

Fusion de biométries sans contact : paume et visage

Audrey POINSOT, Fan YANG, Michel PAINDAVOINE

Laboratoire LE2I CNRS 5158

Faculté des Sciences Mirande, Université de Bourgogne, 21078 Dijon, France

audrey.poinsot@u-bourgogne.fr, fanyang@u-bourgogne.fr, paindav@u-bourgogne.fr

Résumé – La biométrie est une application majeure de la reconnaissance de forme. Dans cet article, une chaîne de reconnaissance biométrique basée sur la paume et le visage est présentée. Les traitements utilisés sont adaptatifs, peu coûteux et facilement implémentables sur un système embarqué pour une utilisation grand public. Ils respectent de plus des contraintes de faible connaissance a priori. Le système proposé est totalement sans contact, apportant confort et hygiène à l'utilisateur. L'extraction de paramètres par filtrage de Gabor est étudiée dans le cas de la paume et du visage. Les possibilités de fusion sont aussi discutées et testées sur une base multimodale de 40 personnes. De bonnes performances ont été obtenues grâce à la paume seule, mais aussi grâce à la fusion : l'apport de la multimodalité est confirmée. Cette étude préliminaire valide la faisabilité d'un système biométrique embarqué robuste et fiable.

Abstract – Biometric science is an important application of pattern recognition. This paper proposes an adaptive, contactless, biometric system which combines two modalities: palmprint and face. The processing chain has been designed to overcome embedded system constraints and small sample set problem. After a palmprint is extracted from a hand image, Gabor filters are applied to both the palmprint and face in order to extract parameters, which are then used for classification. Fusion possibilities are also discussed and tested using a multimodal database of 40 persons designed by the authors. High recognition performance has been obtained with palmprint only and with fusion of palmprint and face. This preliminary study shows the feasibility of a robust and efficient multimodal hardware biometric system.

1 Introduction

La biométrie est une branche de la reconnaissance de forme qui suscite toujours plus d'intérêt. Les systèmes biométriques permettent en effet d'identifier automatiquement les personnes à l'aide de caractéristiques physiques ou comportementales telles que le visage, l'empreinte, la signature ou encore la démarche. Au vu des besoins grandissants en matière de sécurité, leur fiabilité est un enjeu essentiel.

Chaque caractéristique biométrique possède ses avantages et ses inconvénients. L'utilisation d'une seule modalité implique parfois certains désagréments auxquels il faut faire face comme la mauvaise qualité ponctuelle des données, la non-universalité de la caractéristique choisie, la faiblesse des résultats ou encore la possibilité de l'usurpation d'identité. C'est pourquoi de plus en plus de recherches se focalisent sur la fusion de plusieurs sources d'information biométrique : la biométrie multimodale permet de palier certains inconvénients de la monomodalité, mais surtout d'augmenter les performances [1].

Le but de notre travail est de proposer un système biométrique multimodal destiné à des applications grand public qui respecte des contraintes de confort et d'embarquabilité (calculs peu coûteux, robustesse, temps réel, utilisation mémoire réduite, faible encombrement, basse consommation, ...). Dans ce contexte, le problème de la faible connaissance a priori devient incontournable (« small sample set problem ») : afin de réduire l'utilisation mémoire et la durée d'acquisition, seuls quelques exemples peuvent être capturés lors de la phase d'enregistre-

ment de l'utilisateur.

Dans cet article, nous proposons un système adaptatif, utilisant la paume et le visage. Les biométries basées sur la main demandent traditionnellement un contact avec le capteur, mais de plus en plus d'études s'attachent à la rendre plus confortable et plus hygiénique en supprimant cette obligation de contact [2]. Il existe déjà quelques résultats en fusion de paumes et de visages [3, 5], mais aucune de ces méthodes n'est adaptée aux contraintes de coût et de faible connaissance a priori que l'on retrouve dans un système embarqué grand public. Dans les sections suivantes, nous décrivons la chaîne complète de traitement, puis nous présenterons les résultats expérimentaux, et discuterons de nos perspectives de travail.

2 Système biométrique complet

2.1 Reconnaissance de la paume sans contact

Nous travaillons sur des images de mains gauches acquises sur fond vert, sans aucun contact, et dans des conditions d'éclairage naturelles, à l'aide d'une webcam Logitech QuickCam Pro 9000. Au cours de l'expérience, les sujets ont eu pour seules consignes de placer leur main horizontalement, et de façon à ce que leurs doigts ne se touchent pas entre eux. Un exemple d'image acquise selon ce protocole est visible FIG. 1.

