

# CARTES DE DISTANCE : Prétraitement robuste à l'illumination pour les Modèles Actifs d'Apparence.

Sylvain LE GALLOU\*, Gaspard BRETON\*, Christophe GARCIA\*, Renaud SÉGUIER\*\*

\* : France Telecom R&D - TECH/IRIS

4 rue du clos courtel, BP 91226, 35 512 Cesson Sévigné

\*\* : Supélec - IETR, SCEE Team

Avenue de la boulaie, BP 81127, 35 511 Cesson Sévigné

sylvain.legallou@orange-ft.com, gaspard.breton@orange-ft.com, christophe.garcia@orange-ft.com,  
renaud.seguier@supelec.fr

**Résumé :** Les Modèles Actifs d'Apparence (AAM) peuvent être très utiles dans la conception des systèmes d'interface Homme-Machine intégrant de la vision par ordinateur car ils permettent une bonne localisation des visages. Les applications sont nombreuses : codage vidéo pour la visiophonie, animation de visages synthétiques, reconnaissance visuelle de la parole, analyse d'expressions et d'émotions, suivi et reconnaissance de visages. Par contre les AAM sont peu robustes aux variations des conditions d'éclairage, ce qui peut poser problème dans le cas d'environnements non contraints. En effet, un visage peut avoir des illuminations variées au cours des différentes communications avec la machine. Nous proposons donc un prétraitement robuste permettant d'améliorer les performances des Modèles Actifs d'Apparence dans des conditions d'illumination changeantes. Pour cela, les images en niveaux de gris en entrée de la chaîne de traitement des AAM sont remplacées par des cartes de distance. Ces cartes de distance contiennent des informations de distance entre les différents contours, ce qui permet de s'affranchir fortement des effets de variations lumineuses qui rendent instable l'utilisation des modèles AAM.

**Mots-clés :** Imagerie, Illumination, Modèle Actif d'Apparence, Analyse de visage, Interface Homme-Machine.

## 1 INTRODUCTION

Dans le contexte des interfaces Homme-Machine, l'alignement de visage (détection précise des éléments caractéristiques : les yeux, le nez et la bouche) est une étape nécessaire et importante pour l'analyse de la gestuelle. Les Modèles Actifs d'Apparence (AAM) proposés par Cootes dans [Cootes, 1998] permettent une localisation fine de ces éléments clés. Pour ce faire, les AAM utilisent conjointement l'information de forme avec l'information de texture. En effet, les AAM sont basés sur la connaissance a priori des formes (points d'intérêts reliés entre eux) et des textures projetées sur une forme normalisée.

Ils peuvent alors générer un ensemble de représentations plausibles de formes et de textures pour les objets appris. Ils permettent aussi de rechercher ces mêmes objets dans de nouvelles images à l'aide d'un processus d'optimisation sur les paramètres du modèle. Il s'agit alors d'adapter le plus finement possible le modèle sur la zone de l'image contenant l'objet recherché.

Les applications visées nécessitent une bonne robustesse face aux variations d'éclairage. Malheureusement, les AAM, qui sont basés sur une analyse de la texture, sont sensibles aux changements d'illumination.

Pour résoudre ce problème d'illumination, plusieurs méthodes ont déjà été proposées. Certains travaux utilisent des méthodes basées réseau d'ondelettes, comme [Hu, 2003] qui propose de remplacer la partie d'analyse de la texture dans les AAM par une représentation en réseau d'ondelettes. D'autres travaux s'appuient sur des approches basées contours ou filtrage, dans lesquels la composante illumination est minimisée grâce à des modèles d'illumination. Huang et al. [Huang, 2004] ont présenté ces approches pour les Modèles Actifs de Forme (Active Shape Model) [Cootes, 1995].

Dans cet article, nous proposons une méthode plus simple à mettre en œuvre et permettant d'améliorer les performances des AAM dans des illuminations variées. Notre contribution consiste à remplacer les images en niveaux de gris en entrée de la chaîne de traitement des AAM par des cartes de distance. Ce prétraitement produit des images contenant une information de distance entre les différents contours de l'objet présents dans l'image originale en niveaux de gris.

