

Cartes Orientées et Modèles Actifs d'Apparence

Renaud SÉGUIER, Sylvain LE GALLOU

SUPELEC / IETR / SCEE

Avenue de la Boulaie, CS 47601, 35576 Cesson-Sévigné Cedex, France

tel : 0299844500, fax : 0299844599, Renaud.Seguier@supelec.fr, Sylvain.LeGallou@supelec.fr

Résumé – Les Modèles Actifs d'Apparence (MAA) sont particulièrement performants en alignement de visages, c'est à dire pour localiser précisément les yeux, le nez et la bouche dans des images représentant un visage préalablement détecté. Nos travaux consistent à rendre ces modèles plus robustes aux variations lumineuses par le biais du prétraitement. Nous proposons de remplacer les textures de visage modélisées dans les MAA par des cartes d'orientations basées sur des textures égalisées de façon adaptative. Ces textures sont de plus prises en compte d'une manière adaptée au travers d'une métrique particulière dans les MAA. Des tests sur des bases publiques (CMU-PIE et BioId) illustrent le gain significatif en robustesse apporté par les cartes orientées à la fois dans des conditions difficiles et normales d'illumination.

Abstract – Active Appearance Models (AAM) are particularly efficient in face alignment which consist to localise precisely the eyes, nose and mouth in a face previously detected. Our work consist of making them more robust in real conditions. We propose a specific pretreatment robust to lighting variations for deformable models methods. It consists in replacing AAM input textures by oriented maps extracted from the original textures specifically equalized. In addition, we use an adapted metric to evaluate those new textures during the AAM convergence. Tests were conducted on public data bases (BioId and CMU-PIE) which shows that our proposition enhance AAM efficiency in standard and rough illumination conditions.

1 Introduction

Dans le cadre de la radio logicielle, nous compressons des vidéo-communications en nous adaptant au contenu. Nous localisons ainsi les yeux, le nez et la bouche d'un visage afin de le compresser "intelligemment" : les yeux seront mieux compressés que la peau du visage mais moins bien que la bouche et le fond sera fortement compressé. Nous utilisons pour ce faire des modèles actifs d'apparence [3] (MAA) qui permettent d'aligner des objets mais qui sont peu robustes aux changements d'illumination. Les MAA s'appuient sur la minimisation d'une erreur entre l'image de l'objet analysée et une image synthétisée par le modèle statistique de l'objet. L'efficacité de l'alignement dépend donc de la cohérence entre les conditions d'illumination des images de l'objet ayant servis à générer le modèle et celles de l'image dans laquelle nous analysons l'objet.

1.1 Robustesse à l'illumination

Les équipes visant à améliorer la robustesse en illumination des MAA empruntent les voix suivantes [14] :

Caractéristique invariante [13] [15] [8] [6]. La texture du modèle en niveau de gris est remplacée par une caractéristique invariante ou moins sensible à l'illumination comme le gradient de l'image ou des cartes de distance.

Mise en forme canonique [16] [9]. Les variations d'illumination sont normalisées en appliquant par exemple une égalisation d'histogramme [4]. On peut également synthétiser les images dans une forme "canonique" en évaluant les ombres sur l'objet. Ces dernières approches restent encore aujourd'hui fragiles, et peu robustes en conditions réelles.

Modélisation de la variation de lumière [7] [11] [10]. On tente d'apprendre les variations de la lumière en introduisant dans la base d'exemple des images différemment illuminées ou de construire un sous-espace de paramètres contrôlant les variations d'illumination. Ces paramètres viennent ensuite s'ajouter aux paramètres des MAA à optimiser. Cette approche n'est pas adaptée à notre problématique de temps-réel et d'images 2D.

1.2 Proposition

Nous proposons dans cet article un traitement particulier des textures et une métrique spécifique afin de rendre les MAA robustes aux conditions d'illumination. Ce prétraitement est basé sur une égalisation d'histogramme adaptative (Contrast Limited Adaptive Histogram Equalization -CLAHE-) [17] et des cartes orientées [5]. La métrique intègre l'un des paramètres du vecteur de pose (l'orientation de l'objet dans le plan perpendiculaire à la caméra). Cette métrique permet par ailleurs d'accélérer la vitesse d'exécution des MAA en évitant une normalisation en luminance de la texture de l'objet synthétisé à chaque itération durant l'optimisation des paramètres du modèle.

Après avoir brièvement présenté les Modèles Actifs d'Apparence (section 2), nous détaillerons notre proposition dans la section 3. Des tests (section 4) illustreront les performances de notre proposition sur des bases publiques et précéderont une rapide conclusion (section 5).

