Animation faciale basée données : un état de l'art

Clément Reverdy, Sylvie Gibet, Caroline Larboulette

Animation faciale basée données : un état de l'art

Clément Reverdy, Sylvie Gibet, Caroline Larboulette

IRISA
Université de Bretagne Sud

Résumé

Cet article dresse un panorama des différentes problématiques liées à l'animation faciale basée données et présente les dernières avancées et solutions proposées par l'état de l'art. Le but de l'animation basée données est d'animer des personnages virtuels reproduisant les actions effectuées par des acteurs humains. Dans ce contexte, le visage joue un rôle prépondérant puisqu'il est l'un des principaux vecteurs de l'émotion et de la communication chez l'humain. Par ailleurs, contrairement au reste du corps dont les mouvements sont contraints par des articulations et des os, les déformations du visage suivent une autre forme de dynamique ce qui en fait un cas d'étude à part. Les applications sont diverses, par exemple, la création d'avatars virtuels ou l'animation de personnages présentant un comportement naturel. Dans ce papier, nous aborderons la question depuis la capture des données faciales (les différents dispositifs et méthodes de capture) jusqu'aux méthodes de synthèse exploitant ces données.

Mots clé : Animation faciale, Blendshapes, Modèles à couches fines, Capture de mouvement (*MoCap*)

1. Introduction

L'animation faciale basée données est un sujet qui a gagné en intérêt au cours des dernières années. Il s'agit d'animer des avatars virtuels à partir de données réelles. Les enjeux sont divers. Les industries audiovisuelles et vidéoludiques notamment sont demandeuses de ce type de technologies car elles permettent de générer des animations d'autant plus crédibles aux yeux du public qu'elles sont produites à partir de comportements humains réels. Parmi les applications possibles on peut citer la communication en temps réel d'humains à humains anonymisée par le biais de ces avatars, ou bien encore la création d'agents virtuels présentant un comportement humain pour des interfaces humains-machine.

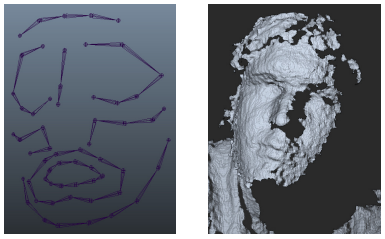
Les quinze dernières années ont vu se développer successivement deux principaux types de technologies de capture de données $3D$: (i) La *Motion Capture (MoCap)* qui consiste à suivre par triangulation les positions $3D$ d'un nombre limité de marqueurs disposés sur le corps d'un acteur (et, dans le cas qui nous intéresse, le visage) ; (ii) et plus récemment, les technologies de types *RGB-D* (ou caméras à capteurs de profondeur) reposant sur la vision binoculaire ou l'émission de lumière structurée.

Dans la mesure où les dispositifs de type *RGB-D* ont émergé sur le marché grand publique à des coûts relativement faibles (e.g., Kinect), nous avons fait le choix de ne pas explorer davantage les méthodes reposant sur l'acquisition de données via des vidéos *2D* classiques. Néanmoins, il est intéressant de remarquer un certain nombre de travaux [CHZ14,CWZ*14,CWLZ13] visant à inférer des formes $3D$ à partir de vidéos *2D* via des méthodes d'apprentissage statistiques.

Parmi les méthodes d'animation faciale, celles basées *blendshapes* restent très populaires de part leur facilité d'utilisation. Elles s'appuient sur des formes de base qui étaient initialement définies par des graphistes. Aujourd'hui, comme nous le verrons dans cet article, des méthodes permettent d'automatiser cette étape. Plus récemment, d'autres méthodes (notamment les modèles à couches fines) ont vu le jour et se sont développées. Elles permettent d'appliquer au maillage des déformations avec davantage de degrés de libertés que ceux permis par les modèles de type *blendshapes*. Quel que soit le choix du type de modèle, le lien entre les données sources et les modèles est essentiellement assuré par des techniques d'optimisation.

En dehors des modèles strictement géométriques, des méthodes dites physiques (e.g., reposant sur des systèmes masse-ressort) ainsi que des techniques (aussi bien géométriques que physiques) ayant vocation à représenter des modèles anatomiques [SNF05,HWHM15,LTW95] ont aussi été développées. Néanmoins, ce type de méthode n'est pas le plus utilisé.

En dépit du fait que l'animation faciale basée données ait gagné en maturité notamment grâce à ces dernières innovations, il n'existe que très peu d'états de l'art traitant spécifiquement de ce sujet. À notre connaissance, l'étude la plus complète et pertinente est proposée par [DN08]. Néanmoins, les dernières avancées technologiques (en particulier les dispositifs de type *RGB-D*) et les possibilités qu'elles offrent n'y sont pas évoquées.



(1) (2)

Figure 1: *Représentation des données capturées via (1)* MoCap *ou (2) dispositif* RGB-D *(nuage de points uniquement).*

Cet article vise à fournir une introduction à l'animation faciale basée données ainsi qu'une source de références vers les dernières tendances du domaine. Dans un premier temps nous présenterons les principaux dispositifs de capture de type *MoCap* et *RGB-D*, les principes généraux sur lesquels ils reposent et leurs principaux problèmes (section 2). Dans un second temps, les principales méthodes d'animation à partir de données sources capturées sont présentées (section 3). Enfin une discussion concernant les avantages et inconvénients liés à ces différentes techniques est proposée (section 4).

2. Acquisition des données sources

Dans cette section, nous discuterons de deux principales méthodes de capture (figure 1) : la capture de mouvement basée marqueurs (*MoCap*) et la capture sans marqueur via des dispositifs de type *RGB-D*.

2.1. Capture de mouvement 3*D* basée marqueurs

La *MoCap* est une technique couramment utilisée (tant pour la capture du visage que du corps de l'acteur) par l'industrie audiovisuelle. Il s'agit de suivre les déplacements de marqueurs placés sur le corps d'un acteur à l'aide d'un ensemble de caméras ; chaque marqueur doit être visible par au moins deux caméras pour que sa position exacte puisse être connue à un instant donné. Cette technologie permet de capturer le mouvement de ces marqueurs avec une précision spatiale et une fréquence temporelle élevées (respectivement ≤ 1 mm et ≥ 120 Hz). Néanmoins la résolution spatiale de ce type de capture est faible dans la mesure où le nombre de marqueurs disposables sur le visage de l'acteur est limité. Les données obtenues sont donc décrites par un nuage de K points évoluant au cours du temps.

