Une texture polynomiale pour les modèles actifs d'apparence

Cristina BORDEI[1], Pascal BOURDON[1], Bertrand AUGEREAU[2], Philippe CARRÉ[2]

[1]Technicolor R&I, 975 avenue des Champs Blancs, CS 17616, 35576 Cesson-Sevigné, France

[2]XLIM-SIC (UMR CNRS 7252), Bât. SP2MI, Téléport 2,
Bvd Marie et Pierre Curie, BP 30179 86962 Futuroscope Chasseneuil Cedex France

[1]Prénom.Nom@technicolor.com, [2]Prénom.Nom@univ-poitiers.fr

Résumé – Dans cet article, nous proposons une nouvelle approche pour la représentation de texture dans les modèles actifs d'apparence (AAM). Celle-ci est basée sur l'utilisation de coefficients issus de projections des intensités lumineuses sur une base polynomiale complète. Parce qu'elle propose une représentation compacte et hiérarchique des images, la décomposition polynomiale est une alternative efficace aux représentations trop globales telles que l'ACP, ou trop redondantes telles que les ondelettes de Gabor. De plus, elle apporte une certaine souplesse par rapport aux représentations en ondelettes dans les paramètres de la décomposition. Nous décrirons comment des coefficients de projection sur bases polynomiales peuvent être utilisés dans un modèle AAM en fournissant des résultats expérimentaux dans un contexte d'alignement de visages. Ceux-ci illustreront la capacité de notre approche à améliorer la robustesse face aux changements de pose et d'expression faciale.

Abstract – In this paper, we propose a new polynomial texture representation method for Active Appearance Models. While many texture representations have been proposed over the years to improve the accuracy and reliability of computer vision applications such as object tracking or image alignment, most descriptors are usually unable to both provide precise multi-scale and multi-orientation analysis and handle the redundancy problem effectively. We will explain how coefficients obtained from polynomial projections of pixel intensities on a complete basis can be used for compact, hierarchical image approximation and structural analysis, providing experimental results in face alignment that will demonstrate their ability to improve robustness against pose and facial expression changes.

1 Introduction

Les modèles actifs d'apparence sont un outil puissant proposé initialement par Cootes *et al.* [1], permettant de recaler de façon robuste des objets visuels à partir de modèles déformables alliant géométrie et texture issus d'un apprentissage statistique. Si les méthodes telles que l'ACP sont très performantes dans l'étude de variations globales, une analyse plus locale, ainsi que plus scalable, présente un intérêt non négligeable dans la création et l'alignement du modèle. A titre d'exemple, les avantages d'une représentation en ondelettes de l'apparence ont déjà été démontrés dans la littérature [2, 3]. Le succès de ce type d'approche est fortement lié à la capacité que le mode de représentation aura de fournir à la fois une analyse multi-échelle/multi-orientations rigoureuse, tout en traitant efficacement le problème de la complexité calculatoire et de la redondance des coefficients.

Nous proposons donc une nouvelle représentation de texture pour les modèles actifs d'apparence, basée sur un modèle polynomial, semblable à une décomposition en paquets d'ondelettes à reconstruction parfaite pour une échelle donnée mais avec une souplesse accrue dans la décomposition. Nous verrons comment les coefficients obtenus à partir de projections des intensités lumineuses d'une image sur une base polynomiale complète peuvent être utilisés pour une approximation hiérarchique et compacte du signal image, et pour son analyse structurelle. Des résultats expérimentaux, obtenus en alignement de visages, seront comparés à l'état de l'art et montreront la capacité de la représentation polynomiale à améliorer la robustesse contre les changements d'expression faciale et de pose.

