Une nouvelle transformation pour la localisation des yeux dans une image de visage monochrome

M. MILGRAM L. PREVOST R. BELAROUSSI

[1] Université Pierre & Marie Curie, Laboratoire des Instruments & Systèmes d'Ile de France Groupe PARC

4 Place Jussieu, 75252 Paris Cedex 5, BC252

Résumé – Nous présentons une méthode de localisation des yeux dans des images de visages et les résultats sur environ 3000 visages de la base ECU. Cette méthode fonctionne sur des images monochromes à l'aide de la Transformation Chinoise (TC). Celle ci est un outil original capable de détecter certaines formes par accumulations de votes. Après la TC, un modèle géométrique Gaussien est utilisé ainsi qu'une technique de projection des niveaux de gris afin de raffiner la localisation. Une notion de disimilarité entre les images des 2 yeux fournit un critère de validation. De nombreux résultats sont présentés et commentés.

Abstract *– A method for the localization of the eyes in a facial image and the results on nearly 3000 faces from the ECU database are presented. This method works on grayscale images, applying the Chinese Transformation (CT) on edge pixels. This transformation is an original tool able to detect patterns by accumulating votes. After the CT , we apply a Gaussian model to find eyes location and we refine the model with a grayscale profile approach. The dissimilarity between the 2 eyes is then introduced as a validation criterion. Experimental results on a large database is presented and commented.*

1. Introduction

Nous présentons dans cet article un nouvel algorithme de traitement d'images permettant d'estimer la position des yeux dans une image de visage. Cette image peut-être indifféremment monochrome ou couleur car nous n'utilisons pas l'information colorimétrique comme attribut pertinent ; l'appartenance ethnique ne modifie donc pas les performances du système. Les artefacts comme le port de lunette ne sont pas un obstacle majeur pour notre méthode. Cette approche est basée sur une technique inspirée de la Transformée de Hough (la « Transformée Chinoise ») qui utilise la disposition des pixels de contour ainsi que l'orientation du gradient en ces pixels. Elle est associée à des modèles probabilistes et à deux experts de raffinement et de validation. Les résultats expérimentaux, estimés sur une base de 2500 images de visages, sont très encourageants avec un taux de détection supérieur à 98% et une erreur de localisation relative supérieure à 92%.

La detection/localisation de visages est devenue un enjeu très important pour des applications comme la vidéo-conférence, la biométrie, la détection de présence, l'indexation d'images, etc. Cette détection est reliée à plusieurs sous-problèmes :

- détection d'éléments caractéristiques (bouche, yeux)
- détection et suivi du regard
- estimation de la pose
- analyse de l'expression faciale

Nous présentons ici une solution au sous-problème de la détection/ localisation des yeux. Pour cela, on trouve dans la littérature deux approches génériques, parfois combinées :

- l'approche par décisions binaires qui analyse un ensemble de sous-images et prend une décision pour chacune d'elles (exemple : *pattern matching*)
- l'approche globale qui traite directement le visage en entier (exemple : techniques de votes dans le style Transformée de Hough)

Dans les 2 cas, un codage préalable est parfois utilisé comme dans Huang et Weshler [10] où ce sont les ondelettes associées à un réseau RBF. Daugman [8] utilise aussi les ondelettes de Gabor pour localiser l'iris. D'autres comme Braathen, Bartlett et Littleworth-Ford [7] s'attaquent aux expressions faciales avec ce même type de codage. Les points caractéristiques (contours, coins) sont combinés à des traitements de morphologie mathématique par Gu, Su et Du [9] dans leur approche complètement ascendante du problème. La dimension fractale a été mise à contribution par Lin, Lam et Siu [11] qui l'emploie, combinée avec des contraintes géométriques. D'autres chercheurs proposent de localiser la partie blanche de l'œil, la sclera, et non l'iris à l'aide d'un modèle multi-gaussien [6].

Notre contribution comprendra une description des pré traitements suivie par la présentation de la TC. Les résultats de cette TC seront ensuite renforcés par notre modèle probabiliste et raffinés par une technique de profil. Enfin, nous présenterons et analyserons nos résultats sur la base ECU.