En général, pour la capture de mouvement non-faciale, le positionnement des marqueurs est choisi afin de pouvoir inférer les paramètres de rotation et les positions des différentes articulations du squelette selon des modèles biomécaniques. En revanche, en ce qui concerne le visage, le choix du nombre de marqueurs et surtout leurs positionnements n'ont à ce jour fait l'objet que de très peu d'études [LZD13] [RGL15]. Dans les faits, le nombre de marqueurs varie globalement d'une trentaine de marqueurs (disposition Face Robot du logiciel commercial Softimage d'Autodesk) à une centaine [DCFN06, HCTW11]. Augmenter le nombre de marqueurs aura pour effet d'augmenter la quantité d'information spatiale capturée concernant la déformation du visage de l'acteur à chaque *frame* et donc de rendre plus efficace et plus fidèle la synthèse des expressions faciales. Néanmoins cela a un coût, que ce soit en ce qui concerne le post-traitement des données (e.g., inversion de marqueurs dans le suivi des trajectoires) ou la préparation des séances de capture (e.g., risque d'oubli / de chute / mauvais placement d'un marqueur). Il faut également tenir compte du fait que l'information n'est pas uniformément répartie sur l'ensemble du visage et que celle portée par certains points du visage peut-être redondante avec celle portée par d'autres points ; une répartition intelligente des marqueurs est donc susceptible d'améliorer significativement la qualité des données capturées. À cela s'ajoute le besoin réel de ces informations, par exemple, si le modèle d'animation ciblé est relativement simple, s'il n'a qu'un nombre de degrés de liberté limité comme c'est le cas pour un modèle basé *blendshapes* (voir section : 3) où l'espace de représentation est celui des combinaisons linéaires de ses bases (une cinquantaine dans le cas où ses bases sont calquées sur le FACS), alors un nombre restreint de marqueurs bien placés peut suffire.

Par ailleurs, si la *MoCap* permet de capturer efficacement les déformations à large échelle liées aux expressions faciales et aux mouvements labiaux, du fait de sa faible résolution spatiale, elle échoue à capturer à elle-seule les déformations plus fines telles que celles liées aux rides. Il est néanmoins possible de contourner ce problème en faisant appel à des solutions hybrides. Dans [BBA*07] les auteurs ont, d'une part, incorporé à leur modèle facial cible un modèle permettant de représenter les déformations liées aux rides et, d'autre part, ajouté à leur dispositif de capture des caméras haute-résolution afin de détecter les rides et en inférer les paramètres pour leur modèle. Dans [HCTW11], les détails fins sont aussi pris en charge par le modèle (basé *blendshapes*) dont les bases ont été générées à partir de captures haute résolutions effectuées via un scanner laser.

2.2. Capture de mouvement 3*D* sans marqueur type *RGB-D*

Nous parlerons ici des méthodes de capture 3*D* sans marqueurs qui reposent principalement sur deux technologies, la lumière structurée et la stéréovision. Les enjeux de ce type de capture sont d'une part de capturer un nuage de points autour de la surface du visage, d'autre part d'acquérir un maillage 3*D* représentant le visage de l'acteur avec une haute résolution à partir de ce nuage de point à un instant donné, et enfin de suivre les déformations de ce maillage au cours du temps de façon à préserver une structure (ensemble de sommets + liens de connexité) commune au cours du temps.

2.2.1. Obtention d'un nuage dense de points

La lumière structurée [RHHL02] [ZH04] consiste à projeter un motif lumineux ayant des caractéristiques connues, l'analyse de la déformation de ce motif sur la surface des objets de la scène permet d'obtenir des informations sur la géométrie de ces objets. La figure 2 illustre un exemple de ce type de dispositif. Un état de l'art sur les différents motifs pouvant être utilisés a été réalisé par Salvie et al. [SFPL10],

Zhang Song [Zha10] a réalisé un état de l'art spécifique aux motifs de franges (e.g., sinusoïdes).

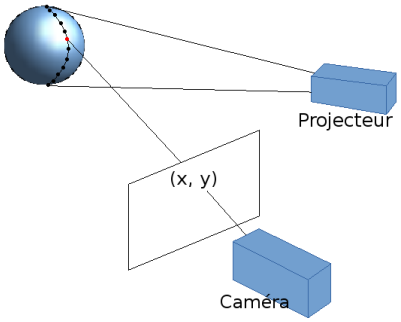


Figure 2: *Illustration de l'estimation de la position d'un point dans l'espace par projection d'une lumière structurée (ici le motif est une bande). La position du point sera déterminée par l'intersection entre la droite partant de la caméra et le plan de la bande de lumière projetée.*

La stéréovision consiste à capturer la scène sous plusieurs angles de vue simultanément. La figure 3 illustre ce type de dispositif, sachant qu'un couple de pixels P_1 et P_2 sont les projections d'un même point X sur deux caméras différentes dont on connait les centres de projections O_1 et O_2. La position $3D$ du point X est à l'intersection des droites O_1P_1 et O_2P_2. Pour un pixel P_1 donné, l'intégralité des pixels P_2 pouvant être un projeté de X sont situés sur la droite épipolaire, il s'agit alors de trouver le pixel le plus similaire parmi ces derniers. Un exemple de mesure de similarité pouvant être employée pour évaluer la correspondance possible entre deux pixels est la corrélation croisée des pixels P_1 et P_2 sur leurs voisinages. Scharstein et Szeliski [SS02] ont réalisé une taxonomie complète des différentes stratégies pour la mise en correspondance de couples d'images.