2 Représentation de texture dans les AAM

Un modèle actif d'apparence (AAM) est un modèle statistique permettant de faire conjointement l'analyse et la synthèse d'une classe d'objets à partir d'un ensemble d'apprentissage comprenant différentes vues d'un objet. A partir d'une base d'images annotées manuellement avec des points d'intérêt, un modèle de forme $\mathbf{x}$ et un autre de texture $\mathbf{g}$ sont créés, par une analyse en composantes principales (ACP), pour représenter les variations de la position des points et des intensités des pixels dans les données d'entrée :

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_{\mathbf{x}}\mathbf{b}_{\mathbf{x}} \qquad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_{\mathbf{g}}\mathbf{b}_{\mathbf{g}} \tag{1}$$

où $\mathbf{P}$ est une matrice contenant les principaux modes de variation de forme ou texture et $\mathbf{b}$ un vecteur contrôlant la forme/texture reconstruite.

Au cours de l'étape d'ajustement du modèle AAM, qui consiste à trouver les paramètres de texture et de forme synthétisant de façon optimale de nouvelles images, l'algorithme va minimiser

la distance entre l'image réelle projetée dans la forme moyenne calculée et la texture du modèle générée par le vecteur d'apparence [1, 4]. La précision de la mise en correspondance des modèles dépendra donc fortement de la représentation de la texture.

Les AAMs standards s'appuyant sur les intensités des pixels dans les calculs de texture, de nombreuses représentations alternatives ont été proposées pour rendre l'algorithme plus robuste à des changements d'illumination, de pose, et, dans le cas de l'analyse faciale, d'identité ou d'expression. Stegmann et Larsen [5] montrent qu'un mélange de caractéristiques (niveaux de gris, teinte et contours) donne des résultats meilleurs à toute représentation individuelle, Séguier et Le Gallou [6] utilisent des cartes d'orientations basées sur des textures égalisées de façon adaptative, tandis que Su et Tao [3] proposent une représentation combinant ondelettes de Gabor avec des motifs binaires locaux (LBP). Les ondelettes de Gabor ont été également utilisées par Davoine *et al.* [7] pour construire un modèle AAM hiérarchique.

Parce qu'elles fournissent un formalisme théorique efficace pour l'analyse multi-échelles et multi-orientations, les ondelettes sont efficaces pour traiter les problèmes de changements d'éclairage et de pose, et sont largement utilisées dans des applications d'analyse faciale. Nous proposons d'étudier et d'utiliser une représentation similaire à la représentation en ondelettes, mais plus souple et adaptative pour la texture des AAM : la transformée polynômiale.

3 Décomposition polynomiale de texture par bases complètes

Notre motivation pour utiliser une représentation polynomiale pour la texture dans les AAM vient du fait que les polynômes orthogonaux présentent certaines propriétés liées au système visuel humain [8], notamment une représentation multi-échelle /multi-résolution de l'information. De plus, une image pourrait être approximée à partir des coefficients polynomiaux en ne conservant qu'un nombre défini de coefficients assurant une certaine énergie cumulée, similaire à l'analyse en composantes principales.

Le polynôme bi-variable de degré d est la fonction de $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ définie comme :

$$P(\mathbf{x}) = \sum_{\substack{(d_1,d_2)\in[0;d]^2 \\ d_1+d_2\leq d}} a_{d_1,d_2} x_1^{d_1} x_2^{d_2} \tag{2}$$

où $d_1 \in \mathbb{N}^+$ et $d_2 \in \mathbb{N}^+$ sont les degrés respectifs des variables x_i et $\{a_{d_1,d_2}\} \in \mathbb{R}$ les coefficients du polynôme. Le degré du polynôme est alors donné par le maximum de $d_1 + d_2$.