Les images obtenues ne sont pas directement exploitables : elles doivent subir un prétraitement au cours duquel une ré-

gion d'intérêt (la paume) est extraite. Cette première étape débute par une segmentation de la main basée sur la couleur : le fond étant vert, un seuillage de la composante rouge de l'espace RGB permet de reconnaître les pixels de peau. Elle se poursuit par le calcul du contour de la main binarisée par un codage de *Freeman*. Ce dernier nous permet de déterminer 9 points caractéristiques - les 5 bouts des doigts et les 4 vallées correspondantes - localisés grossièrement grâce aux maxima et aux minima de l'abscisse du contour, puis réajustés par un algorithme itératif inspiré par [4]. De cette localisation, une fenêtre paramétrée est définie : le libre placement de la main fait apparaître des différences en translation, en rotation et en échelle entre les images, mais la position et les dimensions de la fenêtre sont calculées à partir des coordonnées et des distances entre les points de références afin d'adapter l'extraction au placement de la main. Cette première étape se termine par la normalisation de la fenêtre extraite par rotation et sous-échantillonnage. Les images issues de ce prétraitement ont une taille fixe de 64×64 pixels.

La paume présente une texture assez riche en lignes et en stries pour différencier les individus. C'est pourquoi beaucoup d'études sur la reconnaissance de paume la traitent comme une image texturée et lui appliquent des techniques classiques de reconnaissance de forme telles que les filtres directionnels [5] et l'Analyse en Composantes Principales (ACP) [6]. Mais grâce à ses bonnes performances et à ses qualités spécifiques, le filtre de Gabor compte parmi les outils les plus répandus. Nous nous sommes donc servis de ce filtre pour extraire les paramètres de la paume. Ce choix reste en accord avec notre contexte : les calculs réguliers comme la convolution sont facilement implémentables et réduisent la consommation du circuit.

Le filtre de Gabor est le produit d'une sinusoïde complexe et d'une enveloppe gaussienne, il permet de filtrer les motifs ayant une orientation et une fréquence donnée. Il existe une grande variété d'implémentations de ce filtre. Au vu de ses performances, et dans l'optique de réduire le temps de calcul, nous avons choisi un filtre de Gabor ellipsoïdal réel comme celui proposé par [2] :

$$G(x, y) = \exp\left[-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right] \times \cos\left(2\pi \frac{x'}{\lambda}\right) \quad (1)$$

Avec :

$$\begin{cases} x' = (x - \xi) \cos(\Theta) - (y - \nu) \sin(\Theta) \\ y' = (x - \xi) \sin(\Theta) + (y - \nu) \cos(\Theta) \end{cases} \quad (2)$$

Où le couple (ξ, ν) définit le centre de la fonction, Θ contrôle l'orientation, et λ la période de l'onde. Cette dernière est corrélée à la déviation standard de l'enveloppe gaussienne σ par le rapport constant $\sigma/\lambda = 0.56$. γ est le ratio de l'aspect visuel fixé à 0.5.

L'image de paume extraite à l'étape précédente est donc convoluée par un filtre de Gabor dont les coefficients sont choisis empiriquement. Le résultat de cette convolution est ensuite seuillé afin de limiter la taille de la matrice paramètre, ainsi que le temps de calcul de l'étape de comparaison. La définition du

filtre permet de conserver l'information nécessaire avec un seuil de 0.

La mesure de similarité utilisée pour comparer les matrices paramètres entre elles est basée sur la *distance de Hamming normalisée*. Cette dernière est une comparaison pixel à pixel utilisant l'opérateur booléen \oplus (ou exclusif). Nous lui avons ajouté une certaine élasticité : l'extraction de la région d'intérêt n'étant pas forcément idéale, une tolérance en translation (de 5 pixels verticalement et horizontalement) a été introduite. La distance est calculé pour chacun des décalages, et le minimum donne la valeur finale.