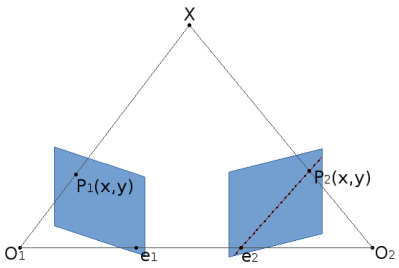


Figure 3: *Illustration de l'estimation de la distance d'un point par vision binoculaire. O_1 et O_2 sont les centres de projection des deux caméras, e_1 et e_2 sont les intersections entre la droite O_1O_2 et les plans de projection de chacune des caméras. Le point P_1 correspond à la projection du point X sur le plan de la caméra* 1. *La droite e_2P_2 (appelée ligne épipolaire) est l'intersection entre le plan XO_1O_2 et le plan de projection de la caméra* 2. *Le point P_2 est le pixel le plus similaire à P_1 situé sur la ligne épipolaire. Les droites O_1P_1 et O_2P_2 sont situées sur un plan commun et s'intersectent au point X.*

Les techniques basées sur la projection de lumière structurée ont longtemps permis d'obtenir des résultats plus robustes (l'éclairage permet d'injecter de l'information connue) que les méthodes basées strictement sur la stéréovision. Néanmoins les travaux récents de Bradley et al. en 2008 [BBH08] puis de Beeler et al. [BBB*10] ont montré qu'il est possible d'obtenir des résultats comparables avec des dispositifs de capture totalement passifs (c.à.d. sans émission de lumière). L'un des points forts de [BBB*10] est sa robustesse ; le système est conçu selon un mode pyramidal, la mise en correspondance des pixels est réalisée de façon progressive, d'abord à basse résolution puis en l'augmentant à chaque niveau, des contraintes de cohérence sont utilisées à chacun de ces étages ainsi qu'un système de raffinement progressif.

D'autres approches hybrides ont été proposées. Dans Zhang et al. [ZSCS04] la projection de motifs sert essentiellement à ajouter des variations de couleurs sur le visage et donc faciliter l'évaluation de la correspondance entre les pixels de chacune des caméras. Weise et al. [WLVG07] ont proposé une méthode d'acquisition tirant parti à la fois de la projection de motifs lumineux et de la stéréovision.

2.2.2. Post-traitements

L'inférence des coordonnées $3D$ associées à chaque pixel peut être sujet à des erreurs. Un certain nombre de traitements peuvent être effectués afin d'améliorer la qualité des résultats, réduire la redondance des informations, filtrer les erreurs aberrantes.

Aligner différents nuages de points est une tâche qui peut s'avérer nécessaire. Pour une modélisation complète du visage, il est utile de pouvoir le capturer sous plusieurs angles de vue. Soit en ayant plusieurs dispositifs capturant simultanément la scène sous différents angles de vue [BBB*10], soit en ayant un appareil unique capturant le visage sous différentes poses [RHHL02] [WBLP11]. Les différences (angle et translation) entre chacune de ces prises de vue ou poses ne sont pas forcément connues. L'algorithme le plus fréquemment utilisé pour retrouver ces paramètres de transformation rigide est l'ICP (*Iterative Closest Points*). Brièvement, il s'agit d'un algorithme itératif visant à aligner un nuage de point de référence avec un autre nuage de points en estimant successivement pour chaque point du nuage de référence le plus proche voisin dans l'autre nuage connaissant une estimation des paramètres de transformation, puis à réestimer ces paramètres afin d'aligner ces couples jusqu'à convergence. Dans [RL01], les auteurs proposent différentes variantes de cet algorithme.

Le filtrage des anomalies est une tâche qui revient régulièrement dans la littérature. Suivant le dispositif de capture et la méthode de détermination de la profondeur choisie, la génération de points aberrants peut avoir des origines différentes. Par exemple avec une méthode par décalage de phase dans un motif sinusoïdal projeté, on peut avoir des erreurs générées par l'ambiguïté liée à la périodicité [WLVG07] [ZH04]. Dans le cas d'une reconstitution par stéréovision, les erreurs peuvent être dues à un calcul de disparité erronné à cause d'une mauvaise mise en correspondance de pixels.

Avant même de filtrer le nuage de points, le nombre de points aberrants peut être fortement réduit par l'ajout de contraintes lors de sa génération assurant ainsi sa cohérence.

Les anomalies restantes doivent être détectées, pour cela il convient de définir ce qui caractérise un point aberrant et définir des contraintes afin de les discriminer. Dans [WLVG08] et [MAW*07], les auteurs tirent parti des informations collectées depuis les différents points de vue afin de détecter des incohérences. Par exemple une erreur survient lorsqu'une caméra détecte un point qui devrait normalement avoir été masqué par un autre point détecté par une autre caméra. Dans [BBH08] un critère basé sur la distance par rapport au plan approximant ses voisins (estimable par la méthode des moindres carrés) est utilisé afin de détecter les anomalies (dépassant un certain seuil), d'autres critères sont proposés dans [WPK*04].

Sous-échantillonner le nuage points peut parfois être utile, afin de réduire la redondance d'informations [BBH08] ou pour accélérer les temps de traitement (pour permettre un rendu temps réel par exemple) [RHHL02]. Dans le premier cas une structure arborescente tirée de [BHGS06] a été utilisée, dans le second cas les points redondants ont été fusionnés via une grille de voxels.

3. Animation

Dans cette section, nous explicitons le formalisme général utilisé dans l'article puis nous décrivons les principales méthodes utilisées dans la littérature.

3.1. Formalismes

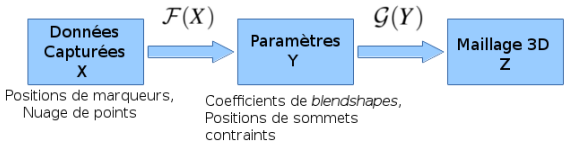


Figure 4: *Des données sources au maillage final.*

Le moyen le plus commun de représenter le visage comme la plupart des objets $3D$ est sous la forme d'un maillage $3D$ de polygones (souvent triangulaires). L'ensemble des sommets et leurs liens de connexité forment la structure géométrique du maillage. En règle générale, cette structure reste constante, en revanche les coordonnées de chaque sommet peuvent être modifiées afin de déformer le maillage à des fins d'animation.

Le but de l'animation basée données est de produire une séquence $Z = (Z_1, \cdots, Z_n, \cdots, Z_N)$ de N "poses", où chaque pose $Z_n = (Z_n^1, \cdots, Z_n^v, \cdots, Z_n^V)$ est le vecteur de dimension $3 \times V$ des positions $3D$ des sommets du maillage de la $n^{ième}$ pose.

