
Contrôle du regard et des mouvements des paupières d'une tête parlante virtuelle

Alix Casari, Frédéric Elisei, Gérard Bailly & Stephan Raidt

*Institut de la Communication Parlée, 46 av. Félix Viallet, 38031 Grenoble - France
Correspondance: frederic.elisei@icp.inpg.fr*

RÉSUMÉ. Cet article présente un modèle de contrôle et de déformation géométrique des paupières d'une tête parlante virtuelle prenant en compte la direction du regard. L'analyse de données vidéo d'un sujet montre que les paupières subissent une déformation importante permettant à la fois de dégager le champ de vision tout en protégeant de manière optimale le globe oculaire. Le modèle issu de ces données prédit de manière satisfaisante la déformation observée à partir de paramètres tels que la direction du regard et l'ouverture/fermeture volontaire propre des paupières.

ABSTRACT. We present here a control model of shape deformation for the eyelids of a virtual talking head that takes into account direction of eye gaze. The analysis of video data for one target subject shows large deformations of the eyelids according to gaze direction that free the vision field while optimally protecting the eyes. The model extracted from this data predicts satisfactorily the observed deformations from few parameters such as gaze direction and proper voluntary opening/closing movements of the eyelids.

MOTS-CLÉS: agents conversationnels animés, regard, paupières, tête parlante.

KEYWORDS: embodied conversational agents, eye gaze, eyelids, talking face.

1. Introduction

Le but de ce travail est de doter une tête parlante virtuelle d'un regard suffisamment convaincant pour capter l'attention d'un interlocuteur humain en situation d'interaction face-à-face. Ce travail s'inscrit dans un projet plus ambitieux visant à doter un Agent Conversationnel Animé (ACA) de capacités à percevoir l'environnement dans lequel il est plongé afin de le rendre sensible aux changements de cet environnement induits ou non par ses propres actions. L'environnement inclut évidemment le ou les interlocuteurs : l'objectif de cet ancrage sur le monde réel, de cette attention (portée à l'environnement de l'interaction, à l'interlocuteur, à ce qu'il regarde, désigne ou dit) est de donner aux partenaires humains des gages de présence et de mesurer l'impact de ces gages de présence sur le dialogue (compréhension de l'information linguistique et paralinguistique échangée, charge cognitive, croyance en l'information, etc.).

Nous sommes très sensibles à la direction du regard de l'autre et notre capacité à inférer des intentions et des états mentaux à nos partenaires de communication s'appuie fortement sur cette information. De nombreux travaux ont souligné l'influence de la direction du corps et de la tête dans la perception de la direction du regard. Quelques auteurs ont déjà souligné la forte corrélation entre mouvement des paupières et des yeux dans leurs données. Lee, Badler et al (2002) notent ainsi : « During the analysis of eye-tracking images, we noticed a high correlation between the eyes and the eyelid movement which could be incorporated ». Tateno, Takemura et al (2005) démontrent que le respect de cette corrélation est bénéfique à un système de rectification du regard dans un dispositif de réalité mixte. De même, Gemmel et al. (2000) montrent que la manipulation seule de la position de l'iris pour changer la direction du regard d'un visage peut conduire à des changements drastiques de l'expression faciale perçue. Nos travaux antérieurs sur l'impact du contrôle du regard d'un ACA dans un dispositif de réalité partagée (Raidt, Bailly et al. 2006) confirment l'insuffisance du seul contrôle de l'iris pour la désignation de cibles précises.

Après un rapide état de l'art sur la modélisation des paupières utilisées couramment en infographie, nous décrivons les données sur lesquelles nous avons travaillé, les traitements effectués et le modèle de contrôle et de déformation des paupières qui en est issu.

