

# Estimation discrète de l'angle Pan de la tête

A. Dahmane<sup>1,2</sup>

S. Larabi<sup>2</sup>

I. M. Bilasco<sup>1</sup>

C. Djeraba<sup>1</sup>

<sup>1</sup> Laboratoire d'Informatique Fondamentale LIFL

Université des Sciences et Technologies Lille 1

<sup>2</sup> Laboratoire de Recherche en Intelligence Artificielle LRIA

Université des Sciences et Technologies H. Boumediene

{afifa.dahmane, marius.bilasco, chabane.djeraba}@lifl.fr,

slarabi@usthb.dz

## Résumé

*Dans cet article, nous présentons une méthode géométrique d'estimation de la pose de la tête à partir d'images monoculaires. Nous allons démontrer que la taille des régions symétriques du visage est un bon indicateur du mouvement Pan de la tête. L'approche proposée ne nécessite pas la détection de points spécifiques sur le visage ce qui la rend robuste aux occlusions partielles ainsi qu'à la résolution des vidéos. Les dimensions de la région du visage qui contient une symétrie bilatérale est utilisée pour décider de l'orientation de la tête. Nous utilisons un modèle de classification à base d'arbre de décision. L'approche a été validé et les résultats évalués sur deux bases de tests (FacePix et Boston University head pose video).*

## Mots clefs

Estimation de la pose, orientation du regard, symétrie.

## 1 Introduction

L'estimation de la pose de la tête à partir d'images consiste à estimer l'orientation des trois degrés de liberté de la tête (*Pan*, *Tilt* et *Roll*). Les mouvements de la tête sont d'importants paramètres pour beaucoup d'applications ou d'interfaces homme-machine. divers domaines sont concernés tels que la surveillance de la vigilance des conducteurs, l'apprentissage à distance ou la vidéo conférence.

L'estimation de la pose de la tête est une des tâches importante en vision par ordinateur. Le problème crucial de toutes les approches proposées est de trouver les caractéristiques du visage les plus discriminantes qui permettent de satisfaire un ensemble de critères : estimation de la pose avec précision, système opérationnel avec une seule caméra (vision monoculaire), système autonome qui ne nécessite pas une intervention manuelle, présence de plusieurs personnes dans l'image, l'invariance à la lumière, l'indépendance à la résolution, intervalle complet du mouvement et l'estimation de la pose en temps réel. Une contrainte vient s'ajouter aux critères cités, qui est la disponibilité de matériel de pointe afin de pouvoir donner une

bonne estimation de la pose. Effectivement, dans la plus part des situations, nous ne disposons que d'une caméra standard.

Nous présentons dans cet article une approche pour l'estimation de la pose de la tête qui rallie entre la précision des méthodes géométriques et la généralité des méthodes globales. L'approche utilise des caractéristiques faciles à extraire et qui ne nécessite pas la localisation de points spécifiques du visage. Elle se base sur les régions du visage qui contiennent une symétrie bilatérale.

L'article est organisé comme suit. Premièrement, nous présentons dans la section 2 les travaux déjà réalisés. Ensuite, nous fournissons la méthodologie utilisée pour l'estimation de la pose de la tête à partir des régions symétriques du visage dans les sections 3 et 4. Dans la section 5, l'approche est évaluée avec la base d'images FacePix [16]. Des tests sont également effectués avec des séquences vidéos de la base de l'Université de Boston [17]. Finalement, nous concluons et discutons des travaux à venir dans la section 6.

## 2 État de l'art

Plusieurs approches ont été proposées et peuvent être classifiées en trois catégories :

- **Approches basées sur la forme** ; Construisent un modèle à partir de caractéristiques spécifiques du visage telles que les yeux, le nez et la bouche. Le modèle construit peut être soit géométrique [1], soit flexible comme les AMMs [2].
- **Approches basées sur l'apparence globale** ; modélisent l'apparence globale de la tête à partir d'une base d'apprentissage. Ces méthodes peuvent être basées sur un patron (Template Based Method), tel que l'utilisation de la distribution gaussienne multi-dimensionnelle [3], ou basées sur un ensemble de détecteurs (Detectors Arrays) tel que l'utilisation des SVM [4] ou bien basées sur la régression non-linéaire comme l'utilisation d'un réseau neuronal de régression généralisé [5].
- **Approches hybrides** ; combinent deux ou plusieurs des

méthodes issues des deux approches précédentes [6]. L'approche la plus courante consiste à compléter l'estimation statique par un système de suivi.