Ces positions peuvent être déterminées à partir de paramètres représentés par le vecteur $Y = (Y_1, \cdots, Y_n, \cdots, Y_N)$ où chaque vecteur $Y_n = (Y_n^1, \cdots, Y_n^p, \cdots, Y_n^P)$ de dimension P est le vecteur des paramètres du modèle à la $n^{ième}$ pose. La relation entre Y et Z est donnée par le modèle représenté par la fonction :

$$\mathcal{G}(Y) = Z : \mathbb{R}^P \mapsto \mathbb{R}^V \tag{1}$$

Dans le cadre de l'animation basée données, on cherche à déterminer automatiquement les paramètres Y à partir des données $X = (X_1, \cdots, X_n, \cdots, X_N)$ avec $X_n = (X_n^1, \cdots, X_n^d, \cdots, X_n^D)$ le vecteur de dimension D des données capturées à la $n^{ième}$ pose. La fonction recherchée est :

$$\mathcal{F}(X) = Y : \mathbb{R}^D \mapsto \mathbb{R}^P \tag{2}$$

Donc par composition on peut relier les données "sources" X aux données "cible" Z (figure 4) :

$$(\mathcal{G} \circ \mathcal{F})(X) = Z : \mathbb{R}^D \mapsto \mathbb{R}^V \tag{3}$$

3.2. Création d'un maillage $3D$

Il est possible, si l'on dispose d'un nuage suffisamment dense de points représentant la surface de l'objet de générer un maillage $3D$.

Ce maillage peut être construit directement. Bradley et al. en 2008 [BBH08] ont utilisé une triangulation de Delaunay, néanmoins cette méthode n'étant efficace que sur de petits ensembles de points, une classification non supervisée (*clustering*) a dû au préalable être effectuée suivant un octree dirigé par la densité ; la triangulation a ensuite été effectuée localement sur chaque cluster. [BBB*10] ont utilisé une méthode (*Poisson Surface Reconstruction*) proposée en 2006 par Kashdan et al. [KBH06].

Une autre solution consiste à déformer un maillage générique existant afin de coller au nuage de points. Classiquement, le problème est posé comme une optimisation visant à minimiser un terme d'ajustement des sommets du maillage par rapport au nuage de points et un terme de régularisation visant à éviter des déformations incohérentes. C'est le choix qui a été fait notamment dans [ZSCS04] mais aussi dans [WLVGP09] .

L'usage de scanner laser est assez fréquent. En demandant à un acteur de conserver une expression fixe le temps d'un scan, il est possible d'obtenir un modèle fidèle de son visage. Si cette méthode ne permet d'obtenir qu'une modélisation statique du visage de l'acteur, le maillage obtenu peut ensuite être déformé à partir de données capturées.

Le *Morphable Model* proposé dans [BV99] est fréquemment utilisé afin de générer des maillages 3D. Il s'agit d'un système basé Analyse en Composante Principale (ACP) avec une segmentation éventuelle des différentes parties du visage. Une ACP est calculée sur une base de données de maillages $3D$ correspondant aux visages de personnes acquis via un scan $3D$. Les paramètres du modèle sont les coordonnées du maillage $3D$ dans le sous-espace défini par l'ACP. De nouveaux maillages peuvent être générés par optimisation de ces coefficients en minimisant l'écart entre la projection du maillage généré dans l'espace des données (X) et les données effectivement capturées.

3.3. Animation à base de formes clés (*Blendshapes*)

Dans le cadre de l'animation basée *blendshapes*, la fonction $\mathcal{G}$ est une combinaison linéaire de P formes de bases,

chacune pondérée par un coefficient w_p :

$$\mathcal{G}(Y) \equiv b_0 + \sum_{p=1}^{P} w_p b_p \tag{4}$$

avec b_0 représentant l'expression neutre (c.à.d. les coordonnées initiales des sommets) et b_p la $p^{ième}$ base du modèle. Avec une notation matricielle, on aura donc :

$$\mathcal{G}(Y) \equiv WB \tag{5}$$

avec W le vecteur de dimension P des coefficients w_p (avec $w_0 = 1$) et B la matrice de dimension $(P \times V)$ des sommets des bases b_p.

Les paramètres (Y) du modèle sont donc d'une part l'ensemble des bases b_p et d'autre part les coefficients w_p associés. Néanmoins, les bases sont en général déterminées une fois pour toute et restent inchangées par la suite. Les seuls paramètres évoluant au cours du temps sont donc les coefficients w_p.

Ce type de modèle présente deux principaux avantages. D'une part il offre un niveau d'abstraction supplémentaire facilitant le transfert d'animations d'un modèle à l'autre. En effet, pour deux maillages cibles différents, même topologiquement, si les bases associées à chacun d'entre eux représentent des expressions identiques, alors le transfert des animations peut se faire directement en prenant le même ensemble de coefficients. D'autre part, il s'agit d'un modèle linéaire relativement simple, ce qui permet des temps de calcul compatibles avec des applications en temps réel [WBLP11].

Les deux principales problématiques liées aux *blendshapes* seront traitées dans les sous-sections suivantes.

3.3.1. Choix et génération des bases

Le choix des bases b_p doit répondre à plusieurs problématiques. En premier lieu, elles doivent être aussi indépendantes que possible les unes des autres de façon à faciliter l'éventuelle édition [LD08] ultérieure par des animateurs et ainsi éviter que lors du processus d'optimisation des coefficients w_p certains prennent des valeurs inférieures à 0 ou supérieurs à 1 pour compenser des déformations liées à d'autres bases (en général, les bases sont construites de telle sorte que le coefficient qui lui est associé doit être compris entre 0 et 1). En second lieu, l'espace de représentation défini par ces bases doit couvrir autant que possible celui de l'expressivité humaine. Enfin, il est préférable que les déformations liées à chacune de ces bases aient une signification particulière (e.g., *sourire*, *fermeture d'un oeil*, *ouverture de la bouche*, etc) afin d'être plus facilement manipulables par un animateur.

Le *Facial Action Coding System* (FACS) est un choix de bases assez fréquent. Originellement proposé par Ekman en 1978 [EF78] afin de décrire les expressions faciales humaines, il propose une cinquantaine d'Action Unitaires (AU) correspondant chacune à l'activation d'un muscle ou d'un groupe de muscles. Créer une base correspondante à chacune de ces AU [WBLP11] [CWZ*14] est un choix qui présente l'avantage d'être assez réaliste tout en facilitant l'édition.

Dans [HCTW11] [XCLT14], les bases sont sélectionnées parmi les *frames* d'une séquence capturée via *MoCap* ou d'une animation existante. L'idée étant de sélectionner un nombre minimal de *frames* à partir desquelles il est possible de minimiser sur l'intégralité des *frames* de la séquence l'erreur de reconstruction par rapport aux données capturées.