Dans la partie suivante (partie 2), nous présentons brièvement les Modèles Actifs d'Apparence (AAM). La partie 3 explique le processus de création des cartes de distance utilisées comme prétraitement pour les AAM. Dans la partie 4, nous exposons nos résultats expérimentaux : les images utilisées dans la méthode des AAM, la création du modèle AAM et la comparaison

entre l'utilisation standard des AAM et l'utilisation avec le prétraitement des cartes de distance qui montre le gain en robustesse apporté par notre contribution. Enfin, la partie 5 conclut ce papier.

## 2 LES MODÈLES ACTIFS D'APPARENCE (AAM)

Le Modèle Actif d'Apparence est un modèle déformable qui permet à la forme et à la texture d'être conjointement synthétisées. Les AAM proposés par Cootes, Edwards et Taylor en 1998 [Cootes, 1998], se basent sur la connaissance a priori des formes et des textures projetées sur une forme normalisée. Les trois étapes de cette méthode sont brièvement présentées ici :

- La première étape est la phase d'apprentissage dans laquelle le modèle et ses paramètres de déformation sont créés.

Une Analyse en Composante Principale (ACP) est appliquée successivement sur une base d'apprentissage de formes puis sur une base d'apprentissage de textures affranchies de leur forme afin de créer les modèles statistiques de forme et de texture donnés par :

$$x_i = x_{moy} + \Phi_x * b_x \quad (1)$$

$$g_i = g_{moy} + \Phi_g * b_g \quad (2)$$

où  $x_i$  et  $g_i$  sont respectivement la forme et la texture synthétisée,  $x_{moy}$  et  $g_{moy}$  la forme moyenne et la texture moyenne,  $\Phi_x$  et  $\Phi_g$  les matrices de vecteurs propres des matrices de covariance des formes et des textures et,  $b_x$  et  $b_g$  les vecteurs contrôlant la déformation de la forme synthétisée et de la texture synthétisée.

Une autre ACP est ensuite appliquée sur le paramètre  $b$ , concaténation de  $b_x$  et de  $b_g$ , afin d'obtenir le paramètre d'apparence  $c$  :

$$b = \Phi * c \quad (3)$$

où  $\Phi$  est la matrice de vecteurs propres de l'ACP.  $c$  est un vecteur contrôlant à la fois  $b_x$  et  $b_g$  (équation 3), c'est-à-dire la forme (équation 1) et la texture (équation 2) du modèle.

- La seconde étape est la phase de création des matrices d'expériences dans laquelle est créée, à partir de plusieurs expériences, une relation entre les variations du paramètre de contrôle ( $c$ ) du modèle ainsi que l'image résiduelle entre le modèle déformé et son image originale.

En effet, chaque image de la base d'apprentissage contient un objet synthétisé par une valeur particulière du paramètre  $c$ . Soit  $c_0$  la valeur particulière de  $c$  de l'image  $i$  de la base d'apprentissage. En modifiant le paramètre  $c_0$  de  $\delta c$  ( $c = c_0 + \delta c$ ), une nouvelle forme  $x_m$  et une nouvelle texture  $g_m$  (équation 3) sont alors synthétisées. Considérons maintenant la texture  $g_i$  de l'image originale  $i$  se trouvant à l'intérieur de la forme  $x_m$ . La différence de pixels  $\delta g = g_i - g_m$  (appelée aussi image résiduelle) et une régression linéaire à multiples variables sur un certain nombre d'expériences (modification des objets synthétisés dans la base d'apprentissage de  $\delta c$ ), nous donnent une relation entre  $\delta c$

et  $\delta g$  :

$$\delta c = R_c * \delta g \quad (4)$$

$R_c$  est appelé matrice d'expériences.

- La dernière étape est la phase de recherche qui permet au modèle déformable de s'ajuster aux objets dans de nouvelles images en utilisant la relation de l'équation 4.