2. Modèles de forme du globe oculaire et des paupières

Il est remarquable de constater qu'un modèle générique aussi détaillé que celui proposé par Moriyama et al (2006), qui vise à rendre compte de la diversité de l'apparence des yeux tant du point de vue de la variabilité structurelle individuelle que du mouvement des yeux, traite de manière indépendante et peu réaliste les

mouvements du globe oculaire et des paupières : trois paramètres de mouvement des paupières (fermeture/ouverture indépendantes des deux paupières et obliquité gauche/droite de l'ouverture) sont manipulés indépendamment de deux paramètres de mouvement haut/bas et gauche/droite de l'iris.

Ces quelques paramètres contrôlent par contre une géométrie finement maillée composée de trois structures : cornée, pupille et paupières. Dans la plupart des modèles 3D, chaque œil est une sphère quasi-blanche dont une portion est texturée différemment pour rendre l'iris et la pupille. Les paupières, quant à elles, sont figurées soit par deux demi-sphères¹, soit par un certain nombre de segments (ou quadrilatères). Pour un rendu plus réaliste, les yeux sont modélisés à l'aide de deux sphères imbriquées, l'une représentant le globe oculaire, l'autre la cornée sans oublier la pupille (cf. Figure 1), le tout permettant un meilleur rendu des reflets des yeux (dont notamment les multiples réflexions de Purkinje de l'environnement sur l'œil). Notons que des modèles physiques, intégrant les propriétés biomécaniques des diverses composantes de l'œil, ont été aussi développés, notamment pour la simulation d'opérations chirurgicales virtuelles (Sagar, Bullivant et al. 1994).

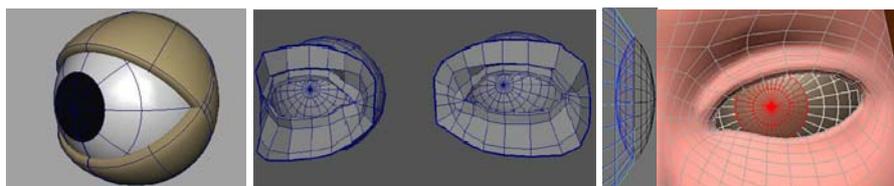


Figure 1 : Maillages génériques des yeux et des paupières des plus simples aux plus complexes qui distinguent globe oculaire, iris et cornée.

L'animation de ces modèles géométriques génériques est d'autant plus difficile que leur structure est complexe et proche de l'anatomie humaine. Les modèles de contrôle sont confrontés généralement à la définition des zones d'action des divers paramètres de contrôle opérant sur des organes petits et très mobiles ainsi qu'aux choix d'un mode de superposition de ces actions. Les mouvements des sourcils, des yeux et des paupières ont par ailleurs une forte valeur communicative. La redondance et la cohérence des déformations induites sur la zone des yeux par ces divers paramètres sont donc essentielles à la perception robuste de ces gestes intentionnels.

Au lieu d'ajuster à des données réelles des modèles génériques imposant aux formes des degrés de liberté prédéfinis, nous avons suivi la méthode de modélisation basée-données que nous avons développée pour le contrôle des mouvements faciaux (Revéret, Bailly et al. 2000; Badin, Bailly et al. 2002). Cette méthode de modélisation, appelée Analyse en Composantes Principales (ACP) guidée, cherche les corrélations statistiques existantes entre des paramètres

¹ Voir <http://www.jawa9000.com/technical/eyes/eyes.htm>

articulatoires imposés *a priori* et des sous-ensembles de points de chair dont le mouvement est collecté sur des sujets accomplissant des tâches spécifiques visant à recruter ces articulations.