Considérant un ensemble fini de couples $D = \{(d_1, d_2)\} \subset \mathbb{N}^2$, nous représentons par $\mathbb{E}_D$ l'espace de tout polynôme réel bi-variable tels que $a_{d_1,d_2} \equiv 0$ si $((d_1, d_2) \notin D)$ et par $\mathcal{K}_D$ le sous-ensemble de monômes réels :

$$\mathcal{K}_D = \left\{ K_{d_1,d_2}(\mathbf{x}) = x_1^{d_1} x_2^{d_2} \right\}_{(d_1,d_2)\in D} \tag{3}$$

$\mathcal{K}_D$ est une famille libre et génératrice donc, $\mathcal{K}_D$ est une base de $\mathbb{E}_D$, la base canonique. Dans notre contexte, nous recherchons des bases avec des propriétés appropriées telles que l'orthogonalité ou la normalité. A cet effet,nous devons considérer le domaine discret des points de collocation sous-jacents :

$$\Omega = \left\{ \mathbf{x}_{(u,v)} = \left(x_{1,(u,v)}, x_{2,(u,v)} \right) \right\}_{(u,v)\in D} \tag{4}$$

où D représente l'ensemble des paires associées à Ω. Diverses voies sont alors envisageables afin de générer une base polynômiale orthonormée à partir de $\mathcal{K}_D$. Nous choisissons de le faire en appliquant le processus de Gram-Schmidt. Cela implique que nous avons besoin d'un produit et d'une norme pour les fonctions bi-variables réelles définies sur Ω. En tenant compte des charges calculatoires et du contexte discret nous avons choisi le produit scalaire suivant :

$$\langle F | G \rangle_w = \sum_{(u,v)\in D} F\left(\mathbf{x}_{(u,v)}\right) G\left(\mathbf{x}_{(u,v)}\right) w\left(\mathbf{x}_{(u,v)}\right) \tag{5}$$

où F et G sont deux fonctions définies sur Ω et où apparait une fonction de pondération w définie positive sur Ω (Legendre, Chebichev, Hermite, ...). Alors, le processus réel de construction d'une base orthonormée :

$$\mathrm{B}_{D,w} = \{B_{d_1,d_2}\}_{(d_1,d_2)\in D} \tag{6}$$

est une récurrence sur (d_1, d_2).

Un cas particulier est la *base complète* où l'ensemble D est homomorphe à Ω dans le sens où D représente exactement l'ensemble des paires associées à Ω :

$$D = [0; n_1] \times [0; n_2] \tag{7}$$

et où la base est donc définie par la famille : $\{B_{d_1,d_2}(\mathbf{x})\}_{\substack{d_1=0..n_1 \\ d_2=0..n_2}}$

Le nombre de polynômes de la base polynomiale complète est donc déterminé par la taille du domaine $(n_1+1)\times(n_2+1)$.

Pour analyser une image, nous avons besoin de sa représentation spatio-fréquentielle. Si certains paramètres caractéristiques d'une texture sont récupérés directement sur les niveaux d'intensité des pixels de l'image, d'autres caractéristiques importantes ne peuvent être dévoilées qu'à travers un filtrage de l'image. Les polynômes orthonormés discrets de la base complète peuvent être considérés comme une décomposition multi-échelle discrète, ce qui nous permet de représenter les informations de texture de manière compacte et précise. Soit U une fonction définie sur un domaine Ω. Le processus de décomposition à une étape L se décrit comme tel :

−Pavage du domaine discret Ω^L avec une partition en sous-domaines $\Omega^L_{k_1,k_2}$, avec (k_1, k_2) les indices des sous-domaines, et tels que $\cup_{k_1,k_2}\Omega^L_{k_1,k_2} = \Omega$;

− Pour tout $\Omega^L_{k_1,k_2}$, calcul des coefficients polynomiaux par projection : $b^{L,k_1,k_2}_{d_1,d_2} = \left\langle U^{L,k_1,k_2} I \,\middle|\, B^L_{d_1,d_2} \right\rangle$ où U^{L,k_1,k_2} est la restriction de U au sous-domaine $\Omega^L_{k_1,k_2}$ et où les $B^L_{d_1,d_2}$ sont les polynômes d'une base complète définie sur ce même sous-domaine ;

−Regroupement des coefficients b par degré de l'opérateur de projection en respectant l'ordonnancement spatial des sous-domaines ceci afin de créer une structure multi-résolution.

