

Modèles actifs d'apparences adaptés

Adapted active appearance models

**Renaud Séguier¹, Sylvain Le Gallou², Gaspard Breton²,
Christophe Garcia²**

¹ SUPÉLEC/IETR, Avenue de la Boulaie, 35511 Cesson-Sévigné

² Orange Labs – TECH/IRIS, 4 rue du clos courtel, 35 512 Cesson Sévigné

Manuscrit reçu le 21 décembre 2007

Résumé et mots clés

Les Modèles Actifs d'Apparence (MAA) sont efficaces lorsqu'il s'agit d'aligner (détecter les contours des yeux, du nez et de la bouche) des visages connus dans des espaces contraints (illumination et pose contrôlées). Nous proposons des Modèles Actifs d'Apparence Adaptés afin d'aligner des visages inconnus dans des poses et illuminations quelconques. Notre proposition repose d'une part sur une transformation des textures du modèle actif en carte d'orientation, ce qui impacte l'opération de normalisation des MAA; et d'autre part sur une recherche dans une banque de modèles pré-calculés du MAA le plus adapté au visage inconnu. Des tests sur des bases publiques et privées (Biold, CMU-PIE) montrent l'intérêt de notre approche: il devient possible d'aligner en temps réel des visages inconnus dans des situations où la lumière et la pose sont non contrôlées.

Modèle Actif d'apparence, Interface Homme-Machine, Analyse de visage.

Abstract and key words

Adapted Active Appearance Models

Active Appearance Models (AAM) are able to align known faces in an efficient manner when face pose and illumination are controlled. The AAM exploit a set of face examples in order to extract a statistical model. There is no difficulty to align a face with the same type (same morphology, illumination and pose) which constitute the example data set. Unfortunately, the AAM are less outstanding from the moment when the illumination, pose and face type changes. AAM robustness is link to the variability introduced in the learning base. The more the AAM will contain variability, the more it will be able to adapt itself to variable faces with the following drawback: the data represented in the reduced parameters space will then form different classes letting appear holes, regions without any data (see Fig. 1). It is therefore very difficult to make the AAM converge in this scattered space.

We propose in this paper a robust Active Appearance Models allowing a real-time implementation. To increase the AAM robustness to illumination changes, we propose Oriented Map AAM (OM-AAM). Adapted AAM will be presented after to increase the AAM robustness to any other types of variability (in identity, pose, expression etc.).

OM-AAM

We propose a specific transformation of the active model texture in an orientation map, which change the normalization process of the AAM. First of all, we apply systematically an adaptive histogram equalization from CLAHE [Zuiderweld 1994] on the images. We then evaluate the horizontal and vertical gradient and simply use the gradient angle on each pixel instead of its gray level. This angle is evaluated on N_a values. In practice we quantify it on eight bits so $N_a = 255$. The new texture is then made up from an image representing the orientation of each pixel, that we call an oriented map. To overcome a discontinuity problem between 0 and 2π associated to similar edge orientations, we

realize a mapping (Eq. 11) from $[0..2\Pi]$ to $[0..\frac{\Pi}{2}]$. In order to reduce the noise in uniform regions as illustrated in the background of Figure 3c, we emphasize the signal correlated with the region with high gradient information as it is proposed by [Cooste 2001] in using a non-linear function f (see Eq. 12). During the modelization, the oriented texture from image I_4 (Eq. 13) will replace the texture usually used by the AAM.

We compare the OM-AAM performance to those of the DM-AAM and classical AAM on two public databases. The first one is dedicated to illumination problems (CMU-PIE: 1386 images of 66 faces under 21 different illuminations [SIM 2002]) and the other one is composed of different faces with several expressions taken in different backgrounds (BIOD: 1521 images [Res01]) under variable light exposition (see Fig. 5). These comparisons will be made in a generalization context: the faces used to construct the model (18 persons from the M2VTS database [Pigeon 1996]) do not belong to the testing databases.

We normalize the error found on four relevant points (gravity centers of the eyes, nose and mouth) by the distance between the eyes (see Eq. 15). Figure 6 represents the percentage of the images which have been aligned with the error e . On the CMU-PIE database, the OM-AAM are able to align 94% of the faces with a precision less or equal to 15% where the DM-AAM and classical ones are less efficient: their performances are of 88% and 79% respectively. But when the faces are acquired in real situations, our proposition overcome other methods: on BIOD database, the OM-AAM can align 52% of the faces with a precision less or equal to 15% which represents a gain of 27 and 42% in the performance relatively to classical AAM and DM ones respectively.

Adapted AAM

As said above, the AAM robustness is related to the face variability in the learning base. Instead of using a very generic model containing a lot of variability, we propose to use an initial model M_0 which contains only a variability in identity and then use a specific model M_{adapt} containing variability in pose and expression.

Let a general database contains three types of variability: in expression, identity and pose (see Fig. 7). It is made of several different faces, holding four distinct expressions. Each of the face presents each expression in five poses. The initial model M_0 is realized from a database BDD_0 containing different frontal faces in neutral expression (see Fig. 8). This initial model will be used to perform a rough alignment on the unknown face. Let C_0 be the appearance vector after the alignment of the model M_0 on the unknown analyzed face. In the reduced space of the model parameters, we seek for the k nearest parameters vectors of C_0 belonging to the learning initial database BDD_0 . These k nearest neighbors correspond to the k nearest face of the analyzed one. For example in figure 9, the vector C_p will identify the face number p as the most resemble to the analyzed one. From this set of k nearest identities, we generate an adapted database BDD_{adapt} containing the corresponding faces in different expressions and poses. BDD_{adapt} is a subset of the general database (Fig. 7) from which we generate the adapted model M_{adapt} . When $k = 1, 2$ or 3 it is possible to evaluate beforehand the adapted model. When we need to align an unknown face in a static image, we then simply align the face with the initial model M_0 and apply the pre-computed model which corresponds to the k nearest faces.

We test the adapted AAM on the static images of the general database BDD_0 (Fig. 7). A test sequence is then constituted of one unknown person showing four expressions under five different poses, the learning base associated to this testing base being constituted of all the other persons. A cross-validation of type *Leave-one-out* is used: every face is tested separately, using all the other one for the learning base. All the faces of the database have been tested, representing at the end a set of 580 images.

We compare the performance of our system when $k = 2$ ("Adapted AAM") to the two other different AAM. The first one ("AAM 28") get identity as the only variability and is constructed from the 28 faces (the twenty-ninth being tested) in frontal view and neutral expression. The second one ("AAM 560") is full variability since it is based on 560 images representing 28 faces showing four expressions under five different poses. Figure 11 shows the superiority of the "Adapted AAM" on the two other models. If we look at the performances at the reference error of 15% our proposition is ten times more rapid and 16% more effective than a model constructed from a more rich database ("AAM 560"). If now we compare the "Adapted AAM" to the "AAM 28" which has the same computational complexity, it is more effective in 55% of the cases (94% versus 49%).

As a conclusion our model is more rapid and effective compared to the other models because it focuses on a relevant database relatively to the testing face.

Active Appearance Model, Human Machine Interface, Face analysis.

1. Introduction

Les applications liées à l'analyse et à la synthèse de visage (interaction homme-machine, compression en vidéo-communication, réalité augmentée) nécessitent toutes de détecter puis d'aligner le visage de l'utilisateur. Cette dernière opération consiste à localiser précisément le centre de gravité des yeux, du nez et de la bouche pour les traitements ultérieurs. L'opération de détection peut être réalisée à présent en temps réel et de façon assez efficace [Garcia 2004] [Viola 2004]; le verrou technologique réside désormais dans l'alignement du visage en conditions réelles; c'est précisément l'objet de cet article.

Depuis les Modèles Actifs d'Apparence (MAA) de [Cootes 1998] il devient possible d'aligner des visages en temps réel. Les MAA exploitent une base d'exemples de visages afin d'en extraire un modèle statistique. Lorsqu'il s'agit d'aligner un visage inconnu dans une image, il faut adapter les paramètres du modèle afin qu'il corresponde au mieux aux caractéristiques du visage analysé. Aligner un visage du même type (même morphologie, illumination et pose) que ceux contenus dans la base d'exemples ne pose pas de problèmes. Malheureusement, ces modèles actifs sont beaucoup moins performants dès lors que l'illumination, la pose ou le type de visage changent. Nous proposons dans cet article des Modèles Actifs d'Apparence robustes et implémentables en temps réel. Dans la section suivante nous nous attacherons à dresser un état de l'art des différentes techniques visant à rendre les MAA robustes. Nous verrons qu'aucune d'entre elles n'adresse à la fois les trois types de robustesses qui nous intéressent: pose, illumination et identité. Après avoir rapidement introduit les Modèles Actifs d'Apparence (section 3), nous proposerons dans la section 4 un traitement particulier des textures prises en compte dans les MAA ainsi qu'une nouvelle normalisation afin de les rendre robustes à l'illumination. Dans la section 5, nous détaillerons notre proposition consistant à rendre les MAA robustes à la pose et à l'identité avant de conclure en proposant de nouvelles pistes de recherche dans la dernière section.