Cette phase consiste à trouver une texture et une forme particulière dans de nouvelles images. Les modifications à apporter au modèle d'apparence  $c$  à partir de l'équation 4 permettent au modèle de s'ajuster à l'objet recherché dans de nouvelles images. L'algorithme de recherche est décrit ici :

- 1- Générer  $g_m$  et  $x$  à partir des valeurs du paramètre  $c$  (initialement mis à 0).
- 2- Calculer  $g_i$ , la texture à l'intérieur de la forme  $x$  et dans l'image dans laquelle se trouve l'objet recherché.
- 3- Calculer  $\delta g_0 = g_i - g_m$  et  $E_0 = |\delta g_0|$ .
- 4- Prédire  $\delta c_0 = R_c * \delta g_0$ .
- 5- Trouver le 1er coefficient d'atténuation  $k$  (de façon dichotomique parmi [1.5, 0.5, 0.25, 0.125, 0.0625]) donnant  $E_j < E_0$ , où  $E_j = |\delta g_j|$ ,  $\delta g_j = g_{ij} - g_{mj}$  où  $g_{mj}$  est la texture donnée par  $c_j = c - k * \delta c_0$  et  $g_{ij}$  la texture de l'image à l'intérieur de  $x_{ij}$  (forme donnée par  $c_j$ ).
- 6- Tant que l'erreur  $E_j$  n'est pas stable, recommencer à l'étape 1 avec  $c = c_j$ .

Lorsque la convergence de cette troisième étape est atteinte, les représentations de la texture et de la forme de l'objet recherché sont alors respectivement synthétisées à travers le modèle par  $g_m$  et  $x$ . La figure 1 est un exemple de recherche de visage à l'aide des AAM.

## 3 PRÉTRAITEMENT DES AAM : LES CARTES DE DISTANCE

Afin d'obtenir une méthode plus robuste aux variations d'illumination, nous ne prenons plus directement en compte les images originales en niveaux de gris en entrée des AAM mais des images transformées (appelées cartes de distance). Ces cartes de distance permettent de tenir compte des relations de distance existant entre les différents contours composant l'objet recherché. Le prétraitement des AAM que nous proposons ici est donc la transformation des images originales en cartes de distance.

### 3.1 Création d'une carte de distance

La création d'une carte de distance associée à une image originale (voir Figure 2) se déroule en quatre étapes :

- L'image originale est divisée en un certain nombre de régions rectangulaires dans lesquelles une égalisation d'histogramme est mise en œuvre. Cette égalisation d'histogramme adaptative permet en fait d'améliorer

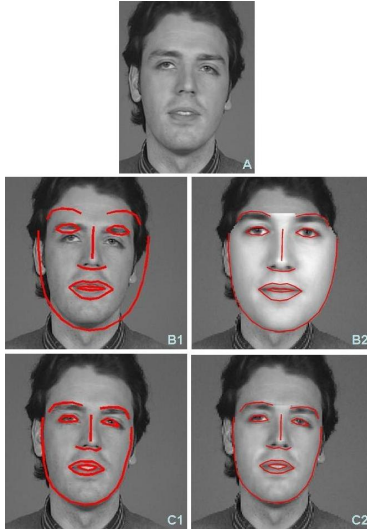


FIG. 1 – Exemple de recherche de visage à l’aide des AAM : Image originale (Image A), initialisation du modèle (Images B1 et B2), convergence du modèle (Images C1 et C2). Sur les images B1 et C1 seule la forme du modèle est affichée, tandis que sur les images B2 et C2, la texture et la forme sont affichées.

les contrastes de l’objet présent dans l’image originale. Nous avons utilisé la méthode d’égalisation adaptative de CLAHE (Contrast-Limited Adaptive Histogram Equalization) [Zuiderveld, 1994], car celle-ci permet d’augmenter les contrastes sur les contours. Plus précisément, l’image est divisée en  $8 \times 8$  régions homogènes (i.e. 64 régions dans une image), ensuite, dans chacune de ces régions, est appliquée la méthode de CLAHE selon une distribution de Rayleigh (Figure 2-B).