Peu de travaux ont été consacrés à l'utilisation de la symétrie pour l'estimation de la pose de la tête. La discrimination entre les orientations de la tête est basée sur deux indices : la déviation de la forme de la tête de la symétrie bilatérale, et la déviation de l'orientation du nez de la verticale [7]. Le modèle d'illumination basé sur la symétrie proposé dans [8] est basé sur trois caractéristiques (les yeux et le bout du nez). Pour chaque combinaison de deux yeux et un nez, la pose de tête est calculée en utilisant une projection géométrique. Dans [9] la propriété de symétrie de la tête a été utilisé comme un indicateur de l'intention visuelle pour les personnes handicapées. La pose de la tête (Pan et Roll) est estimée dans [10] en exploitant la structure symétrique du visage humain et en utilisant l'image miroir d'un visage test comme une seconde vue virtuelle. La méthode est basé sur l'extraction de points caractéristiques faciaux à partir de l'image et de son miroir et leurs correspondance. Une étude récente sur l'estimation de la pose de la tête peut être trouvée dans [11]. Cette étude conclue que les méthodes géométriques sont simple et rapide mais l'inconvénient principal se trouve dans la détection précise des points caractéristiques dans les situations où les points de repère faciaux (coins des yeux, bout du nez et coins de la bouche) sont obscurcis. En outre, l'imagerie à faible résolution rend difficile de déterminer précisément les emplacements de ces points caractéristiques.

Puisque notre approche n'utilise pas de points spécifiques du visage, la détection de la symétrie bilatérale sur le visage est suffisante. Plusieurs approches ont été proposées pour la détection de l'axe de symétrie dans une image [12, 13, 14]. Ces méthodes se basent sur des contraintes ou des hypothèses initiales ou sur des points caractéristiques difficiles à extraire lorsque la caméra est éloignée. Pour cela, nous avons choisi une méthode de détection d'axe de symétrie [15] qui ne nécessite pas d'intervention manuelle ni de connaissance préalable sur les caractéristiques de la symétrie.

### 3 Détection et analyse des parties symétriques du visage

Nous allons démontrer que la dimension des régions du visage qui contiennent une symétrie bilatérale est une bonne caractéristique pour l'estimation de la pose de la tête. Cette caractéristique va nous permettre de construire un modèle de prédiction de poses discrètes suivant le mouvement horizontal de la tête (le Pan).

#### 3.1 Propriété des parties symétriques du visage

La symétrie procure des connaissances de haut niveau sur les images. Nous allons utiliser la symétrie bilatérale de la tête afin d'estimer la pose de la tête. Lorsque le visage est face à la caméra, la symétrie entre ses deux parties appa-

rait clairement et la ligne qui passe entre les yeux et sur le bout du nez définie l'axe de symétrie du visage. Cependant, lorsque la tête effectue un mouvement vers la gauche ou vers la droite, cette symétrie disparaît.

Nous démontrons que la taille des régions symétriques du visage sont de bons et robustes caractéristiques pour l'estimation de la pose de la tête. On considère que le visage est parallèle à la caméra.  $a$  et  $b$  deux points symétriques sur le visage et  $m$  le milieu du segment  $[ab]$ . Soit  $a_i, b_i, m_i$  la projection de  $a, b$  et  $m$  sur le plan image. Puisque  $(ab)$  est parallèle au plan image,  $[a_i m_i]$  et  $[m_i b_i]$  sont symétriques par rapport à  $m_i$  comme le montre la figure 1 (a).

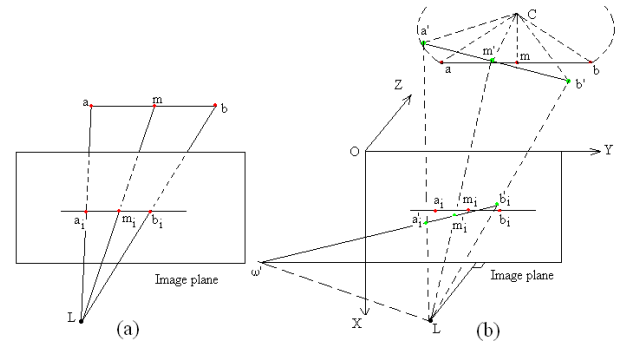


Figure 1 – (a) Projection du centre de symétrie  $m$  d'un segment  $(ab)$  parallèle à une image plane, (b) Projection du centre de symétrie d'un segment après un mouvement Pan.