2. État de l'art

Nous proposons de classer les méthodes qui visent à rendre les modèles déformables plus robustes de la façon suivante. En italique est indiqué le type de robustesse adressé.

- Prétraitement
 - Caractéristique invariante (*illumination*)
 - Mise en forme canonique (*illumination*)
- Extension de l'espace des paramètres
 - Modélisation de la lumière (*illumination*)
 - Modélisation 3D (*pose*)
- Extension du nombre de modèles
 - Spécification non-automatique des classes (*pose/expression*)
 - Spécification automatique des classes (*pose/expression*)
- Spécialisation de la base d'apprentissage
 - Approche hiérarchique (*pose/expression*)
 - Spécification de l'identité (*identité*)

Les méthodes dites de **Prétraitement** cherchent à remplacer l'image de texture en entrée de l'algorithme par une image pré-traitée. Cette approche envisage seulement de minimiser la dépendance des modèles d'apparence déformable aux variations de l'illumination. L'axe **Caractéristique invariante** consiste à utiliser une caractéristique de l'image qui est invariante ou moins sensible à la variation de l'illumination comme le gradient de l'image [Cootes 2001], des détecteurs de motifs particuliers comme les coins pour les yeux ou la bouche [Scott 2003], la concaténation de plusieurs composantes de couleur (H et S du code HSV et gradient par exemple) [Stegmann 2002], des réseaux d'ondelettes [Hu 2003a] ou des cartes de distances [Giri *et al.* 2006]. Hormis cette dernière, ces méthodes ont le désavantage d'augmenter la complexité de l'algorithme en concaténant les différentes caractéristiques invariantes, ce qui a pour effet d'augmenter la taille de la texture traitée par les MAA. Dans l'axe **Mise en forme canonique**, on tente de «normaliser» les variations d'illumination [Zhu 2003] ou de les atténuer [Huang 2004]. On est également amené à évaluer les ombres [Zhao 2000] afin de retrouver le modèle 3D du visage et de reproduire une texture sans effets d'ombrage. Toutes ces approches restent encore aujourd'hui fragiles.

Les méthodes dites de **l'Extension de l'espace des paramètres** tentent d'augmenter le nombre de paramètres du modèle déformable à optimiser afin de modéliser la variabilité introduite dans l'espace des paramètres. Dans l'axe **Modélisation de la lumière**, on apprend et construit un sous-espace de paramètres contrôlant les variations d'illumination. On est amené à proposer une modélisation par cônes d'illumination [Georghiades 2000] [Lee 2005] ou champs lumineux [Gross 2002] [Christoudias 2004]. On peut également estimer la direction de l'illumination en construisant une base d'apprentissage de visages pris chacun sous un certain nombre d'illuminations dif-

férentes créées par la variation de position d'une seule source de lumière [Kittipanya-ngam 2006]. Les variations de l'illumination sont ensuite modélisées par une analyse en composantes principales, partie intégrante des MAA. Toutes ces méthodes complexifient notablement la mise en œuvre des MAA en augmentant le nombre de paramètres à optimiser et en morcelant l'espace des paramètres. L'optimisation, conduite dans un espace des paramètres de plus grande dimension et non compact, est alors plus difficile à mener. Dans l'axe **Modélisation 3D**, on déplace la variabilité en pose de l'ensemble de l'espace des paramètres d'apparence au sous-espace des paramètres contrôlant la pose. [Romdhani 1999] introduit un nouveau paramètre à optimiser en exploitant l'information de pose de chaque visage constituant la base d'apprentissage. On peut également utiliser un MAA 3D soit à partir de formes et textures acquises à l'aide d'un scanner [Banz 1999], soit à l'aide de vues 2D de face et de profil de chaque visage de la base d'apprentissage [Li 2001] [Xiao 2004] [Sattar 2007]. Ces techniques de modélisation 3D du visage sont très pertinentes pour réduire la fragilité des MAA en variabilité de pose. Cependant, le modèle 3D étant plus conséquent que le modèle 2D, une implémentation temps réel reste difficile.

Les méthodes dites d'**Extension du nombre de modèles** consistent à spécifier les classes formées dans l'espace réduit des paramètres du modèle et de faire converger un modèle déformable dans chacune de ces classes spécifiques. Dans l'axe **Spécification non-automatique des classes** on spécifie dès le départ la variabilité de la base d'apprentissage à traiter, les classes composant l'espace réduit sont connues : les différentes vues du modèle pour la variabilité en pose [Cootes 2000a] [Cootes 2000b] [Li 2002] [Hu 2003b] et les différentes expressions pour la variabilité en expressions [FENG 2006]. Dans l'axe **Spécification automatique des classes** on propose de trouver automatiquement les classes existantes dans l'espace réduit ; il devient ainsi possible de faire converger les modèles en se restreignant aux classes apprises. Pour trouver automatiquement les classes auxquelles seront associés les modèles actifs, on utilise un K-means [Chang 2004] ou des mixtures de gaussiennes [Cootes 1999] [Hu 2004] [Christoudias 2005]. Dans toutes ces méthodes on multiplie le nombre de modèles actifs. Il faudra donc en parallèle optimiser l'ensemble de ces modèles afin de décider lequel correspond le mieux au visage analysé. Ceci est irréalisable en temps réel dans notre contexte applicatif.

Les méthodes dites de **Spécialisation de la base d'apprentissage** permettent de restreindre l'espace de recherche des paramètres à des espaces spécifiques contenant la seule variabilité de chaque partie du visage ou d'une seule identité. L'axe **Approche hiérarchique** regroupe les méthodes divisant la recherche des éléments faciaux en 2 étapes : la recherche approximative des points clés du visage, et la recherche précise des éléments faciaux par modèle spécialisé de chaque élément facial (yeux, nez et bouche). L'avantage de cette solution est de prendre en compte seulement les variabilités extrinsèques du visage pour la première étape puis seulement les variabilités

intrinsèques pour la seconde étape [Zalewski 2005] [Xu 2005] [Langs 2005] [Tong 2006]. Comme dans les méthodes précédentes, ces approches consistent à multiplier le nombre de modèles à optimiser, ce qui alourdit considérablement l'exécution des MAA. Dans l'axe **Spécification de l'identité**, on élimine de la base la variabilité extrinsèque (identité) partant du constat [Gross 2005] qu'un MAA générique (comprenant de la variabilité en identité, pose, expression et en illumination) est moins performant qu'un MAA spécifique à une identité dans sa variabilité en pose, illumination et expression. [Canzler 2004] propose de s'adapter en ligne à l'identité de la personne en construisant un modèle MAA 3D à partir d'une première vue de face et sans expression de la personne. Cette méthode reste très fragile, du fait de la construction 3D du visage à partir d'une seule vue 2D.

Aucune de ces méthodes ne répond pleinement à notre problématique car aucune d'entre elles ne considère dans son ensemble le suivi des éléments faciaux sur des images 2D de visages inconnus dans des poses et illuminations variées. Rappelons que notre objectif est de conserver les bonnes performances d'exécution des MAA tout en améliorant leur robustesse. À ce titre, nous nous sommes inspirés de l'axe **Caractéristique invariante**, en ce qui concerne la robustesse à l'illumination, dans lequel la texture prise en compte dans les MAA est remplacée par une texture prétraitée. Nous proposerons une technique qui n'augmente pas la complexité des MAA (section 4). Relativement à la robustesse en pose et identité, partant des travaux issus de l'axe **Spécification de l'identité**, nous proposons d'adapter le modèle actif à la personne à partir de modèles pré-calculés (section 5).