- Une image lissée est obtenue par un filtrage passe-bas (Figure 2-C).
- Une extraction de contours est effectuée dans les régions rectangulaires de l’image lissée. Cette extraction de contours adaptative permet au seuil du filtre de contours de s’adapter au contexte local de l’image. Pour mettre en œuvre cette étape, nous avons utilisé le filtre de Sobel dans les 64 mêmes régions de l’étape d’égalisation d’histogramme adaptative. A la suite de cette étape est créée une image de contours (Figure 2-D).
- Enfin, pour chaque pixel de l’image de contours, la distance euclidienne entre ce pixel et le pixel contour le plus proche est calculée. Cette dernière étape engendre la carte de distance (Figure 2-E), associée à une image originale (Figure 2-A), classiquement utilisée par les AAM.

### 3.2 Utilisation du prétraitement dans les AAM

La répercussion du prétraitement des cartes de distance dans les AAM se fait dans les trois étapes (voir partie 2). Le modèle AAM, communément composé d’une forme et d’une texture en niveaux de gris, devient alors une forme et une texture en carte de distance.

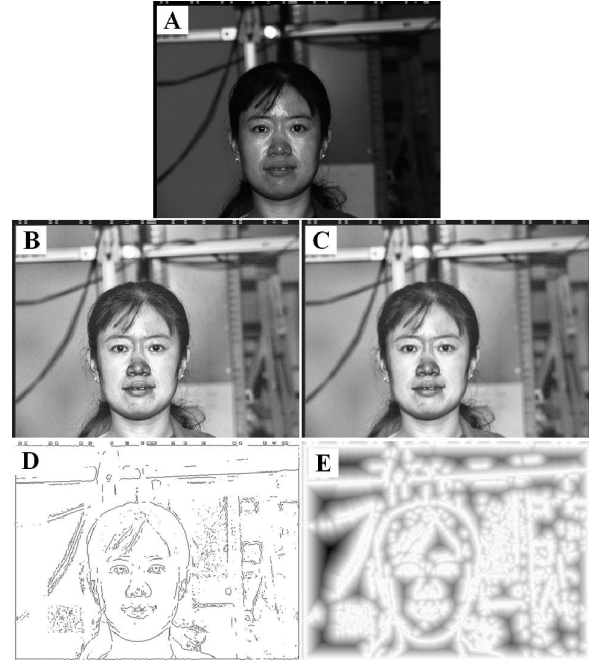


FIG. 2 – Exemple de création d’une carte de distance : A- Image originale, B- Égalisation d’histogramme adaptative, C- Lissage de l’image, D- Extraction de contours, E- Carte de distance.

Dans la phase d’apprentissage, l’ACP sur les textures en niveaux de gris est remplacée par une ACP sur les cartes de distance associées. L’ACP suivante se fait donc sur la concaténation des vecteurs  $b_x$  et  $b_g$  contrôlant la déformation de la forme et de la texture en carte de distance du modèle.

La seconde étape permet toujours de créer les matrices d’expériences, mais dans ce cas elle se base sur les cartes de distance constituant la base d’apprentissage.

Enfin, dans l’étape de recherche, le modèle (forme et texture en carte de distance) va s’ajuster à l’objet recherché non plus dans une nouvelle image en niveaux de gris mais dans sa carte de distance associée.

## 4 RÉSULTATS EXPÉRIMENTAUX

### 4.1 Base de donnée

Notre algorithme d’AAM basé sur les travaux de Cootes [Cootes, 2004] et de Stegmann [Stegmann, 2000] crée un modèle de visage dans une fenêtre de  $64 \times 64$  pixels de telle façon que la texture du visage moyen contienne environ 1400 pixels.

