

Reconstruction 4D de maillages d'acteurs par suivi pseudo-rigide

L. Blache¹, C. Loscos¹, L. Lucas¹ et M. Desbrun²

¹Université de Reims Champagne-Ardenne

²Caltech

Résumé

Nous présentons dans cet article une méthode de reconstruction spatio-temporelle de maillages à partir de séquences de volumes. Les studios de tournages emploient de plus en plus de systèmes de capture pour modéliser des acteurs en 3D. Les technologies de reconstruction multi-vues utilisent un ensemble de caméras pour générer une séquence d'enveloppes visuelles via les silhouettes des prises de vues. Cependant, ces séquences manquent de cohérence temporelle car chaque pose est modélisée indépendamment des autres, ce qui rend difficile leur traitement par les outils de post-production habituels. Nous proposons une méthode de suivi automatique pour convertir une séquence de volumes reconstruits en un unique maillage animé. Un maillage de référence est déformé de manière globale par une transformation pseudo-rigide préservant les détails et guidée par un flot de mouvements estimé à partir de deux poses consécutives. Une étape d'optimisation locale permet d'ajuster la surface du modèle avec l'enveloppe visuelle afin d'obtenir une animation robuste de maillage 4D.

We present a method to reconstruct a time-evolving triangle mesh representation from a sequence of binary volumetric data representing an arbitrary motion. Movie studios are increasingly relying on markerless capture systems to turn an actor's performance into a 4D representation. Typically, a multi-view camera rig captures the motion which is then converted into a time series of visual hulls via volumetric silhouette-based reconstruction. However, the reconstructed sequence lacks temporal coherence as each frame is generated independently, preventing easy post-production editing with off-the-shelf modeling tools. We propose an automated tracking approach to convert the raw input sequence into a single, animated mesh. A template mesh is globally deformed via as-rigid-as-possible, detail-preserving transformations guided by a motion flow estimated from consecutive frames. Local deformation is added to better match the mesh surface to the current visual hull, leading to a robust 4D mesh reconstruction.

Mots clé : Maillage dynamique, flot de mouvements, modélisation, appariement de voxels, déformation ARAP.

1. Introduction

Le mélange d'images réelles et virtuelles est de plus en plus utilisé dans le cadre de la production de contenus destinés à l'industrie de l'image et du multimédia. Afin d'adapter cette technologie aux contraintes de la production télévisuelle, différentes techniques telles que la capture de mouvements ou le *matte painting* ont été utilisées. Récemment, plusieurs projets tels que GrImage [AFM*06], RECOVER3D [LSI*13] ou KINOVIS [kin15] ont été développés, avec pour objectif de mettre en place un système de *vidéo virtuelle* pour la production de contenus audiovisuels en utili-

sant la technologie de *reconstruction multi-vues*. Un modèle 3D représentant un *clone virtuel* d'un acteur est construit à partir d'un ensemble de captures vidéo multi-points de vues. Contrairement à la capture de mouvements (MoCap) traditionnelle, ces systèmes s'affranchissent de l'utilisation des marqueurs libérant ainsi les acteurs de contraintes vestimentaires et de mouvements tout en permettant une capture plus détaillée et générique.

Dans la plupart des studios de production cinématographiques, les modèles 3D sont générés à chaque pas de temps indépendamment les uns des autres. Les résultats de la reconstruction sont ainsi produits sous forme d'une séquence de poses statiques, ce qui limite leur utilisation en post-production et entraîne des incohérences dues aux approxi-

mations de reconstruction d'une image à l'autre. Notre objectif est, au contraire, de créer un unique modèle animé représentant la séquence de mouvements capturée. A terme, l'approche proposée doit être aussi générique que possible afin d'apporter des solutions à différents types de scènes : un ou plusieurs acteurs, portant des costumes ou des accessoires, éventuellement filmés en plans rapprochés. Ces contraintes nécessitent ainsi une méthode qui ne soit pas limitée seulement à la reconnaissance des mouvements rigides ou articulés.

