

Modélisation de visages par un réseau d'ondelettes de Gabor

Ahmed CHAARI¹, François CABESTAING¹, Dorra SELLAMI-MASMOUDI²

¹Laboratoire LAGIS, UMR CNRS-8146, université de Lille
Bât. P2, Cité Scientifique, 59655 Villeneuve d'Ascq, France

²Unité de recherche CIELS, université de Sfax
ENIS, B.P.W 3038 Sfax, Tunisie

Ahmed.Chaari@univ-lille1.fr, Francois.Cabestaing@univ-lille1.fr
Dorra.Masmoudi@enis.rnu.tn

Résumé – Nous proposons une nouvelle méthode de codage d'une image d'un visage par un réseau d'ondelettes de Gabor. Ce type de réseau permet d'approcher toute fonction par une combinaison linéaire d'ondelettes de Gabor auxquelles sont associées des poids. Les paramètres du réseau, à savoir les poids et les coefficients intervenant dans la définition des ondelettes de Gabor, sont déterminés par apprentissage. Dans notre méthode, nous associons un indice de confiance à chaque ondelette, lequel permet d'accorder plus d'importance aux ondelettes les plus significatives. Le coefficient de confiance est déterminé en étudiant les évolutions des paramètres de l'ondelette au cours de l'apprentissage, étape qui dans notre cas fait intervenir plusieurs images du visage d'un même individu avec différentes expressions faciales. Le masque obtenu par combinaison des amplitudes des ondelettes pondérées par les coefficients de confiance indique les zones du visage contenant des indices caractéristiques fiables pour la reconnaissance.

Abstract – We propose a new method for coding a face image with a Gabor wavelet network. This type of network allows for approximating any function with a linear combination of weighted Gabor wavelets. Network parameters, i.e. weights and coefficients defining the wavelet functions, are obtained through a training process. In our method, we associate a confidence coefficient to each wavelet, which indicates its level of significance in the face coding. The confidence coefficient is determined by studying the variations of wavelet parameters during the training step, that in our case uses several face images of the same person with different facial expressions. The resulting mask, obtained by weighting the wavelet amplitudes by the confidence coefficients, indicates the face areas that provide accurate and consistent features for face recognition.

1 Introduction

Depuis les années 1980, la reconnaissance de visage ne cesse de susciter l'intérêt de la communauté scientifique. Cette technique de biométrie, qui se veut efficace, sans contact et sans aucune coopération de l'utilisateur, représente maintenant un axe de recherche très important pour les groupements de recherche académiques, mais aussi et surtout pour les entreprises privées spécialisées dans la biométrie et la sécurité. Les systèmes de reconnaissance de visages qui ont été développés intègrent en général la même procédure de traitement composée de trois étapes : acquisition et pré-traitement des images, apprentissage et extraction de signature (étape hors ligne) et enfin reconnaissance et décision (étape en ligne).

Plusieurs catégories de méthodes ont été proposées et décrites dans la littérature en vue de reconnaître une image d'un visage, à savoir les techniques statistiques, les techniques géométriques, les techniques reposant sur une modélisation 3D du visage. Parmi les techniques jugées efficaces figurent en bonne place celles qui modélisent la fonction image, notamment par l'intermédiaire de fonc-

tions élémentaires comme les ondelettes. Dans cette catégorie, les réseaux approchant la fonction image par une combinaison linéaire d'ondelettes de Gabor ont prouvé leur grande efficacité et leur robustesse vis-à-vis des changements qui peuvent survenir lors de l'acquisition des images. En effet, l'apparence de l'image du visage peut être modifiée par des facteurs intrinsèques (grimaces, expressions faciales variées) et des facteurs extrinsèques (port de lunettes, éclairage, vêtements, etc...).

Dans cet article, nous présentons une méthode de modélisation de visage par un réseau d'ondelettes de Gabor. Cette méthode permet d'extraire une signature discriminante d'une personne à partir de plusieurs images représentant différentes expressions faciales. Une analyse des paramètres des ondelettes lors de la phase d'apprentissage permet d'associer un indice de confiance à chaque ondelette. L'article comporte trois parties, la première est dédiée à la présentation du réseau d'ondelettes de Gabor, la deuxième décrit l'algorithme d'apprentissage et la troisième est consacrée à la description de la démarche d'extraction des indices de confiance associés aux ondelettes.