Les tests que nous présentons pour comparer la méthode standard des AAM et les AAM avec le prétraitement, se sont déroulés dans un contexte de généralisation. En effet, pour constituer le modèle AAM, nous avons utilisé seulement dix visages illuminés de face de la base M2VTS [Pigeon, 1996] comme base d’apprentissage (Figure 3). La base M2VTS comprend 37 visages différents et 5 prises de vues (une semaine d’intervalle entre chaque prise). Puis, pour les tests, nous avons utilisé dix-sept visages (différents de la base d’apprentissage) provenant

de la base PIE [Sim, 2002] du CMU (Carnegie Mellon University) disponibles sous vingt-et-une illuminations différentes soit 357 images de test (Figure 4). La base d'images PIE a été créée par le CMU de telle façon que chaque visage (68 au total) soit acquis sous 21 illuminations différentes à l'aide d'un système de 21 flashes disposés de gauche à droite et plus ou moins en hauteur autour du visage (Figure 5).



FIG. 3 – Base d'apprentissage (10 visages de M2VTS).



FIG. 4 – Les 17 visages utilisés de la base PIE pour les tests vus suivant un angle d'illumination différent (illumination forte de droite pour le visage en haut à gauche jusqu'à une illumination forte de gauche pour le visage en bas à droite).

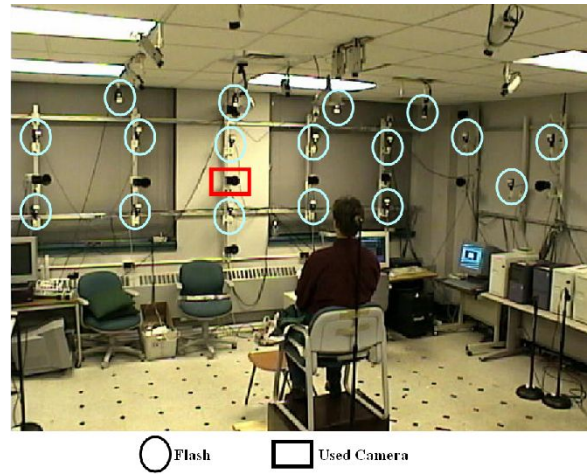


FIG. 5 – Dispositif d'acquisition de visage du CMU : positions des 17 flashes (4 des 21 flashes ne sont pas visibles dans cette vue) et position de la caméra.

## 4.2 Résultats

La figure 6 présente un résultat de convergence des AAM dans le cas de son utilisation standard et dans le cas de son utilisation avec le prétraitement des cartes de distance. Par la suite, afin d'avoir une meilleure représentation des résultats à comparer, les formes trouvées après convergence des AAM dans les deux expériences (standard et avec prétraitement) sont affichées sur les images originales en niveaux de gris sans la texture trouvée. Il est à noter que pour chaque test, le nez du modèle est initialisé sur le nez du visage recherché.

La figure 7 montre quelques résultats obtenus (résultats de l'utilisation standard des AAM sur la colonne de gauche et résultats de l'expérience des AAM avec le prétraitement des cartes de distance dans la colonne de droite) dans 4 illuminations différentes (d'une illumination forte de la droite : images du haut, à une illumination forte de la gauche : images du bas). Les résultats des deux expériences sont équivalents en illumination de face mais la figure 7 montre l'intérêt du prétraitement des cartes de distance lorsque la direction de l'illumination varie.

Le graphe de la figure 8 présente les courbes de convergence obtenues pour l'utilisation normale des AAM (courbe reliée par les ronds) et pour l'utilisation avec le prétraitement des cartes de distance (courbe reliée par les carrés). Ce graphe représente le pourcentage de convergence en fonction de la direction de l'illumination. Le modèle est considéré comme ayant bien convergé lorsque l'erreur de localisation du centre de gravité des yeux, du centre de gravité de la bouche, et du nez ne dépassent pas le seuil de  $D_{yeux}/4$ , où  $D_{yeux}$  représente la distance entre les yeux. L'axe des ordonnées représente donc le pourcentage d'images sur lesquelles le modèle a bien convergé. L'axe des abscisses représente la direction de l'illumination : une direction d'illumination de 1 correspond à une illumination venant de la gauche, une direction d'illumination de 21 correspond à une illumination venant de la droite du visage. Les illuminations