2. Pré-traitement et normalisation

L'étude a été mené sur la base ECU [5] qui comporte 2997 images RGB de visages. Parmi ces images, seules 2640 contiennent un visage où les deux yeux sont visibles. L'étiquetage de ces images (position des yeux) a été réalisé manuellement. Nous avons supposé que les visages étaient cadrés de manière relativement uniforme tout en nous efforçant d'accepter certaines variations dans le cadrage. Les visages ayant une orientation (angle entre l'horizontal et la droite joignant les yeux) supérieure à 30° ont été écartés automatiquement. Finalement, 2450 images ont passé cette étape de filtrage. C'est sur ces données qu'ont été estimés les moyennes empiriques mx_1, mx_2, my_1 et my_2 des abscisses (x_1 et x_2)et ordonnées (y_1 et y_2) des deux yeux et la matrice de variance-covariance géométrique Σ (supposée diagonale).

Les images RGB sont ramenées en niveau de gris par simple moyenne des 3 plans couleurs, sans prise en compte de connaissance colorimétrique *a priori* comme la teinte chair [11]. On détermine ensuite l'image *ORIENX* des orientations de gradient en évaluant les deux composantes *Gx* et *Gy* du gradient par un opérateur de Roberts, et sa direction $\alpha=arctan(Gy/Gx)$ discrétisée sur N valeurs (N=8). Le module du gradient permet de sélectionner les points de contour par seuillage absolu (le seuil a été déterminé empiriquement et validé *a posteriori* en fonction des résultats du système complet de localisation).

3. Transformée chinoise

La Transformation Chinoise (TC) est une transformation globale qui travaille sur les orientations de gradient des points de contour d'une image et fournit en sortie deux tableaux de vote. Le principal tableau de vote (*VOTE*), dans notre application, donne une information sur la position du centre des yeux. L'idée de base est que les yeux occupent dans le visage, en incluant les sourcils, une zone qui est plus sombre que son environnement immédiat et possède une forme elliptique. Du fait de la quasi symétrie centrale de l'ensemble, le centre de l'œil se trouve au milieu de nombreux segments joignant 2 points ayant des orientations de gradient opposés (figure 1). La TC peut rappeler la THG (qui utilise le gradient) mais elle en diffère fortement car elle peut détecter à la fois des ellipses de toute taille et excentricité. Elle repose sur les propriétés de symétrie centrale de l'ellipse.

Le principe général de la TC consiste d'abord à déterminer, pour chaque paire de pixels de contour *(M',M")* d'orientations respectives *(O',O")* (*O =0,1,..,N-1)* s'ils font ou non partie de l'ensemble des pixels votants. Le critère d'appartenance est basé sur la position relative de *M'* et *M"* et sur la condition *O'*=2 et *O"*=6 (soit Nord et Sud). Les points *M(x',y')* et *M"(x",y")* doivent vérifier les contraintes géométriques suivantes :

(1) $y'-y'' < \delta_{VERT}$

(2) $45° < \beta < 135°$ avec $\beta = arctan[(y'-y'') / (x'-x'')]$

(3) $| M' M'' | < \delta_{DIST}$

Des ajustements ont conduit à prendre δ_{VERT}=0.08xHauteur_image et δ_{DIST} =0.25xHauteur_image.

Deux tableaux de vote *VOTE* et *DIAM* (figure 2), de même taille que l'image *ORIENX* des orientations de gradient sont crées. Pour chaque paire *(M',M")* de l'ensemble des votants, on va incrémenter la valeur de *VOTE (M)* où *M* est le milieu du segment *[M'M"]*. (d'où le nom de Transformée Chinoise, en référence à la Chine, « Empire du Milieu »). Pour conserver une information sur la distance entre deux points votants pour un point *M*, on stocke dans *DIAM(M)* (comme diamètre) la moyenne des distances *d(M',M")*. Cette information permettra de déterminer un rectangle englobant chaque œil.