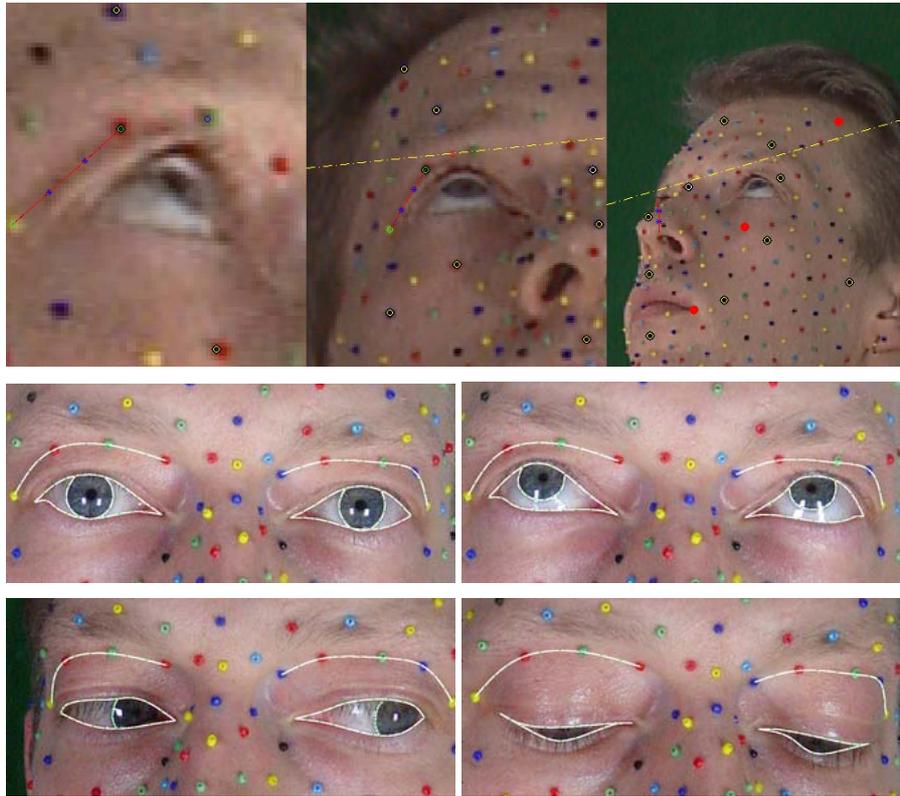


Figure 2. Données photogrammétriques. En haut : capture vidéo multi-vues (profil droit, vue de face et profil gauche à divers niveaux de zoom) ; en bas : tracé des paupières et de l'iris sur des photographies numériques.

3. Données, tracés et modélisation géométrique

Le modèle de visage présenté a été développé pour un sujet sur le visage duquel ont été collées plus de 400 billes (cf. Figure 2). Des billes ont été notamment positionnées sur la base des paupières et les sourcils. Les données photogrammétriques utilisées ont été capturées par deux dispositifs : des images haute définition (1200x700 pixels) prises par un appareil photo numérique alors que le sujet maintenait une configuration précise (regard de face ou en butée suivant diverses orientations du regard) pour trois orientations de la chaise pivotante où il était assis et une vidéo multi-vues (3x720x576) PAL permettant d'avoir accès à des mouvements en plus grand nombre et plus naturels.

Un logiciel interactif développé au laboratoire, exploitant modèles de caméra et géométrie épipolaire, permet d'obtenir des données 3D de toutes les billes dans un référentiel lié à la tête ainsi que la position de points de contrôle de courbes de Bézier positionnées à la main sur le contour des yeux (cf. Figure 2). Ces courbes de Bézier sont ensuite échantillonnées à intervalles réguliers sur une abscisse curviligne de manière à obtenir une définition de la surface de l'œil de l'ordre de 1 point par 10mm².

Un modèle générique complet d'œil intégrant globe oculaire, iris et cornée dont les paramètres morphologiques ont été ajustés au sujet (diamètre du globe, position relative et diamètre de la cornée, centres de rotation des yeux supposés constants, etc.) a été utilisé pour calculer la direction du regard par rapport à la tête. Le mouvement de la tête est quand à lui estimé grâce à l'ajustement d'un modèle complet du visage intégrant déformations du cou et mouvements faciaux produits par le mouvement de la tête (6 paramètres de roto-translation) et l'articulation (7 paramètres liés aux mouvements de mâchoire, des lèvres et du larynx). 24 visèmes présentant des configurations de regard échantillonnant de manière assez large le champ de vision du sujet ont été étiquetés complètement (yeux et visage).

