

# Détection de l'iris dans des visages de pose quelconque : approche multi-textures et Modèles Actifs d'Apparence 2.5D

Hanan SALAM<sup>1</sup>, Renaud SEGUIER<sup>1</sup>, Nicolas STOIBER<sup>2</sup>

<sup>1</sup>Supélec, IETR, Equipe SCEE  
Avenue de la Boulaie - CS 47601 ,35576 Cesson Sévigné Cedex, France

<sup>2</sup>Dynamixyz  
80 avenue des Buttes de Coesmes, 35700 Rennes, France  
hanan.salam@supelec.fr, renaud.seguier@supelec.fr  
nicolas.stoiber@dynamixyz.com

**Résumé** – Nous proposons un système pour l'évaluation de la direction du regard dans un visage d'orientation quelconque, visage filmé par une simple webcam. L'originalité de ce système est la fusion de deux modèles actifs d'apparence (AAM) qui modélisent l'oeil et son environnement (paupière et sourcil). En Outre, la système utilise une optimisation multi-objectif pour analyser conjointement l'oeil droit et gauche avec un même vecteur de pose relatif à l'orientation du visage en particulier.

Les résultats sur deux bases publiques (HPEG [11] et UlmHPG [10]) illustrent l'intérêt de notre proposition. Notre système permet de localiser le centre de l'iris dans respectivement 90% et 81% des images des deux bases contre 51% et 21% des images pour un système plus classique modélisant les yeux par un AAM 2.5D.

**Abstract** – We propose a system for the detection of gaze direction in the face for any head pose and filmed by a simple webcam. The originality of this system is the fusion of two active appearance models (AAM), which model the eye and its environment (eyelid and eyebrow). In addition, the system uses a multi-objective optimization to jointly analyze the left and right eyes, with the same vector, relative to the head pose in particular.

The results on two public databases (HPEG [11] and UlmHPG [10]) illustrate the interest of our proposal. Our system allows a detection of the center of the iris in respectively 90% and 81% of the images of the two bases against 51% and 21% of the images for a more conventional system modeling the eyes by a 2.5D AAM.

## 1 Introduction

La direction de regard est une information cruciale dans les interactions homme-machine. Son estimation nécessite trouver la position de l'iris dans l'oeil. Malheureusement, cette tâche est difficile lorsque de simples webcams sont utilisées : (i) l'iris a une très faible résolution (de l'ordre de  $20 \times 15$  pixels) ; (ii) les clignements des yeux, la position de la tête et parfois le port de lunettes provoquent des occlusions qui gênent l'analyse. En outre, la variabilité de la couleur de l'iris et de sa position dans l'oeil le rend plus difficile à analyser à l'aide d'un modèle global du visage.

De nombreuses méthodes ont été proposées pour détecter la direction du regard, et peuvent être classées en trois catégories :

**i) Les méthodes basées IR (infrarouge)** [1] ont l'inconvénient d'être fortement dépendantes de la luminosité des pupilles qui peuvent être influencées par la fermeture des yeux et de l'occlusion, l'éclairage extérieur et la distance de l'utilisateur de l'appareil photo.

**ii) Les méthodes basées sur des modèles déformables** [2, 3] ont besoin de la définition d'un ensemble adéquat de pa-

ramètres initiaux pour le modèle qui doit être initialisé près de l'oeil. En outre, elles enregistrent généralement un échec lorsque la tête présente de fortes orientations de la tête. Certaines méthodes utilisent une modélisation de l'iris par une ellipse [4] et réussissent à trouver l'emplacement de l'iris ou de la pupille à condition que les images soient de haute résolution. Un autre inconvénient est leur incapacité à faire face aux différents états de fermeture des paupières.

**iii) Les méthodes basées sur la synthèse** sont plus robustes. Elles s'appuient sur la synthèse d'une image en utilisant un modèle. Le calcul de la position de l'iris est fait en comparant l'image de synthèse à la image captée de l'oeil. L'intérêt de ces méthodes réside dans leur capacité à modéliser l'apparence et la forme de l'oeil simultanément contrairement aux modèles déformables ou basés sur des ellipse qui modèlent la forme seulement. Pour éviter la formalisation manuelle du modèle de l'oeil, les Modèles Actifs d'Apparence (AAM) [5] sont adaptés. Construire un AAM pour la localisation d'iris nécessite la constitution d'une base d'exemple de visages annotés et proposant différentes direction du regard. Ainsi, l'apparence de l'oeil est apprise en même temps que celle du visage [6, 7]. Il

est nécessaire de constituer une très large base de données afin de considérer toute la variabilité concernant la pose (si l’AAM est 2D) et la direction du regard.