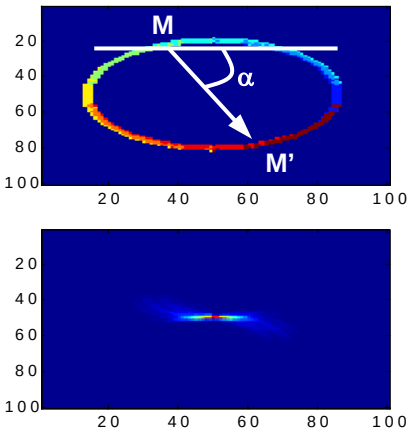

FIG. 1 : Principe de la Transformée Chinoise appliquée à une ellipse et tableau VOTE.

Une analyse *a posteriori* montre que la position des yeux correspond de manière assez fiable aux pixels ayant reçus le plus de votes. Malheureusement, les 2 premiers maxima (maximum absolu et second maximum) ne sont pas toujours les bons.

4. Modèle probabiliste

Pour renforcer le précédent détecteur, nous avons mis en œuvre un modèle probabiliste sur les positions des yeux dans le visage. Nous devons intégrer 2 types d'informations :

- les résultat du tableau de *VOTE* (information radiométrique)
- les positions relatives des paires candidates dans le visage (information géométrique)

Afin de normaliser les données issues du tableau de *VOTE* et pour ne pas être trop dépendant des variations inter-images, nous avons décidé d'utiliser pour quantifier la première information une loi géométrique (inspirée des « *neural gas* » [3]). La log-vraisemblance du $n^{ième}$ candidat est donc donnée par :

$LR(n) = \log(Prob(n)) = - rang(n)$

où *rang(n)* désigne la position du n^{ième} candidat dans la liste des candidats ordonnées par score (résultat du tableau *VOTE*) décroissant. *LPR(n)* ne prend donc en compte que le score lié à la radiométrie.

Le second terme *LA* (*A* pour abscisse) prend en compte la distribution statistique conjointe des abscisses des deux yeux (x_1, x_2) (normalisées entre 0 et 1 en divisant par la largeur de l'image) :

$$LA(x_1, x_2) = \tfrac{1}{2}\ X^T \Sigma^{-1} X \quad \text{où } X=[x_1-mx_1\ x_2-mx_2]^T$$

Finalement, nous supposons classiquement l'indépendance des 2 types d'évènements pour mesurer la log-vraisemblance du couple (x_1, x_2):

$$LC(n, m) = LA(x1(n,m) , x2(n,m)) + LR(n) + LR(m)$$

où $x_1(n,m)$ désigne l'abscisse du candidat n ou m selon leur position et $x_2(n,m)$ l'abscisse du candidat restant.

Dans un premier temps, on sélectionne le couple (n^*,m^*) qui maximise la log-vraisemblance $LC(n^*,m^*)$. Si les 2 positions ainsi déterminées correspondent effectivement à 2 yeux, la vraisemblance $LC(n^*,m^*)$ doit être élevée (en fait supérieure à un seuil δ_{LPC} déterminé empiriquement à la valeur –100). Sinon , on va rejeter ce couple et :

- chercher la détection la plus probable (et l'œil correspondant) selon un modèle gaussien pour chaque œil
- prédire la position du second œil en utilisant un prédicteur linéaire d'ordre 1 :

$x_2 = mx_2 + (x_1 - mx_1).C_{12}$ et $y_2 = y_1$ (C_{12} coefficient de corrélation empirique centré entre x_1 et x_2)

Si nous n'avons qu'un seul candidat (maximum local dans le tableau de VOTE), ce qui précède nous permet aussi de prédire la position du second œil.

5. Amélioration de la précision

A ce stade, l'erreur selon l'axe vertical est nettement dominante car la TC a tendance à générer des confusions avec le sourcil dont l'apparence est parfois très proche de celle de l'œil. Nous avons donc développé un expert de raffinement et un expert de validation.