Lorsque la tête effectue un mouvement Pan de gauche à droite, les points caractéristiques  $(a, b, m)$  bougent et deviennent  $(a', b', m')$ , leur projection est  $(a'_i, b'_i, m'_i)$ .

Soit  $\omega'$  le point de fuite associé à  $(a', b')$ . En appliquant l'invariance du cross-ratio [18], on obtient :

$$(a', b', m', \infty) = (a'_i, b'_i, m'_i, \omega') \quad (1)$$

$$\frac{m'a'}{m'b'} = \frac{m'_i a'_i}{m'_i b'_i} \div \frac{\omega'_i a'_i}{\omega'_i b'_i} \quad (2)$$

Comme les deux membres de l'équation 2 sont égaux à 1,  $m'_i$  n'est pas le milieu de  $a'_i b'_i$  mais sa position dépend de celle de  $a'_i$  et de  $b'_i$  par rapport à  $\omega'$ . Sur la figure 1 (b), comme  $a'_i$  est proche de  $\omega'$ ,  $m'_i$  sera aussi proche de  $a'_i$  que de  $b'_i$ . Nous concluons alors que  $m'_i a'_i$  n'est pas symétrique à  $m'_i b'_i$ .

Puisque  $m$  et  $m'$  sont respectivement les centres de symétrie de  $ab$  et  $a'b'$  alors une symétrie partielle existe sur la ligne  $a'_i b'_i$ . Dans ce cas, le centre de symétrie n'est pas le milieu de  $[a'_i b'_i]$  mais  $m'_i$ . Cette symétrie concerne les segments  $m'_i a'_i$  et  $m'_i d'_i$  où  $d'_i$  se situe entre  $m'_i$  et  $b'_i$ . (Voir figure 2).

Lorsque la tête effectue un mouvement Pan plus grand, la largeur de la région qui contient une symétrie bilatérale décroît pour deux raisons :

- $a'$  n'apparaît plus sur l'image plane.

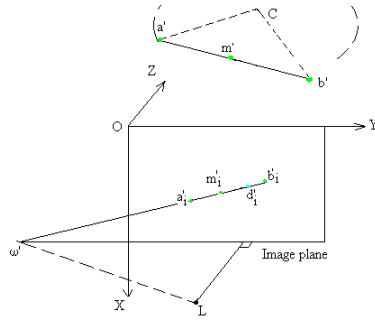


Figure 2 – Projection des segments symétriques après un mouvement Pan.

–  $m'_i$  devient plus proche de  $a'_i$  parce que le point de fuite  $\omega'$  devient plus proche de  $a'_i$ .

La figure 3 représente le changement dans le ratio des pixels symétriques par rapport à l'angle Pan. Chaque courbe de la figure correspond à une personne différente. L'ordonnée indique le pourcentage de symétrie par rapport au visage tout entier selon différents angles entre  $-45^\circ$  et  $+45^\circ$ .

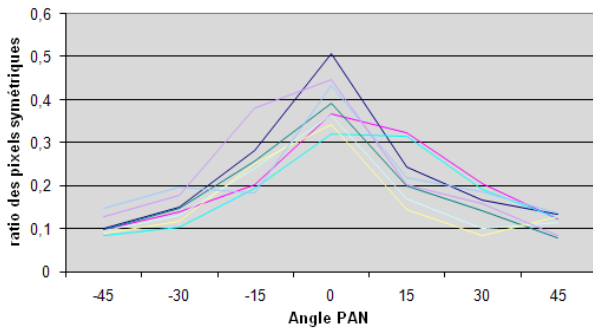


Figure 3 – Le changement dans le ratio des pixels symétriques par rapport à l'angle Pan

Nous concluons que plus l'ampleur du mouvement est grande, plus la taille de la région du visage qui contient une symétrie bilatérale devient petite.

### 3.2 Extraction de la symétrie du visage

Nous avons implémenté l'algorithme proposé par Stentiford [15] avec quelques modifications visant à augmenter la précision dans la détection de l'axe de symétrie ainsi que la réduction du temps de calcul. Alors que l'approche de Stentiford cherche l'angle auquel correspond le plus grand nombre d'axes de symétrie, nous fixons l'angle d'inclinaison à  $90^\circ$  puisque nous traitons le mouvement Pan. Ceci nous permet d'obtenir une plus grande précision quand à la position de l'axe de symétrie vertical du visage. La figure 4 montre les résultats obtenus pour cinq personnes. Chaque personne se trouve sous un angle tilt précis et effectue un mouvement Pan.