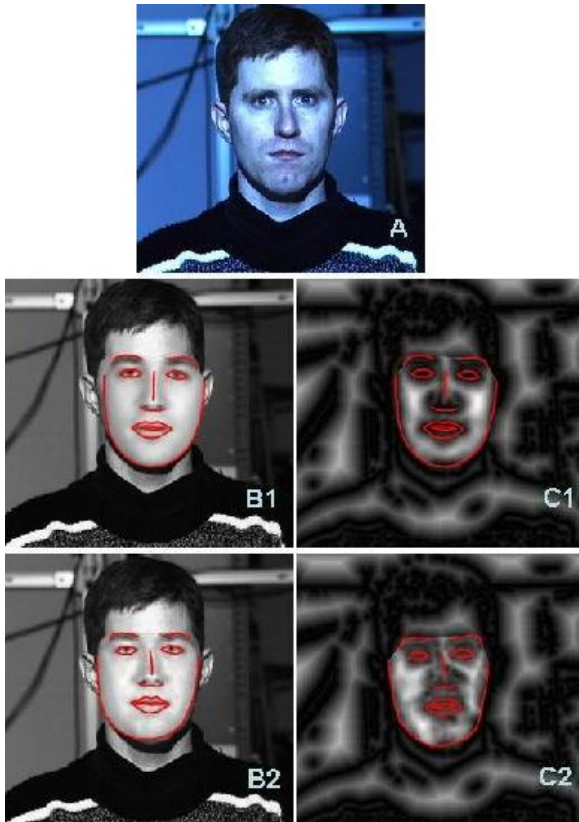


FIG. 6 – Résultat de convergence des AAM : A- Image Originale, B- Expérience standard, C- Expérience avec carte de distance. Sur les images du haut (B1 et C1) les initialisations du modèle sont affichées, et sur les images du bas (B2 et C2) les convergences du modèle sont affichées.

centrales se trouvent dans les directions d'illuminations numérotées de 10 à 12.

Ces courbes révèlent tout d'abord la robustesse des AAM face à la généralisation : malgré un apprentissage sur seulement 10 visages de M2VTS, les AAM peuvent détecter les éléments faciaux des 17 autres visages très différents (dans des illuminations de face : directions d'illumination de 7 à 15).

Ces courbes de convergence attestent de la robustesse du prétraitement des cartes de distance face aux variations d'illumination. La courbe de convergence de l'expérience avec le prétraitement se trouve toujours au-dessus de la courbe de convergence de l'expérience standard des AAM lorsque l'illumination est fortement de côté. Les résultats sont similaires pour les deux expériences lorsque l'illumination est frontale. De plus, contrairement à la courbe de convergence de l'expérience standard, la courbe de convergence de l'expérience avec prétraitement est toujours supérieure à 85%, ce qui atteste donc d'une plus grande robustesse à l'illumination. En effet, nous pouvons constater la baisse très nette de convergence des AAM standards dans les illuminations de côté (directions 1 à 6 et 16 à 21), tandis que la courbe de l'expérience avec le prétraitement des cartes de distance reste stable dans toutes les directions de l'illumination.



FIG. 7 – Résultats de convergences d'AAM obtenus dans l'utilisation normale des AAM (colonne de gauche) et dans l'utilisation avec le prétraitement des cartes de distance (colonne de droite). Quatre illuminations différentes sont ici représentées : illumination forte de droite (images du haut) à une illumination forte de gauche (images du bas).