2 Modèle Actif d'Apparence

Proposés par Edwards, Cootes et Taylor [3], les Modèles Actifs d'Apparence (MAA) permettent de synthétiser une forme et une texture conjointement à l'aide de trois ACP. La modélisation consiste à appliquer, sur une base de données d'exemples de l'objet à modéliser, une Analyse en Composante Principales (ACP) sur les formes s_i (ensemble de points définissant l'objet i) et une autre sur les textures g_i normalisées dans la forme moyenne.

$$\begin{aligned} s_i &= \bar{s} + \Phi_s * b_s \\ g_i &= \bar{g} + \Phi_g * b_g \end{aligned} \quad (1)$$

\bar{s} et \bar{g} sont les formes et textures moyennes, Φ_s et Φ_g sont les matrices des vecteurs propres de chaque ACP et b_s et b_g sont des vecteurs représentant les paramètres de forme et texture. En appliquant une troisième ACP sur le vecteur $b \begin{bmatrix} b_s \\ b_g \end{bmatrix}$ nous obtenons :

$$b = \Phi * c \quad (2)$$

Avec ϕ la matrice des d_c vecteurs propres issus de la troisième ACP.

Le modèle de l'objet est finalement représenté par un vecteur d'apparence c . La modification de ce vecteur permet donc de générer une nouvelle forme et texture de l'objet. Lorsque l'on souhaite segmenter l'objet dans une image inconnue i , on positionne le modèle suivant un vecteur de pose $t = [\theta, scale, t_x, t_y]^t$. θ représente l'orientation de l'objet dans le plan perpendiculaire à la caméra, $scale$ un facteur d'agrandissement et t_x et t_y la position dans l'image de l'objet recherché. On ajuste alors pas à pas le modèle de l'objet en faisant varier les paramètres du vecteur d'apparence c , ce qui a pour effet de générer une nouvelle forme (x_m) et une nouvelle texture (g_m) normalisée en forme et luminance. Considérons la texture normalisée en forme et luminance (g_i) de l'image i originale se trouvant à l'intérieur de la forme x_m et notons δ_g la différence de pixels $g_i - g_m$. C'est cette différence que est minimisée à l'aide de matrice de régression [3], simplex [1] ou descente de gradient [12]. Au bout d'un certain nombre d'itérations l'erreur a convergé : le modèle se superpose à l'objet dans l'image i et épouse sa forme et sa texture.

3 Cartes orientées et Modèles Actifs d'apparence

3.1 Travaux antérieurs

Remplacer les textures prises en compte dans les MAA par des textures égalisées de façon spécifique (CLAHE [17]) permet, dans des conditions difficiles d'illumination, d'aligner efficacement les visages [6].

Les cartes orientées quand à elles n'ont jamais été utilisées en alignement de visage. Elles consistent à remplacer les niveaux de gris par l'orientation des contours de l'image. En détection de visage, [5] et [2] proposent de générer deux images à partir de l'image originale en niveaux de gris. Ces images représentent le sinus et le cosinus de l'angle du contour détecté en chaque pixel.

3.2 Cartes orientées

Après avoir égalisé (CLAHE) les images $I(x, y)$, nous proposons d'en extraire le gradient dans la direction des lignes et des colonnes en appliquant des filtres de Sobel G_x et G_y , afin d'évaluer un angle entre 0 et $\pi/2$ traduisant l'orientation du contour sur lequel se trouvent les pixels de l'image. Considérer le sinus et le cosinus des angles des contours en chaque pixel comme [5] et [2] reviendrait à multiplier par deux la taille des textures : il faut alors effectuer à chaque itération deux fois plus de warping. Ceci est incompatible avec la nécessité d'exécuter les MAA en temps réel. Pour cette raison, nous remplaçons tout simplement la valeur de chaque niveau de gris par la valeur de l'angle détecté en chaque pixel.