On constate ici que la technique offre une réelle souplesse, notamment vis-à-vis du choix des facteurs de résolution, qui peuvent être indépendants entre niveaux de décomposition. Par conséquent, la transformée polynomiale multi-échelle sera plus compacte qu'une représentation en ondelettes de Gabor, permettant de faire disparaître la plupart des problèmes d'échantillonnage, comme le compromis entre l'échantillonnage fréquentiel et d'orientations. De plus, les bases complètes permettent d'obtenir, pour un ensemble donné de valeurs, une fonction interpolatrice qui est un polynôme d'osculation du premier ordre (*ie* tel que $\forall \mathbf{x}_{(u,v)} \in D, P_I\left(\mathbf{x}_{(u,v)}\right) = I\left(\mathbf{x}_{(u,v)}\right)$). Par ailleurs, on peut considérer la projection sur $B_{i,j}$ comme un opérateur de différences finies multi-échelle relatif à la différentiation $\partial_1^i \partial_2^j$.

Quant aux représentations classiques temps-fréquence, à l'instar des ondelettes, la décomposition polynomiale n'est pas soumise à une décomposition dyadique, ce qui la rend plus adaptative. Deux exemples de décompositions de niveau 1 d'une même image sont présentés en figure 1 avec à gauche une décomposition utilisant une base de Chebichev complète à support3×3 et à droite une base complète d'Hermite à support 5×4.

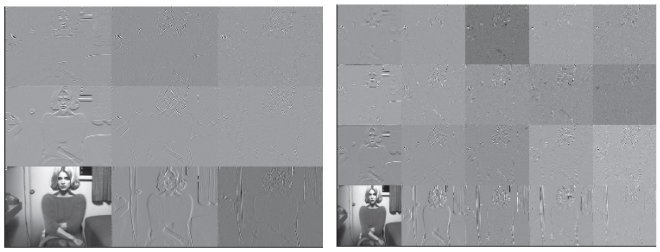


FIGURE 1 – Décompositions au premier ordre de Chebichev 3×3 (g) et Hermite 5×4 (d)

Afin d'améliorer la robustesse du processus d'ajustement des AAM, nous proposons de remplacer le mode de représentation de texture du modèle de référence par des projections polynomiales sur une base complète orthonormée. Ceci revient à calculer un modèle d'apparence en remplaçant le vecteur des intensités pixels en entrée de l'ACP par un vecteur de coefficients obtenus par projections polynomiales dans la base complète sur des textures alignées. Nous allons donc remplacer dans l'équation 1 le modèle de texture $\mathbf{g}$ par $\mathbf{gp}$, un vecteur des coefficients polynomiaux d'approximation obtenus sur une texture alignée.

Deux possibilités se présentent pour le calcul du vecteur de coefficients : il pourra être effectué soit sur des régions d'intérêts situées autour de points annotés, soit à partir d'une décomposition polynomiale multi-résolution de la texture, suivie d'une étape éventuelle de quantification.

4 Résultats expérimentaux

Pour évaluer les performances de notre méthode, nous avons réalisé des expériences d'alignement de modèle sur des images appartenant à deux bases d'images faciales : IMM (qui contient une variabilité en pose et expression faciale) et CMU Multi-Pie (comportant une variabilité en pose, expression faciale et illumination). Pour chaque base d'images nous avons sélectionné un sous-ensemble de 40 sujets, 10 étant utilisés pour l'apprentissage et les 30 restants pour les tests. Les bases d'images que nous utilisons présentent un défi certain car elles contiennent des variabilités multi-utilisateur/multi-expression faciale et multi-utilisateur/multi-pose.