3. Modèles Actifs d'Apparence

3.1. Modélisation

Les Modèles Actifs d'Apparence créent un modèle conjoint de la forme et de la texture d'un objet à partir d'une base d'images I_i de cet objet. La texture contenue dans la forme s_i est normalisée en forme (par un warping dans la forme moyenne) et en luminance (en prenant en compte la moyenne et la variance des niveaux de gris de la texture) et conduit à une texture g_i . Deux Analyses en Composantes Principales (ACP) sont réalisées sur les exemples de formes et de textures de la base (voir Eq.1).

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

Avec \bar{s} et \bar{g} les formes et textures moyennes, Φ_s et Φ_g les matrices constituées des vecteurs propres et b_{si} et b_{gi} les para-

mètres de projection de la forme s_i et de la texture g_i dans les sous-espaces définis par les deux ACP. Une troisième ACP est conduite sur les vecteurs $b_i = [b_{si}|b_{gi}]$ (voir Eq.2).

$$b_i = \Phi * c_i \quad (2)$$

Avec Φ la matrice des vecteurs propres et c_i le vecteur d'apparence. À partir d'un vecteur d'apparence c , il est donc possible de synthétiser une image de l'objet. Les ACP, parce qu'elles associent une valeur propre à chaque vecteur propre, permettent de ne conserver qu'un certain nombre de vecteurs propres parmi les plus significatifs. Dans les faits, nous gardons 99% de la variance expliquée pour limiter le nombre de composantes du vecteur c .

3.2. Segmentation

En phase de segmentation, lorsque l'on souhaite aligner l'objet contenu dans une nouvelle image, on positionne le modèle défini par le vecteur c dans l'image à l'aide d'un vecteur de pose t :

$$t = [\theta, S, t_x, t_y]^t \quad (3)$$

θ est la rotation appliquée au modèle dans le plan de l'image, S le facteur d'échelle et t_x et t_y les coordonnées du centre de gravité du modèle dans l'image. On ajuste pas à pas les composantes du vecteur c en générant à chaque itération une nouvelle forme x_m et une nouvelle texture g_m normalisée en forme et luminance. Considérons à présent la texture g_{iraw} associée à la région de l'image I_i contenue dans la forme x_m . On « warpe » cette texture dans la forme moyenne \bar{s} (Eq.1) au moyen d'une fonction W (Eq.4) et on applique une normalisation photométrique (Eq.5) en utilisant la moyenne $\overline{g_{iraw/\bar{s}}}$ et la variance $\sigma(g_{iraw/\bar{s}})$ de la texture « warpée » $g_{iraw/\bar{s}}$. L'erreur résiduelle δ entre la texture g_i contenue dans l'image et la texture g_m générée par le modèle est alors minimisée en agissant sur les paramètres du modèle à l'aide du Jacobian pré-calculé liant l'erreur aux variations des paramètres d'apparence [Cootes 1998], ou en appliquant des techniques d'optimisation plus classiques telles que le simplex [Aidarous 2007] ou la descente de gradient [Stegmann 2000].

$$g_{iraw/\bar{s}} = W(g_{iraw/s_i}, c) \quad (4)$$

$$g_i = \frac{g_{iraw/\bar{s}} - \overline{g_{iraw/\bar{s}}}}{\sigma(g_{iraw/\bar{s}})} \quad (5)$$

$$\delta = g_i - g_m \text{ avec } \delta = [\delta_1 \dots \delta_i \dots \delta_N]^t; \quad (6)$$

Avec N le nombre de pixels dans la texture. Après un certain nombre d'itérations (typiquement une centaine), l'erreur e_{pix} (Eq.7) converge vers une valeur assez faible : le modèle s'est aligné sur l'objet dans l'image I_i et propose une estimation de sa

forme et de sa texture. Toutes ces étapes sont résumées dans l'Algorithme 1.

$$e_{pix} = \frac{1}{N} \sum_{i=1}^N \sqrt{\delta_i^2} \quad (7)$$

Algorithme 1. MAA classiques : segmentation.

1. Acquisition de l'image
2. Optimisation. Répéter de façon itérative (a) à (e)
 - (a) À partir du modèle, générer la forme x_m et la texture g_m
 - (b) Récupérer la texture non normalisée g_{iraw} dans l'image
 - (c) Normaliser g_{iraw} pour produire g_i :
 - i. « Warper » g_{iraw} dans la forme moyenne (Eq.4)
 - ii. Normaliser la luminance de g_{iraw} (Eq.5)
 - (d) Évaluer l'erreur $g_i - g_m$ (Eq.6)
 - (e) Adapter les paramètres du modèle

$$N_{optim} \approx N \left(\frac{8}{3} N_c + 12 \right) + N_{PtsCles} (2N_c + 17) + 4N_c^2 \quad (8)$$

Le nombre d'opérations N_{optim} nécessaires à l'optimisation (voir Eq.8) est évalué à partir du nombre de pixels N contenus dans la texture, des N_c composantes du vecteur d'apparence c et des $N_{PtsCles}$ points clés annotés sur le visage et qui constituent le vecteur de forme. N_{optim} ne prend pas en compte le coût algorithmique du warping (Alg.1.c.i) : celui-ci est réalisé directement sur la carte graphique et consomme 50% du temps de calcul (l'implémenter sur le cpu revient à multiplier d'un facteur cent le temps de traitement).

3.3. Robustesse

La robustesse des MAA est donc liée à la variabilité introduite dans la base d'apprentissage. Plus celle-ci contient de variabilité, plus le modèle déformable pourra s'adapter à la variabilité des visages. Malheureusement, il est impossible de faire converger un modèle déformable créé à partir d'une base d'apprentissage contenant beaucoup de variabilité. En effet, plus la variabilité de la base d'apprentissage est grande, plus les données représentées dans l'espace des paramètres du modèle déformable forment des « classes » laissant apparaître des « trous », des zones sans données. Il est alors très difficile de faire converger l'algorithme dans cet espace morcelé. La Figure 1 illustre ce

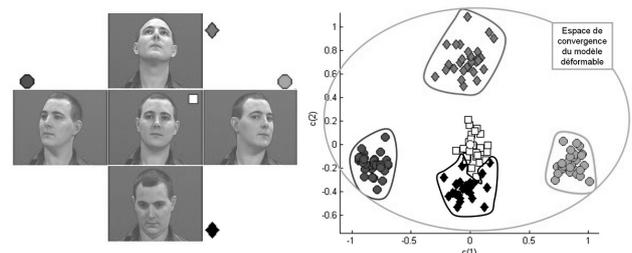


Figure 1. Espace réduit morcelé après ACP.

problème. La base d'apprentissage est constituée d'une trentaine de visages dans 5 poses différentes. La projection de ces exemples sur les deux premiers paramètres d'apparence indique clairement l'existence de 4 regroupements (seuls les visages de face et orientés vers le bas ne forment pas des régions disjointes dans le plan défini par les deux premières composantes du vecteur d'apparence), chacun correspondant à une pose bien particulière. La variété sous-jacente dans l'espace de départ est bien ici morcelée.

4. MAA robustes à l'illumination

4.1. MAA et cartes orientées

Des comparaisons empiriques faites en reconnaissance de visage [Du 2005] montrent que parmi les méthodes transformant une image I en $I' = T(I)$ (incluant donc les approches dites de « caractéristique invariante » et de « forme canonique », voir section 2), ce sont les égalisations d'histogramme (uniforme ou spécifique) qui conduisent aux meilleurs taux de reconnaissance. C'est pourquoi nous proposons d'appliquer systématiquement sur toutes les images traitées une égalisation adaptative de Clahe [Zuiderveld 1994]. Elle consiste à découper l'image en soixante-quatre blocs (huit par huit) et à réaliser à l'intérieur de chacun de ces blocs une égalisation spécifique d'histogramme (respectant la distribution de Rayleigh). À chaque bloc correspond donc une fonction particulière d'égalisation. Afin d'éliminer les effets de bords entre blocs, la valeur finale en chaque pixel est le résultat d'une interpolation bilinéaire prenant en compte les valeurs des fonctions d'égalisation des quatre blocs voisins appliquées sur le pixel considéré.

$$I_1(x, y) = CLAHE(I_0(x, y)) \tag{9}$$

Une étude comparée ([Cristinacce 2003]) des détecteurs de visages de Viola et Jones [Viola 2004] et Froba [Froba 2002] indique que leurs performances sont comparables lorsque le fond de l'image est uniforme, sachant que le premier détecteur est plus efficace en présence de fonds complexes mais plus lourd à implémenter. Dans notre cadre applicatif, un détecteur de visage localise grossièrement le centre du visage à aligner, de sorte que le fond ne perturbe pas trop l'alignement. Pour cette raison, nous nous sommes inspiré du prétraitement mis en œuvre par [Froba 2002] et [Belaroussi 2005] qui utilisent le sinus et le cosinus de l'angle détecté en chaque pixel de l'image et de ceux de [Cootes 2001] qui exploitent le gradient horizontal et vertical de l'image. Nous proposons de remplacer la valeur des pixels qui constituent les textures par l'angle du contour évalué en chaque pixel.

