

Reconnaissance faciale basée sur les ondelettes robuste et optimisée pour les systèmes embarqués

Pascal COTRET¹, Stéphane CHEVOBBE², Mehdi DAROUICH²

¹CentraleSupélec/IETR - Équipe SCEE (Signal, Communication et Électronique Embarquée)
Avenue de la Boulaie CS 47601, 35576 Cesson Sévigné Cedex, France

²CEA, LIST, L3A (Laboratoire Adéquation Algorithme Architecture)
91191 Gif-sur-Yvette, France

`pascal.cotret@centralesupelec.fr, stephane.chevobbe@cea.fr, mehdi.darouich@cea.fr`

Résumé – Dans le domaine du traitement d’images, la reconnaissance faciale est une technique appliquée dans de nombreuses applications : télésurveillance, accès à des zones restreintes, déverrouillage de systèmes électroniques, etc. Dans ce contexte, cette contribution propose une méthode rapide de reconnaissance faciale basée sur la transformée en ondelettes robuste aux variations de position et de luminosité pour des applications temps réel. La méthode proposée a une tolérance de +/- 10% aux variations de position avec des conditions de luminosité variables. Sur une plateforme embarquée type RaspberryPi, le temps de reconnaissance moyen est de 26 ms par visage avec une empreinte mémoire 64 fois plus faible que l’approche de référence et des taux de reconnaissance équivalents.

Abstract – In the image processing field, face recognition is a technique applied in several domain: CCTV, restricted areas access, electronic devices delocking and so on. In this context, this contribution proposes a wavelet-based face recognition method robust against face position and light variations for real-time applications. Method proposed here has a +/- 10% tolerance in terms of spatial variations with several lightning conditions. On an embedded platform such as a RaspberryPi, mean recognition time for a single face is 26 ms with a memory footprint 64 times lower than the reference method and with equivalent face recognition rates.

1 Introduction

Un des enjeux majeurs dans le développement de systèmes embarqués est de maintenir un compromis entre la puissance de calcul disponible et les ressources système utilisées (consommation, surface de silicium, quantité de mémoire...). Dans le domaine du traitement d’images, et plus spécifiquement dans celui de la reconnaissance faciale utilisant des méthodes basées sur les eigenfaces[7], la puissance de calcul est directement proportionnelle à la résolution des images d’entrée. Par conséquent, en limitant la résolution des images en entrée de tels algorithmes, le temps de traitement peut diminuer de manière significative. En utilisant les coefficients d’approximation d’une transformée en ondelettes, la quantité de données à traiter est divisée d’un facteur 2^{2K} (K étant le niveau de décomposition de la transformée). La tolérance de cette méthode avec les variations de lumières peut être améliorée avec une égalisation d’histogramme par régions [1].

2 État de l’art

Les méthodes classiques de reconnaissance faciale sont sensibles aux variations de luminosité et de qualité de l’image (position du sujet dans le cadre, image bruitée...). Dans les travaux existants, plusieurs contributions proposent de surmon-

ter l’alignement du sujet en repositionnant la zone de cadrage (ROI, *Region of Interest*) [8] ou en utilisant des filtres de corrélation [5]. Cependant, dans un contexte embarqué, la puissance limitée nous empêchent d’utiliser ces méthodes complexes.

Shan et al. [6] proposent une méthode insensible aux variations de luminosité et d’expression implémentée sur un DSP et un processeur NIOS II. Malgré plusieurs optimisations, la reconnaissance d’une personne parmi 16 est réalisée en 1 seconde, ce qui est inacceptable dans un contexte embarqué. Endluri et al. [2] proposent une implémentation d’une méthode PCA (*Principal Component Analysis*) sur un TSK3000a (processeur RISC 32-bits embarqué) et qui réalise la reconnaissance sur une base de 2 sujets avec des images de résolution 320x240 pixels. Ces études, qui utilisent une classification par PCA, nécessitent des temps de calcul trop élevés pour pouvoir être implémentée sur des architectures embarquées.

Malgré tout, aucune contribution récente n’a étudiée la robustesse de ces méthodes envers les variations de l’image ; de plus, les données relatives aux performances réelles et à la taille de la ROI sont rarement communiquées. Pavan et al. [3] utilisent une méthode PCA modulaire pondérée qui améliore la robustesse face aux variations du visage. Le système proposé, basé sur un processeur NIOS couplé à un coprocesseur PCA matériel, fonctionne à 26 fps sur un FPGA Stratix II avec une base maximum de 5 sujets (mémoire interne limitée).