Dans cet article, nous nous concentrons sur des scènes composées d'un seul acteur animé de mouvements libres, présentant une complexité sur les accessoires et vêtements portés. Nous proposons une nouvelle méthode basée sur le suivi des volumes au cours du temps afin d'identifier les mouvements de l'acteur pour ensuite les appliquer à un modèle de surface dynamique. Les données d'entrée de notre méthode sont des séquences de volumes discrets générés grâce à un algorithme *d'enveloppe visuelle* [Lau94] volumétrique à chaque image des vidéos multi-vues, indépendamment les uns des autres. Ce type de données est classique dans la génération de contenus 3D animés à partir de vidéos. Notre méthode (voir figure 1) comporte les étapes suivantes :

1. Extraction du mouvement : Nous estimons les mouvements de la scène en calculant un champ de mouvements entre les poses (*frames*) de cette séquence en utilisant un algorithme basé sur l'appariement des voxels entre deux volumes successifs.
2. Animation du maillage :
 - Le flot de mouvements est utilisé pour animer un maillage 3D de référence représentant l'acteur et lui faire adopter la pose définie par le volume suivant. Ce processus d'animation est basé sur un algorithme de déformation de maillage pseudo-rigide guidé par une sélection automatique de points d'ancrages, la difficulté principale résidant dans la pertinence du choix de ces points.
 - Cet alignement global est suivi d'un algorithme de déformations locales basé sur un système mécanique itératif.

L'objectif est ainsi d'obtenir un maillage unique, animé au cours de la séquence. Cette méthode fonctionne sur des données génériques, quel que soit l'objet ou l'acteur reconstruit.

Après une revue des différentes techniques d'animation de maillage (voir section 2), nous détaillerons notre méthode. Les sections 3, 4 et 5 présentent respectivement les étapes de calcul du champ de mouvements, d'animation et de régularisation des maillages. Des résultats sont enfin présentés en section 6.

2. Etat de l'art

Un certain nombre de méthodes ont été proposées pour générer des maillages animés à partir de reconstructions multi-vues. Cette section présente un tour d'horizon des différentes méthodes d'animation qui pourraient être utilisées dans notre contexte, à savoir déformer un maillage conformément à une succession de poses obtenues par reconstruction multi-vues.

2.1. Reconstruction basée modèle

Les méthodes de reconstruction multi-vues basées modèles utilisent une géométrie de référence de l'objet (un maillage 3D la plupart du temps) construite manuellement ou générée par un système d'acquisition spécifique, par exemple un scanner 3D, comme présenté par De Aguiar *et al.* [dAST*08]. La reconstruction consiste alors à faire évoluer cette forme de référence en fonction de données issues des vidéos (poses de références ou flots optiques). Dans ce contexte, Vlastic *et al.* [VBMP08] ou encore Gall *et al.* [GSdA*09] utilisent une animation basée squelette afin d'animer globalement un modèle de référence (*template*) avec une pose définie par une silhouette ou une enveloppe visuelle. Des déformations locales sont ensuite appliquées sur le maillage afin de l'ajuster au mieux avec les détails de surface issus de mouvements non rigides (vêtements ou chevelures par exemple). L'avantage de ces méthodes est de générer des données dynamiques (*i.e.* un objet qui évolue au cours du temps). Néanmoins, l'utilisation d'une forme de référence limite fortement la généralité du processus et les techniques basées modèles sont le plus souvent utilisées pour reconstruire un humain seul, même si certaines méthodes comme celle de Liu *et al.* [LSG*11] permettent de représenter plusieurs acteurs. Cagniart *et al.* [CBI10] utilisent un maillage dynamique basé sur un ensemble de *patches* modélisés à partir de la première pose de la séquence. Ce maillage est ensuite déformé pour obtenir les poses successives de la séquence.