4. Contrôle de la direction du regard et des paupières

À partir des paupières tracées et des pupilles ajustées, on a créé un modèle de paupières à six paramètres de contrôle. Ce modèle comporte deux types de paramètres dont les actions sur la géométrie 3D des yeux s'ajoutent linéairement. Le premier jeu lie la déformation des paupières à la direction du regard. Le deuxième jeu modélise le mouvement résiduel, qui correspond principalement à une fermeture/ouverture des paupières.

Le premier jeu vise à capturer l'incidence de la direction du regard sur la déformation des paupières. Les paramètres de contrôle imposés sont les angles d'élévation (α_2) et d'azimut (α_1) de chaque œil². Une simple régression linéaire ne permettant pas de rendre compte de manière satisfaisante de la variance des données, nous avons utilisé un modèle déjà expérimenté sur la main (Gibert, Bailly et al. 2005). Ce modèle consiste à estimer une régression linéaire non pas sur les paramètres angulaires mais sur les sinus et cosinus de ces angles, de leur somme et de leur différence. En multipliant les deux matrices de rotation et développant les produits, on obtient 8 prédicteurs en sinus et cosinus. Vu le faible volume de données d'apprentissage (24 configurations), seuls les plus efficaces – en termes d'explication de la variance des points de paupières – sont retenus: $\sin(\alpha_1+\alpha_2)$,

² α_1 et α_2 sont estimés dans un repère tête. Il était donc nécessaire d'estimer de manière précise l'orientation de la tête, découplée de l'articulation grâce à l'inversion du modèle de visage. La vergence des yeux sera assurée par un modèle de plus haut niveau à trois paramètres (coordonnées du point de l'espace regardé).

$\sin(\alpha_1-\alpha_2)$, $\cos(\alpha_1)$ et $\cos(\alpha_2)$. Ce modèle explique respectivement 66% et 61% de la variance des coordonnées 3D des points de paupières droite et gauche.

Le deuxième jeu de paramètres permet le contrôle de l'ouverture/fermeture propre des paupières (permettant notamment le clignement volontaire, la lubrification lacrymale du cristallin par clignement ou la protection des yeux vis à vis d'un choc ou d'une forte intensité lumineuse). Nous avons pour ceci effectué une simple ACP sur le résidu non encore expliqué des coordonnées 3D des points de paupières. La variance des données résiduelles expliquée par la première composante principale est respectivement de 40.9 et 30.8% pour les paupières droite et gauche (ce qui nous amène à 79,89% pour l'œil droit et 73% pour l'œil gauche pour la reconstitution totale des données). Ces deux paramètres correspondent à un mouvement des paupières majoritairement sur l'axe des y. L'action des autres composantes est difficile à expliquer autrement que par un bruit de tracé et ne saurait être intégrée au vu de la faible quantité de données (24 configurations) dont nous disposons pour l'instant.

Les actions des divers paramètres de contrôle sur la géométrie de l'œil sont présentées Figure 3. On peut noter que la déformation des paupières vise globalement à libérer le champ de vision tout en protégeant la partie non opératoire du globe oculaire. Les paupières basses et hautes sont toutes actives dans ce processus : ce recrutement est particulièrement sensible dans les déplacements verticaux. Il est en outre intéressant de noter que la base des paupières hautes n'est que très faiblement affectée par ces mouvements.