Cet angle sera calculé sur N_{angle} niveaux. En pratique nous le quantifions sur huit bits, soit donc $N_{angle} = 255$. En dessous d'une quantification sur 6 bits, les résultats se détériorent. Une carte de contours orientés va donc remplacer la texture de l'image traitée par les MAA.

Si G_x et G_y représentent les images gradients (horizontaux et verticaux) évalués sur l'image I_1 alors la carte d'orientation dont les valeurs évoluent entre 0 et 2π est calculée de la façon suivante.

$$I_2(x, y) = \frac{N_{angle}}{2} \cdot \left(1 + \frac{1}{\pi} \cdot atan2 \left(\frac{G_y(x, y)}{G_x(x, y)} \right) \right) \tag{10}$$

Avec $atan2$ la tangente inverse à quatre quadrants. Comme on peut le constater sur la figure 2, il y a une discontinuité entre 0 et 2π : des contours approximativement verticaux génèrent des valeurs à la fois élevées et très faibles. On observe l'effet de cette discontinuité sur le contour droit du visage contenu dans la figure 3b qui oscille entre le noir (partie supérieure du visage) et le blanc (partie inférieure).

Nous proposons d'appliquer un mapping (Eq.11) de $[0 \dots 2\pi]$ vers $[0 \dots \frac{\pi}{2}]$. Avec $mod_{\frac{N_{angle}}{2}}$ l'opération modulo $\frac{N_{angle}}{2}$ et abs la valeur absolue.

$$I_3(x, y) = \frac{N_{angle}}{4} - abs \left(mod_{\frac{N_{angle}}{2}} (I_2(x, y)) - \frac{N_{angle}}{4} \right) \tag{11}$$

Comme on le voit sur la figure 2, après mapping, les contours proches de l'horizontale (angle du gradient du contour proche

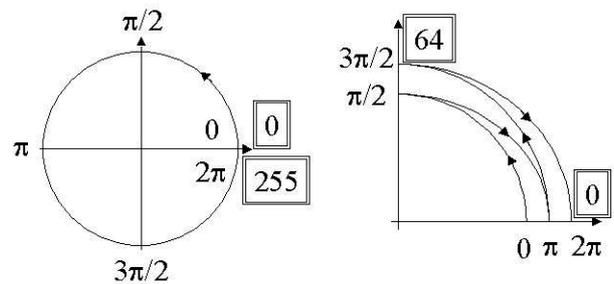


Figure 2. Mapping de $[0 \dots 2\pi]$ à $[0 \dots \frac{\pi}{2}]$.

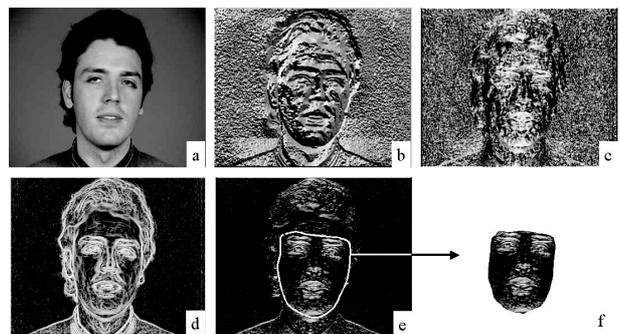


Figure 3. a: I_0 , b: I_2 , c: I_3 , d: $f(G)$, e: I_4 , f: texture orientée.

de $\frac{\pi}{2}$ ou $\frac{3\pi}{2}$) auront des valeurs élevées, ceux proches de la verticale (angle du gradient nul ou proche de π ou 2π) des valeurs très faibles.

Pour réduire le bruit particulièrement présent dans les régions uniformes (voir Fig.3c), nous proposons d'accentuer le signal contenu dans les zones à fort gradient comme cela est proposé dans [Cootes 2001] en appliquant la fonction non linéaire f à l'image :

$$f(G) = \frac{G}{G + \overline{G}} \text{ avec } G = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (12)$$

\overline{G} étant la moyenne de l'image G (voir Fig.3d).

$$I_4(x,y) = f(G) \cdot *I_3(x,y) \quad (13)$$

Avec $*$ l'opérateur produit élément par élément. Durant la modélisation, les textures *orientées* des images I_4 remplaceront les textures classiques exploitées par les MAA.

En phase de segmentation, on évalue la différence entre l'image synthétisée par le modèle et la texture analysée dans l'image (voir Fig.3f). Dans les MAA classiques, la texture est normalisée en forme et en luminance avant comparaison. Cette normalisation en luminance n'est plus nécessaire ici puisque les niveaux de la texture représentent à présent un angle et non plus une intensité lumineuse. Si l'objet à analyser dans l'image est orienté d'un angle θ , on positionne le modèle relativement au vecteur t (Eq.3) et on évalue une différence entre la texture de l'image contenue dans la forme générée et positionnée par le modèle et la texture générée par ce même modèle. Cette comparaison se fait dans le référentiel du modèle : une forme d'orientation $\theta = 0$. Cela ne pose pas de problème lorsque la texture représente des niveaux de gris. Dans notre cas, puisque nous avons remplacé les niveaux de gris par l'angle des contours, il n'y a plus d'invariance en rotation.

À titre d'exemple, considérons l'ellipse de la figure 4 avec un pixel P_{model} sur un contour de 45 degrés. Dans une carte orientée (Fig.4a), 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.4b), le pixel P_{image} correspondant sur l'objet aura une valeur de 0 puisque les filtres à partir desquels on extrait les gradients 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 quelle que soit son orientation dans l'image, nous retranchons simplement un offset (Eq.14) aux valeurs de la texture avant de la comparer à celle de l'image analysée.

$$Offset = floor(N_{angle} * \theta / (2 * \pi)); \quad (14)$$

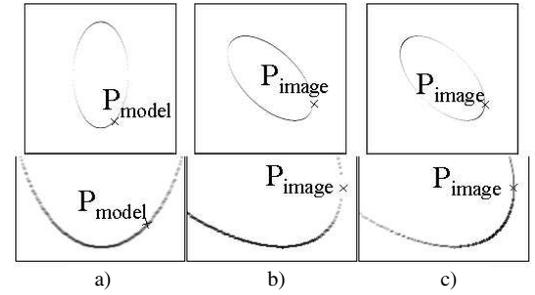


Figure 4. a : Modèle, b : Objet dans l'image testée sans offset, c : Objet dans l'image testée avec prise en compte de l'offset.

On voit bien sur la figure 4c que cette opération permet de comparer les niveaux de gris du modèle et ceux de l'image analysée quelle que soit l'orientation de l'objet dans l'image.

Pour soustraire la valeur de l'offset, il est nécessaire de conserver les valeurs originales des angles des contours évalués dans l'image. Nous proposons d'évaluer, en phase de segmentation, les cartes d'orientation entre $[0 \dots 2\pi]$ durant une étape de pré-traitement (Alg.2.2.b) puis de réaliser le mapping à chaque itération de la phase d'optimisation (Alg.2.3.c.ii) et le produit (Alg.2.3.c.iii) par la fonction non linéaire f calculée au préalable (Alg.2.2.c). La nouvelle segmentation proposée est résumée dans l'Algorithme 2 suivant.