2.2. Reconstruction libre et suivi de mouvements

Différentes approches permettent d'établir la continuité d'une forme géométrique générique au cours d'une succession de poses. Li *et al.* [LLV*12] utilisent un graphe de déformation pour établir une correspondance entre les maillages d'une séquence reconstruite par un scanner 3D. Popa *et al.* [PSDB*10] appliquent également un algorithme de *mapping* entre les poses successives. Cependant, ces méthodes s'appliquent sur des maillages reconstruits par scanner qui fournissent, malgré les trous issus du scan, des surfaces de haute qualité, ce qui n'est pas le cas des enveloppes visuelles sur lesquelles nous travaillons. Une manière d'introduire une cohérence temporelle dans une séquence de maillage est d'utiliser une méthode de suivi de maillage qui établit une correspondance entre deux maillages grâce

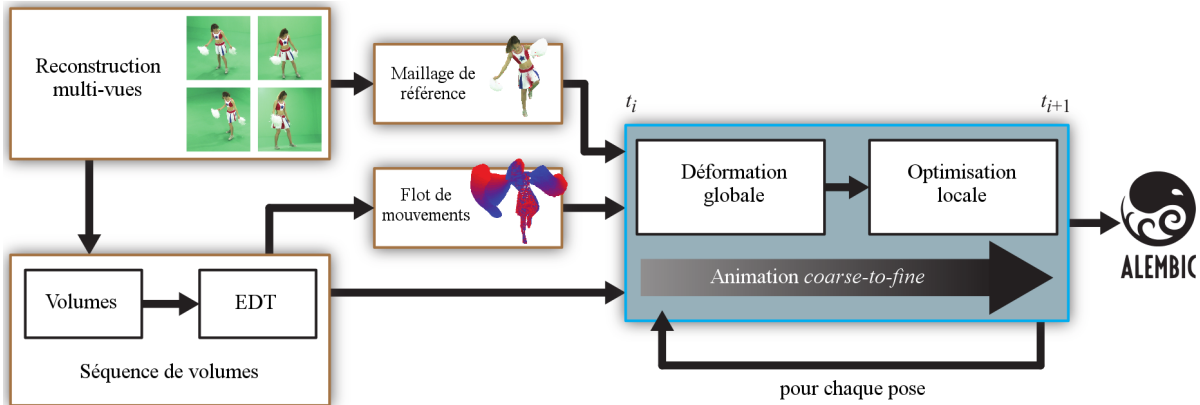


Figure 1: **Pipeline général.** Après une reconstruction libre par enveloppe visuelle, on extrait une séquence de volumes et un maillage de référence (sections 3.1). Un champ de mouvements est calculé à partir des poses successives (section 3.2). Le maillage de référence est ensuite animé, grâce aux vecteurs de mouvements, selon une approche en deux étapes : alignement global puis optimisation locale (sections 4 et 5).

à des critères tels que la courbure ou la texture [SH07, VZBH08, TM10]. Ce suivi d'objets peut être utilisé pour calculer le champ de mouvements qui décrit les gestes d'un personnage entre deux poses [PLBF11]. L'extraction des mouvements d'une scène 3D est généralisée par la notion de *flot de scène* [VBR*99]. Il s'agit de l'équivalent tridimensionnel du flot optique, obtenu par la fusion des flots optiques d'un système multi-points de vue. Anuar et Guskov [AG04] proposent une méthode qui adapte un algorithme de flot optique traditionnel à des espaces 3D discrets obtenus par reconstruction multi-vues. Ce champ de mouvements est ensuite utilisé pour animer un maillage. Dans le cas de la reconstruction par enveloppe visuelle, les maillages obtenus contiennent souvent de trop nombreux artefacts pour pouvoir être utilisés dans une méthode de suivi de maillage. Ainsi, une approche volumique semble plus appropriée, comme par exemple la méthode proposée par Nobuhara et Matsuyama [NM04] qui calcule un champ de mouvements grâce à un algorithme de mise en correspondance de voxels appliqué sur des séquences de volumes issues de reconstructions basées silhouettes. Ces champs de mouvements sont ensuite utilisés pour animer un maillage de référence obtenu par un algorithme de *marching cube* appliqué sur le volume de la première pose. Cependant, les champs de mouvements utilisés dans cette méthode sont obtenus par une mise en correspondance des voxels basée uniquement sur la distance Euclidienne, ce qui produit des vecteurs de mouvements qui manquent de précision. La méthode décrite dans [BLNL14] propose également d'utiliser un appariement de voxels pour calculer un flot de mouvements dans une séquence de volumes. Toutefois la fonction d'appariement prend également en compte l'orientation de la surface et la couleur des voxels, issue de la texture du personnage, afin d'assurer une meilleure correspondance entre les volumes.