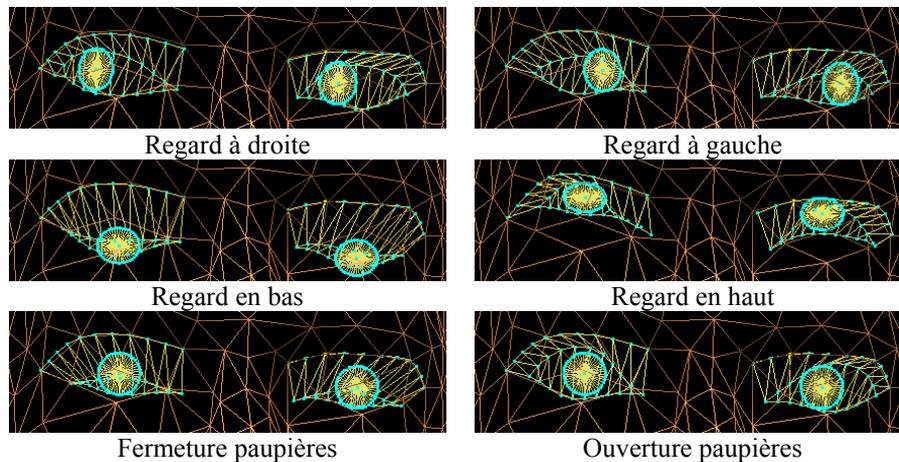


Figure 3. Action des divers paramètres de contrôle sur la géométrie de l'œil.

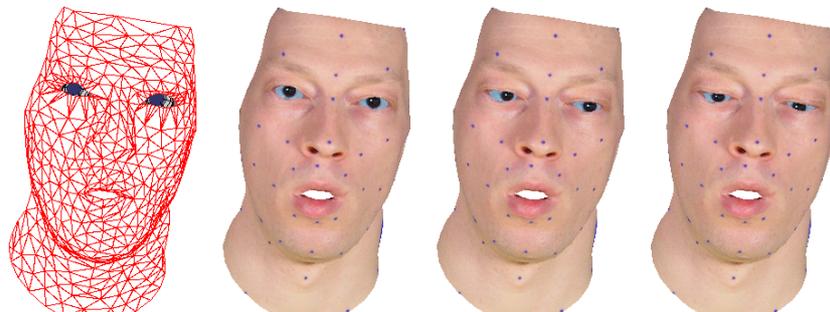


Figure 4. Synthèse (fil-de-fer puis texturée), pour différentes directions du regard, du modèle de visage incluant le modèle de paupières.

5. Tête parlante virtuelle

Ce modèle de contrôle est alors greffé au modèle de contrôle des mouvements de la tête et du visage préalablement construit (cf. Figure 4). Au système de suivi automatique de billes de couleur permettant de suivre les déformations du visage dues à la parole et aux principales expressions faciales (Bailly, Elisei et al. 2006) a été intégré un module spécifique de suivi du regard et du mouvement des paupières :

1. Le mouvement du visage est d'abord estimé très précisément, grâce au modèle articulatoire du visage et aux positions des billes.
2. Le contraste entre le ton chair vs la zone inter-paupières est donné par la couleur rouge du pixel normalisée par sa luminance $p=R/(R+G+B)$. Pour estimer la valeur du troisième paramètre de chaque œil (ouverture/fermeture propre), on minimise la somme des p des points prédits paupières/peau et des $1-p$ sinon (corps vitré + iris/pupille).
3. Le contraste iris vs corps vitré permet de retrouver les angles de la direction du regard en minimisant l'erreur de reconstruction de la seule zone entre les pupilles (les pixels de teinte chair étant évidemment exclus de cette minimisation).
4. La valeur obtenue à l'étape 2 est remise en cause pour avoir une erreur globale minimale (contours+direction de regard, cf. Figure 5).

Ce module se comporte correctement dès lors que les yeux ne sont pas presque fermés ou vus comme tels. Sinon, il faut recourir à une interpolation des angles non mesurables avec les valeurs précédentes/futures. Les poils des cils entraînent un biais sur l'estimation de l'ouverture/fermeture. L'implémentation *via* OpenGL sur une carte 3D PCI x16 estime les 6 paramètres pour une image multi-vues (face + deux profils) en environ 800ms.