2 Réseau d'ondelettes de Gabor

Le réseau d'ondelettes de Gabor (GWN: Gabor Wavelet Network) a été introduit par Krüger pour la représentation d'images [1]. Son architecture est similaire à celle des réseaux RBF (Radial Basis Function), les fonctions radiales d'activation étant remplacées par des ondelettes de Gabor. Dans un GWN, l'image en niveaux de gris $I(\mathbf{v})$, avec $\mathbf{v} = (x \ y)^T$, est considérée comme une fonction \mathcal{L}^2 définie sur un support compact de \mathbb{R}^2 [2]. Cette fonction peut être approchée par la superposition de N ondelettes de Gabor ψ_g dont chacune est définie par un vecteur regroupant 5 paramètres $(t_x \ t_y \ \theta \ d_x \ d_y)^T$ — deux coefficients de translation t_x et t_y , un angle θ et deux coefficients de dilatation d_x et d_y — et pondérée par un poids w . Plus précisément, l'approximation est obtenue par la combinaison linéaire suivante :

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

dans laquelle \mathbf{p}_i et w_i sont respectivement le vecteur des 5 paramètres et le poids de la $i^{\text{ème}}$ ondelette, \mathbf{p} et \mathbf{w} désignant respectivement l'ensemble des paramètres (vecteur à $5N$ composantes) et des poids. L'expression analytique de l'ondelette de Gabor est la suivante [3] :

$$\begin{aligned} \psi_g(\mathbf{v}, \mathbf{p}_i) &= \sin[d_{x_i} \cos \theta_i (x - t_{x_i}) - d_{y_i} \sin \theta_i (y - t_{y_i})] \\ &\times \exp \left[-\frac{1}{2} \left(d_{x_i} \cos \theta_i (x - t_{x_i}) \right. \right. \\ &- \left. \left. d_{y_i} \sin \theta_i (y - t_{y_i}) \right)^2 - \frac{1}{2} \left(d_{x_i} \sin \theta_i (x - t_{x_i}) \right. \right. \\ &+ \left. \left. d_{y_i} \cos \theta_i (y - t_{y_i}) \right)^2 \right]. \end{aligned} \quad (2)$$

Cette ondelette a également été utilisée dans la technique d'EBGM (Elastic Bunch Graph Matching), laquelle a donné de bons résultats dans le cadre de la reconnaissance de visage [4]. L'idée de base consiste à représenter le visage par un ensemble d'ondelettes codant ses principales caractéristiques. Chaque point est représenté par un filtre regroupant 40 ondelettes de Gabor [5].

Dans la phase d'apprentissage, l'ajustement des paramètres du réseau pour une image discrète $I(\mathbf{v})$ est assuré par la minimisation, selon l'algorithme de Levenberg-Marquardt [6], de la fonction énergie suivante :

$$E(\mathbf{p}, \mathbf{w}) = \sum_{x=1}^X \sum_{y=1}^Y (I(\mathbf{v}) - \hat{I}(\mathbf{v}, \mathbf{p}, \mathbf{w}))^2, \quad (3)$$

dans laquelle X et Y sont respectivement le nombre de lignes et le nombre de colonnes de l'image $I(\mathbf{v})$.

À l'issue de cette phase, nous obtenons les paramètres optimisés $(\hat{\mathbf{p}}, \hat{\mathbf{w}})$ du réseau codant au mieux l'image originale.

$$(\hat{\mathbf{p}}, \hat{\mathbf{w}}) = \arg \min_{(\mathbf{p}, \mathbf{w})} E(\mathbf{p}, \mathbf{w}). \quad (4)$$

Dans ce cas, le codage n'est valable que pour une seule expression faciale de la personne [7, 8]. Dans [9], nous avons proposé une amélioration de la méthode de Krüger, qui permet d'obtenir un modèle à partir de plusieurs images d'un même individu (*cf.* section 3). Dans la section 3, nous montrons qu'il est possible d'associer un indice de confiance à ce modèle, ce qui permettra par la suite de pondérer les ondelettes durant la phase de reconnaissance.