Nous proposons un système qui consiste à fusionner un Modèle Actif d’Apparence (AAM) de l’iris et un AAM de la peau (paupière et sourcil) autour de l’oeil. La texture de l’iris glisse sous la texture de la peau qui entoure l’oeil permettant de synthétiser de nouvelles directions du regard. Les deux yeux sont pris en compte grâce à une optimisation multi-objectifs qui exploite explicitement l’angle de rotation de la tête. Par rapport à l’état de l’art, le système proposé fonctionne avec des images basse résolution, il ne contraint pas l’utilisateur à des exigences particulières (illumination IR, l’équipement hardware ...) et il exploite l’apparence et la forme de l’oeil tout en évitant la conception explicite manuelle du modèle. Notre solution présente l’avantage de limiter la base d’apprentissage de l’AAM aux personnes en vue frontale et en regardant en face d’eux.

Les contributions de ce papier sont :

- Combinaison de deux AAMs (un AAM pour la peau et un pour l’iris) pour détecter l’emplacement de l’iris ;
- Utilisation explicite de l’orientation de la tête dans l’algorithme pour rendre le système plus robuste aux larges orientations de la tête ;
- Détection du regard sans la nécessité d’inclure des personnes avec différentes directions du regard dans la base d’apprentissage de l’AAM.

Ce papier est organisé comme suit : dans la section 2 nous décrivons le système globale. Dans la section 2.1, nous présentons notre algorithme. Ensuite, dans la section 3, nous montrons quelques résultats expérimentaux. Enfin, dans la section 4, nous tirons nos conclusions.

## 2 Système général

Considérons qu’une personne est en face de l’écran où une webcam est installée. La première étape est la détection de l’orientation de la tête à l’aide d’un modèle 2.5D AAM [8]. Selon l’orientation du visage détecté, l’oeil gauche et l’oeil droit sont inégalement représentés dans l’image de la webcam. Nous proposons donc d’analyser la direction du regard en utilisant une optimisation multi-objectif : la contribution de chaque oeil à la direction du regard final est pondérée en fonction de l’orientation du visage détectée. Dans la section suivante, nous décrivons comment la position de l’iris pour un oeil est calculée en utilisant le modèle AAM multi-texture proposé. Puis dans la section 2.2, nous présentons l’optimisation multi-objectif.

### 2.1 AAM Multi-texture pour la localisation de l’iris

L’idée de base de l’AAM Multi-Texture (MT-AAM) est que l’intérieur de l’oeil est considéré comme une texture distincte de celle du visage. En combinant cette texture avec la texture de surface de la peau et en lui permettant de bouger à partir d’une

vecteur des paramètres, nous réussissons à synthétiser toutes les directions du regard. Donc, nous fusionnons deux AAMs, un AAM pour la texture de l’iris (AAM Iris) et un autre pour la peau environnante (AAM locale de peau).

**AAM locale de peau** – Ce modèle est construit en utilisant 22 annotations qui décrivent le contour de l’oeil entier, y compris les sourcils et la texture entourant l’oeil. La figure 1(a) est une illustration de la texture moyenne de notre modèle montrant le trou à l’intérieur de l’oeil droit avec les annotations pour obtenir ce modèle.

**AAM Iris** – La figure 1(b) est une illustration de la texture moyenne du modèle de l’iris avec les annotations correspondantes. Le modèle comporte à la fois l’iris et une texture blanche autour. Celle-ci modélise le sclera. Pour la formation de ce modèle nous utilisons un modèle de 13 annotations dont 8 décrivent la forme circulaire de l’iris et 1 décrit la position approximative de son centre ; à fin d’apprendre la texture blanche autour de l’iris, 4 points supplémentaires formant une forme rectangulaire autour de l’iris sont placés.