3 Algorithme de reconnaissance faciale basé sur les ondelettes

La méthode utilisée dans cette contribution est basée sur les eigenfaces [7]. Le travail de Turk et al. décrit un procédé complet pour la reconnaissance faciale fondé sur des calcul en deux dimensions là où les méthodes complexes peuvent être tridimensionnelles et aller jusqu'à la reconstruction complète du visage. En utilisant les eigenfaces, la méthode est divisible en deux étapes : un apprentissage et la reconnaissance.

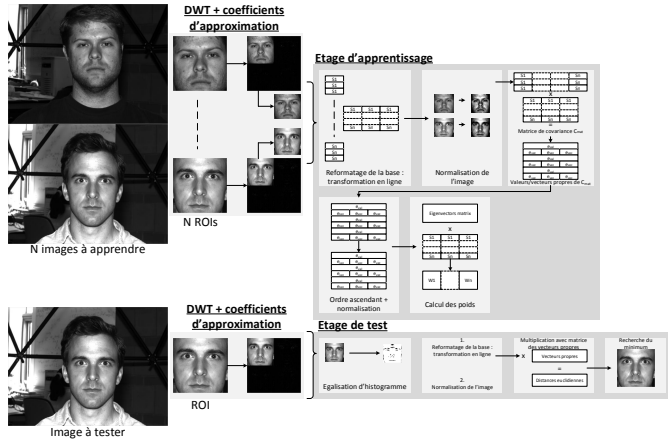


FIGURE 1 – Flot de reconnaissance faciale simple basé sur les eigenfaces

La figure 1 présente la méthode de reconnaissance faciale complète utilisant la transformée en ondelettes. La transformée en ondelettes est réalisée juste avant l'étape d'apprentissage et on ne conserve que les coefficients d'approximation [1] (on divise la quantité de donnée injectée d'un facteur 2^{2K} , K étant le niveau de décomposition en ondelettes). La phase d'apprentissage est identique aux travaux de Turk et al. [7]. Dans cette approche, on considère que les visages sont bien centrés sur les clichés (le fond n'est pas obligatoirement uniforme). En supposant que l'on dispose d'une base de N ROIs (*Region of Interest*, le visage cadré) de taille $M * M$, l'apprentissage est réalisée en 5 étapes :

- **Étape 1.** Une transformation en ligne est réalisée sur l'ensemble de la phase de test. Pour une unique ROI, cela consiste en la concaténation de ses lignes. On obtient une matrice D_{inline} de taille $M^2 * N$.
- **Étape 2.** Pour améliorer la qualité de l'image, on effectue une normalisation (avec la moyenne globale et l'écart-type). La dimension du résultat $D_{inline, norm}$ n'est pas affectée.
- **Étape 3.** Premièrement, on réalise une multiplication entre $D_{inline, norm}$ et sa transposée pour obtenir une matrice de covariance C_{mat} de taille $N * N$. Ensuite, on calcule les valeurs propres e_{val} et les vecteurs propres e_{vec} de cette matrice de covariance.
- **Étape 4.** Cette étape consiste en un tri croissant des valeurs propres (et des vecteurs propres associés).

- **Étape 5.** La dernière étape donne les poids de chaque visage dans la base d'apprentissage ; Cette opération est réalisée par une multiplication entre la matrice des vecteurs propres et la base en ligne (étape 1). Le résultat, W_{mat} , est une matrice de taille $N * N$.

Les différentes données récupérées sont traitées dans l'étape de reconnaissance, cet étage est réalisé en trois étapes :

- **Étape 1.** Une égalisation d'histogramme par région qui permet de mettre en avant les détails et les bords du visage.
- **Étape 2.** Une transformation en ligne et une normalisation comme lors de l'apprentissage.
- **Étape 3.** Un produit entre ce résultat et la matrice des eigenfaces. En comparant avec W_{mat} , on obtient N distances euclidiennes $E_{d,n}$: l'indice de la plus faible distance euclidienne donne l'identifiant du visage détecté.

4 Expérimentations et résultats

4.1 Robustesse

4.1.1 Décalage en $[x,y]$

Par exemple, la figure 2 montre les diagrammes des scores de deux visages. Les coordonnées représentent la position du centre de la ROI ($[0, 0]$ quand elle est centrée sur le visage de l'utilisateur).

