

which corresponds to a sine wave along a single dimension (the first one in this case) weighted by an isotropic gaussian. This mother function, like any other wavelet, can be dilated, rotated and translated according to a set of parameters, denoted by \mathbf{p} . \mathbf{p} contains $3n-1$ elements: n dilation parameters, $n-1$ angles and n coordinates of the translation vector. A wavelet derived from the mother function for the set \mathbf{p} of parameters is expressed as:

$$\psi_{gp}(\mathbf{v}, \mathbf{p}) = \psi_g(D(\mathbf{p})R(\mathbf{p})(\mathbf{v} - T(\mathbf{p}))), \quad (3)$$

where $D(\mathbf{p})$ and $R(\mathbf{p})$ are matrices defined by the dilation and rotation parameters and $T(\mathbf{p})$ is the translation vector.

B. GWN for Image Representation

A GWN, representing functions defined on \mathbb{R}^n , is defined by N Gabor wavelets, therefore by $N \times (3n-1)$ wavelet parameters and N weights. In the special case of grey level images, the image function $I(\mathbf{v})$ — with $\mathbf{v} = (x, y)^T$ — is usually defined on a rectangular subset of \mathbb{R}^2 . Each Gabor wavelet is then defined by 5 parameters: 2 dilation coefficients (d_x, d_y), 1 angle θ and 2 translation parameters (t_x, t_y). In this case, the whole GWN is characterized by $5N$ parameters and N weights.

If we assume that the average (DC) value of the image is null, the approximation $\hat{I}(v, p, w)$ of the image function, for the set \mathbf{p} of $5N$ parameters and the set \mathbf{w} of N weights, is given by:

$$\hat{I}(\mathbf{v}, \mathbf{p}, \mathbf{w}) = \sum_{i=1}^N w_i \psi_{gp}(\mathbf{v}, \mathbf{p}_i), \quad (4)$$

where \mathbf{p}_i denotes the 5 parameters of the i^{th} wavelet.

C. GWN training

The aim of GWN training is to estimate the set of parameters $\hat{\mathbf{p}}$ and weights $\hat{\mathbf{w}}$ that allow for an optimal representation of the original image $I(\mathbf{v})$ by the approximation $\hat{I}(\mathbf{v}, \hat{\mathbf{p}}, \hat{\mathbf{w}})$. The parameters and weights are determined by minimizing an energy function — also called residual — defined as the squared difference between both image functions summed over the definition rectangle of \mathbb{R}^2 .

In the discrete case, we assume that the image is defined for pixels with integer coordinates, i.e. for \mathbf{v} in $[1 \dots X] \times [1 \dots Y]$ if Y denotes the number of rows in the discrete image and X the number of pixels in each row. In this case, the energy function is given by:

$$E(\mathbf{p}, \mathbf{w}) = \sum_{x=1}^X \sum_{y=1}^Y (I(x, y) - \hat{I}(x, y, \mathbf{p}, \mathbf{w}))^2. \quad (5)$$

As said previously, the best coding of the original image by the GWN is obtained with the parameters and weights given by:

$$(\hat{\mathbf{p}}, \hat{\mathbf{w}}) = \arg \min_{(\mathbf{p}, \mathbf{w})} E(\mathbf{p}, \mathbf{w}). \quad (6)$$

This energy function cannot be minimized analytically, but using well known iterative techniques, such as gradient descent. At each iteration, the parameters and weights are updated according to the values of partial derivatives of the energy. In more complex optimization techniques such as Levenberg-Marquardt, which has been used by Krüger and Sommer in [3], second order derivatives allow for a more rapid convergence.

III. OUR APPROACH

In [3], Krüger and Sommer have shown that GWNs including a rather small number of Gabor wavelets — from 16 to 216 — could efficiently represent face images. They have used such GWNs for face coding, face tracking, pose estimation, etc, with great success. However, in their approach, they train each GWN on a single face image, i.e. for a single facial expression. Then, to deal with different facial expressions in their face recognition application, they train several GWNs in order to get many face templates for each individual.

In this paper, we propose a different approach to face coding for the specific case of face recognition. In the previously described approach, each wavelet in the network has the same significance in the coded image. The significance of a wavelet is not its weight, since face features with different contrasts will be coded by wavelets with different weights in the sum. Here, we define the significance of a wavelet as a coefficient telling if this part of the face model remains valid among various face images of the same person.