2.3. Animation de maillage

Parmi les différents types d'animation de maillage, les techniques basées squelette sont couramment utilisées dans le domaine de l'animation et de la capture de mouvements. Un squelette topologique simplifié est associé au maillage. Les mouvements des sommets du maillage sont ensuite interpolés à partir des mouvements du squelette (*skinning*). On peut citer diverses méthodes comme le *Linear Blend Skinning* (LBS) ou *Skeletal Subspace Deformation* (SSD) [LCF00], le *Multi-Weight Enveloping* (MWE) [WP02] ou encore l'*implicit skinning* [VBG*13]. L'usage d'un squelette n'est cependant pas toujours adapté aux modèles animés de mouvements trop complexes. Dans ce cas, on peut utiliser des méthodes d'animation basées sur une *cage*, c'est à dire un ensemble de polygones simples qui englobent le modèle et dont les mouvements sont interpolés à la manière d'un squelette externe [TTB12] [JSW05] [LLCO08]. Les modèles de surfaces actives sont un autre modèle de surfaces déformables, comme par exemple les *δ-snake* [LM99] qui peuvent être appliqués au suivi de formes au cours du temps [BBPL06]. Cependant ces modèles modifient la structure et la connexité du maillage de manière à s'adapter au mieux à la forme de l'objet, ce qui nous empêche de conserver un maillage stable. Récemment, les méthodes de déformation de maillage basées sur la minimisation d'une énergie globale sont de plus en plus utilisées [BTP14]. Dans ce contexte, les méthodes d'animation *as-rigid-as-possible* [SA07] offrent une solution intéressante car elles consistent à déformer un maillage en conservant au maximum sa structure et notamment la connexité et les angles du maillage. La déformation est dirigée par un ensemble de points d'ancrage et le reste du maillage est ensuite guidé dans sa déformation par la minimisation d'une énergie. Cependant, cette méthode est habi-

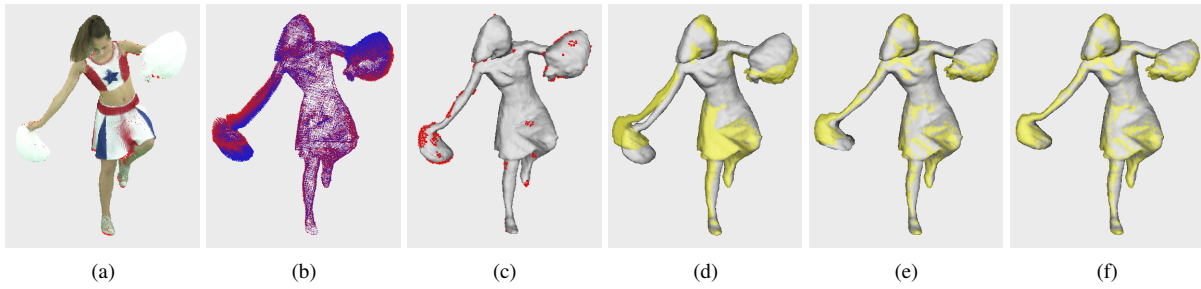


Figure 2: A partir d'une séquence de volumes (a), on extrait un flot de mouvements (b). Puis on échantillonne un ensemble de points d'ancrages (c) sur le maillage initial (d) représenté en gris. Après application de la déformation globale (e), on effectue une optimisation locale du maillage afin de se rapprocher au maximum de la pose ciblée représentée en jaune (f).

tuellement utilisée dans une approche de modélisation géométrique, avec des points d'ancrage sélectionnés et déplacés manuellement. Dans notre cadre d'utilisation, cet algorithme doit être adapté de manière à intégrer une sélection automatique de ces points. La minimisation d'énergie qui guide la déformation du maillage doit également être adaptée aux mouvements parfois irréguliers de ces points d'ancrage.