Pendant la phase d'apprentissage, nous avons généré 8 modèles différents d'AAM : 4 sur 60 images de 10 individus choisis au hasard de la base IMM et 4 sur 60 images frontales de 10 sujets de la base Multi-Pie avec diverses expressions faciales. Pour chaque base de données, les 4 modèles calculés sont : le modèle standard de Cootes *et al.* [1] , la méthode compositionelle inverse de Matthews et Baker (ICIA) [4], notre modèle d'AAM polynomial, dont les coefficients d'apparence sont calculés sur des régions situées autours de points fiduciaux (PAAM), et enfin un modèle utilisant le premier niveau d'une décomposition polynomiale multi-résolution de toute la texture alignée (FT-PAAM).

Pour le modèle PAAM, les coefficients de l'approximation polynomiale sont obtenus via des projections sur une base complète d'Hermite 15×15, taille raisonnable pour modéliser les changements locaux de texture, tandis que pour le FT-PAAM nous utilisons une base complète d'Hermite à support 3×3 suffisante pour les calculs de gradient. Pour les deux approches nous avons utilisé la méthode de collocation de Chebychev.

Pour évaluer la précision d'ajustement, nous avons utilisé l'erreur point par point définie comme la distance euclidienne entre les points de modèle estimé $\mathbf{x}$ et les points annotés manuellement $\mathbf{x}_{hl}$, présentée en équation 8.

$$E(\mathbf{x}_{hl}, \mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(x_i - x_{hl,i})^2 + (y_i - y_{hl,i})^2} \qquad (8)$$

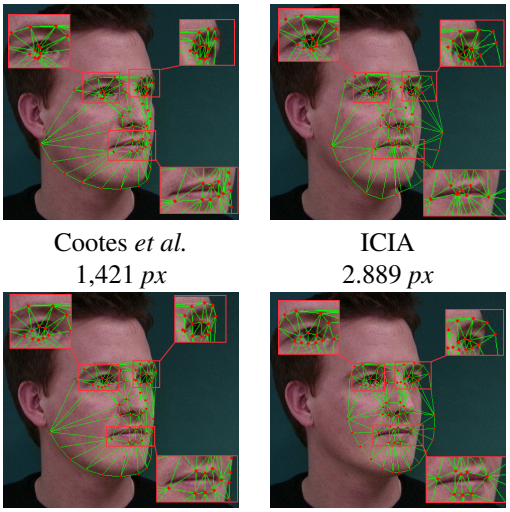
L'erreur moyenne et son écart-type calculée avec les quatre méthodes différents est montrée dans le tableau 1. Comme on peut le constater, nos deux modèles de texture proposés permettent d'obtenir une meilleure précision. Nos premiers résultats sont très satisfaisants et montrent que par ses propriétés - sa paramétrisation simple et sa souplesse, la représentation polynomiale est un substitut prometteur aux représentations classiques de texture.

Une différence notable est que la méthode PAAM est plus robuste aux changements de pose. Dans nos expériences, nous constatons des améliorations significatives de notre méthode à la fois sur celle proposée par Cootes *et al.* et l'ICIA(voir Fig. 2). Dans cette approche, le modèle AAM est construit sur des régions autour des points d'intérêt, donc il peut donner des résultats plus précis que celui calculé sur l'ensemble des pixels du modèle AAM, en particulier dans le cas de la variation d'expression faciale ou pose. Comme nous utilisons des modèles de texture spatialement localisés autour des points d'intérêt, notre méthode offre obligatoirement plus de robustesse aux modification locales de texture.

	MultiPie	IMM
Cootes et al.	1.372 ± 0.452	1.590 ± 0.413
ICIA	1.195 ± 0.659	1.600 ± 0.788
PAAM	1.380 ± 0.433	$\mathbf{1.535 \pm 0.407}$
FT-PAAM	$\mathbf{1.152 \pm 0.369}$	1.552 ± 0.663

TABLE 1 – Erreur moyenne et écart-type

Sur la figure 3, nous présentons les résultats obtenus sur un individu de la base MultiPie. Nous observons qu'en utilisant le modèle FT-PAAM, les points d'intérêt sont déterminés avec une meilleure précision par rapport aux autres modèles, en particulier pour les points sur le menton, qui sont assez difficiles à situer et qui ne sont généralement pas pris en compte dans les calculs d'erreurs. Cela est dû au fait que l'on obtient une représentation hiérarchique de l'information lors de la transformation des coefficients de texture via des projections polynômiales.