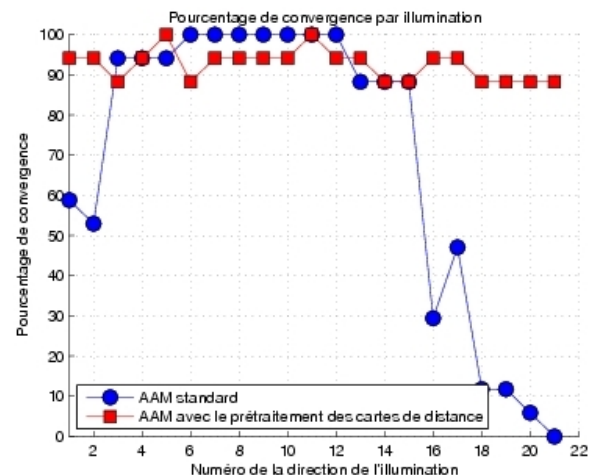


FIG. 8 – Pourcentage de convergence des AAM standards et des AAM avec le prétraitement des cartes de distance, par direction d'illumination sur les 17 images de PIE.

## 5 CONCLUSION

Nous avons décrit un nouveau prétraitement utile aux Modèles Actifs d'Apparence pour la détection des éléments faciaux sous des illuminations variées. La méthode utilise une approche basée contours dans laquelle l'information concernant les distances entre les contours des objets de la scène est contenue dans une image appelée "Carte de Distance". La contribution permet de tenir compte des relations de distance entre les contours d'une forme recherchée. Les tests ont montré la robustesse de cette méthode dans un contexte de généralisation forte. En effet, les tests fait sur les images de la base PIE du CMU, avec un modèle créé à partir de visages totalement différents (images de la base M2VTS), ont montré que lorsque le prétraitement des cartes de distance était appliqué, c'est-à-dire lorsque les images en niveaux de gris à l'entrée de la chaîne de traitement des AAM sont remplacées par leur carte de distance associée, la détection des éléments faciaux était beaucoup moins dépendante de la direction de l'éclairage par rapport à la méthode standard.

Actuellement nous travaillons sur une optimisation des Modèles Actifs d'Apparence avec les cartes de distance. Nous essayons aussi d'améliorer la finesse de détection des traits caractéristiques du visage en utilisant notre prétraitement.

## BIBLIOGRAPHIE

[Cootes, 1995] Cootes T. F., Cooper D., Taylor C. J. and Graham J. : Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding*, Vol.61, No.1, p.38-59.

[Cootes, 1998] Cootes T. F., Edwards G. J. and Taylor C. J. : Active Appearance Models. *Proceedings of ECCV'98, European Conference on Computer Vision*, Vol.2, p.484-498, Freiburg, Germany (1998).

[Cootes, 2004] Cootes T. F. and Taylor C. J. : Statistical Models of Appearance for Computer Vision. *Technical report from Imaging Science and Biomedical Engineering*, University of Manchester, [www.isbe.man.ac.uk](http://www.isbe.man.ac.uk) (2004).

[Hu, 2003] Hu C., Feris R. and Turk M. : Active wavelet networks for face alignment. *Proceedings of BMVC'03, British Machine Vision Conference*, Norwich, United Kingdom (2003).

[Huang, 2004] Huang Y., Lin S., Li S. Z., Lu H. and Shum H. Y. : Face Alignment Under Variable Illumination. *Proceedings of FGR'04, Automatic Face and Gesture Recognition*, Seoul, Korea (2004).

[Pigeon, 1996] Pigeon S. : M2VTS Database. [www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html](http://www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html).

[Sim, 2002] Sim T., Baker S. and Bsat M. : The CMU Pose, Illumination, and Expression (PIE) Database. *Proceedings of FGR'02, IEEE International Conference on Automatic Face and Gesture Recognition*, Washington D.C., U.S.A (2002).

[Stegmann, 2000] Stegmann M. B. : Active Appearance Models : Theory, Extensions and Cases. *Master thesis from Informatics and Mathematical Modelling*, Technical University of Denmark, DTU.

[Zuiderveld, 1994] Zuiderveld K. : Contrast Limited Adaptive Histogram Equalization. *Graphics Gems IV*, p.474-485.