For example, consider the four images of figure 1, where the first one corresponds to a neutral facial expression. The region corresponding to the nose remains almost constant among all images. Therefore, features like edges or contrast changes modeled in this region can be considered as reliable. On the other hand, regions near the mouth sides or near eyes vary greatly according to the facial expression. Parts of the model describing these regions should be considered as less reliable than the previous ones.

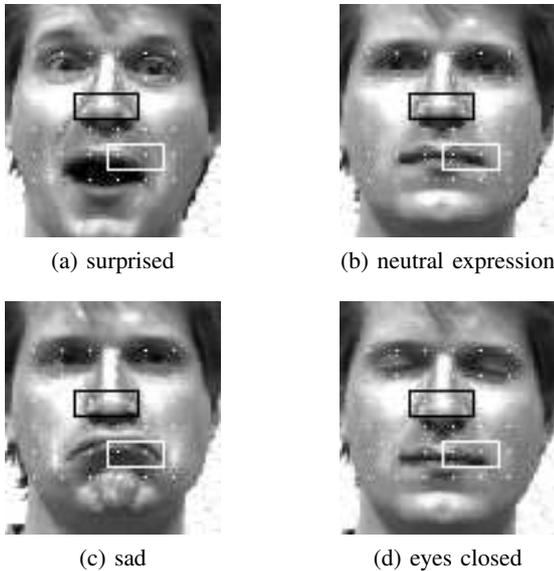


Fig. 1. 4 images of the same person with different facial expressions

A. Modified training process

In order to determine the significance of every wavelet in the network, we have implemented a modified training process. It starts with the same technique as Krüger and Sommer, using

the face image corresponding to the neutral expression to train the GWN, which yields a reference face model.

After this, additional energy minimization steps are performed, but using different face images of the same person as reference function for the GWN rather than the original one. Since the residual is different, the minimum of the energy function moves in the parameter space and additional iterations move the parameters toward the new optimum. In fact, to allow larger displacements in the parameter space, a face image is randomly selected among all available images, used as reference during a fixed — and rather small — number of iterations, then replaced by another randomly selected image, and so on.

In the database, each individual is represented by a *gallery* of face images. Let $G = \{F_i \mid i = 1 \dots n\}$ denote the database, with n the number of individuals. Each set F_i contains m different face images of the same individual: $F_i = \{f_{ij} \mid j = 1 \dots m\}$. In the training process, these m face images are used randomly. For each person F_i , a GWN with parameters $(\mathbf{p}_i, \mathbf{w}_i)$ is optimized and stored. Each individual is therefore represented by a specific GWN in the gallery of models $G' = \{(\mathbf{p}_i, \mathbf{w}_i) \mid i = 1 \dots n\}$. Each GWN $(\mathbf{p}_i, \mathbf{w}_i)$ is considered to be the representation of the individual F_i . GWNs are optimized using algorithm 1.

```

for  $i \leftarrow 1$  to  $n$  do
  for  $j \leftarrow 1$  to  $m$  do
     $I[j] \leftarrow \text{LoadImage}(f_{ij});$ 
  end
   $(p[i], w[i]) = \text{InitializeGWN}();$ 
  for  $k \leftarrow 1$  to  $\text{ItersPerIndividual}$  do
     $f = \text{random}(m);$ 
     $(p[i], w[i]) = \text{OptimizeGWN}(p[i], w[i], I[f]);$ 
     $P[i][k] = p[i];$ 
     $W[i][k] = w[i];$ 
  end
end

```

Algorithm 1: GWN training

In order to determine if a specific wavelet of a GWN is significant in terms of face coding, we analyze the variations of its parameters during the training iterations. This analysis is based on the values of vectors \mathbf{p}_i and \mathbf{w}_i stored during the training. If a wavelet models a part of the face which remains almost constant among all face images, its parameters also remain almost constant. On the other hand, if a wavelet models a feature which varies greatly among images, its parameters will also vary.

B. Analysis of parameter variations

The variation of wavelet parameters during the training depends on the face region in which this wavelet is located. In the images of figure 1, two regions are outlined: one near the nose (black frame) and one near the mouth (white frame). It appears clearly that the region near the nose does not change a lot with the facial expression, but that the region near

the mouth is very different among all the images. Therefore, variations of wavelet parameters will be very different in these two regions.