Figure 4 – Pixels symétriques localisés pour un mouvement Pan de  $-45$  à  $+45$  degrés avec un pas de  $15$  degrés et un angle tilt de (a)  $+30^\circ$ , (b)  $+15^\circ$ , (c)  $0^\circ$ , (d)  $-15^\circ$  et (e)  $-30^\circ$ .

## 4 Calcul de l'angle Pan

Afin d'estimer l'ampleur du mouvement Pan, un arbre de décision est construit moyennant les caractéristiques relatives extraites des régions symétriques pour chaque pose. Pour accroître les performances de la prédiction, nous utilisons l'arbre de décision alternatif (Alternating Decision tree) [19] qui est basé sur le boosting. L'arbre alterne entre des nœuds de prédiction et des nœuds de décision. La racine est un nœud de prédiction et contient une valeur. La somme des valeurs de prédiction croisées en parcourant tous les chemins pour lesquels tous les nœuds de décision sont vrais, est utilisée pour classifier une instance donnée.

Nous avons utilisé la base de données FacePix [16] pour l'apprentissage et l'évaluation. L'ensemble d'images représente les poses de la tête pour lesquelles l'axe de symétrie est bien détecté. Nous avons sélectionné les poses qui varient de  $-45^\circ$  (gauche) à  $+45^\circ$  (droite). Nous commençons par extraire les caractéristiques du visage. Telles que le ratio des pixels symétriques dans la région qui contient une symétrie bilatérale et la largeur de cette région. Afin de localiser les caractéristiques avec précision, nous délimitons la région d'intérêt au haut du visage. Le visage est détecté manuellement lors de la phase de l'apprentissage. Par contre pour les tests, le processus est complètement automatique. Nous utilisons pour cela le détecteur de Viola Jones [20]. Enfin, nous construisons le modèle à partir des vecteurs de caractéristiques dérivés des images de plusieurs personnes prises dans des poses différentes. Ces images représentent sept poses associées aux angles suivants :  $-45^\circ$ ,  $-30^\circ$ ,  $-15^\circ$ ,  $0^\circ$ ,  $15^\circ$ ,  $30^\circ$  and  $45^\circ$ . Le classificateur qui estime l'orientation de la tête a quatre classes distinctes : classe 1 :  $-45^\circ$  et  $45^\circ$ , classe 2 :  $-30^\circ$  et  $30^\circ$ , classe 3 :  $-15^\circ$  et  $15^\circ$  et classe 4 :  $0^\circ$ . Les poses droite et gauche sont réunis dans la même classe car elles contiennent la même symétrie et, par conséquent, les mêmes informations. La figure 5 montre les quatre classes de pose avec leurs régions caractéristiques.

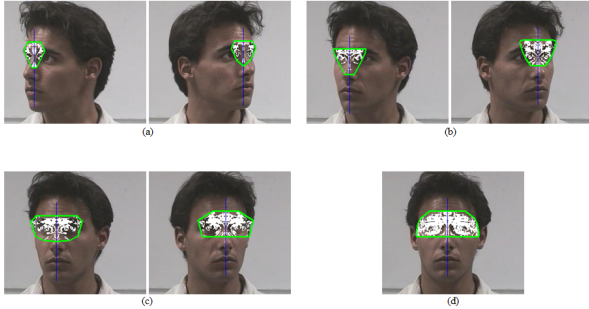


Figure 5 – (a) classe 1 :  $-45^\circ$  et  $45^\circ$ , (b) classe 2 :  $-30^\circ$  et  $30^\circ$ , (c) classe 3 :  $-15^\circ$  et  $15^\circ$ , (d) classe 4 :  $0^\circ$

## 5 Résultats expérimentaux

Nous utilisons la base de données FacePix [16] pour construire l'arbre de décision et pour l'évaluer. La base FacePix se compose de trois ensembles d'images de visages. L'ensemble contenant des poses avec variations d'angle est composée de 181 images de visage, représentant l'intervalle de  $-90^\circ$  à  $+90^\circ$  avec un pas de 1 degré, de 30 personnes différentes. Parmi les 181 poses, nous utilisons 7 poses (de  $-45^\circ$  à  $+45^\circ$  avec un pas de  $15^\circ$ ) parce que quand cet intervalle est dépassé, la symétrie bilatérale disparaît de l'image plane.