2.2 Reconnaissance du visage

L'acquisition des images de visage a été réalisée dans des conditions équivalentes à celle des images de main. Une webcam identique à la première a été utilisée dans des conditions d'éclairage naturelles. Au cours de l'expérience, les sujets devaient, grâce à une prévisualisation du flux vidéo, placer leur visage dans un cadre de 360×480 pixels dessiné à l'écran. La position du visage entre deux prises peut donc légèrement varier en translation, en rotation et en échelle, mais l'étape de prétraitement par localisation et mise à l'échelle n'est cette fois pas nécessaire. L'expression et le port d'accessoires comme les lunettes ne sont, de plus, pas contrôlés.

La reconnaissance du visage est une biométrie mature qui a été beaucoup étudiée. Néanmoins, les méthodes les plus classiques comme les techniques *Eigenface*, ou encore *Fisherface*, ne sont pas adaptées à notre travail car leurs performances s'effondrent dans le cas de la faible connaissance a priori [7]. Dans ce contexte, l'utilisation d'un banc de filtre de Gabor suivie d'une réduction de l'espace des paramètres permet d'améliorer les résultats [3], mais le coût calculatoire reste élevé. Quelques études se sont intéressées à l'utilisation d'un ou de quelques filtres de Gabor [8] et ont obtenu des résultats satisfaisants. Aussi, il semble pertinent d'appliquer la même méthode sur nos deux modalités paume et visage, dans le but de faciliter les implémentations futures.

En pratique, chaque image de visage est sous-échantillonnée pour atteindre la taille fixe de 64×64 pixels. Puis elle est convoluée à un filtre de Gabor paramétré de façon empirique. Pour finir, le résultat de convolution est seuillé, ce qui permet d'utiliser la mesure de similarité définie précédemment.

2.3 Fusion

Ajouter une modalité à un système biométrique, c'est ajouter une nouvelle source d'information. C'est pourquoi les systèmes multimodaux permettent d'obtenir de meilleurs résultats que les systèmes unimodaux correspondants. Ajouter des modalités permet également d'augmenter l'universalité du système. Car si l'utilisateur est réfractaire à une caractéristique biométrique donnée, l'information récoltée sur les autres modalités peut compenser la faible qualité de l'acquisition.

Le processus de reconnaissance biométrique se compose de

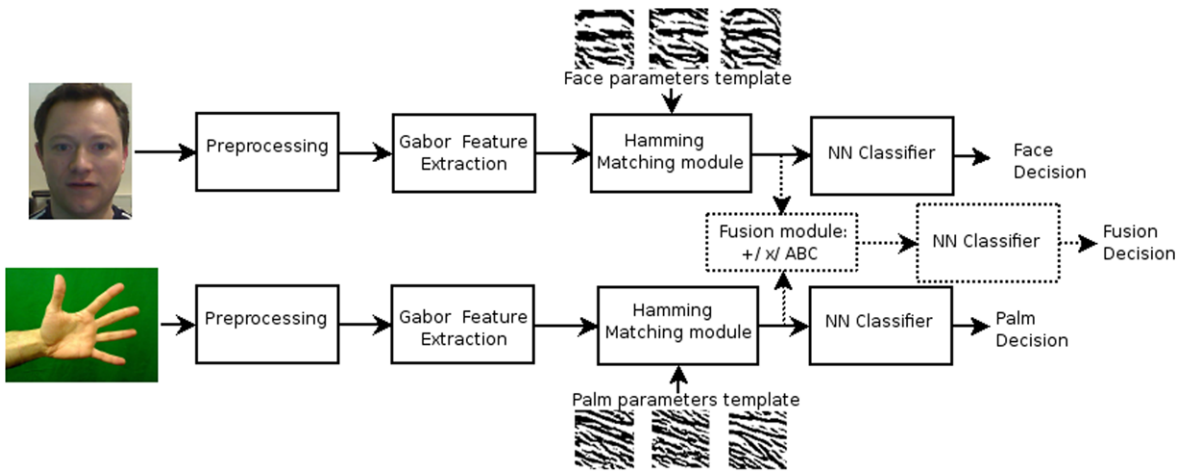


FIG. 1 – Chaîne de traitement complète avec possible fusion des scores.

quatre étapes principales : la capture des données, l'extraction de paramètres, la comparaison à une base d'exemples, et la prise de décision. La fusion biométrique est possible à chacun de ces niveaux : aux deux premiers niveaux, elle consiste souvent en une normalisation des données suivie d'une concaténation [3] ; au niveau de la comparaison, elle se traduit par une combinaison des scores ; enfin, au niveau de la décision, elle consiste en une confrontation des résultats de classification qui aboutit à une décision finale [9].