For example, in figure 2, we have represented the different positions of the center of two wavelets (translation parameters) during the training. The one initially located in the nose region remains at the same position (figure 2(a)). On the other hand, the wavelet initially located near the mouth moves during the iterations and locks on very different features located on the cheek or even near the eye (figure 2(b)).



(a) Wavelet near the nose (b) Wavelet near the mouth

Fig. 2. Variations of wavelet center position during the training process

To analyze more precisely parameter variations, let's consider the curves of figures 3, 4 and 5. Each figure represents the evolution during iterations of the same parameter — translation on figure 3, orientation on figure 4 and dilation on figure 5 — of two different wavelets: one in the nose region and the other near the mouth side (corresponding to regions outlined in black and white in figure 1). It appears clearly that the parameters remains more constant for the wavelet approximating a feature near the nose.

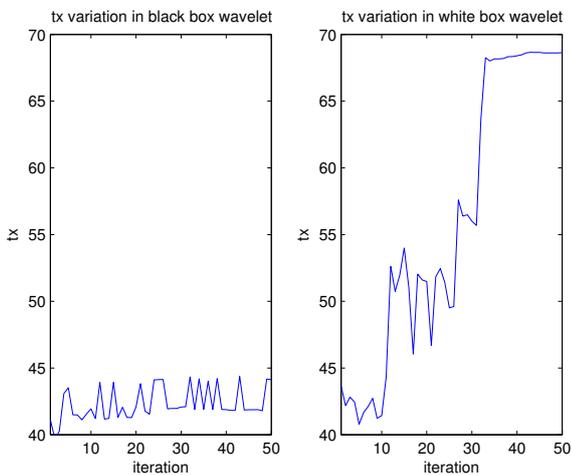


Fig. 3. Variation of translation t_x

To associate a significance index to each wavelet, we compute the variances of all its parameters, which are stored during iterations of the training process. This yields a single value for each parameter, inversely proportional to its significance. For example, we have computed the variances of the 5 parameters

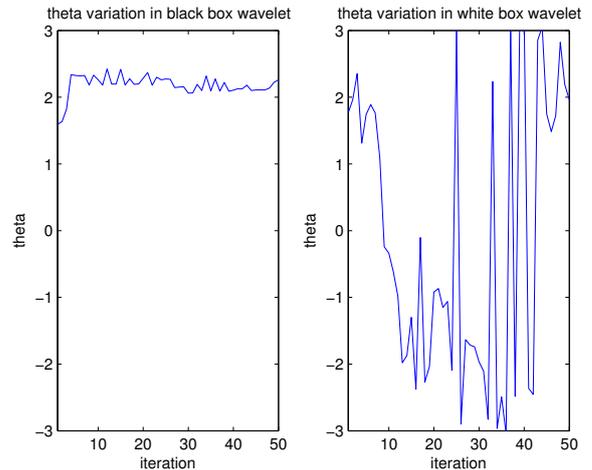


Fig. 4. Variation of orientation θ

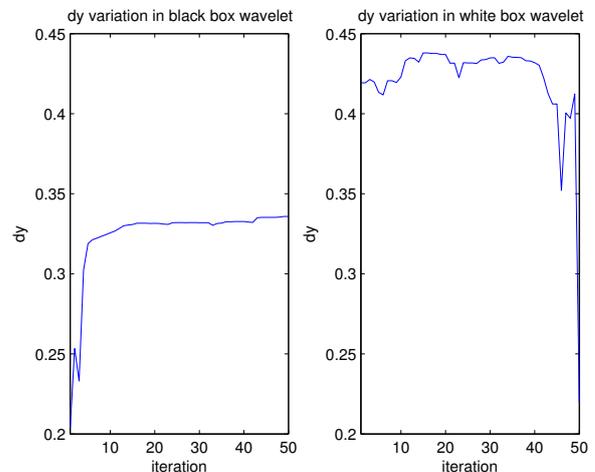


Fig. 5. Variation of dilation d_y

and of the weight in the network for the two wavelets already selected for the example of figures 3, 4 and 5. These variances are given in table I.