**Fusion des 2 AAM (modèles)** – Pour chercher la position de l’iris dans l’oeil, nous combinons le modèle de la peau et le modèle de l’iris. On retrouve les paramètres optimaux des yeux (selon le modèle de la peau) dans une étape préalable. Nous utilisons ensuite les paramètres trouvés pour reconstruire l’image décrivant la peau autour des yeux. Le modèle d’iris glisse au dessous avec le vecteur de pose de l’iris  $T^{iris}$  décrivant la position de l’iris dans l’oeil.

$$T^{iris} = [S^{iris} t_x^{iris} t_y^{iris}] \quad (1)$$

$S^{iris}$  est l’échelle de l’iris,  $t_x^{iris}$  et  $t_y^{iris}$  sont les paramètres de translation horizontale et verticale décrivant la position des points de l’iris par rapport au centre de l’oeil.

Les deux modèles, une fois fusionnés, constituent le modèle final décrivant la région des yeux (cf. figure 1 (c)). Un problème de discontinuité entre les deux modèles se pose lorsque les textures sont surajoutées. Afin de résoudre ce problème, nous appliquons un filtre passe-bas sélectif sur la peau et le blanc de l’oeil, tout en préservant l’iris. Il lisse la discontinuité entre la paupière et l’iris, et reproduit également l’effet d’ombrage induit par la paupière sur le blanc de l’oeil.

**Algorithme** Les Étapes de construction du MT-AAM pour un oeil sont les suivantes.

1. Localisation de l’oeil en utilisant le modèle AAM de la peau autour des yeux ;
2. Extraction de la texture de l’oeil ( $g_i$ ), à partir de l’image de la webcam (figure 1 (d)) ;
3. Synthèse de la peau ( $g_m^{oeil}$ ) en utilisant les paramètres optimaux trouvés par le modèle de la peau (figure 1 (a)) ;
4. Jusqu’à ce que le critère d’arrêt soit atteint :
  - (a) Création de la texture du modèle de l’iris ( $g_m^{iris}$ ) basé sur la pose et les paramètres d’apparence du modèle de l’iris (figure 1 (b)).
  - (b) Fusion des deux textures  $g_m^{oeil}$  et  $g_m^{iris}$  pour obtenir la texture  $g_m$  (figure 1(c)) ;

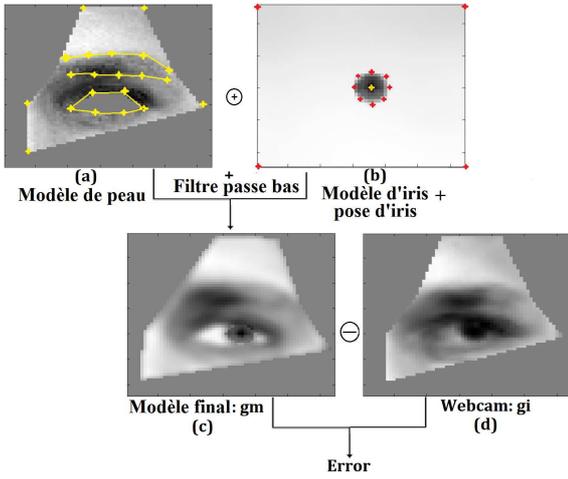


FIGURE 1 – Fusion du modèle de la peau et du modèle de l'iris.

- (c) Application du filtre passe bas sélectif ;
- (d) Evaluation de  $E : E = g_i - g_m$  ;
- (e) Réglage des paramètres de pose et de l'apparence du modèle de l'iris.

## 2.2 Optimisation multi-objective

Il existe une forte corrélation entre le mouvement et l'apparence des iris des deux yeux. Pour cette raison, nous utilisons un vecteur de pose et un vecteur d'apparence communs pour décrire la pose et l'apparence des deux iris. Avoir des informations sur les deux yeux permet au système de rester robuste lorsque la personne fait des grands mouvements de tête autour de l'axe vertical où que les yeux sont partiellement ou complètement fermés. Nous proposons un AAM multi-objectif [9]. Dans notre système, un modèle unique de l'iris est fusionné avec les modèles de peau des deux yeux. Les textures résultantes sont comparées à celles des yeux dans l'image afin de produire deux erreurs liées aux deux yeux. Celles-ci sont pondérées en fonction de l'orientation de la tête et additionnées pour obtenir une erreur globale qui est minimisée en utilisant un algorithme génétique (GA). Cette erreur devient :

$$E = \alpha E^{gauche} + \beta E^{droit} \quad (2)$$

où  $E^{gauche}$  et  $E^{droit}$  sont les erreurs correspondant respectivement à l'oeil gauche et à l'oeil droit.  $\alpha$  et  $\beta$  sont des facteurs de pondération. Ils sont déterminés en fonction de la rotation de la tête autour de l'axe z, et évalué juste après la détection de visage par un AAM 2.5D global sur le visage en entier. Les deux suivent une loi double logistique (cf. figure 2).