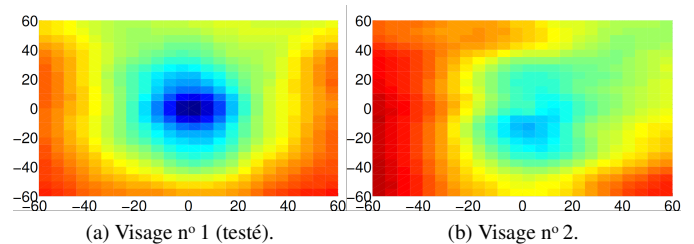


FIGURE 2 – Diagramme des scores de 2 visages.

En appliquant la méthode décrite précédemment sur l'ensemble de la base d'apprentissage, le score minimum est obtenu quand la ROI est centrée sur le visage du sujet (c'est également le cas sur les autres projections). La figure 3 présente des diagrammes de scores cumulés pour l'utilisateur n° 1.

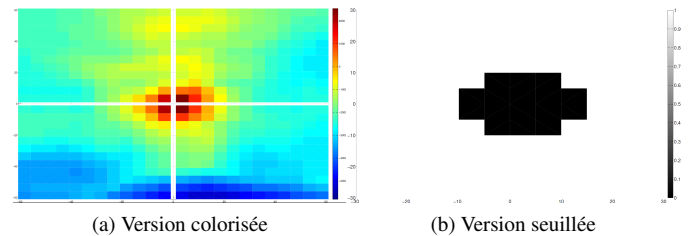


FIGURE 3 – Diagrammes cumulés des minimum pour le sujet #1.

Sur les deux diagrammes, le pas de décalage est fixé à 5 pixels. D'après la figure 3b, on observe une tolérance de +/- 10% en

position avec des taux de reconnaissances supérieurs à 75%.

4.1.2 Décalage en $[z]$

La figure 4 montre la probabilité d’obtenir une reconnaissance correcte en fonction du niveau de décomposition et du facteur d’échelle (plus ou moins éloigné du capteur par rapport à la position utilisée à l’apprentissage).

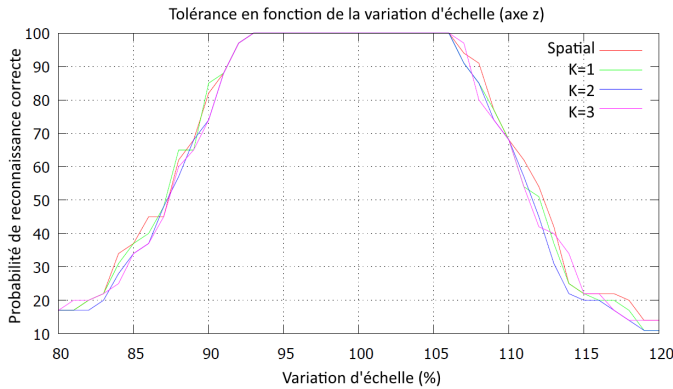


FIGURE 4 – Probabilité d’une reconnaissance correcte (visage reconnu) en fonction de la variation d’échelle sur un visage centré.

On remarque que le niveau de décomposition influe peu sur les résultats de reconnaissance, l’algorithme supporte une variation d’échelle de +/- 10% (taux de reconnaissances supérieurs à 75%). Cette tolérance passe à +/- 5% pour un taux de reconnaissance maximal.

4.2 Performances

4.3 Cadre des expérimentations

Les mesures ont été effectuées sur trois architectures différentes :

- *arch1* : processeur x86 Intel Core i7 à 3,4 Ghz et 8 Mo de cache.
- *arch2* : un SoC iMX6 de SECO¹ avec un quad-coeur Cortex-A9 à 1 Ghz et 32 Ko de cache.
- *arch3* : un RaspberryPi B avec un coeur simple Broadcom à 700 Mhz et un cache de 32 Ko.

L’architecture *arch1* donne une référence pour le pipeline de reconnaissance et pour le temps d’exécution. Les deux architectures ARM donnent des résultats sur des plateformes plus embarquées. Les programmes ont été compilés avec des options standard (-o3) pour être exécutés sur un seul coeur dans chaque configuration.