Les études s'accordent à dire que plus le niveau de fusion est bas (pixels, paramètres, ...), plus les performances sont bonnes. Dans notre cas, ce principe ne s'applique pas aux deux plus bas niveaux de fusion, car le décalage introduit dans la mesure de similarité est calculé de façon optimale sur chaque biométrie séparément. La fusion des pixels obligerait à calculer un décalage commun, ce qui induirait une perte d'information. L'augmentation des performances induite par la multimodalité ne serait donc plus assurée. La fusion des scores semble être le meilleur choix. D'autant plus que la fusion des décisions demande souvent beaucoup de connaissance a priori et de puissance calculatoire, ce qui n'est pas adapté à notre problème.

La chaîne de traitement présentée dans la Figure 1 illustre notre choix de fusion des scores. Il est à noter que cette chaîne est totalement adaptative, puisqu'elle permet d'utiliser la paume seule, le visage seul et aussi leur fusion à l'étape de comparaison. Il peut même être envisagé d'ajouter d'autres modalités texturées comme l'oreille ou l'empreinte des articulations des doigts sans la modifier. Nous avons testé trois types de fusion : l'addition, la multiplication des deux scores et une dernière technique nommée *Averaged Bayes Classifier* (ABC) [9]. L'ABC est également une fusion des scores par somme, mais elle travaille sur des distances normalisées de telle sorte qu'elles puissent être manipulées comme des probabilités.

Dans tous les cas, la décision finale est prise grâce à l'algorithme du plus proche voisin (NN). En pratique, l'image testée est confrontée à une base de comparaison constituée de quelques exemples de chaque individu à reconnaître. Les scores

entre l'image testée et chacun des exemples de la base sont calculés. La décision du classifieur correspond alors à la classe de l'image de la base ayant obtenu le score le plus faible (respectivement le plus fort pour l'ABC).

3 Résultats expérimentaux

La validation de nos algorithmes a été effectuée sur une base de 40 personnes ayant chacune fourni 9 images de mains et 9 images de visages. Les acquisitions ont été réalisées en 3 sessions réparties sur un mois. Comme nous nous plaçons dans un contexte de faible connaissance a priori, la base de comparaison ne contient qu'un, deux ou trois exemples par personne et par modalité. Toutes les autres images sont utilisées pour tester l'algorithme. Les résultats sont qualifiés grâce au taux de bonne reconnaissance, qui correspond au rapport du nombre de bonnes classifications sur le nombre total de tests. Nous avons de plus effectué plusieurs expériences en changeant aléatoirement les exemples utilisés dans la base de comparaison, puis moyenné les résultats. L'indicateur obtenu est appelé ARR (pour Average Recognition Rate) dans la suite du document.

Comme nous l'avons vu en Section 2.1 et 2.2 les coefficients de chacun des deux filtres ont été choisis empiriquement. Pour cela, nous nous sommes placés dans le cas où la base de comparaison ne contient qu'un exemple par personne, et nous avons testé 24 filtres (Θ a été choisi dans $\pi/8 \times \{0; 1; 2; 3; 4; 5; 6; 7\}$, et λ dans $\{5, 47; 8, 20; 10, 93\}$). Nous avons observé des résultats très similaires sur la gamme de filtres testés, mais nous avons choisi les paramètres permettant d'obtenir les meilleurs résultats sur chacune des modalités séparément : pour le visage, comme pour la paume, les meilleures performances ont été obtenues pour $\lambda = 10.93$ et $\Theta = 3\pi/8$. Ces valeurs ont donc été choisies pour effectuer d'autres tests, dans le cas d'une base contenant 2 et 3 exemples par personne et aussi dans un contexte de vérification. La taille des filtres utilisés est fixée à 9×9 pixels.

TAB. 1 – Comparaison des taux de reconnaissance moyennés dans le cas d'un, deux et trois exemples par personne.