$$\alpha(R_{yaw}) = \begin{cases} 0.5 & \text{if } -d \leq R_{yaw} \leq d \\ 0 & \text{if } -90 \leq R_{yaw} < -22 \\ 1 & \text{if } 22 \leq R_{yaw} < 90 \\ 0.5(1 + l(1 - \exp(\frac{-(R_{yaw}-ld)^2}{\sigma^2}))) & \text{else} \end{cases} \quad (3)$$

$$\beta(R_{yaw}) = 1 - \alpha$$

$l = \text{sign}(R_{yaw})$ ,  $\sigma$  est le facteur de pente et  $d$  est la bande de telle sorte que les deux fonctions  $\alpha$  and  $\beta$  sont égal à 0.5.

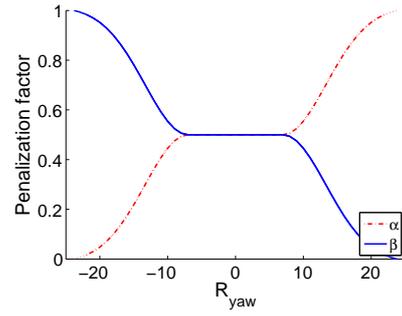


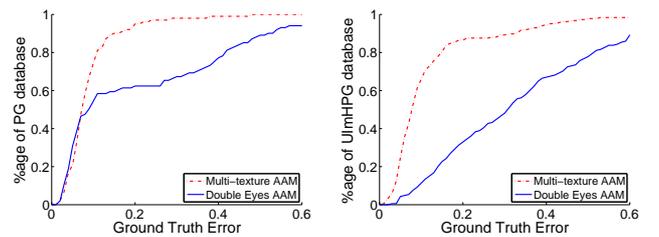
FIGURE 2 – Fonction double logistique function,  $d = 7$

De cette manière, l'orientation du visage est prise en compte dans l'optimisation.

En ce qui concerne l'algorithme génétique, la population initiale est générée aléatoirement entre les limites supérieure et inférieure des paramètres en utilisant une distribution uniforme. La fonction de fitness est la somme pénalisée des erreurs relatives aux deux yeux.

Nous prenons en compte l'emplacement de l'iris et la taille de l'oeil pour contraindre l'espace de recherche des AG. Ces contraintes sont fondées sur des moyennes anthropométriques. La largeur moyenne de l'iris est évaluée à 1/1.6 de la largeur de l'oeil. L'échelle varie autour d'une échelle initiale calculée à partir de la largeur de l'iris et celle de l'iris moyen. Les paramètres de translation horizontale et verticale ne doivent pas dépasser la moitié de la largeur de l'oeil et de celle de la hauteur des yeux en prenant le point de référence le point au centre des coins des yeux.

## 3 Résultats



(a) GTE of HPEG database (b) GTE of UImHPG database

FIGURE 3 – Multi-texture AAM vs. Double Eyes AAM

Afin de valider notre algorithme, nous avons mené des tests sur les bases de données UIm Head Pose and Gaze (UImHPG) [10] et Head Pose-Eye Gaze (HPEG) [11]. Nous comparons la méthode MT-AAM avec une méthode classique qui consiste à modéliser les yeux avec un AAM 2.5D, le 2.5D AAM Double Eyes AAM (DE-AAM). L'apprentissage du DE-AAM est réa-

lisé en utilisant un certain nombre de visages avec différentes directions du regard (10 personnes de la base HPEG avec cinq images par personne : regard frontal, regards à l'extrême gauche et droit, regards dans des directions intermédiaires). L'apprentissage du MT-AAM est réalisé sur la même base mais en considérant seulement une image par personne (regard frontal) et la base d'iris [12].

Dans la figure 3, nous présentons pour les deux bases le pourcentage d'images alignées en fonction de l'erreur d'alignement (Groud Truth Error - GTE). Le GTE est la moyenne de la distance euclidienne entre la vérité terrain (position réel de l'iris marqué manuellement) et le centre de l'iris donnée par la méthode de détection du regard, normalisée par la distance interoculaire [13].

La figure 3(a) donne les résultats obtenus sur la base (HPEG), tandis que la figure 3(b) montre les résultats obtenus en généralisant sur les nouvelles données (base UlmHPG).