Sachant que l'orientation des pixels est quantifié sur $NbOrientation$ valeurs, nous évaluons cette orientation entre 0 et 2Π de la façon suivante.

$$I_3(x, y) = \frac{NbOrientation}{2} \cdot \left(1 + \frac{1}{\Pi} \cdot atan2 \left(\frac{G_y(x, y)}{G_x(x, y)} \right) \right) \quad (3)$$

La fonction $atan2$ étant la tangente inverse à quatre quadrants. La direction de l'angle n'est pas pertinente en ce qui nous concerne. A titre d'exemple, sur un fonds clair ou foncé les contours extérieurs du visage doivent avoir la même signature sur la carte d'orientation. Pour cette raison, nous réalisons un mapping de $[0..2\Pi]$ vers $[0..\frac{\Pi}{2}]$. Avant mapping, les pixels sont codés entre 0 et 2Π de sorte qu'il existe une discontinuité en 0 : les contours pratiquement verticaux génèrent des valeurs à la fois nulles et de forte intensité. Comme on peut le voir sur la figure 1 cette discontinuité est supprimée avec le mapping. Les contours proches de la verticale ont des valeurs faibles dans les cartes orientés et ceux proches de l'horizontale ont des valeurs élevées. Les contours horizontaux sont donc rehaussés par ce type de transformation (voir figure 2) alors que les contours verticaux peu nombreux à l'intérieur du visage se confondent avec les pixels appartenant à des plages uniformes.

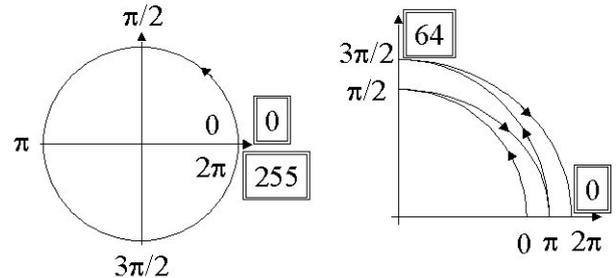


FIG. 1 – Mapping from $[0..2\Pi]$ to $[0..\frac{\Pi}{2}]$

En phase de segmentation, les MAA cherchent à positionner le modèle à l'aide du vecteur t et à mesurer une différence entre l'image source à l'intérieur de la forme proposée par le modèle et la texture du modèle. Dans les MAA classiques la texture segmentée dans l'image analysée est normalisée en forme et en luminance à chaque



FIG. 2 – Image source, carte orientée avant mapping et carte orientée après mapping

itération avant d'être comparée à la texture du modèle. Dans les applications temps réel, il est fondamental de limiter les calculs effectués à chaque itération. C'est cette opération de normalisation de l'image analysée qui est la plus coûteuse en temps de calcul : elle représente 80% du temps de traitement global.

Normalisation de la luminance. Cette normalisation n'est plus nécessaire dans notre cas puisque la dynamique de l'image a été normalisée par le CLAHE et que les cartes orientées traduisent un angle : un rapport entre les gradients de luminance. Il en résulte un gain en complexité non négligeable.

Normalisation de la forme. Lorsque l'objet est orienté d'un angle θ , l'algorithme cherche à positionner le modèle à l'aide du vecteur t et à mesurer une différence entre l'image source à l'intérieur de la forme proposée par le modèle et la texture du modèle. La différence entre ces deux textures se fait dans le référentiel du modèle : une forme normalisée c'est à dire d'orientation $\theta = 0$. Ceci ne pose pas de problème lorsque les textures sont constituées des niveaux de gris. Dans notre cas cependant, puisque nous avons remplacé les niveaux de gris par les orientations des contours portés par chaque pixel, il n'y a plus d'invariance en rotation. A titre d'exemple, considérons l'ellipse de la figure 3 avec un pixel P_{model} sur un contour de 45 degrés. Dans une carte orientée (Fig.3a), ce pixel, dans le référentiel du modèle, aura une valeur de 45 (si les niveaux de gris vont de 0 à 90). Si on recherche la même ellipse mais orientée de -45 degrés dans une image test (Fig.3b) : le pixel P_{image} correspondant sur l'objet aura une valeur de 0 puisque les filtres de Sobel travaillent dans les mêmes directions que l'objet soit ou non orienté. Après le warping de l'objet prenant en compte le paramètre de pose $\theta = -45$, la texture de cet objet ayant subi la rotation aura les mêmes niveaux de gris qu'avant rotation. Le pixel P correspondant dans le modèle ($P_{model} = 45$) sera comparé au pixel P de l'image ($P_{image} = 0$). Afin de rendre la texture du modèle comparable à celle de l'objet quelque soit son orientation dans l'image, nous retranchons simplement, avant de comparer la texture du modèle à celle de l'image analysée, un offset aux niveaux de gris de la texture de l'image correspondant à l'angle θ du paramètre de pose. On voit bien sur la figure 3c que cette opération permet de comparer les niveaux de gris du modèle et ceux de l'image analysée quelque soit l'orientation de l'objet dans l'image.