3 Modélisation via plusieurs images

L'algorithme d'apprentissage proposé par Krüger est exécuté sur une seule image et ne permet donc le codage que d'une seule expression faciale dans les paramètres du GWN [3]. Dans notre approche, nous proposons d'optimiser les paramètres du GWN en utilisant plusieurs images du visage d'un même individu, représentant par exemple différentes expressions faciales. L'image servant de référence est changée périodiquement durant l'apprentissage, sans remise à zéro des paramètres du réseau, ce qui fournit en définitive un modèle valable pour l'ensemble des images [9].

En effet, après n itérations, nous changeons l'image de référence I dans le processus d'optimisation (*cf.* équation 3) tout en gardant les mêmes paramètres. Le choix de l'image se fait de manière aléatoire parmi les m images choisies préalablement de la même personne avec différentes expressions faciales. Durant cette phase d'apprentissage, nous enregistrons pour chaque itération les valeurs des différents paramètres (\mathbf{p}, \mathbf{w}) obtenues ainsi que le numéro de l'image choisie comme référence.

Dans les exemples présentés, nous avons choisi de modéliser le visage avec quatre images faciales (*cf.* figure 1), et de changer l'image de référence tout les dix itérations ($m = 4, n = 10, \text{nombre total d'itérations} = 500$).

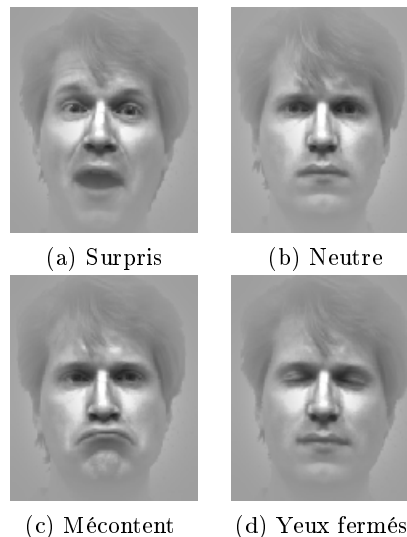


Figure 1: 4 images d'un même individu

Il est important de signaler que ces images ont subi une phase de pré-traitement qui consiste à repositionner le visage au centre de l'image et à appliquer un filtre (Hamming) pour effacer les bords présents entre le visage et le fond. Ce pré-traitement a un impact considérable sur l'efficacité de la phase d'optimisation. Par exemple, la figure 2 présente deux résultats de codage de la même image d'un visage par un GWN, avec et sans prétraitement. Dans l'image (a), on constate que quelques ondelettes se sont déplacées vers le bord du visage pour coder des éléments du fond, alors que dans (b) toutes les ondelettes sont restées à l'intérieur du visage pour coder ses éléments caractéristiques.

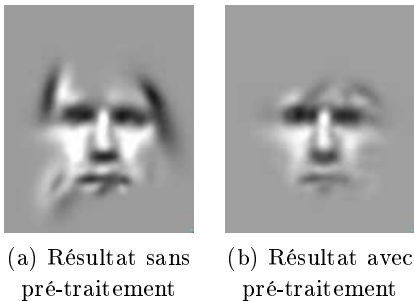


Figure 2: Utilité du pré-traitement des images

La méthode que nous avons proposée initialement présente un inconvénient lié au fait qu'elle ne tient pas compte explicitement de la variabilité de l'image selon l'expression faciale [9]. Considérons par exemple les positions occupées dans l'image par les centres (paramètres t_x et t_y) de deux ondelettes particulières durant le processus d'apprentissage, l'une située dans la région du nez, l'autre dans celle de la bouche. Dans la figure 3, ces positions sont repérées par des points noirs dans les deux images.