Nous pouvons voir comment le MT-AAM surpasse le DE-AAM. Par exemple, la figure 2(a) montre que, pour une erreur inférieure ou égale à 15% le MT-AAM a détecté la bonne position de l'iris dans 90% des images, alors que le DE-AAM donne la bonne position dans 59% seulement des images pour le même niveau d'erreur. La figure 2(b) illustre les capacités de généralisation des deux approches : les visages analysés n'ont pas été utilisés au moment de l'apprentissage des modèles. Dans cette situation également, notre approche est plus efficace puisque nous affichons un taux de bonne détection de 81% (pour un niveau d'erreur de 15%) contre 22% pour le DE-AAM. Notons que le MT-AAM n'a utilisé que les visages de face de la base HPEG (soit 10 images de visages regardant droit devant eux) alors que pour des résultats inférieurs, le DE-AAM nécessite 50 images avec quatre images par personnes supplémentaires pour modéliser la direction du regard.

## 4 Conclusion

Nous avons présenté une nouvelle approche pour la détection de l'iris qui consiste à modéliser séparément la texture de l'iris et celle relative à la zone entourant l'oeil : paupière et sourcil. Nous avons ensuite paramétré le mouvement de l'iris dans l'oeil en combinant la texture de l'iris avec le modèle de la peau. En conséquence, chaque direction peut être synthétisée et ensuite détectée. Par ailleurs une optimisation multi-objectifs prenant en compte l'angle de rotation de la tête permet de localiser le centre de l'iris même lorsque le visage est dans une pose quelconque. Les résultats expérimentaux montrent que la méthode proposée surpasse la méthode classique de détection de l'iris par AAM en particulier dans un contexte de généralisation lorsque le sujet est inconnu.

## Références

[1] S. Asteriadis, D. Soufleros, K. Karpouzis, and S. Kollias, "A natural head pose and eye gaze dataset," in *Proceedings of the*

*International Workshop on Affective-Aware Virtual Agents and Social Robots*, 2009, p. 1.

- [2] U. Weidenbacher, G. Layher, P. M. Strauss, and H. Neumann, "A comprehensive head pose and gaze database," in *3rd IET International Conference on Intelligent Environments (IE 07)*, pp. 455–458.
- [3] J. P. Hansen, J. S. Augustin, and H. Skovsgaard, "Gaze interaction from bed," in *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 2011, p. 11.
- [4] Qiu Chen, Koji Kotani, Feifei Lee, and Tadahiro Ohmi, "An accurate eye detection method using elliptical separability filter and combined features," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 9, no. 8, August 2009.
- [5] H. Wu, Y. Kitagawa, T. Wada, T. Kato, and Q. Chen, "Tracking iris contour with a 3D eye-model for gaze estimation," in *Proceedings of the 8th Asian conference on Computer Vision Part I*, 2007, pp. 688–697.
- [6] W. J. Ryan, D. L. Woodard, A. T. Duchowski, and S. T. Birchfield, "Adapting starburst for elliptical iris segmentation," in *IEEE second international conference on biometrics : Theory, Applications and Systems*, Sept. 2008, pp. 1–7.
- [7] 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, 2001.
- [8] I. Bacivarov, *Advances in the Modelling of Facial Sub-Regions and Facial Expressions using Active Appearance Techniques*, Ph.D. thesis, National University of Ireland, 2009.
- [9] P. Ivan, "Active appearance models for gaze estimation," M.S. thesis, Vrije University, Amsterdam, 2007.
- [10] A. Sattar, Y. Aidarous, S. Le Gallou, and R. Seguier, "Face alignment by 2.5d active appearance model optimized by simplex," in *International Conference on Computer Vision Systems (ICVS)*, 2007.
- [11] A. Sattar and R. Seguier, "Facial feature extraction using hybrid genetic-simplex optimization in multi-objective active appearance model," in *Fifth International Conference on Digital Information Management (ICDIM)*, 2010, pp. 152–158.
- [12] M. Dobes, L. Machala, P. Tichavski, and J. Pospisil, "Human eye iris recognition using the mutual information," *Optik-International Journal for Light and Electron Optics*, vol. 115, no. 9, pp. 399–404, 2004.
- [13] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz, "Robust face detection using the hausdorff distance," in *Third International Conference on Audio and Video-based Biometric Person Authentication*, 2001, pp. 90–95.