Algorithme 2. MAA orientés : segmentation

1. Acquisition de l'image
2. Pré-traitement
 - (a) Égalisation d'histogramme (CLAHE) (Eq.9).
 - (b) Création de la carte orientée : les angles évoluent de 0 à 2π (Eq.10).
 - (c) Calcul de la fonction non-linéaire $f(G)$ (Eq.12)
3. Optimisation. Répéter de façon itérative (a) à (e)
 - (a) À partir du modèle, générer la forme x_m et la texture g_m
 - (b) Récupérer la texture non normalisée g_{iraw} dans l'image
 - (c) Normaliser g_{iraw} pour produire g_i :
 - i. Ajouter l'offset à la texture (Eq.14).
 - ii. «Mapper» la carte d'orientation de $[0 \dots 2\pi]$ à $[0 \dots \frac{\pi}{2}]$ (Eq.11).
 - iii. Multiplier chaque valeur de la carte d'orientation par la fonction non-linéaire évaluée à l'étape 2.c
 - iv. «Warper» la nouvelle texture dans la forme moyenne pour produire g_i
 - (d) Évaluer l'erreur $g_i - g_m$
 - (e) Adapter les paramètres du modèle

Le surcoût algorithmique généré par les contours orientés (étapes 3.c.i à 3.c.iii) est de l'ordre de $9N$ opérations. Sachant que dans des conditions réelles d'utilisation, nous exploitons une texture de $N = 1756$ pixels annotés sur $N_{ptsCles} = 68$ points pour une dizaine de paramètres du vecteur d'apparence ($N_c = 10$, voir Eq.8); le surcoût de l'optimisation est de l'ordre de 11 % compte tenu du fait que le warping consomme la moitié du temps de calcul. Sur nos implémentations, nous observons effectivement une augmentation du même ordre (13.5 %

pour être précis) lorsque nous comparons le temps total de traitement de l'algorithme classique à celui de notre proposition, étape de prétraitement comprise (Alg.2.2).

4.2. Performance des MAA orientés

Rappelons que notre objectif est d'améliorer la robustesse des MAA sans augmenter leur complexité. À ce titre, seuls les DM-AAM de [Giri *et al.* 2006] respectent notre cahier des charges. Nous nous comparerons donc à eux ainsi qu'aux MAA classiques à titre de référence. Nous utiliserons la même métrique que [Cristinacce 2006] pour évaluer l'erreur :

$$e = \frac{1}{M \cdot D_{eye}} \sum_{j=1}^M e_j \tag{15}$$

Où e_j est l'erreur commise sur l'un des $M = 4$ points représentant les centres des yeux, du nez et de la bouche et D_{eye} est la distance entre les deux yeux du visage analysé. Dans le cadre de la robustesse à l'illumination, la pose et l'identité, ces quatre points sont suffisants pour illustrer les performances de notre algorithme. La précision de la vérité terrain est de l'ordre de 10% de la distance entre les deux yeux des visages annotés ; au delà de $e = 25\%$, soit un quart de D_{eye} , nous considérerons que l'algorithme a divergé. Pour cette raison nous évaluerons l'erreur e entre 10 et 25%. Les MAA sont paramétrés de la façon suivante : une texture de 1756 pixels associée à un modèle de forme de 68 points clés et une variance expliquée de 99%. Concernant les cartes orientées, aucun paramétrage n'est néces-

saire si ce n'est la quantification du nombre de niveaux d'orientations N_{angle} sur 8 bits : ce paramètre est constant quelles que soient les bases testées.

Nous proposons d'illustrer les performances des cartes d'orientation dans un contexte de généralisation : les visages ayant servi à la création du modèle (18 personnes de la base M2VTS [Pigeon 1996]) sont différents des visages testés. Très souvent, un traitement améliorant la robustesse d'un algorithme dans un cas spécifique, le dégrade dans des conditions standards [Du 2005]. Pour cette raison, nous testerons notre proposition sur une base dédiée à la problématique de l'illumination (CMU-PIE : 1386 images de 66 visages pris sous 21 illuminations différentes [SIM 2002]) et une autre base contenant des visages très différents sous des éclairages quelconques (BIOID : 1521 images [Res01]). Comme l'illustre la figure 5, les visages de la bases du CMU sont illuminés de droite à gauche tandis que ceux de la base BioId résultent d'un éclairage ambiant non contrôlé (néon plus lumière naturelle) les visages étant alors différemment exposés suivant la lumière extérieure.

La figure 6 permet de visualiser les performances des outils d'alignement (en terme de pourcentage de visages de la base de tests ayant été alignés) en fonction de la précision de l'alignement (erreur e). Un point de référence dans la littérature est souvent le point d'abscisse 0.15. Sur la base CMU-PIE, notre méthode permet d'aligner les visages dans 94% des cas avec une précision supérieure ou égales à 15% alors que lorsque des DM-AAM ou des MAA classiques sont utilisés, les performances chutent respectivement de 8% et 15%. Mais c'est lorsque les conditions sont totalement quelconques que notre



Figure 5. Exemples d'images des bases BIOID (en haut) et CMU-PIE (ligne du bas).

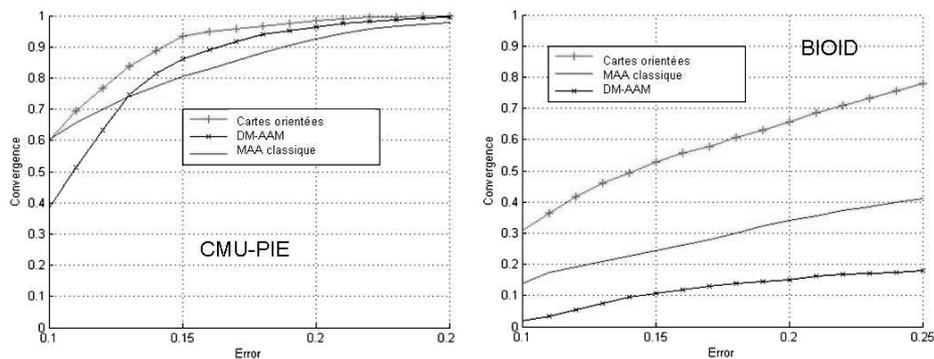


Figure 6. Performances de trois algorithmes d'alignement sur les bases CMU-PIE et BIOID.

proposition affiche le meilleur gain. Sur la base de test BIOID, dans 52% des cas, nous parvenons à aligner les visages avec une précision au moins égale à 15%, ce qui représente un gain de 27% et 42% par rapport aux MAA classiques et aux DM-AAM respectivement. Cette base est plus difficile à traiter car les visages ne sont pas systématiquement de face, peuvent avoir des expressions très différentes et porter des barbes ou des lunettes. Les DM-AAM sont certes plus efficaces que les MAA classiques sur des visages de faces de taille normalisée mais différemment éclairés (base CMU-PIE), mais ne sont plus du tout compétitifs par rapport aux MAA classiques dans des conditions standards (base BIOID). Les DM-AAM utilisent des cartes de distances évaluées à partir d'une extraction de contours. Le seuil utilisé pour extraire ces contours est alors crucial. Il est nécessaire que toutes les images de la base testée aient la même dynamique. Ca n'est pas le cas des images de la base BioID dont les contrastes varient fortement. À l'inverse, notre méthode n'utilise aucun seuil puisque nous ne faisons pas une extraction de contours mais une évaluation du gradient en chaque pixel de l'image.

5. MAA robustesse à la pose et à l'identité

5.1. MAA adaptés

Nous avons vu au paragraphe 3.3 que la robustesse des MAA est liée à la variabilité des visages introduite dans la base d'apprentissage. Cette variabilité conduit à un espace réduit morcelé qui gêne la convergence des MAA. Plutôt que d'utiliser un seul modèle très général contenant beaucoup de variabilité,

nous proposons d'exploiter d'abord un seul modèle M_0 présentant une forte variabilité en identité, puis d'utiliser un modèle particulier M_{adapte} présentant une forte variabilité en expression et en pose.

5.1.1. Modèle initial

Considérons une base de données générale contenant trois types de variabilité: en expression, en identité et en pose. La figure 7 illustre une telle base. Elle est constituée de plusieurs visages différents, présentant quatre expressions distinctes: « neutre », « A », « I » et « O ». Chacun de ces visages exprime chacune de ces expressions dans cinq poses différentes: visages de face, inclinés vers le bas et le haut et orientés vers la droite et la gauche.

Le modèle initial M_0 est réalisé à partir d'une base de données BDD_0 contenant des visages différents tous de face, dans des illuminations normales et des expressions neutres (voir Fig.8). Cela revient à n'exploiter que les images de la base de données générale sur l'axe horizontal. Ce modèle va être utilisé pour aligner grossièrement dans un premier temps le visage inconnu.

5.1.2. Identification du type de visage analysé

Soit C_0 le vecteur d'apparence après alignement du modèle M_0 sur le visage inconnu. Dans l'espace de représentation des paramètres de contrôle du modèle déformable d'initialisation, nous recherchons les k plus proches vecteurs de paramètres de C_0 appartenant à la base d'apprentissage d'initialisation BDD_0 . Ces k plus proches vecteurs d'apparence sont appelés les « k plus proches visages » par analogie à la méthode des k -plus-proches-voisins, ou $kppv$. La métrique utilisée est simplement la distance euclidienne dans l'espace réduit après ACP.