L'Analyse en Composantes Principales (ACP) est une méthode parfois utilisée pour créer les bases du modèle basé *blendshapes* [CB02] [WLVGP09] [LD08]. Elle vise à déterminer un ensemble de P vecteurs orthogonaux (appelés composantes principales) m_p, tels que la variance des projections des observations dans le sous-espace défini par les vecteurs m_p est maximisé. Ainsi toute observation m peut être représentée sous la forme :

$$m = m_0 + \sum_{p=1}^{P} \alpha_p \times m_p \tag{6}$$

avec α_p les coordonnées dans l'espace de l'ACP et m_0 le vecteur moyen des observations. Cette représentation est assez proche du modèle de *blendshapes* puisque chaque observation est représentée par une combinaison linéaire de vecteurs. Dans [CB02] les auteurs ont choisi pour bases les observations les mieux représentées sur chacun des axes de l'ACP. Dans [LD08], une méthode de création et manipulation de *blendshapes* basée ACP permettant de faciliter l'édition est proposée.

Lorsqu'elles ne sont pas obtenues directement par ACP, les bases peuvent être créées manuellement ou générées à partir d'expressions humaines réelles capturées via un dispositif de capture $3D$ [CBK*06] [HCTW11] [WBLP11] [CWZ*14]. On ne dispose pas toujours d'une capture par base, par exemple, dans le cas du FACS la plupart des humains ne parviennent pas à contrôler chacune des AU indépendamment. La méthode proposée dans [LWP10] permet d'inférer les bases à partir d'une série d'exemples d'expressions faciales quelconques pour lesquelles les degrés d'activation w_p de chacune des bases sont connus. Cette méthode a notamment été exploitée dans [WBLP11] et [CWZ*14]. La méthode de transfert de déformations de [SP04] (section 3.6) a été utilisée dans [XCLT14] afin de générer un jeu de bases associées à un maillage cible à partir d'un jeu de bases associées à un premier maillage source. Dans [WLVGP09], les bases sont déterminées linéairement par résolution au sens des moindres carrés à partir d'un ensemble d'expressions pour lesquelles les déformations du maillage et les coefficients de *blendshapes* étaient connus.

3.3.2. Détermination des coefficients

Étant donné un maillage et un ensemble de bases du modèle *blendshapes* associé, l'obtention des coefficients w_p à partir de données sources X revient à un problème d'optimisation. Il s'agit de minimiser une énergie E_{match} telle que :

$$\mathcal{F}(X) = Y^* = arg\min_{Y} E_{match} \tag{7}$$

Le plus souvent le problème est formulé au sens des moindres carrés :

$$E_{match} = \sum_{d=0}^{D} \|WB(v_d) - x_d\|^2 \tag{8}$$

avec v_d, le sommet associé à la donnée capturée x_d et $B(v_d)$ le vecteur de dimension P des déformations de chaque base appliquée au vertex v_d correspondant à la donnée capturée $x_d \in X$. Il s'agit d'un problème d'optimisation linéaire pouvant être résolu de façon analytique.

Formaliser le problème au sens des moindres carrés n'est pas la seule option. Par exemple le problème peut être abordé avec une approche bayésienne [WBLP11]. Dans ce cas le problème est formulé comme une MAP (maximisation de l'estimation *a posteriori* [Her04]) :

$$p(Y|X) = \frac{p(X|Y) \times p(Y)}{p(X)} \tag{9}$$

avec $p(Y|X)$ la probabilité a posteriori, $p(X|Y)$ la vraisemblance des données X connaissant les paramètres Y et $p(Y)$ la fonction de distribution *a priori* des paramètres Y. La probabilité $p(X)$ étant indépendante de Y, on peut se contenter de chercher :

$$Y^* = arg\max_Y p(X|Y) \times p(Y) \tag{10}$$

Cela peut se ramener à un problème de minimisation d'énergie tout en présentant l'avantage de permettre une interprétation bayésienne :

$$Y^* = arg\min_Y -ln(p(X|Y)) - ln(p(Y)) \tag{11}$$

D'une manière générale, le problème n'est pas toujours posé de façon linéaire et doit être résolu par des méthodes d'optimisation non-linéaires itératives (par exemple : *descente de gradient*, *Gauss-Newton*, *gradient conjugué*, etc).

Un problème qui peut se poser vient du fait que l'optimisation posée en l'état ne tient compte ni de la cohérence temporelle des coefficients trouvés (les coefficients doivent être cohérents les uns par rapport aux autres au cours des *frames* successives) ni de la cohérence spatiale (les coefficients doivent rester autant que possible entre 0 et 1). En effet, il n'est pas rare d'obtenir des coefficients très grands compensés par d'autres coefficients inférieurs à 0 [CB02]. Une première méthode permettant d'éviter cet effet indésirable est d'ajouter une contrainte de non-négativité dans le processus d'optimisation. Ainsi [XCLT14, HCTW11, CB02] ont utilisé le *Non-Negative Least Square Solver* (NNLS) proposé par [LH74]. On peut aussi adjoindre une seconde énergie E_{reg} de régularisation à minimiser, l'équation (7) devient alors :

$$\mathcal{F}(X) = Y^* = arg\min_Y(E_{match} + \alpha E_{reg}) \tag{12}$$

avec α un paramètre permettant de pondérer le degré de régularisation et E_{reg} une énergie pénalisant les valeurs de Y improbables. Dans [SLS*12] [CWLZ13] [WBLP11], cette énergie de pénalisation est assimilée à une fonction de densité de probabilité a priori $p(Y)$ dont les paramètres ont déjà été calculés sur des données d'entraînement.

Deng et al. ont proposé en 2006 [DCFN06] une approche consistant à entraîner à partir de L couples d'entraînement (X_l, W_l) un régresseur RBF (Radial Basis Function network) inférant les coefficients w_i de données sources quelconques X. Il s'agit alors de construire la fonction $\mathcal{F}(X)$ sous cette forme :

$$Y* = \mathcal{F}(X) = \sum_{l=0}^{L} c_l \phi_l(X) \tag{13}$$

avec c_l les paramètres de $\mathcal{F}(X)$ que l'on va chercher à estimer et ϕ_l une fonction noyau (par exemple gaussien $\phi_l(X) = exp(-\|X - X_l\|^2/2\sigma_l^2)$), les paramètres $C = (c_0, c_1, ..., c_L)$ seront optimisés selon le principe des moindres carrés à partir des exemples d'entraînement :

$$C^* = arg\min_C \sum_{l=0}^{L} (w_l - \mathcal{F}(X_l))^2 + \alpha \sum_{l=0}^{L} c_l^2 \tag{14}$$

$\alpha \sum_{l=0}^{L} c_l^2$ étant un terme de régularisation pénalisant les paramètres trop importants. Néanmoins pour appliquer cette méthode, il est nécessaire de se procurer au préalable les couples (X_l, W_l).