2.4. Notre approche

La méthode que nous avons mise en place se base sur l'utilisation d'un maillage de référence. Nous proposons d'utiliser pour cela la surface générée de manière automatique à partir de la première pose de la séquence. Les flots de déplacements qui définissent les mouvements de l'acteur entre les poses successives sont calculés par une méthode d'appariement de voxels basée sur [BLNL14] (voir section 3, figure 2b). Ces mouvements sont ensuite appliqués au maillage initial par une adaptation de l'animation *As-Rigid-As-Possible* (ARAP) [SA07] prenant en compte l'incertitude liée au champ de mouvements (voir section 4, figure 2c,2e). Enfin, une étape d'optimisation locale assure la correspondance entre le maillage et la pose ciblée (voir section 5, figure 2f).

3. Extraction du mouvement

3.1. Séquences de volumes

Nos données d'entrée sont des séquences de volumes discrets issus d'une phase de reconstruction multi-vues par enveloppe visuelle. Chaque volume de cette séquence représente l'acteur à une pose t définie par les images des vidéos multi-points de vue au même instant. L'animation du personnage dans la scène est ainsi représentée par une succession de n volumes qui constituent les poses successives de l'acteur de l'instant t_0 à t_{n-1} . Les volumes utilisés sont des volumes discrets binaires, c'est à dire une grille régulière de *voxels* définis par des valeurs binaires (0 pour un voxel vide, 1 pour un voxel plein). Nous commençons par

calculer à partir de chaque volume une carte de distances Euclidiennes, appelée *EDT* (*Euclidean Distance Transform*) comme décrite par Saito et Toriwaki [ST94]. On obtient ainsi pour chaque voxel une valeur correspondant à sa distance Euclidienne par rapport à la surface de l'objet décrit dans le volume, ce qui peut être représenté par un volume en niveaux de gris (voir figure 3b). La dérivée de cette nouvelle grille de voxels peut ensuite être utilisée pour estimer le gradient (voir section 5). Pour calculer cette dérivée, on utilise des filtres de Sobel qui estiment, dans une fenêtre de dimension $3 \times 3 \times 3$ autour de chaque voxel, les variations des valeurs de l'EDT selon chaque axe de l'espace.

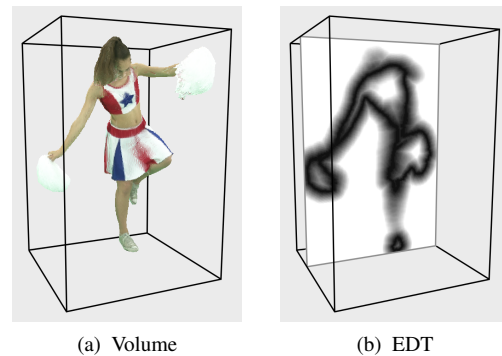


Figure 3: (a) Volume texturé issu de la reconstruction multi-vues. (b) Représentation en coupe de l'EDT correspondante.