À titre d'exemple, seul le plus proche vecteur (C_p) est sélectionné sur la figure 9 ($k = 1$).

Ces vecteurs correspondent à des exemples particuliers de visages (identités particulières) dans la base d'apprentissage

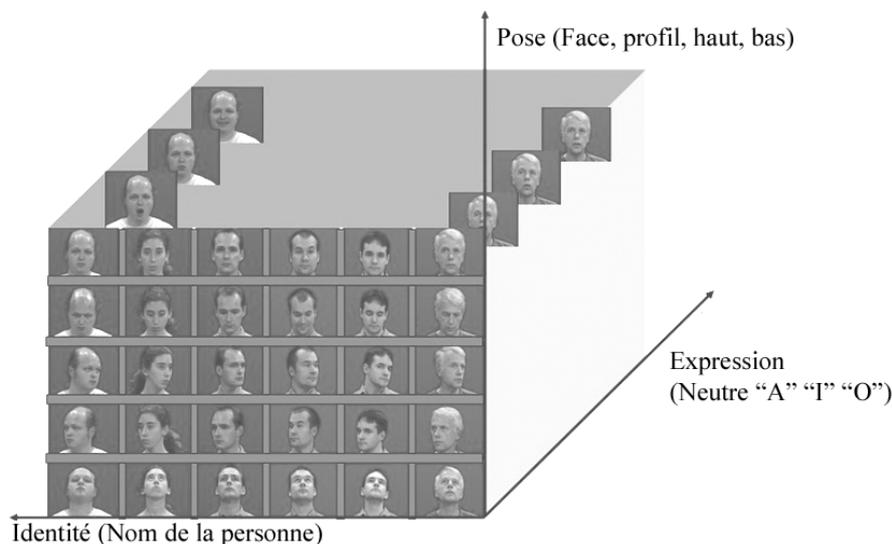


Figure 7. Base de données générale.



Figure 8. Base de données initiale BDD_0 .

d'initialisation BDD_0 . Ce sont les visages qui ressemblent le plus au visage inconnu analysé.

5.1.3. Modèle adapté

À partir de cette collection d'identités différentes (les k plus proches visages), nous générons une base BDD_{adapte} contenant les images des personnes ayant été identifiées comme ressemblant au visage analysé, prises dans des poses et avec des expressions différentes. La figure 10 illustre une telle base dans le cas où $k = 1$. Cela revient à constituer une nouvelle base à partir de la base de données générale, en en extrayant des tranches (autant de tranches que de plus proches visages) verticales et en profondeur (voir Fig.7).

À partir de la base BDD_{adapte} , nous générons le modèle M_{adapte} . Lorsque k est inférieur à quatre, il est possible de calculer des modèles au préalable avec une, deux ou trois personnes ayant des expressions et des poses différentes.

L'intérêt de prendre un nombre restreint de plus proches visages est de limiter le nombre de composantes du vecteur d'apparence c . En diminuant la variabilité en terme d'identité et en conservant celle relative à la pose et à l'expression, nous restreignons la variabilité globale de la base d'apprentissage. Les ACP ne conservent que les vecteurs propres parmi les plus significatifs : limiter la variabilité de la base d'apprentissage revient donc à limiter le nombre de composantes du vecteur d'apparence. Ce nombre restreint de composantes d'une part facilite la convergence et améliore donc l'efficacité des MAA, et d'autre part accélère le temps de traitement. Nous avons observé un facteur dix entre les dimensions

du vecteur d'apparence issu d'un MAA construit sur la base de données générale (Fig.7, $N_c = 25$) et celui construit sur une base adaptée ($k = 2$, $N_c = 250$). Dans des conditions réelles d'utilisation (texture de 1756 pixels annotés sur 68 points clés), cela revient à diminuer la complexité algorithmique (Eq. 8) d'un facteur dix.

5.1.4. Mise en œuvre

Lorsqu'il faut aligner un visage inconnu dans une image fixe, il suffit donc de faire converger le modèle M_0 , de trouver les k plus proches visages et d'appliquer le modèle pré-calculé correspondant pour parvenir à aligner correctement le visage.

Dans le cas où il s'agit de traiter un flot vidéo, nous pouvons utiliser la première seconde de la vidéo pour rendre plus robuste l'opération de sélection du modèle focalisé. Sur les premières images, nous faisons converger le modèle initial M_0 . Nous évaluons pour chaque image le résultat de cette convergence en évaluant l'erreur moyenne e_{pix} (voir Eq.7) sur l'ensemble de la texture reconstruite. Cette erreur est remarquablement stable du fait de l'utilisation des cartes d'orientation. Il est donc possible de la comparer à un seuil afin de décider si le modèle a convergé ou non. Nous évaluons alors, à partir des images sur lesquelles l'alignement a été correct, les plus proches visages dans la base générale afin de choisir le modèle pré-calculé « focalisé ». Ce modèle est alors exploité sur l'ensemble des images constituant le flot vidéo pour aligner les visages.

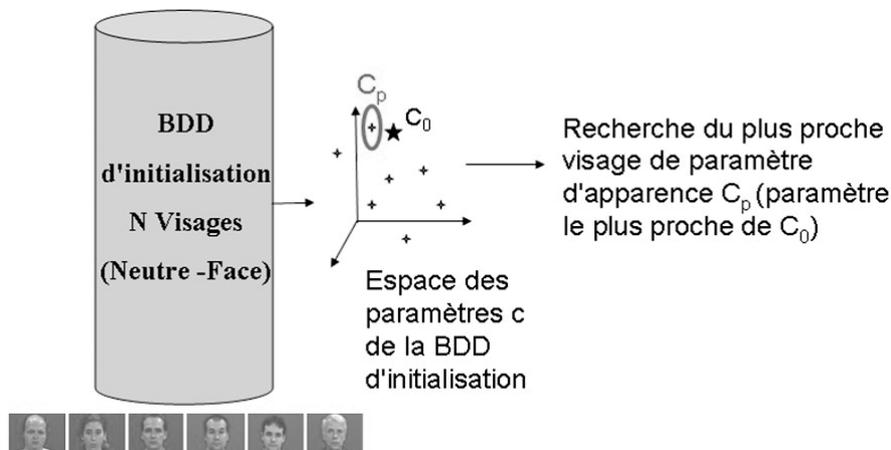


Figure 9. Recherche des plus proches visages.



Figure 10. Base de données focalisée sur une identité BDD_{adapte} .

5.2. Performance des MAA adaptés

Nous proposons de tester les MAA adaptés sur les images fixes de la base de données générale (Fig. 7) constituée de 29 visages présentant 4 expressions sous cinq poses différentes, soit donc 20 images par personne. Une validation croisée de type *Leave-one-out* est utilisée : le test est réalisé sur un visage dans toutes ses poses et expressions, sachant que les 28 autres visages ont été exploités pour l'apprentissage du modèle. Les 29 visages ont tous été testés, soit une base globale de test de 580 images. Il s'agit donc de localiser les centres des yeux, du nez et de la bouche de 29 visages inconnus dans des poses et expressions variées. La base initiale permettant de générer le modèle M_0 est la même que celle illustrée sur la figure 8, mis à part que nous avons retiré la personne testée ; elle contient donc 28 visages différents.

Le paramétrage des MAA est identique à celui de la section 4.2 : une texture de 1756 pixels associée à un modèle de forme de 68 points clés et une variance expliquée de 99 %. Concernant le modèle adapté le seul paramètre à régler est la valeur de k .

Nous proposons de comparer les performances de notre système lorsque $k = 2$ («MAA adapté») à deux autres MAA différemment construits. Le premier («MAA 28») est un MAA bâti sur une base contenant 28 exemples : les 28 visages de face et d'expression neutre ; cette base contient peu de variabilité. Le second («MAA 560») est un MAA bâti sur une base très riche en variabilité contenant 560 exemples : les 28 visages sous cinq poses et quatre expressions.

La complexité algorithmique du «MAA adapté» est du même ordre que celle du «MAA 28» puisque la dimension de leur vecteur d'apparence est comparable (25 en moyenne). À l'inverse, la complexité du «MAA 560» est bien plus grande (vecteur d'apparence de dimension 250 environ) et exclut une implémentation en temps réel.