3.4. Modèles à couches fines

Le but de ce type de méthode consiste à trouver à partir d'un sous-ensemble de sommets d'un maillage les coordonnées $3D$ de chacun de ses autres sommets en minimisant la déformation de la surface initiale. La déformation est décrite par deux termes, l'énergie d'étirement et l'énergie de torsion. Le problème est détaillé dans [BS08], brièvement il est montré que minimiser cette énergie revient à résoudre l'équation aux dérivées partielles d'Euler-Lagrange suivante :

$$-k_s \Delta d + k_b \Delta^2 d = 0 \tag{15}$$

avec k_s, k_b des paramètres permettant de contrôler la raideur de la surface, d la déformation de la surface par rapport à la position au repos et Δ et Δ^2 respectivement l'opérateur laplacien et bilaplacien ($\Delta^2 d = \Delta(\Delta d)$). D'un point de vue numérique, l'opérateur laplacien doit être discrétisé afin de s'appliquer à un maillage triangulaire. Cette discrétisation revient à calculer le laplacien $\mathcal{L}$ du maillage : $\mathcal{L} = \mathcal{M}^{-1}\mathcal{L}^*$ avec $\mathcal{M}$ diagonale et $\mathcal{L}^*$ symétrique les matrices de normalisation associées respectivement à chaque sommet et à chaque arête (en général calculés selon la discrétisation par cotangeantes [MDSB03]). Si on reprend l'équation (15), en prémultipliant par $\mathcal{M}$ on obtient :

$$(-k_s \mathcal{L}^* + k_b \mathcal{L}^* \mathcal{M}^{-1} \mathcal{L}^*) d = 0 \tag{16}$$

En passant les colonnes correspondantes aux sommets dont les positions sont contraintes à droite de l'équation puis en supprimant les lignes correspondantes, on obtient :

$$ad^* = b \tag{17}$$

où d^* est le vecteur des déplacements que l'on souhaite calculer, a, b ne dépendent que du laplacien du maillage et des positions connues de sommets et d^* est le vecteur des positions de sommets inconnues.

Si l'on se rapporte au formalisme proposé précédemment, les paramètres du modèle $\mathcal{G}(Y) \equiv d^* = (a^T a)^{-1} b$ (a^T étant la transposée de a) sont d'une part les matrices $\mathcal{M}$ et $\mathcal{L}^*$ calculées en une seule fois sur le maillage initial, et d'autre part, les positions des sommets contraints évoluant au cours du temps. Si les sommets contraints demeurent toujours les

mêmes, alors il suffit de calculer $(a^T a)^{-1}$ une seule fois, le reste du calcul pouvant quant à lui s'effectuer en des temps compatibles avec du temps réel. Si la liste des sommets contraints change, alors la structure de a est changée et il est nécessaire de recalculer l'inversion de $(a^T a)$ via une décomposition de Cholesky. Lorsque les données sont capturées via *MoCap* et que le maillage est similaire au visage de l'acteur, on peut associer à chaque marqueur le sommet qui lui correspond sur le maillage, ces sommets correspondront aux sommets contraints dont les positions sont déterminées par les positions des marqueurs.

Le modèle à couches fines a notamment été employé dans [BBA*07] et [LZD13] afin d'animer directement les maillages 3*D* à partir des positions des marqueurs *MoCap*. [WLVGP09] s'est également servi de ce type de modèle comme énergie de régularisation dans son processus d'optimisation.

3.5. Flux optiques

Lorsque l'on travaille avec des dispositifs de type *MoCap* sans marqueurs (stéréovision, lumière structurée), il est possible d'utiliser des méthodes de flux optique afin de propager les déformations d'un maillage de référence au cours du temps.

Dans [BHPS10] , un maillage initial G_t est d'abord calculé pour chaque *frame*. Ces différents maillages ne partagant pas une structure commune, le maillage de référence G'_t, calculé initialement comme étant G_0, est déformé à chaque *frame* t afin de s'approcher au plus près du maillage G_t puis repris afin de calculer G'_{t+1}. À chaque instant t et pour chaque sommet v du maillage G'_t, le flux du pixel $pi_{t,v,c}$ correspondant à la projection de ce sommet sur l'image de chaque caméra c est calculé afin d'obtenir une nouvelle position de ce pixel $pi_{t+1,v,c}$ à l'instant suivant. Cette nouvelle position est ensuite reprojetée sur la mesh G_{t+1} afin d'obtenir la position du sommet v du maillage G'_{t+1} à l'instant suivant.

Dans [ZPS04], à chaque *frame* un maillage générique est déformé afin de correspondre au visage de l'acteur. Cette déformation est calculée en minimisant une énergie $E = E_s + \alpha E_r + \beta E_m$, où E_s est une énergie de fitting liée à la distance entre chaque sommet du maillage générique et la *carte de profondeur* ; E_r est une énergie de régularisation visant à limiter les grands déplacements entre sommets voisins (les sommets voisins doivent présenter des déformations similaires), enfin, E_m est une énergie limitant la différence entre le déplacement du pixel $pi_{c,v,t}$ au pixel $pi_{c,v,t+1}$ (projections respectives du sommet v à l'instant t et $t+1$ sur la caméra c) et le déplacement du pixel $pi_{c,v,t}$ au pixel $pi'_{c,t+1}$ calculé par flux optique.

L'un des problèmes principaux lié à l'emploi de flux optique concerne la dérive numérique qui est liée à l'accumulation successive des erreurs au cours du temps. [BHB*11] propose de résoudre ce problème en "ancrant" les déformations temporelles successives sur un ensemble de poses de référence.

3.6. Transfert d'expressions

Le but du transfert d'expressions consiste à transférer les déformations appliquées à un maillage 3*D* source vers un autre maillage cible dont les proportions et la structure peuvent être différentes. Le premier papier à introduire ce terme est [NN01], qui propose une mise en correspondance des deux maillages par *interpolation RBF*. Une autre méthode fréquemment utilisée est celle proposée par [SP04] visant à transférer les déformations affines appliquées aux triangles du maillage source vers ceux du maillage cible tout en conservant la cohérence de ce dernier [WLVGP09, XCLT14, CWLZ13].