Methode	ARR (%)	ARR (%)	ARR (%)
	1 exemple	2 exemples	3 exemples
Visage	78,13	86,43	87,78
Paume	93,33	97,86	97,92
Fusion par Somme	94,38	97,98	98,75
Fusion par Multiplication	95,52	98,1	98,75
Fusion par ABC	95,63	98,76	99,7

Le tableau 1 présente l'ensemble des résultats de reconnaissance. Il confirme que la paume est une modalité vraiment performante, même utilisée seule, puisqu'avec seulement 2 exemples par personne on obtient 97.9% de bonne reconnaissance. Le visage n'atteint pas d'aussi bons résultats, mais ces derniers sont déjà élevés au vu du très faible coût calculatoire des algorithmes de traitement. Quand à la fusion, elle permet de toujours surpasser les performances de la paume. La multiplication et la méthode ABC apportent une amélioration des performances encore supérieure à celle apportée par la somme.

D'autres tests confirment ces résultats : dans une optique d'identification de l'individu et non plus de reconnaissance, nous avons calculé l'EER (pour Equal Error Rate) de notre système. Pour cela nous avons conservé tous les scores obtenus lors de la comparaison des images de test avec la base de comparaison. Nous avons ensuite seuillé chacun de ces scores afin de donner une décision binaire d'identification : les deux personnes testées sont identiques ou bien différentes. Ce nouveau système a été qualifié par deux indicateurs. Premièrement, le taux de faux positifs (FAR pour False Acceptance Rate) a été calculé par le ratio du nombre de couple d'échantillons différents considérés par le système comme identiques et le nombre total de couples d'échantillons différents testés. Ensuite, le taux de faux négatifs (FRR pour False Rejection Rate) a été calculé par le ratio du nombre de couple d'échantillons identiques considérés par le système comme différents et le nombre total de couples d'échantillons identiques testés.

Chacun de ces taux a été calculé dans le cas d'une base de comparaison contenant 1 exemple par personne (comme pour l'AAR, 9 bases différentes ont été testées et les résultats ont été moyennés). En faisant varier le seuil, différentes valeurs du FRR et du FAR ont été obtenues. L'EER est le taux pour lequel le FAR est égal au FRR. Les EER calculés confirment donc l'apport de la multimodalité, puisque l'EER obtenu grâce à la paume seule est de 5,2% tandis que celui de la fusion des scores par multiplication est de 3,4%.

4 Conclusion et perspectives

Cette étude nous permet de valider la faisabilité d'un système biométrique multimodal embarqué grand public. Les traitements développés respectent les contraintes de faible coût calculatoire puisque les paumes et les visages sont traités sous Matlab en seulement 120 ms dans le cas de deux exemples par personne.

Une grande base de données (environ 140 sujets) est en cours de construction, afin d'affiner les résultats. Notre prochaine étape de travail consiste à implémenter ces traitements sur des SoC (System On Chip). Nous voudrions réaliser la chaîne de reconnaissance en parallélisant les opérations sur des FPGA contenant des coeurs de processeur.

Références

- [1] A. Ross and A. K. Jain, "Multimodal biometrics : An overview," in *Proceedings of 12th European Signal Processing Conference (EUSIPCO)*, (Vienna, Austria), pp. 1221–1224, 2004
- [2] J. Doublet, O. Lepetit, and M. Revenu, "Contact less hand recognition using shape and texture features," in *Proceedings of 8th International Conference on Signal Processing*, vol. 3, (China), 2006.
- [3] X.-Y. Jing, Y.-F. Yao, D. Zhang, J.-Y. Yang, and M. Li, "Face and palmprint pixel level fusion and kernel dcw-rbf classifier for small sample biometric recognition," *Pattern Recognition*, vol. 40, no. 11, pp. 3209–3224, 2007.
- [4] X. Jiang, W. Xu, L. Sweeney, Y. Li, R. Gross, and D. Yurovsky, "New directions in contact free hand recognition," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, (San Antonio), pp. 389–392, 2007.
- [5] A. Kumar and D. Zhang, "Integrating Palmprint with Face for User Authentication" in *Proc. Multi Modal User Authentication Workshop*, (Santa Barbara, CA, USA), pp. 107–112, 2003.
- [6] H. Dutagaci, B. Sankur, and E. Yorum, "Comparative analysis of global hand appearance-based person recognition," *Journal of electronic imaging*, vol. 17, no. 1, 2008.
- [7] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person : A survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.
- [8] A. Noore, R. Singh, and M. Vatsa, "Robust memory-efficient data level information fusion of multi-modal biometric images," *Information Fusion*, vol. 8, no. 4, pp. 337–346, 2007.
- [9] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.