La figure 11 montre la supériorité du «MAA adapté» sur les deux autres modélisations. Le fait que les performances du «MAA 560» sont inférieures à celle du «MAA adapté» est cohérent, bien que la base ayant servi à générer le «MAA 560» soit la plus riche : l'espace réduit de ce dernier modèle devient morcelé du fait de sa trop grande variabilité et l'optimisation est alors plus difficile à mener (paragraphe 3.3).

Lorsque l'on se place à une erreur de 15 % de la distance entre les yeux, niveau d'erreur habituellement utilisé dans la littérature pour comparer les performances des MAA, notre proposition est à la fois 10 fois plus rapide qu'un modèle évalué sur une base très riche, du fait de la faible dimension de son vecteur d'apparence comparée à celle du vecteur d'apparence du «MAA 560», et 16 % plus efficace (94 % de classifications correctes versus 78 %). Si on compare notre proposition à un modèle aussi rapide («MAA 28») alors les «MAA adaptés» sont plus efficaces dans 55 % des cas (94 % versus 49 %).

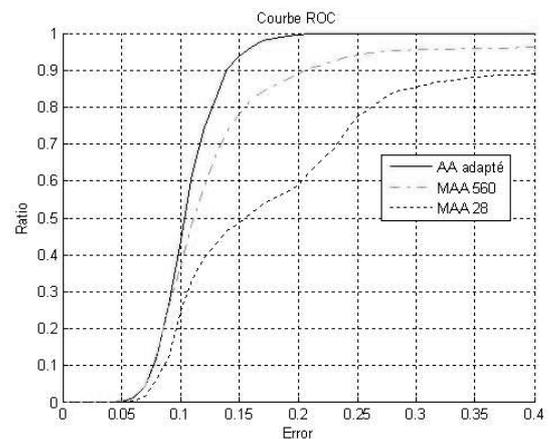


Figure 11. Performances comparées des «MAA adapté», «MAA 28» et «MAA 560».

6. Conclusion et perspectives

Les Modèles Actifs d'Apparence sont efficaces lorsqu'il s'agit d'aligner des visages connus dans des espaces contraints (illumination et pose contrôlées). Afin d'être robuste à l'illumination, nous avons proposé un nouveau prétraitement des textures prises en compte par les MAA avec une nouvelle normalisation durant la convergence itérative de l'algorithme. Pour pouvoir aligner des visages inconnus (détecter les centres des yeux, du nez et de la bouche) de poses et expressions variées, nous avons proposé un système permettant d'adapter le modèle au visage analysé. Cette adaptation se fait en choisissant dans une banque de modèles pré-calculés, les modèles les plus proches du visage inconnu. Des tests effectués sur des bases publiques et privées ont montré l'intérêt de nos propositions : il est possible d'aligner des visages inconnus dans des espaces non contraints avec une

précision suffisante pour la plupart des applications qui nécessitent une opération d'alignement (reconnaissance de visage, analyse de la gestuelle d'un visage, clonage). Contrairement à [Canzler 2004] (section 2) qui génère un modèle spécifique à partir de la première image d'un flot vidéo dans laquelle il fait converger un MAA générique, nous recherchons le modèle qui convient le mieux à l'alignement du visage inconnu. La différence est importante : nous ne répercutons pas de façon définitive les erreurs dues à un alignement imparfait de notre modèle générique. Notre système est de ce fait beaucoup plus robuste eu égard aux performances relatives du modèle générique. Notons enfin que notre MAA adapté et son prétraitement associé par cartes orientées est du même ordre de complexité que les MAA classiques et est donc implémentable en temps réel.

Pour des applications comme l'analyse des émotions ou la lecture labiale, il est important d'avoir un alignement plus précis afin d'être capable de suivre l'évolution du visage. En particulier, les performances du modèle doivent être jugées sur la localisation d'un certain nombre de points autour des yeux, des sourcils et de la bouche, et pas seulement sur leurs centres de gravité. Nous travaillons actuellement à l'élaboration d'un système adapté hiérarchique qui permet de trouver pour chaque caractéristique du visage (yeux et bouche essentiellement) le modèle adapté le plus pertinent.



Références

- [Aidarous 2007] AIDAROUS Y., LE GALLOUS S., SATTAR A. AND SÉGUIER R., «Face Alignment using active appearance model optimized by simplex», *International Conference on Computer Vision Theory and Applications (VISAPP)* Barcelona, 2007.
- [Blanz 1999] BLANZ V., VETTER T., «A morphable model for the synthesis of 3{D} faces», *SIGGRAPH'99, Computer Graphics Proceedings*, pages 187-194. Addison Wesley Longman, 1999.
- [Belaroussi et al. 2005] BELAROUSSI R., PREVOST L. AND MILGRAM M. «Classifier combination for face localization in color images», *International Conference on Image Analysis and Processing (ICIAP)*, 2005.
- [Res01] HUMANSCAN AG BIOD-TECHNOLOGY RESEARCH, «The bioid face database» <http://www.bioid.com/downloads/facepdb/>, 2001.
- [Canzler 2004] CANZLER U. AND B.WEGENER, «Person-adaptive Facial Feature Analysis», *International Conference on Electrical Engineering*, 2004.
- [Christoudias 2004] CHRISTOUDIAS C.M., MORENCY L.P. AND DARRELL T., «Light field appearance manifolds», *European Conference on Computer Vision*, pages 481-493, 2004.
- [Christoudias 2005] CHRISTOUDIAS C.M. AND DARRELL T., «On modelling nonlinear shape-and-texture appearance manifolds» *CVPR'05, Conference on Computer Vision and Pattern Recognition*, volume-2, pages 1067-1074, 2005.
- [Chang 2004] CHANG Y., HU C. AND TURK M., «Probabilistic expression analysis on manifolds», *CVPR'04, Computer Vision and Pattern Recognition*, 2004.
- [Cootes 1998] COOTES T.F., EDWARDS G.J. AND TAYLOR C.J., «Active Appearance Models», *ECCV'98, European Conference on Computer Vision*, 1998.
- [Cootes 1999] COOTES T.F. AND TAYLOR C.J., «A mixture model for representing shape variation», *Image and Vision Computing*, 1999.
- [Cootes 2000] COOTES T.F., WALKER K.N. AND TAYLOR C.J., «View-based active appearance models», *FGR'00, International Conference on Automatic Face and Gesture Recognition*, pages 227-232, 2000
- [Cootes 2000b] COOTES T.F., WHEELER G.V., WALKER K.N. AND TAYLOR C.J., «Coupled-view active appearance models» *BMVC'00, British Machine Vision Conference*, volume-1, pages 52-61, 2000.
- [Cootes 2001] COOTES T.F. AND TAYLOR C.J., «On representing edge structure for model matching», *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001.
- [Cristinacce 2003] CRISTINACCE D. AND COOTES T., «A Comparison of two Real-Time Face Detection Methods», *International Workshop on Performance Evaluation of Tracking and Surveillance*, 2003.
- [Cristinacce 2006] CRISTINACCE D. AND COOTES T., «Feature Detection and Tracking with Constrained Local Model», *Proc. British Machine Vision Conference*, 2006.
- [Du 2005] DU B., SHAN S., QING L. AND GAO W., «Empirical Comparisons of Several Preprocessing Methods for Illumination Insensitive Face Recognition», *ICASSP'05, International Conference on Acoustics, Speech, and Signal Processing*, 2005
- [Feng 2006] FENG X., LV B. AND LI Z., «Automatic facial expression recognition using both local and global information» *Chinese Control Conference*, pages 1878-1881, 2006
- [Froba 2002] FROBA B. AND KULLBECK C., «Robust face detection at video frame rate on edge orientation features», *International Conference on Automatic Face and Gesture Recognition*, 2002.
- [Garcia 2004] GARCIA C. AND DELAKIS M., «Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), Novembre 2004, p. 1408-1423.
- [Georgiades 2000] GEORGIADES A., BELHUMEUR P.N. AND KRIEGMAN D.J., «From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination», *Automatic Face and Gesture Recognition*, 2000.
- [Giri et al 2006] GIRI D., ROSENWALD M., VILLENEUVE B., LE GALLOU S. AND SÉGUIER R., «Scale Normalization for the Distance Maps AAM», *International Conference on Control, Automation, Robotics and Vision (ICARCV)* Singapore, 2006.
- [Gross 2002] GROSS R., MATTHEWS I. AND BAKER S., «Fisher Light-Fields for Face Recognition Across Pose and Illumination», *German Symposium on Pattern Recognition*, 2002.
- [Gross 2002] GROSS R., MATTHEWS I. AND BAKER S., «Generic vs. person specific active appearance models», *Image and Vision Computing*, 2005.
- [Hu 2003a] HU C., FERIS R. AND TURK M., «Active wavelet networks for face alignment», *BMVC'03, British Machine Vision Conference*, 2003.
- [Hu 2003b] HU C., FERIS R. AND TURK M., «Real-time view-based face alignment using active wavelets networks», *AMFG'03, International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
- [Hu 2004] HU C., CHANG Y., FERIS R. AND TURK M., «Manifold based analysis of facial expression», *CVPRW'04, Conference on Computer Vision and Pattern Recognition Workshop*, volume 5, pages 81-87, 2004.
- [Huang 2004] HUANG Y., LIN S., LI S.Z., LU H. AND SHUM H.Y., «Face Alignment Under Variable Illumination», *FGR'04, International Conference on Automatic Face and Gesture Recognition*, 2004.
- [Kittipanya-ngam] KITTIPANYA-NGAM P. AND COOTES T.F., «The effect of texture representations on AAM performance», *ICPR'06, International Conference on Pattern Recognition*, 2006.
- [Langs 2005] LANGS G., PELOSCHKE P., DONNER R. AND BISCHOF H., «A clique of active appearance models by minimum description length» *BMVC'05, British Machine Vision Conference*, pages 859-868, 2005.
- [Lee 2005] LEE K.C., HO J. AND KRIEGMAN D.J., «Acquiring Linear Subspaces for Face Recognition under Variable Lighting», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.
- [Li 2001] LI Y., GONG S. AND LIDDELL H., «Modelling faces dynamically across views and over time» *International Conference on Computer Vision*, pages 554-559, 2001.
- [Li 2001] LI Y., SHUICHENG Y., ZHANG H.J. AND CHENG Q.S., «Multi-view face alignment using direct appearance models», *FGR'02, International Conference on Automatic Face and Gesture Recognition*, pages 324-329, 2002.