3.2. Flot de mouvements

Nous utilisons la méthode décrite dans [BLNL14] pour calculer un champ de mouvements à partir d'une séquence de volumes reconstruits. Pour traiter le flot de mouvements entre les deux premiers volumes de la séquence, on plonge le volume t_0 dans l'EDT de la pose t_1 . La valeur de l'EDT associée à chaque voxel de surface du volume initial indique sa

distance par rapport à la surface de la nouvelle pose. La valeur maximale de cette distance nous renseigne sur le rayon de recherche utilisé pour l'appariement de voxels. Le flot de mouvements initial est ensuite utilisé pour interpoler le mouvement de la pose suivante par prédiction, et ainsi réduire le rayon de recherche lors du traitement des poses suivantes. Ce paramètre est ainsi automatiquement fixé en fonction des données et n'a pas à être renseigné par l'utilisateur. Cette amélioration permet également d'accélérer sensiblement le calcul des champs de mouvements en conférant un gain de 60% de temps de calcul par rapport à la méthode originale décrite dans [BLNL14].

4. Animation de maillage

Nous utilisons comme maillage de référence la pose de l'acteur reconstruite à partir du premier volume de la séquence, à l'état t_0 . Nous partons du principe que la position initiale de l'acteur exclut toute ambiguïté topologique afin de nous assurer une surface de référence adaptée à la morphologie du personnage. La surface de l'acteur est triangulée grâce à un algorithme de *marching cubes* appliqué sur le volume. Ce maillage est ensuite régularisé et simplifié par un algorithme de contraction d'arêtes puis un lissage Laplacien. Ces étapes préliminaires ont pour objectif de fournir un *template* de bonne qualité à notre méthode. Au cours de l'étape d'animation, le maillage initial est plongé dans le champ de mouvements pour appliquer les déplacements. Nous utilisons la méthode de déformation de maillage ARAP décrite par Sorkine et Alexa [SA07] car elle permet de déformer un maillage en restant fidèle à la forme initiale : la surface est déformée tout en minimisant de manière globale la différence entre le maillage de la pose de référence et la nouvelle position du personnage. Une sélection automatique d'un ensemble de sommets jugés pertinents sert de points d'ancrage pour guider la déformation de l'ensemble du maillage.

4.1. Sélection automatique des points d'ancrage

On sélectionne à la pose t_i un ensemble de sommets qui doivent servir de points d'ancrages. Ces sommets sont répartis en plusieurs *patches* sur la surface. Ces points d'ancrage, qui vont guider la déformation du maillage, doivent correspondre aux zones de la surface animées des mouvements les plus importants. On sélectionne donc les sommets qui sont associés aux vecteurs de mouvements de plus forte amplitude. On souhaite également sélectionner certains points correspondant à des déplacements faibles ou nuls de façon à éviter de déformer les zones immobiles du maillage.

On utilise l'EDT à la pose t_{i+1} pour identifier les sommets de la pose t_i associés aux valeurs les plus importantes. Les sommets du maillages à la pose t_i sont ordonnés selon leur valeur associées dans EDT_{i+1} . On sélectionne ensuite les n sommets de plus haute valeur, qui sont donc les plus éloignés de la surface à la pose suivante et correspondent donc

aux mouvements de plus forte amplitude. Dans notre cas, le nombre n de points d'ancrages est empiriquement fixé à 10% du nombre total de sommets. La déformation du maillage peut être plus ou moins contrainte selon le nombre de points d'ancrage. De la même manière, on sélectionne également un ensemble de points immobile ou correspondant à des vecteurs de mouvement de faible amplitude pour éviter la déformation des zones statiques précédemment décrites. Si cette méthode d'échantillonnage des points d'ancrage est adaptée à notre cas d'utilisation, elle est cependant indépendante de l'algorithme de déformation de maillage et peut être remplacée par d'autres approches – automatiques ou manuelles – selon le contexte.

Après avoir sélectionné les points d'ancrage, on récupère le vecteur de mouvement correspondant à chacun de ces points, ainsi que la valeur du *matching* qui lui est associé, selon la fonction décrite dans [BLNL14]. Cette valeur est utilisée comme score pour définir le poids de ce point d'ancrage lors de l'étape de déformation globale suivante.

Les points d'ancrage adjacents sont regroupés en *clusters* où les vecteurs de déplacements et les poids sont interpolés de manière à former un *patch* de mouvements quasi-rigide.