Les interpolations RBF sont fréquemment utilisées pour trouver les positions correspondantes sur une surface donnée de points situés sur une autre surface aux proportions différentes [BBA*07, SLS*12, NN01]. Par exemple, pour déterminer les positions correspondantes de marqueurs *MoCap* situés sur le visage d'un acteur sur un maillage 3*D* aux proportions différentes de celles de l'acteur.

Enfin [SCOL*04] propose un outil (*Laplacian Coating*) permettant de transférer les détails fins d'un maillage vers un autre maillage. Il ne s'agit pas de transfert d'expression à proprement parlé mais ce transfert de détails améliore sensiblement la crédibilité des animations produites lorsque le maillage animé n'est pas une représentation du visage de l'acteur [XCLT14].

4. Bilan et discussion

En ce qui concerne la capture de données faciales, les méthodes reposant sur des dispositifs *RGB-D* (lumière structurée, objectifs multiples) ont gagné de l'ampleur au cours de la dernière décennie, notamment grâce aux travaux de [ZSCS04], [BHPS10] et [BHB*11]. Ce nouveau type de technologie présente l'avantage d'offrir un degré de résolution spatiale très élevé si bien qu'il devient possible de reconstituer et suivre un maillage 3*D* avec un niveau de détails très élevé (de l'ordre des pores de la peau [BHB*11]) tout au long de la capture. Néanmoins ce type de dispositif demeure sensible aux mouvements brusques, ainsi qu'aux modifications de luminosité et le visage de l'acteur doit se trouver à une distance relativement courte de la ou des caméra(s). De plus les méthodes d'inférence de la profondeur des techniques *RGB-D* reposent souvent sur des méthodes de flux optiques ce qui les rend sensibles aux dérives numériques, aussi ne sont-elles pas recommandées pour la capture de longues séquences. Par ailleurs en terme de volume de données, là où la *MoCap* ne capture les positions que d'un nombre limité de marqueurs, les techniques de types *RGB-D* sont nettement plus coûteuses.

La capture de type *MoCap* permet de passer outre ces limitations. De plus, ce dispositif de capture est aussi bien adapté à la capture faciale que corporelle, cela simplifie le travail lorsque l'on souhaite une capture simultanée de ces deux types de mouvements, notamment en ce qui concerne la synchronisation des systèmes. Enfin, la fréquence élevée d'échantillonnage (nombre de *frames* capturées par seconde) des dispositifs de type *MoCap* peut se révéler particulièrement intéressante lorsque l'on cherche à étudier des mouve-

ments rapides tels que les micro-expressions. En définitive, si les méthodes de type *RGB-D* permettent de capturer efficacement et précisément les expressions faciales et que ces techniques continuent d'évoluer, il existe toujours un certain nombre de raisons justifiant l'emploi de la *MoCap*.

Les méthodes d'animation faciale à partir de données sont généralement formalisées à partir de techniques d'optimisation. Deux types de modèles émergent dans la communauté d'informatique graphique, qui sont représentatifs de deux philosophies différentes.

Le modèle à couches fines s'intéresse à la détermination optimale des positions de l'intégralité des sommets du maillage 3*D* à partir des positions d'un sous ensemble de sommets contraints. Ce type de modèle offre une grande liberté de déformation tout en préservant la cohérence du maillage. Par ailleurs, le problème est ramené à un système linéaire ce qui permet des temps de calculs compatibles avec du temps réel même si certaines précautions doivent être prises (à savoir, les sommets contraints doivent toujours rester les mêmes). Cependant, si ce type de méthode est adapté pour animer un maillage similaire au visage de l'acteur (pour de la *MoCap* par exemple, il suffit d'établir une correspondance entre chaque marqueur *MoCap* et un sommet du maillage de destination, puis de contraindre ces sommets afin qu'ils suivent les déplacements des marqueurs), transférer une expression faciale vers un maillage différent n'est pas aussi simple. Il est alors nécessaire d'établir une correspondance entre la structure du visage de l'acteur et celle du maillage cible (par exemple via une interpolation RBF). Une autre méthode consiste à réaliser préalablement l'animation d'un maillage similaire à celui de l'acteur, avant de transférer les déformations qui lui sont appliquées (par exemple via la méthode de [SP04]) ce qui complexifie le problème.

Le modèle basé *blendshapes* s'appuie quant à lui sur une représentation compacte du visage. Le calcul des paramètres et du maillage final à partir de ces derniers est rapide. Mais surtout le transfert d'une expression d'un maillage à l'autre peut se faire directement à condition que les bases du premier modèle représentent les mêmes expressions que les bases du second modèle. Néanmoins, il repose sur l'hypothèse selon laquelle les expressions faciales sont toutes représentables par combinaison linéaire d'un nombre limité d'expressions de base (ce qui n'est pas exact). Ainsi l'espace de représentation défini par ce type de modèle est plus restreint que l'espace défini par l'expressivité faciale réelle, il en résulte que les animations produites par *blendshapes* peuvent paraître légèrement moins naturelles (plus mécaniques) que celles produites via une méthode présentant un plus grand nombre de degrés de liberté.

Références

[BBA*07] BICKEL B., BOTSCH M., ANGST R., MATUSIK W., OTADUY M., PFISTER H., GROSS M. : Multi-scale capture of facial geometry and motion. *ACM Trans. Graph.* (2007).

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M. : High-quality single-shot capture of facial geometry. In *ACM SIGGRAPH 2010 Papers* (2010).

[BBH08] BRADLEY D., BOUBEKEUR T., HEIDRICH W. : Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008).

[BHB*11] BEELER T., HAHN F., BRADLEY D., BICKEL B., BEARDSLEY P., GOTSMAN C., SUMNER R. W., GROSS M. : High-quality passive facial performance capture using anchor frames. In *ACM SIGGRAPH 2011 Papers* (2011).

[BHGS06] BOUBEKEUR T., HEIDRICH W., GRANIER X., SCHLICK C. : Volume-Surface Trees. *Computer Graphics Forum* (2006).

[BHPS10] BRADLEY D., HEIDRICH W., POPA T., SHEFFER A. : High resolution passive facial performance capture. In *ACM SIGGRAPH 2010 Papers* (2010).

[BS08] BOTSCH M., SORKINE O. : On linear variational surface deformation methods. *Visualization and Computer Graphics, IEEE Transactions on* (2008).