Nous avons divisé les données en six sous-ensembles égaux et avons effectué six fois la validation croisée (6-fold cross validation). Dans chaque exécution, cinq ensembles sont utilisés comme ensemble d'apprentissage et l'ensemble restant est utilisé comme ensemble de test. Les sujets des deux ensembles (apprentissage et test) sont distincts puisque chaque personne est prise une seule fois.

Afin de classifier chaque image en entrée en une des quatre poses, nous avons construit un arbre de décision alternatif et nous avons évalué le modèle résultant. Nous avons obtenu une moyenne de bonne classification égale à **81.4%**. La figure 6 montre le taux de bonne classification pour les quatre classes de poses.

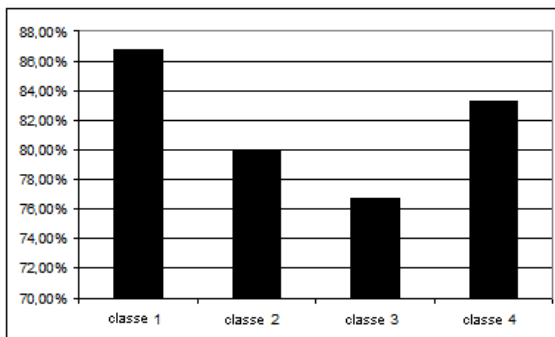


Figure 6 – Taux de classification pour les quatre poses

En plus des expérimentations de validation, nous avons

testé l'arbre de décision sur une autre base de données. Le modèle construit avec les images de la base FacePix a été utilisé pour prédire la pose sur des vidéos de la base de Boston University [17]. Puisque nous n'avons pas d'information à propos de la localisation de la tête sur les séquences vidéo, nous avons utilisé le détecteur de Viola Jones afin de limiter la région d'où extraire les symétries.

Le principal avantage de cette méthode est que le calcul peut commencer à n'importe quelle pose, sans initialisation, puisque la tête et la symétrie du visage sont automatiquement détectées pour les poses entre  $-45^\circ$  et  $+45^\circ$ . Les résultats sur des séquences vidéos fournissent une estimation précise de la pose. Un exemple de frames appartenant à une séquence vidéo test est montré dans la figure 7. La région des caractéristiques symétriques est superposée sur les frames exemple. Un graphe de comparaison entre les poses estimées et la vérité terrain est réalisé. L'erreur moyenne sur cette séquence est de  $4.47^\circ$ , ce qui est très petit en comparaison avec l'intervalle entre les classes de pose qui est  $15^\circ$ . L'erreur moyenne sur toute la base de vidéos (les séquences contenant un mouvement Pan) est de  $6.49^\circ$ .

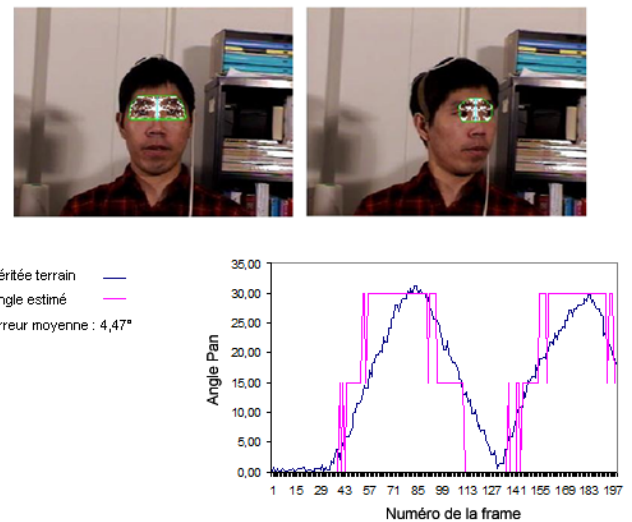


Figure 7 – Exemple d'une séquence vidéo de la base de l'Université de Boston avec ses résultats (pose discrétisée estimée pour chaque frame comparée avec l'angle Pan réel)

## 6 Conclusion

Dans ce papier, nous avons proposé une nouvelle approche d'estimation discrète de la pose de la tête. Nous avons utilisé un ensemble de caractéristiques extraites de la symétrie du visage afin de construire un modèle pour l'estimation de l'angle Pan de la tête. Ces caractéristiques peuvent être extraites même si la caméra est éloignée, que la résolution de l'image est faible ou qu'il y a des occlusions partielles car pas besoin de détecter des points d'intérêt spécifiques. Les résultats obtenus par notre approche ont été évalués avec

la base d'images FacePix et le taux de bonne classification est de 81.4%. Nous avons donc, un modèle que nous pouvons déployé en temps réel pour beaucoup d'applications qui nécessitent une interface homme-machine juste avec une webcam.