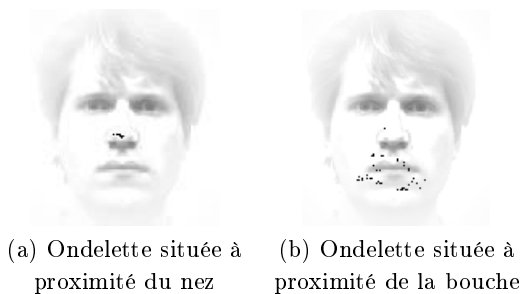


Figure 3: Variations des positions des centres d'ondelettes particulières durant l'apprentissage

On constate que la position du centre de l'ondelette placée à proximité de la bouche est très variable (figure 3(b)), alors que le centre de celle placée sur le nez ne bouge presque pas (figure 3(a)). Cela s'explique aisément par le fait que le nez conserve à peu près la même apparence dans toutes les images, alors que la région proche de la bouche est très variable (*cf.* figure 1). La variabilité

des paramètres d'une ondelette est donc directement liée à celle de l'élément du visage qu'elle modélise. Cette propriété nous a incités à associer à chaque ondelette un indice de confiance, dont la valeur dépend de la variabilité des paramètres de cette ondelette durant l'apprentissage.

4 Indices de confiance

Durant l'apprentissage, nous enregistrons l'évolution de tous les paramètres (\mathbf{p} et \mathbf{w}), ce qui nous permet par la suite d'analyser la variabilité de chacune des ondelettes. À partir de ces variations, notre objectif est de déterminer un indice de confiance associé à chaque ondelette. Si ce coefficient est proche de 1, l'ondelette est considérée comme très significative, ce qui correspond au fait qu'elle code une partie du visage qui dépend peu de l'expression faciale. En revanche, une valeur proche de 0 signifie que l'ondelette code une partie du visage qui varie fortement d'une expression faciale à une autre.

L'indice de fiabilité peut être déterminé selon le principe suivant: si les séries de valeurs successives des paramètres de l'ondelette ont une variance importante — c'est à dire si l'ondelette varie beaucoup durant l'apprentissage —, l'indice de confiance associé à cette ondelette sera proche de 0; inversement, si les séries de valeurs successives des paramètres de l'ondelette ont une variance faible, l'indice de confiance sera proche de 1. Nous proposons d'utiliser la méthode suivante pour déterminer l'indice de confiance :

1. Calcul de la variance des paramètres $[\text{Var}(t_{x_i}), \text{Var}(t_{y_i}), \text{Var}(\theta_i), \text{Var}(d_{x_i}), \text{Var}(d_{y_i})]$ pour chaque ondelette,
2. Calcul de la variance moyenne $[\bar{V}_{t_x}, \bar{V}_{t_y}, \bar{V}_\theta, \bar{V}_{d_x}, \bar{V}_{d_y}]$ pour toutes les images de la base,
3. Calcul de l'écart type des variances $[\sigma_{t_x}, \sigma_{t_y}, \sigma_\theta, \sigma_{d_x}, \sigma_{d_y}]$ pour toutes les images de la base,
4. Normalisation de la variance de chaque paramètre au moyen de ces valeurs. Par exemple, pour le paramètre de translation t_x de la $i^{\text{ème}}$ ondelette, on calcule $\Delta_i^{t_x} = \frac{\text{Var}(t_{x_i}) - \bar{V}_{t_x}}{\sigma_{t_x}}$,
5. Transformation de la moyenne des variances normalisées via une fonction sigmoïde pour recadrer entre 0 et 1 :

$$C_i = \frac{1}{1 + \exp(\tau(\Delta_i^{t_x} + \Delta_i^{t_y} + \Delta_i^\theta + \Delta_i^{d_x} + \Delta_i^{d_y}))} \quad (5)$$

L'indice de confiance associé à chaque ondelette peut être utilisé pour pondérer cette dernière lors de l'utilisation du modèle, par exemple lors d'une phase de reconnaissance de visage. Les indices de confiance associés aux