L'expert de raffinement permet d'augmenter de manière substantielle la précision de la détection selon l'axe vertical en analysant les profils verticaux des deux yeux (valeur moyenne des niveaux de gris pour chaque ligne de la vignette contenant l'œil). La position du centre de l'œil correspond à un minimum local du profil vertical, mais plusieurs minima locaux sont parfois détectés dus aux sourcils ou à l'ombre présente sous l'œil (figure 2). Seuls les deux premiers minima locaux sont conservés, le second pouvant être utilisé en cas de rejet du premier par le critère de dissimilarité présenté dans la suite.

L'expert de validation permet de filtrer une partie importante des fausses détections en comparant l'image de l'œil droit avec l'image miroir de l'œil gauche obtenue par symétrie verticale. On mesure la dissimilarité des candidats en évaluant la distance euclidienne entre les deux vignettes normalisées. Si cette dissimilarité est trop faible, on considère la détection comme peu fiable et on choisit de la rejeter. Ceci provoque une augmentation du taux de rejet et une diminution de l'erreur de localisation. On notera que pour les visages inclinés ou ceux dont les pupilles ne sont pas centrées dans les yeux, cette dissimilarité peut être mise en défaut. Ceci n'est pas trop gênant dans la mesure où ce test ne nous sert qu'à rejeter des détections douteuses.

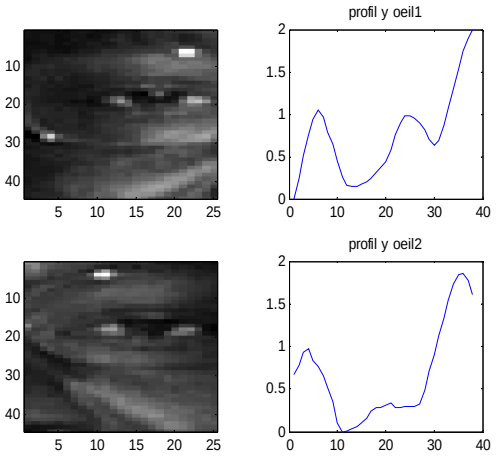

FIG 2 : profils verticaux des deux yeux et résultats de localisation.

6. Résultats expérimentaux

Nos mesures de performance tiennent compte du fait que nous demandons au système de nous fournir deux détections exactement. Dans les cas ou aucun maximum local n'est trouvé dans le tableau de *VOTE* (soit environ 2% des images de la base ECU), nous avons deux détections manquées. Dans les autres cas, le critère d'évaluation est la distance entre la détection (pour chaque œil) et la vérité terrain (normalisée, c'est à dire divisée par la distance entre les deux yeux).

L'utilisation du rejet avec un seuil de 0.1 sur les 2450 visages testés donne moins de 13% de rejet et une diminution significative de l'erreur de localisation, qui passe de 8% à 7.2% pour l'œil 1 et de 7.7% à 7.0% pour l'œil 2. Cette diminution peut sembler faible mais elle provient de l'élimination d'erreurs importantes (*outliers*). Ceci apparaît quand on observe les médianes au lieu des moyennes (tableau 1). Le gain est alors faible (de 6.0% à 5.7% pour l'œil 1 par exemple) ce qui est du au fait que l'on supprime surtout de grandes valeurs ayant une faible incidence sur cette médiane.

On a étudié en détail les cas de mauvaises détections sur les 300 premières images (cas où on obtient, pour au moins un des deux yeux, une erreur relative de position supérieure à 10%). Ceci produit 70 images, ce qui est cohérent avec la performance trouvée de 11% qui concerne la moyenne des distances relatives, d'où la proportion double quand on sélectionne les cas où une seule détection est distante. Nous avons classé ces erreurs par taux d'occurrence décroissant : lunettes (14%), yeux fermés (10% : confusion avec les sourcils), peau très sombre (7%), résolution trop faible (7% : moins de 15 pixels de large pour la vignette contenant un œil), cheveux sur les yeux (6%), bandeau (3%) … divers

(53%). La figure 3 rappelle les différentes étapes du processus de localisation.

