

PuppetFace: Contrôle et animation d'une tête parlante 3D en temps réel

Laboratoire : Inria Nancy Grand Est (LORIA)¹ - **Ville** : Nancy, France. - **Équipe** : Multispeech

Thématique : géométrie algorithmique

Nom et adresse électronique du directeur de stage : Slim Ouni (Slim.Ouni@loria.fr)²

Nom et adresse électronique du directeur du laboratoire : Jean-Yves Marion (Jean-Yves.Marion@loria.fr)

Présentation générale

Dans le domaine de la parole audiovisuelle, nous développons des têtes parlantes pour étudier comment rendre la communication audiovisuelle plus efficace. Nous rappelons qu'une tête parlante ou avatar est une tête virtuelle 3D animée simultanément avec de la parole acoustique. La parole acoustique peut être générée (c'est la synthèse de la parole) ou bien un signal préenregistré. Dans ce dernier cas, il s'agit du lipsync, où la parole acoustique contraint fortement le résultat de l'animation de la tête parlante. Nous pouvons également piloter une tête parlante par les gestes d'un acteur (ou locuteur). Dans ce cas, il s'agit de mimer les gestes faciaux du locuteur et de les reproduire sur la tête parlante. Dans cette dernière situation, il n'y a aucun traitement audio à faire, mais l'aspect temporel est très important, pour avoir un système réactif.

Nous avons développé un algorithme qui permet d'animer une tête parlante avec un nombre très limité de points sur le visage du locuteur. Ces points peuvent correspondre à des marqueurs collés sur le visage, ou bien des points références obtenues grâce à une méthode de suivi de mouvement sans capteurs.

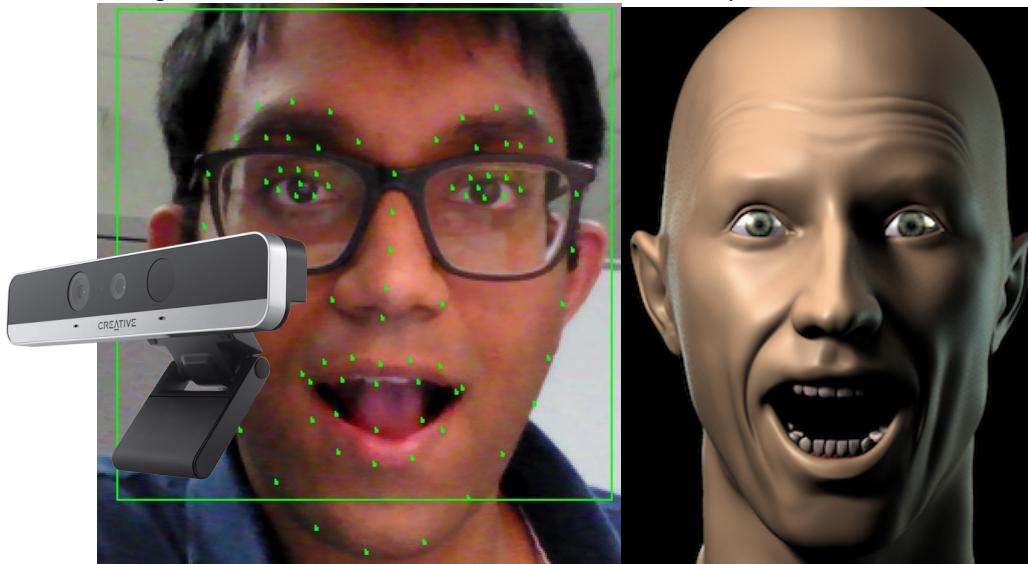


Fig. 1 Système de capture de mouvement Intel RealSense. Suivi des mouvements des gestes du visage sans capteurs. L'objectif du stage est de contrôler le visage d'une tête parlante à partir des gestes d'un vrai locuteur.

Objectifs du stage

L'objectif du travail est de développer un système de « puppetry », c'est à dire, un système qui permet d'animer le visage d'une tête parlante virtuelle 3D en temps réel à partir des gestes du visage d'un locuteur réel. Nous focalisons sur les gestes liés à la parole, c'est à dire les gestes de la bouche et la mâchoire. Nous allons utiliser un système de motion capture sans marqueurs et très accessible, comme la *kinect* (dans ce travail, nous utiliserons, Intel RealSense) pour suivre les mouvements du visage du locuteur (voir Fig. 1). Pour chaque frame, le système fournit un ensemble de points correspondants au visage dans l'espace 3D. Ces points seront projetés sur une tête parlante virtuelle 3D et ils seront utilisés pour l'animer en temps réel grâce à l'algorithme qui a déjà été développé dans l'équipe.

L'aspect « temps-réel » est important pour ce type d'application.

N'hésitez pas à contacter le responsable du stage pour tout complément d'information.

Compétences espérées

¹ <http://www.loria.fr> <http://www.inria.fr/centre/nancy> Nancy - est à 1h30 de Paris par TGV.

² <http://www.loria.fr/~slim>

De bonnes connaissances en géométrie algorithmique, en analyse de données ou/et en machine learning seront appréciés.

Contexte de travail :

Le travail sera fait au sein d'une équipe de recherche dynamique (Equipe de recherche Multispeech). Vous effectuerez votre stage au sein du centre de recherche Inria Nancy Grand Est (LORIA) et vous serez dans un bureau entouré par des chercheurs confirmés et de jeunes chercheurs (doctorants, postdoctorants et ingénieurs) et suivi de près par le tuteur du stage.

Bibliographie

1. Yilong Liu, Feng Xu, Jinxiang Chai, Xin Tong, Lijuan Wang, and Qiang Huo. 2015. Video-audio driven real-time facial animation. *ACM Trans. Graph.* 34, 6, Article 182 (October 2015), 10 pages. DOI=<http://dx.doi.org/10.1145/2816795.2818122>.
2. Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R. W., and Gross, M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics (TOG)* 30, 4, 75.
3. Bouaziz, S., Wang, Y., and Pauly, M. 2013. Online modeling for realtime facial animation. *ACM Transactions on Graphics (TOG)* 32, 4, 40.
4. Cao, Y., Tien, W. C., Faloutsos, P., and Pighin, F. 2005. Expressive speech-driven facial animation. *ACM Trans. Graph.* 24, 4 (Oct.), 1283--1302.
5. Hsieh, P.-L., Ma, C., Yu, J., and Li, H. 2015. Unconstrained realtime facial performance capture. In *Computer Vision and Pattern Recognition (CVPR)*.