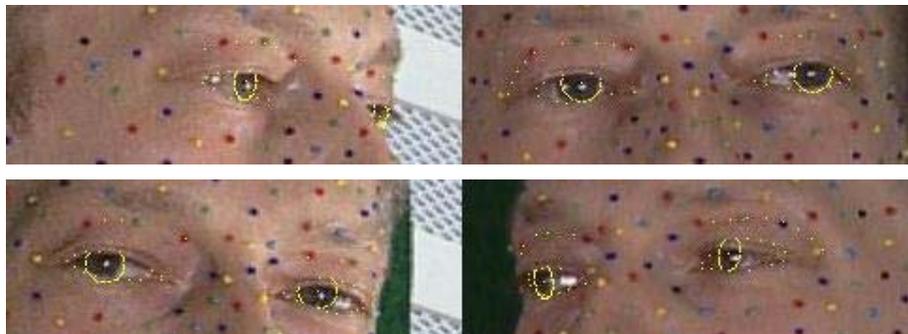


Figure 5. Optimisation du modèle visage + regard sur des images-clés d'une séquence vidéo multi-vues dans une séquence de parole conversationnelle.

6. Conclusions et perspectives

Un système de contrôle et de rendu précis du regard et du mouvement des paupières d'un clone parlant virtuel a été développé pour un sujet à partir de données photogrammétriques. Le modèle de rendu vidéoréaliste en cours de réalisation s'attachera à reproduire fidèlement les changements de texture associés à ces mouvements ainsi que ceux provenant de l'interaction entre ces mouvements et l'éclairage ambiant (synthèse de la réflexion de l'environnement sur la pupille).

Cette technique de clonage virtuel devrait nous permettre à moyen terme d'avoir accès à des ACA plus convaincants, parlant et exprimant des expressions faciales plus naturelles car reproduisant les corrélations entre déplacements de points de chair. Un test des capacités de notre ACA à désigner de manière plus précise des zones de l'espace d'interaction est en cours en utilisant un paradigme d'amorçage déictique déjà expérimenté (Raidt, Bailly et al. 2006).

Remerciements

Ce travail a été financé par le projet «scrutation de scènes multimodales» d'ELESA, le projet «Deixis Multimodale» du GIS PEGASUS, et le projet «Présence» du cluster Infolog de la région Rhône-Alpes.

Bibliographie

Badin, P., G. Bailly, et al. (2002). "Three-dimensional linear articulatory modeling of tongue, lips and face based on MRI and video images." *Journal of Phonetics* **30**(3): 533-553.

- Bailly, G., F. Elisei, et al. (2006). Degrees of freedom of facial movements in face-to-face conversational speech. International Workshop on Multimodal Corpora, Genoa - Italy: 33-36.
- Gemmell, J., C. L. Zitnick, et al. (2000). "Gaze-awareness for videoconferencing: a software approach." IEEE Multimedia 7(4): 26-35.
- Gibert, G., G. Bailly, et al. (2005). "Analysis and synthesis of the 3D movements of the head, face and hand of a speaker using cued speech." Journal of Acoustical Society of America 118(2): 1144-1153.
- Lee, S. P., J. B. Badler, et al. (2002). "Eyes alive." ACM Transaction on Graphics 21(3): 637-644.
- Moriyama, T., J. Xiao, et al. (2006). "Meticulously detailed eye model and its application to analysis of facial image." IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5): 738 - 752.
- Raidt, S., G. Bailly, et al. (2006). Does a virtual talking face generate proper multimodal cues to draw user's attention towards interest points? Language Resources and Evaluation Conference (LREC), Genova - Italy: 2544-2549.
- Revéret, L., G. Bailly, et al. (2000). MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. International Conference on Speech and Language Processing, Beijing - China: 755-758.
- Sagar, M. A., D. Bullivant, et al. (1994). A virtual environment and model of the eye for surgical simulation. SIGGRAPH, Orlando, FL: 205-212.
- Tateno, K., M. Takemura, et al. (2005). Enhanced eyes for better gaze-awareness in collaborative mixed reality. International Symposium on Mixed and Augmented Reality, Vienna, Austria: 100-103.