[BV99] BLANZ V., VETTER T. : A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (1999).

[BV04] BOYD S., VANDENBERGHE L. : *Convex optimization*. Cambridge university press, 2004.

[CB02] CHUANG E., BREGLER C. : *Performance driven facial animation using blendshape interpolation*. Tech. rep., 2002.

[CBK*06] CURIO C., BREIDT M., KLEINER M., VUONG Q. C., GIESE M. A., BÜLTHOFF H. H. : Semantic 3d motion retargeting for facial animation. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (2006).

[CHZ14] CAO C., HOU Q., ZHOU K. : Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Trans. Graph.* (2014).

[CWLZ13] CAO C., WENG Y., LIN S., ZHOU K. : 3d shape regression for real-time facial animation. *ACM Trans. Graph.* (2013).

[CWZ*14] CAO C., WENG Y., ZHOU S., TONG Y., ZHOU K. : Facewarehouse : A 3d facial expression database for visual computing. *Visualization and Computer Graphics, IEEE Transactions on* (2014).

[DCFN06] DENG Z., CHIANG P.-Y., FOX P., NEUMANN U. : Animating blendshape faces by cross-mapping motion capture data. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games* (2006).

[DN08] DENG Z., NOH J. : Computer facial animation : A survey. In *Data-Driven 3D Facial Animation*. 2008.

[EF78] EKMAN P., FRIESEN W. : *Facial Action Coding System : A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978.

[HCTW11] HUANG H., CHAI J., TONG X., WU H.-T. :

Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. In *ACM SIGGRAPH 2011 Papers* (2011).

[Her04] HERTZMANN A. : Introduction to bayesian learning. In *ACM SIGGRAPH 2004 Course Notes* (2004).

[HWHM15] HUNG A., WU T., HUNTER P., MITHRARATNE K. : A framework for generating anatomically detailed subject-specific human facial models for biomechanical simulations. *The Visual Computer* (2015).

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H. : Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing* (2006), vol. 7.

[LD08] LI Q., DENG Z. : Orthogonal-blendshape-based editing system for facial motion capture data. *Computer Graphics and Applications, IEEE* (2008).

[LH74] LAWSON C. L., HANSON R. J. : *Solving least squares problems*. SIAM, 1974.

[LTW95] LEE Y., TERZOPOULOS D., WATERS K. : Realistic modeling for facial animation. In *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques* (1995).

[LWP10] LI H., WEISE T., PAULY M. : Example-based facial rigging. In *ACM SIGGRAPH 2010 Papers* (2010), SIGGRAPH '10.

[LYYB13] LI H., YU J., YE Y., BREGLER C. : Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.* (2013).

[LZD13] LE B., ZHU M., DENG Z. : Marker optimization for facial motion acquisition and deformation. *Visualization and Computer Graphics, IEEE Transactions on* (2013).

[MAW*07] MERRELL P., AKBARZADEH A., WANG L., MORDOHAI P., FRAHM J.-M., YANG R., NISTER D., POLLEFEYS M. : Real-time visibility-based fusion of depth maps. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (2007).

[MDSB03] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. : Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics III*. Springer Berlin Heidelberg, 2003.

[NN01] NOH J.-Y., NEUMANN U. : Expression cloning. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (2001).

[RGL15] REVERDY C., GIBET S., LARBOULETTE C. : Optimal marker set for motion capture of dynamical facial expressions. *Motion In Games* (2015).

[RHHL02] RUSINKIEWICZ S., HALL-HOLT O., LEVOY M. : Real-time 3d model acquisition. *ACM Trans. Graph.* (2002).

[RL01] RUSINKIEWICZ S., LEVOY M. : Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on* (2001).

[SCOL*04] SORKINE O., COHEN-OR D., LIPMAN Y., ALEXA M., RÖSSL C., SEIDEL H.-P. : Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (2004).

[SFPL10] SALVI J., FERNANDEZ S., PRIBANIC T., LLADO X. : A state of the art in structured light patterns for surface profilometry. *Pattern Recognition* (2010).

[SLS*12] SEOL Y., LEWIS J., SEO J., CHOI B., ANJYO K., NOH J. : Spacetime expression cloning for blendshapes. *ACM Trans. Graph.* (2012).

[SNF05] SIFAKIS E., NEVEROV I., FEDKIW R. : Automatic determination of facial muscle activations from sparse motion capture marker data. In *ACM SIGGRAPH 2005 Papers* (2005).

[SP04] SUMNER R. W., POPOVIĆ J. : Deformation transfer for triangle meshes. In *ACM SIGGRAPH 2004 Papers* (2004).

[SS02] SCHARSTEIN D., SZELISKI R. : A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* (2002).

[WBLP11] WEISE T., BOUAZIZ S., LI H., PAULY M. : Realtime performance-based facial animation. In *ACM SIGGRAPH 2011 Papers* (2011).

[WLVG07] WEISE T., LEIBE B., VAN GOOL L. : Fast 3d scanning with automatic motion compensation. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* (2007).

[WLVG08] WEISE T., LEIBE B., VAN GOOL L. : Accurate and robust registration for in-hand modeling. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008).

[WLVGP09] WEISE T., LI H., VAN GOOL L., PAULY M. : Face/off : Live facial puppetry. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2009), SCA '09.

[WPK*04] WEYRICH T., PAULY M., KEISER R., HEINZLE S., SCANDELLA S., GROSS M. : Post-processing of scanned 3d surface data. In *Eurographics symposium on point-based graphics* (2004).

[XCLT14] XU F., CHAI J., LIU Y., TONG X. : Controllable high-fidelity facial performance transfer. *ACM Trans. Graph.* (2014).

[ZH04] ZHANG S., HUANG P. : High-resolution, real-time 3d shape acquisition. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on* (2004).

[Zha10] ZHANG S. : Recent progresses on real-time 3d shape measurement using digital fringe projection techniques. *Optics and Lasers in Engineering* (2010).

[ZPS04] ZHANG Y., PRAKASH E., SUNG E. : A new physical model with multilayer architecture for facial expression animation using dynamic adaptive mesh. *Visualization and Computer Graphics, IEEE Transactions on* (2004).

[ZSCS04] ZHANG L., SNAVELY N., CURLESS B., SEITZ S. M. : Spacetime faces : High-resolution capture for

modeling and animation. In *ACM Annual Conference on Computer Graphics* (2004).