- [Pigeon 1996] PIGEON, «M2VTS Projec», www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html, M2VTS, 1996.
- [Romdhani 1999] ROMDHANI S., GONG S. AND PSARROU A., «A multi-view nonlinear active shape model using kernel pca». *BMVC'99, British Machine Vision Conference*, pages 483-492, 1999.
- [Sattar 2007] SATTAR A., AIDAROUS Y., LE GALLOU S. AND SÉGUIER R., «Face Alignment by 2.5D Active Appearance Model Optimized by Simplex», *ICVS'07, International Conference on Computer Vision Systems*, 2007
- [Scott 2003] SCOTT I.M., COOTES T.F. AND TAYLOR C.J., «Improving Appearance Model Matching Using Local Image Structure» *IPMI'03, Information Processing in Medical Imaging*, 2003.
- [SIM 2002] SIM T., BAKER S. AND BSAT M., «The cmu pose, illumination, and expression (pie) database». *FG'02, IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.
- [Stegmann 2000] STEGMANN M.B., «Active Appearance Models: Theory, Extensions and Cases», *Informatics and Mathematical Modelling*, Technical University of Denmark, DTU, 2000.
- [Stegmann 2000] STEGMANN M.B. AND LARSEN R., «Multi-band modelling of appearance», *GMBV'02, Workshop on Generative Model-Based Vision*, 2002.
- [Tong 2006] TONG Y., WANG Y., ZHU Z. AND Ji Q., «Facial Feature Tracking using a Multi-State Hierarchical Shape Model under Varying Face Pose and Facial Expression», *ICPR'06, International Conference on Pattern Recognition*, 2006
- [Viola 2004] VIOLA P. AND JONES M.J., «Robust real-time face detection», *International Journal of Computer Vision*, 2004.
- [Xiao 2004] XIAO J., BAKER S., MATTHEWS I. AND KANADE T., «Real-Time Combined 2D+3D Active Appearance Models», *CVPR'04, Conference on Computer Vision and Pattern Recognition*, 2004
- [Xu 2005] XU Z., CHEN H. AND HU S.C., «A high resolution grammatical model for face representation and sketching». *CVPR'05, Computer Vision and Pattern Recognition*, volume 2, pages 470-477, 2005.
- [Zalewski 2005] ZALEWSKI L. AND GONG S., «2d statistical models of facial expressions for realistic 3d avatar animation». *CVPR'05, Conference on Computer Vision and Pattern Recognition*, volume~2, pages 217-222, 2005.
- [Zhao 2000] ZHAO W.Y AND CHELLAPPA R., «Illumination-Insensitive Face Recognition Using Symmetric Shape-from-Shading», *CVPR'00, Conference on Computer Vision and Pattern Recognition*, 2000.
- [Zhu 2003] ZHU J., LIU B. AND SCHWARTZ S.C., «General illumination correction and its application to face normalization», *International Conference on Acoustics, Speech, and Signal Processing*, 2003
- [Zuiderveld 1994] ZUIDERVELD K., «Contrast Limited Adaptive Histogram Equalization», *Graphics Gems IV*, 1994.





Renaud Séguier

Renaud Séguier a obtenu un doctorat en Traitement du signal, Image, Radar de l'Université de Rennes I en 1995. Le sujet de sa thèse concernait la détection de visage en temps-réel. Il a travaillé pendant un an chez Philips R\&D dans le domaine de la télévision numérique et de la compression Mpeg2. Il a rejoint l'équipe SCEE (Signal, Communication et Électronique Embarquée) de Supélec en 1997. Il est depuis lors professeur assistant en Traitement d'image, Vie Artificielle et Électronique Numérique. Ses recherches actuelles concernent l'analyse et la synthèse de visage pour la vidéo-compression et les interfaces homme-machine.



Gaspard Breton

Gaspard Breton est ingénieur de recherche à Orange Labs (France Télécom R\&D). Il a reçu un doctorant portant sur l'animation faciale des personnages virtuels en 2002. Depuis, il a continué à travailler dans ce domaine, en élargissant son expertise sur les méthodes d'animation, les systèmes de contrôle des mouvements labiaux pour la parole et l'expressivité, les méthodes de rendu photoréaliste, ainsi que les systèmes de capture de mouvements faciaux. Il a participé à plusieurs projets internes, nationaux et Européens tels que le RNRT Tempovalse, le RNTS Telma ainsi que les IST Danae et Companions et est membre du NoE Humaine.

S



Sylvain Le Gallou

Sylvain Le Gallou a travaillé sur des projets s'appuyant sur la détection des points caractéristiques du visage au sein des laboratoires de Supélec/IETR et France Telecom R\&D entre 2004 et 2007. Il a alors obtenu le grade de Docteur de l'Université de Rennes 1 en Traitement du Signal et Télécommunications pour sa thèse intitulée Détection robuste des éléments faciaux par Modèles Actifs d'Apparence. Depuis 2007, il travaille dans l'équipe de reconnaissance de visage de Sagem Sécurité (Groupe SAFRAN) en tant qu'ingénieur R\&D. Ses recherches sont basées sur la reconnaissance de visage, l'apprentissage et la modélisation de visage.



Christophe Garcia

Christophe Garcia a obtenu le diplôme de doctorat de l'université Lyon I, en Intelligence Artificielle et Analyse d'Image, en septembre 1994. Il a passé deux ans à l'IBM Vision Automation Group, une année au Computer Vision Center de l'université Autonome de Barcelone et 18 mois au German National Research Center.

De 1997 à 2000, il a occupé un poste de chercheur à l'institut FORTH (Foundation for Research and Technology Hellas), en Grèce, et de professeur au Computer Science Department de l'Université de Crète. En 2003, il a rejoint l'IRISA-INRIA de Rennes, pour travailler dans le domaine de la structuration automatique et de l'indexation de contenus vidéo.

Depuis décembre 2003, Christophe Garcia pilote les activités de recherche de France Telecom R\&D - Orange Labs en indexation automatique des images et de la vidéo. Il a déposé 15 brevets d'invention et a publié plus de 80 articles en conférences et journaux internationaux. Il est actuellement éditeur associé de quatre revues internationales: Journal of Visual Communication and Image Representation (Elsevier), Image and Video Processing (Hindawi), Pattern Analysis and Application (Springer-Verlag) et Pattern Recognition (Elsevier).