TAB 1 : Résultats expérimentaux(avec seuil de dissimilarité : 0.1)

	Moyenne	Moyenne + rejet	Médiane	Médiane + rejet	Taux de rejet
Œil 1	0.080	0.072	0.060	0.058	0.13
Œil 2	0.077	0.070	0.059	0.056	0.13

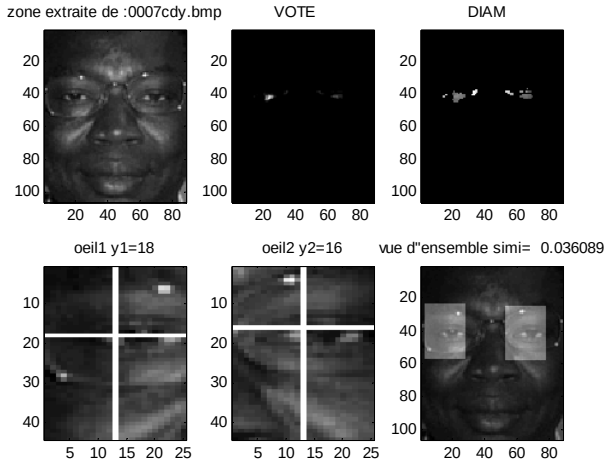

FIG 3 : image initiale, tableaux *VOTE* et *DIAM*, profil verticaux des deux yeux et résultats de localisation.

7. Conclusions

Une méthode globale de localisation des yeux dans une image de visage a été présentée et les résultats sur près de 2500 visages sont concluants. Elle repose sur une nouvelle transformation, dite « chinoise », qui peut servir dans d'autres contextes comme la détection de symétries axiales ou la localisation de patterns [4]. L'utilisation de cette technique pour la détermination de pose ou pour la confirmation d'une hypothèse de pose est la prochaine étape que nous nous sommes assignés. Par ailleurs, cette méthode pourra aussi être utilisée pour assister la détection du visage proprement dite. En effet, dans ce domaine, nous avons obtenu des performances intéressantes [1] mais avec un taux de fausse alarme élevé, donc perfectible. Précisons enfin que l'extension de cette méthode aux images couleurs de visages d'orientation non contrainte est en cours.

Références

[1] Belaroussi R., Prevost L. & Milgram M., Combining model-based classifiers for face localization, MVA'2005. Tsukuba, Japon, May 2005

[2] Hsu R.L. ,Abdel-Mottaleb. M. & Jain A. K., Face detection in color images, IEEE Trans. PAMI, 24(5):696-706, 2002.

[3] Martinetz T. & Schulten K., Topology representing networks, Neural Networks, 7(2), 1994.

[4] Negri P., Clady X. & Milgram M., Visual perception for human grasping gestures, ACIVS Conference, University of Antwerp, Antwerp, Belgium , September 2005

[5] Phung S. L., Bouzerdoum A. & Chai D., Skin segmentation using color pixel classification: Analysis and comparison, IEEE Trans. PAMI, to be published in 2005.

[6] Betke M., Mullally W. & Magee J., "Active Detection of Eye Sclera", *HMAS'2000 (Real Time Proceedings of the IEEE CVPR Workshop on Human Modeling, Analysis and Synthesis), 2000.*

[7] Braathen B. and Bartlett M. S. Littlewort-Ford G. & Movellan J. R., "3-D head pose estimation from video by nonlinear stochastic particle filtering", *UCSD MPLab TR 2001.05, 2001.*

[8] Daugman J.G., "Complete Discrete 2-D Gabor Transform by Neural Networks for Image analysis and Compression", *IEEE Trans. ASSP, 36(7):1169-1179, 1988.*

[9] Gu H, Su G and Du C., "Feature points extraction from face", *IVCNZ'2003 (Image and Vision Computing New Zealand), 2003.*

[10] Huang, J. & Wechsler H., "Eye location using genetic algorithm", *AVBPA'1999 (Audio and Video-Based Person Authentification), 1999.*

[11] Lin K.-H., Lam K.-M. & Siu, W.-C. "Locating the eye in human face images using fractal dimensions", *IEE Proc.Vis.Image Signal Process., 148(6) , 2001.*