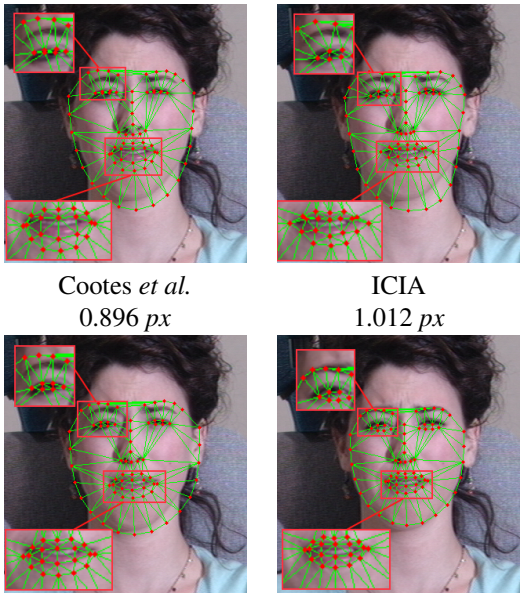
Cootes *et al.* 1,421 *px* — ICIA 2.889 *px*

PAAM 1.265 *px* — FT-PAAM 2.934 *px*

FIGURE 2 – Résultats obtenus sur un sujet de la base IMM

5 Conclusion

Dans cet article, nous avons proposé deux nouvelles approches pour la représentation de texture pour les AAM. Les coefficients résultant des projections polynômiales des valeurs de pixels sur une base complète ont été utilisés pour l'approximation de la texture. Les résultats expérimentaux montrent que nos deux approches fonctionnent très bien dans le cadre des algorithmes d'alignement de visage et qu'en fonction de la méthode choisie, nous obtenons plus de robustesse aux changements de pose ou d'expression faciale. Une approche hybride combinant les deux méthodes proposées est envisagée pour améliorer la précision d'ajustement lorsque la base de données présente des variations de multi-pose/multi expression.



Cootes *et al.* 0.896 *px* — ICIA 1.012 *px*

PAAM 1.014 *px* — FT-PAAM 0.613 *px*

FIGURE 3 – Résultats obtenus sur un sujet de la base MultiPie

Références

[1] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[2] M. B. Stegmann, S. Forchhammer, and T. F. Cootes, "Wavelet enhanced appearance modelling," in *SPIE International Symposium on Medical Imaging*, San Diego, CA, 2004, vol. 5370, pp. 1823–1832, SPIE.

[3] Y. Su, D. Tao, X. Li, and X. Gao, "Texture representation in aam using gabor wavelet and local binary patterns," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, 2009, pp. 3274–3279.

[4] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[5] M.B. Stegmann and R. Larsen, "Multi-band modelling of appearance," *Image and Vision Computing*, vol. 21, no. 1, pp. 61–67, 2003.

[6] S. Renaud, L.G. Sylvain, B. Gaspard, G. Christophe, et al., "Adapted active appearance models," *EURASIP Journal on Image and Video Processing*, vol. 2009, 2010.

[7] F. Davoine, B. Abboud, and D. VAN MO, "Analyse de visages et d'expressions faciales par modèle actif d'apparence," *TS. Traitement du signal*, vol. 21, no. 3, pp. 179–193, 2004.

[8] AS Blaivas, "Visual analysis in unspecialized receptive fields as an orthogonal series expansion," *Neurophysiology*, vol. 6, no. 2, pp. 168–173, 1974.