Tous les profilings de code ont été effectués sur la même séquence vidéo et le même ensemble de visages. La séquence vidéo est composée de 1000 frames de résolution 640x480 avec deux visages dont les positions en x,y,z varient. La séquence

1. <http://www.seco.com/prods/eu/boards/qseven-boards/quadmo747-x-1-mx6.html>

vidéo a été capturée avec une caméra IDS UI-3240CP². La taille de la ROI est de 200x200 pixels. Les tests ont été effectués sur des sous-ensembles de 4 à 40 visages et avec un niveau de décomposition en ondelettes $K \in [0; 5]$. Il est important de préciser que la phase d’apprentissage n’est pas incluse dans les profilings étant donné qu’elle peut-être réalisée hors-ligne.

4.4 Analyse des profilings

Comme on pouvait s’y attendre, le temps d’exécution sur l’architecture *arch1* (0,6 ms à 5,09 ms) est beaucoup plus rapide que ceux sur les deux architectures ARM (de 8 à 157 ms sur *arch2* et de 26 à 380 ms sur *arch3*). De nombreux aspects matériels peuvent expliquer cette accélération : l’organisation du pipeline, l’unité de calcul flottant ou encore la vectorisation. La gestion de la mémoire est un élément clé, les différences dans les temps d’exécution sont principalement dues à ces aspects.

La figure 5 montre l’évolution des temps d’exécution sur les trois plateformes en fonction de K et du nombre de visages dans la base de données.

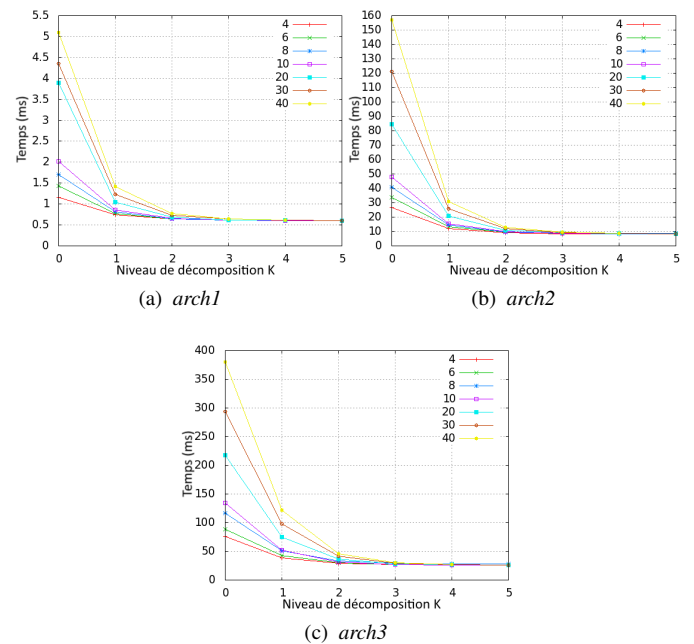


FIGURE 5 – Temps d’exécution par frame en ms pour le pipeline de reconnaissance en fonction de K et de la taille de la base d’apprentissage.

Les trois architectures ont un comportement sensiblement identique en fonction de K . A K constant, le gain peut atteindre un facteur 8 pour l’architecture *arch1*. Le temps d’exécution diminue de manière drastique lorsque le niveau de décomposition augmente. Lorsque $K \geq 3$, les temps d’exécution tendent vers une asymptote et sont quasiment indépendants de la taille de la base d’apprentissage.

Les figures 6 présentent les temps cumulés des trois fonc-

2. <https://en.ids-imaging.com/store/ui-3240cp.html>

tions principales de l'étage de reconnaissance : égalisation d'histogramme (*hist*), décomposition en ondelettes (*DWT*) et classification par eigenfaces (*eigen*). Le temps de classification diminue fortement lorsque K augmente alors que le temps d'exécution de l'égalisation d'histogramme est plus ou moins stable. On remarque également que plus la base d'apprentissage est importante, plus l'accélération en fonction de K est importante.