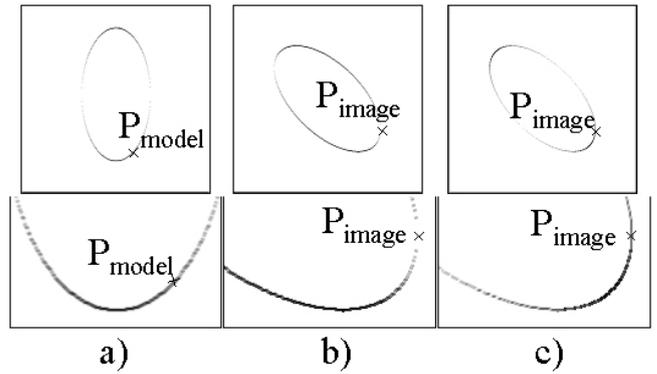


FIG. 3 – Modèle, objet dans l'image testée sans offset, objet dans l'image testée avec prise en compte de l'offset.

4 Tests

Nous illustrons notre proposition dans le cadre de l'alignement de visage. Nous utilisons dix visages de la base M2VTS pour modéliser le visage. Bien souvent un prétraitement permettant de rendre un algorithme robuste dans un cas particulier diminue son efficacité dans des cas différents [4]. Rendre les MAA plus robustes au changement de luminosité peut diminuer leurs performances dans des conditions normales. Pour cette raison nous allons vérifier que notre proposition rend les MAA plus efficaces sur une base où l'illumination est normale (1521 images de la base BioId) et sur une base dédiée entre autres aux problèmes d'illumination (66 visages de la base CMU-PIE pris sous 21 illuminations différentes soit 1386 images tests). Notons que les visages et conditions d'illuminations n'ont rien à voir entre les bases de test et d'apprentissage. Un visage est considéré comme ayant été aligné avec une erreur e si l'erreur moyenne sur quatre points de référence (centre des yeux, du nez et de la bouche) est inférieure à e . L'erreur a été normalisée par la distance entre les deux yeux de chaque visage testé. Les courbes des figures 4 et 5 ont pour abscisse l'erreur e et pour ordonnée le pourcentage de localisation correcte en deçà de cette erreur. Typiquement en dessous de 10% l'erreur n'est plus significative compte tenu de la précision avec laquelle la réalité terrain a été produite. Un bon point de comparaison se situe autour de 15%. Comme prévu, les cartes de distances [6] présentent un bon comportement lorsque l'illumination est difficile (voir Fig.4). Mais dans des conditions normales d'illumination (Fig.5) leur performances s'écroulent devant celles d'un MAA classique basé sur des images prétraitées par un CLAHE. On voit sur ces deux figures l'avantage significatif des cartes orientées : elles surpassent toutes les autres méthodes aussi bien dans des conditions normales que difficiles d'illumination. En particulier sur la base BioId assez difficile quand à la variabilité des visages, les cartes orientées présentent un gain de 25% (à 15% d'erreur) par rapport aux autres méthodes : elles permettent bien, même dans des conditions normales d'illumination, d'améliorer les performances des MAA.

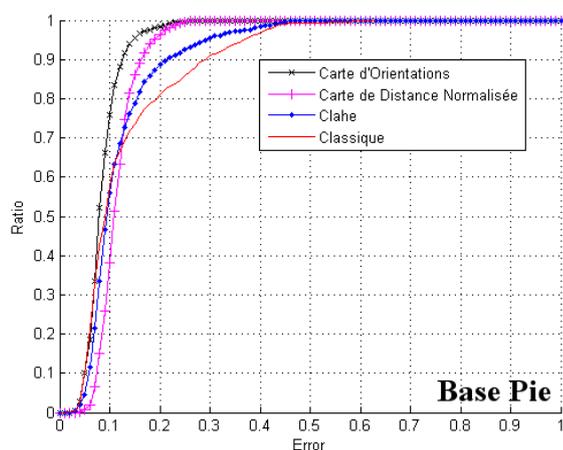


FIG. 4 – Conditions difficiles d'illumination.

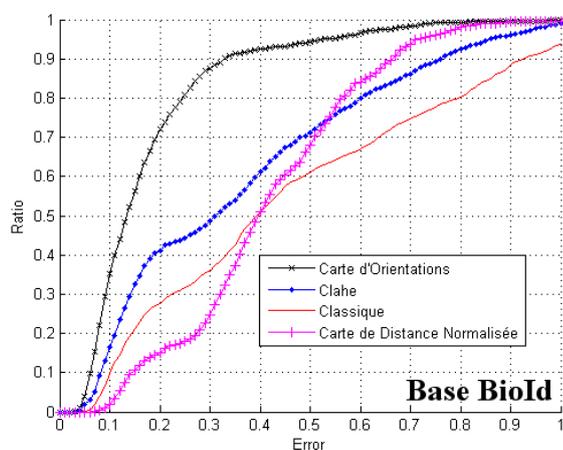


FIG. 5 – Conditions normales d'illumination.