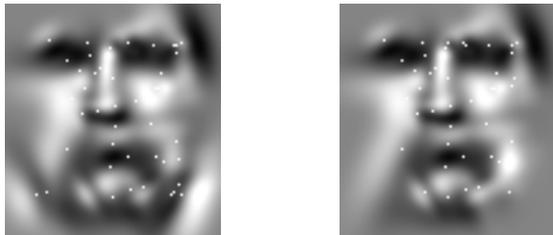
One can verify that the variances of all the parameters for the wavelet in the nose region are much lower than those for parameters of the wavelet in the mouth side region. For example, the variance on the angle θ is more than ten times greater for the wavelet near the mouth side. These values tend to show that the variance of parameters during iterations of GWN training are good indices for measuring the significance of each wavelet in the model.

Figure 6, presents two images reconstructed with different sets of wavelets. We can see that some non significant wavelets, located in unstable regions, can be excluded without really modifying the result: the image reconstructed with 52 wavelets is not very different from that reconstructed with only 43 wavelets.

region	black box	white box
t_x	1.4556	109.832
t_y	0.9425	306.5601
θ	0.0252	4.478
d_x	$4.2 \cdot 10^{-5}$	$5.02 \cdot 10^{-5}$
d_y	$6.3 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$
w	$1.285 \cdot 10^3$	$4.134 \cdot 10^3$

TABLE I

VARIANCES OF PARAMETERS FOR TWO WAVELETS OF DIFFERENT REGIONS (NOSE AND MOUTH SIDE)



(a) 52 wavelets

(b) 43 wavelets

Fig. 6. Images reconstructed with different sets of wavelets

IV. CONCLUSION AND OUTLOOKS

In this paper, we have presented an improved training method for GWNs used to represent face images. In the original approach, proposed by Krüger and Sommer, the GWN is trained on a single face image, yielding a model valid only for the given facial expression. When this modeling technique is used for face recognition, one has to train several GWNs in order to deal with various facial expressions of the same individual.

In the modified technique, we implement a second step in the training process, which allows us to compute statistics on wavelet parameters. To compute these statistics, we change the face image to be coded by the GWN during iterations. The statistics measure the level of significance of each parameter. Parameters remaining stable when the face image is modified are considered to be significant and reliable. Therefore our approach yields a face signature which remains stable across various facial expressions of the same individual.

We are currently implementing a face recognition algorithm which takes advantage of the significance coefficients. In this algorithm, a GWN is trained for each individual of a known set, yielding parameters, weights and significance coefficients. To recognize a person, we first use dual GWNs, to project and reconstructs a face image according to each known model. The person is recognized if the image reconstructed by one of the GWNs is very similar to the original one. The significance coefficients are used in the comparison process.

REFERENCES

- [1] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 889–898, Nov. 1992.
- [2] M. Thuillard, "A review of wavelet networks, wavenets, fuzzy wavenets and their applications," in *European Symposium on Intelligent Techniques, ESIT'2000*, Aachen, Germany, Sep. 2000, pp. 5–16. [Online]. Available: <http://www.erudit.de/erudit/events/esit2000/proceedings/AB-01-1-P.pdf>
- [3] V. Krüger and G. Sommer, "Wavelet networks for face processing," *Journal of the Optical Society of America A*, vol. 19, no. 6, pp. 1112–1119, June 2002.
- [4] R. Feris, R. Cesar, and V. Krüger, "Efficient real-time face tracking in wavelet subspace," in *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, Vancouver, BC, Canada, Jul. 2001, pp. 113–118. [Online]. Available: <http://www.cs.ubc.ca/conferences/ICCV/>
- [5] Y. Oussar, "Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus." these, Université Pierre et Marie Curie, Jul. 1998.
- [6] H. Szu, B. Telfer, and J. Garcia, "Wavelet transforms and neural networks for compression and recognition," *Neural networks*, vol. 9, no. 4, pp. 695–708, 1996.
- [7] R. Feris, J. Gemmell, K. Toyama, and V. Krüger, "Hierarchical wavelet networks for facial feature localization," in *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, Vancouver, BC, Canada, Jul. 2001. [Online]. Available: <http://www.cs.ubc.ca/conferences/ICCV/>
- [8] Y. Zhi and G. Ming, "A SOM-wavelet networks for face identification," in *IEEE International Conference on Multimedia and Expo (ICME'2005)*, July 2005, pp. 852–855.
- [9] L. Wiskott and J.-M. Fellous, "Face recognition by elastic bunch graph matching," *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, pp. 355–396, Jan. 1999.
- [10] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, Jul. 1997.