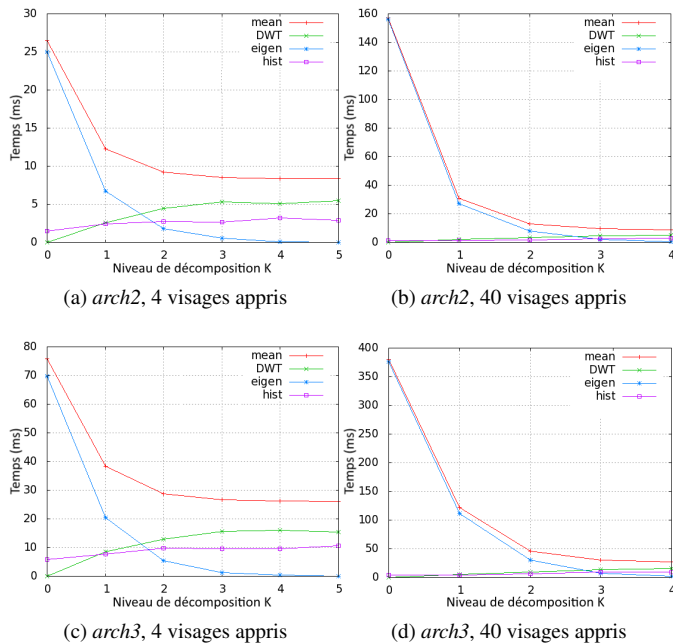


FIGURE 6 – Temps d'exécution par frame pour les principales fonctions de l'étage de reconnaissance.

Les figures 7 présentent les pourcentages des temps d'exécution sur un RaspberryPi (architecture *arch3*) des fonctions principales en fonction de K et de la taille de la base d'apprentissage.

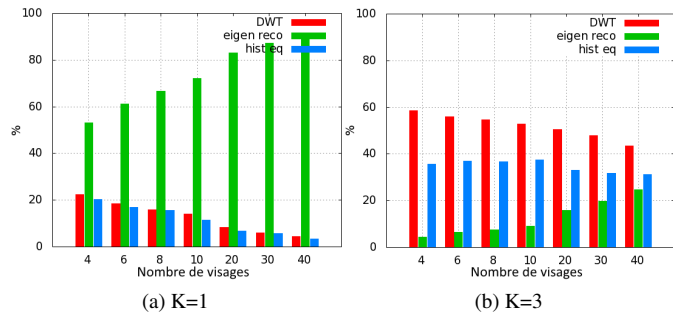


FIGURE 7 – Pourcentages d'exécution par frame pour les fonctions principales de l'étage de reconnaissance sur un RaspberryPi (modèle B).

Pour $K = 1$ et $K = 3$, le temps imputé à la classification augmente par eigenfaces avec la taille de la base d'apprentissage. Pour $K = 1$, cette étape prend 50% pour 4 visages alors qu'elle en prend 90% pour 40 visages. Pour $K = 3$, cette contribution va de 5 à 25% (la décomposition en ondelettes devient alors la

fonction prépondérante).

5 Conclusion

Cette contribution va plus loin que les travaux existants en étudiant la robustesse et en caractérisant les performances sur des plateformes embarquées existantes. Les résultats présentés nous permettent d'envisager une application avec suivi vidéo du visage fluide sur une plateforme avec peu de puissance de calcul et très peu de mémoire embarquée. Dans de futurs travaux, d'autres pistes telle que la caractérisation des visages de la base d'apprentissage ou l'étude poussée du suivi sont envisagées.

Références

- [1] S. Courroux, S. Chevobbe, M. Darouich, and M. Paindavoine. Use of wavelet for image processing in smart cameras with low hardware resources. *Journal of Systems Architecture - Embedded Systems Design*, 59(10-A) :826–832, 2013.
- [2] R. Endluri, M. Kathait, and K. Ray. Face recognition using pca on fpga based embedded platform. In *Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference on*, pages 1–4, Dec 2013.
- [3] A. P. Kumar, V. Kamakoti, and S. Das. System-on-programmable-chip implementation for on-line face recognition. *Pattern Recognition Letters*, 28(3) :342–349, 2007. Advances in Visual information Processing Special Issue of Pattern Recognition Letters on Advances in Visual Information Processing. (ICVGIP 2004).
- [4] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5) :684–698, May 2005.
- [5] M. Sawides, B. Kumar, and P. Khosla. "corefaces" - robust shift invariant pca based correlation filter for illumination tolerant face recognition. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–834–II–841 Vol.2, June 2004.
- [6] T. Shan, A. Bigdeli, B. C. Lovell, and S. Chen. Robust face recognition technique for a real-time embedded face recognition system. In B. Verma and M. Blumenstein, editors, *Pattern Recognition Technologies and Applications : Recent Advances*, pages 188–211. IGI Global, Hershey, PA, USA, 2008.
- [7] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1) :71–86, Jan. 1991.
- [8] S. Yan, H. Wang, J. Liu, X. Tang, and T. Huang. Misalignment-robust face recognition. *Image Processing, IEEE Transactions on*, 19(4) :1087–1096, April 2010.