5 Conclusion

Nous avons montré que remplacer les textures des images prises en compte dans les Modèles Actifs d'apparence par des cartes orientées évaluées sur des images égalisées de façon adaptative permettait d'améliorer notablement la robustesse des MAA dans des conditions normales et difficiles d'illumination. Par ailleurs notre proposition permet d'éviter une normalisation en luminance classiquement effectuée à chaque itération, améliorant ainsi le temps d'exécution. Un problème subsiste cependant : dans nos cartes orientées, des pixels appartenant à des zones uniformes ont la même signature que des pixels proches de contours verticaux. Ceci n'est pas très grave en alignement de visage parce que les caractéristiques du visage sur lesquels sont répartis la majorité des points constituant le modèle (yeux, nez, bouche) sont principalement constitués de contours horizontaux. Il est à craindre cependant que notre prétraitement ne soit pas aussi performant si l'objectif est d'aligner des objets supportant des contours horizontaux et verticaux.

Références

- [1] Y. Aïdarous, S. Le Gallou, A. Sattar et R. Segulier. *Face alignment using active appearance model optimized by simplex*. International Conference on Computer Vision Theory and Applications, 2007.
- [2] R. Belaroussi, L. Prevost et M. Milgram, *Combining model-based classifier for face localization*. IAPR Conference on Machine Vision Applications, 2005.
- [3] T.F. Cootes, G.J. Edwards et C.J. Taylor. *Active Appearance Models*. European Conference on Computer Vision, 1998.
- [4] B. Du, S. Shan, L. Qing et W. Gao. *Empirical Comparison of Several Preprocessing Methods for Illumination Insensitive Face Recognition*. ICASSP, 2005.
- [5] B. Froba et C. Kubleck. *Robust face detection at video frame rate on edge orientation features*. International Conference on Automatic Face and Gesture Recognition, 2002.
- [6] D. Giri, M. Rosenwald, B. Villeneuve, S. Le Gallou et R. Segulier. *Scale normalization for the distance maps AAM*. ICARCV, 2006.
- [7] A. Georghiades, P.N. Belhumeur et D.J. Kriegman. *From Few to Many : Generative Models for Recognition Under Variable Pose and Illumination*. FG'00, Automatic Face and Gesture Recognition, 2000.
- [8] C. Hu, R. Feris et M. Turk. *Active wavelet networks for face alignment*. BMVC'03, British Machine Vision Conference, 2003.
- [9] Y. Huang, S. Lin, S.Z. Li, H. Lu et H.Y. Shum. *Face Alignment Under Variable Illumination*. FGR'04, International Conference on Automatic Face and Gesture Recognition, 2004.
- [10] F. Kahraman et M.B. Stegmann. *Towards Illumination invariant Localization of Faces using Active Appearance Models*. NORSIG'06, Nordic Signal Processing Symposium, 2006.
- [11] C.M. Christoudias, L.P. Morency et T. Darrell. *Light Field Appearance Manifolds*. ECCV'04, European Conference on Computer Vision, 2004.
- [12] S. Baker, R. Gross, and I. Matthews. *Lucas-Kanade 20 Years On : A Unifying Framework : Part 3*. Tech. report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, November, 2003.
- [13] T. Riklin-Raviv et A. Shashua. *The Quotient Image : Class Based Recognition and Synthesis under Varying Illumination Conditions*. CVPR'99, Conference on Computer Vision and Pattern Recognition, 1999.
- [14] T. Sim et T. Kanade. *Combining Models and Exemplars for Face Recognition : An Illuminating Example*. CVPR 2001 Workshop on Models versus Exemplars in Computer Vision, 2001.
- [15] M.B. Stegmann et R. Larsen. *Multi-band modelling of appearance*. GMBV'02, Workshop on Generative Model-Based Vision, 2002.
- [16] W.Y. Zhao et R. Chellappa. *Illumination Insensitive Face Recognition Using Symmetric Shape-from-Shading*. CVPR'00, Conference on Computer Vision and Pattern Recognition, 2000.
- [17] K. Zuiderveld. *Contrast Limited Adaptive Histogram Equalization*. Graphics Gems IV, 1994.