Nous avons en perspective d'exploiter le flux vidéo, et donc l'information temporelle. Le système de suivi servira à obtenir une continuité dans les poses estimées et aussi à étendre l'intervalle du mouvement. La méthode proposée représentera alors la partie statique du système final d'estimation de la pose. Cette dernière sert à initialiser le système de suivi .

## Références

- [1] Y. Pan, H. Zhu, et R. Ji. *3-D Head Pose Estimation for Monocular Image*. Fuzzy Systems and Knowledge Discovery. Springer, 2005.
- [2] J. Xiao, S. Baker, I. Matthews, et T. Kanade. Real-time combined 2d+3d active appearance models. Dans *International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 535–542, 2004.
- [3] Y. Wu et K. Toyama. Wide range illumination insensitive head orientation estimation. Dans *International Conference on Automatic Face and Gesture Recognition (AFGR)*, pages 183–188, 2000.
- [4] J. Huang, X. Shao, et H. Wechsler. Face pose discrimination using support vector machines (svm). Dans *International Conference on Pattern Recognition (ICPR)*, volume 1, pages 154–156, 1998.
- [5] K. Bailly et M. Milgram. Head pan angle estimation by a nonlinear regression on selected features. Dans *International Conference on Image Processing (ICIP)*, pages 3589–3592, Cairo - Egypt, 2009.
- [6] T. Vatahska, M. Bennewitz, et S. Behnke. Feature-based head pose estimation from images. Dans *IEEE-RAS 7th International Conference on Humanoid Robots (Humanoids)*, 2007.
- [7] H. R. Wilson, F. Wilkinson, L. Lin, et M. Castillo. Perception of head orientation. *Vision Research*, 40(5) :459–472, 2000.
- [8] Martin Gruendig et Olaf Hellwich. 3d head pose estimation with symmetry based illumination model in low resolution video. Dans *Lecture Notes in Computer Science*, volume 3175, pages 45–53. Springer, 2004.
- [9] T. Luhadjula, E. Monacelli, Y. Hamam, B.J. van Wyk, et Q. Williams. Visual intention detection for wheelchair motion. Dans *International Symposium on Visual Computing (ISVC)*, pages 407–416, 2009.
- [10] Vinod Pathangay, Sukhendu Das, et Thomas Greiner. Symmetry-based face pose estimation from a single uncalibrated view. *8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, Amsterdam, The Netherlands, 17-19 September, pages 1–8, 2008.
- [11] E. Murphy-Chutorian et M. M. Trivedi. Head pose estimation in computer vision : A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(4) :607–626, 2009.
- [12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60 :91–110, 2004.
- [13] G. Loy et J. Eklundh. Detecting symmetry and symmetric constellations of features. Dans *European Conference on Computer Vision (ECCV)*, pages 508–521, 2006.
- [14] H. Cornelius et G. Loy. Detecting rotational symmetry under affine projection. Dans *International Conference on Pattern Recognition (ICPR)*, pages 292–295, 2006.
- [15] F. Stentiford. Attention based facial symmetry detection. Dans *International Conference on Advances in Pattern Recognition (ICAPR)*, 2005.
- [16] J. Black, M. Gargesha, K. Kahol, P. Kuchi, et S. Panchanathan. A framework for performance evaluation of face recognition algorithms. Dans *ITCOM, Internet Multimedia Systems II*, Boston, 2002.
- [17] R. Valenti et T. Gevers. Robustifying eye center localization by head pose cues. Dans *IEEE conference on Computer Vision and Pattern Recognition*, 2009.
- [18] H.S.M. Coxeter. *Projective Geometry*. Fuzzy Systems and Knowledge Discovery. Springer-Verlag 2nd Revised edition, 2003.
- [19] Geoffrey Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, et Mark Hall. Multiclass alternating decision trees. Dans *ECML*, pages 161–172. Springer, 2001.
- [20] P. Viola et M. Jones. Rapid object detection using a boosted cascade of simple features. Dans *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511 – I–518 vol.1, 2001.