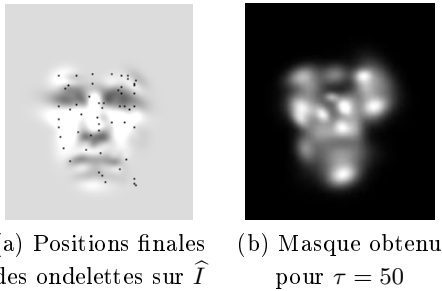
ondelettes du modèle peuvent être représentés par l'intermédiaire d'une image, que nous qualifions de masque $M(\mathbf{v}, \mathbf{p})$, dont le niveau de gris est donné par :

$$M(\mathbf{v}, \mathbf{p}) = \sum_{i=1}^N C_i \psi'_g(\mathbf{v}, \mathbf{p}_i), \quad (6)$$

dans laquelle $\psi'_g(\mathbf{v}, \mathbf{p}_i)$ désigne l'amplitude de l'ondelette de Gabor (sans sa partie oscillante), donnée par :

$$\begin{aligned} \psi'_g(\mathbf{v}, \mathbf{p}_i) = & \exp \left[-\frac{1}{2} \left(d_{x_i} \cos \theta (x - t_{x_i}) \right. \right. \\ & - \left. d_{y_i} \sin \theta (y - t_{y_i}) \right)^2 - \frac{1}{2} \left(d_{x_i} \sin \theta (x - t_{x_i}) \right. \\ & \left. \left. + d_{y_i} \cos \theta (y - t_{y_i}) \right)^2 \right]. \end{aligned} \quad (7)$$

L'image (a) de la figure 4 représente les centres des ondelettes superposés à l'image du modèle codant au mieux les quatre visages de la figure 1. L'image (b) de cette même figure représente le masque associé. Une zone claire dans le masque correspond à une ondelette qui peut être considérée comme significative dans le modèle. On constate par exemple que les quelques ondelettes codant le visage à proximité du nez sont relativement significatives. En conclusion, cette méthode nous permet d'obtenir un codage pertinent des attributs les plus discriminants d'un visage à partir de différentes expressions faciales.



(a) Positions finales des ondelettes sur \hat{I} (b) Masque obtenu pour $\tau = 50$

Figure 4: Modèle du visage et masque associé

5 Conclusion

Les réseaux d'ondelettes de Gabor ont montré une grande efficacité dans différentes applications, en particulier la reconnaissance automatique de visage et le suivi de visage dans une séquence vidéo. Nous avons montré dans cet article l'utilité de suivre l'évolution des paramètres des ondelettes lors de la phase d'apprentissage. Une analyse profonde du comportement des ondelettes durant cette phase, particulièrement quand on change le visage utilisé comme cible du processus d'optimisation, peut renseigner sur l'utilité de chacune d'entre elles et permet donc de définir son importance dans la modélisation finale du visage.

L'algorithme d'apprentissage présenté dans cet article permet de prendre en compte plusieurs expressions faciales de la même personne, ce qui donne plus de généralité au modèle extrait. Les perspectives de ce travail peuvent

consister à analyser l'intérêt de l'utilisation des indices de confiance, ou du masque associé, sur le taux de reconnaissance des visages, notamment sur des bases de données standard.

References

- [1] V. Kruger 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.
- [2] C. Hu, R. Feris, and M. Turk, "Active wavelet networks for face alignment," in *British Machine Vision Conference*, East Eaglia, Norwich, UK, 2003.
- [3] R. S. Feris, R. Cesar, and V. Kruger, "Efficient real-time face tracking in wavelet subspace," in *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, no. 7024549, Vancouver, BC, Canada, Jul. 2001, pp. 113–118.
- [4] 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.
- [5] 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.
- [6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Press Syndicate of the university of Cambridge, 1992, ch. 15 Modeling Data, pp. 681–683.
- [7] R. S. Feris, J. Gemmell, K. Toyama, and V. Kruger, "Hierarchical wavelet networks for facial feature localization," in *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, Vancouver, Canada, Jul. 2001.
- [8] Y. Z. G. Ming, "A SOM-wavelet networks for face identification," in *IEEE International Conference on Multimedia and Expo, ICME 2005*, July 2005, pp. 4–.
- [9] A. Chaari, F. Cabestaing, and D. Sellami-Masmoudi, "A new approach to face image coding using gabor wavelet networks," in *Proceedings of the International Workshops on Image Processing Theory, Tools and Applications (IPTA '08)*, Sousse, Tunisia, Nov. 2008.