4.2. Animation

Etant donné que le flot de mouvements (section 3) peut être fortement bruité, nous utilisons une méthode variationnelle en cherchant à déformer un maillage M' par l'application de transformations localement rigides, tout en conservant autant que possible la position des points clés (figure 2e).

Formulation : Nous cherchons à minimiser l'énergie suivante :

$$E(M') = E_{ARAP}(M') + E_{ANC}(M')$$

E_{ARAP} est l'énergie de la déformation *As-Rigid-As-Possible* :

$$E_{ARAP}(M') = \sum_{i=1}^n \sum_{j \in N(i)} w_{ij} \|(p'_i - p'_j) - R_i(p_i - p_j)\|^2 \quad (1)$$

avec $N(i)$ le voisinage direct de i et R_i la matrice de rotation qui transforme p_i en p'_i . La minimisation de cette énergie permet de conserver les coordonnées différentielles des sommets du maillage au cours de l'animation. On s'assure ainsi que la déformation de la surface n'altère pas la structure des triangles. Notre contrainte de cohérence temporelle est ainsi respectée. E_{ANC} est une énergie quadratique appliquée sur les points d'ancrages, définie par :

$$E_{ANC}(M') = \sum_{i=1}^n w_{a_i} \|p'_i - p_i\|^2 \quad (2)$$

avec w_{a_i} le poids correspondant au degré de confiance accordé à chaque point d'ancrage calculé en section 4.1. On fixe $w_{a_i} = 0$ si p_i n'appartient pas à l'ensemble des points d'ancrages.

Résolution : A partir des conditions d'optimalité des énergies ci-dessus, on déduit des expressions précédentes un système linéaire composé de l'équation de l'ARAP telle qu'elle est décrite dans [SA07] à laquelle sont ajoutés les termes de la forme quadratique (eq. 2) :

$$\sum_{j \in N(i)} w_{ij}(p'_i - p'_j) + w_{a_i} p'_i = \sum_{j \in N(i)} \frac{w_{ij}}{2} (R_i - R_j)(p_i - p_j) + w_{a_i} p_i \quad (3)$$

La partie gauche de cette expression correspond à la matrice Laplacienne du maillage. Le système peut alors être écrit sous la forme matricielle :

$$Lp' = b$$

La matrice L inclut dans sa diagonale les poids w_{a_i} associés à chaque point d'ancrage (eq. 2). Le vecteur p' contient les coordonnées p'_i et b est une matrice colonne contenant la partie droite de l'expression (eq. 3). La résolution du système (eq. 3) s'effectue itérativement : la rotation optimale $R(p_i, p'_i)$ est calculée par SVD à partir des positions initiales de p_i et p'_i fixées. Cette matrice de rotation est ensuite appliquée à p_i pour obtenir la nouvelle valeur de p'_i grâce à un solveur linéaire. Ces deux étapes sont répétées jusqu'à convergence du système.

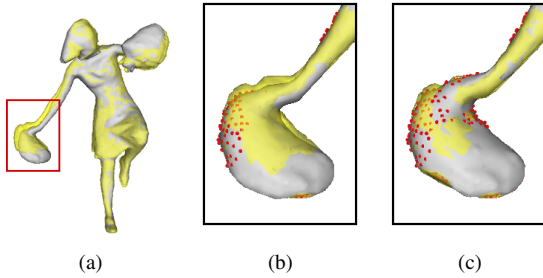


Figure 4: **Ajustement ARAP.** Maillage de référence (gris) et pose ciblée (jaune) (a). Après la première étape de déformation guidée par les vecteurs du flot de mouvements (b), plusieurs étapes d'ajustements peuvent être appliquées, guidées par la grille de l'EDT (c). Les points d'ancrage sont représentés en rouge.

4.3. Ajustement de la pose globale

Les points d'ancrage sélectionnés (section 4.1) permettent de suivre les mouvements les plus importants de l'acteur et de faire adopter grossièrement au maillage la position de la pose ciblée. Cependant dans de nombreux cas de mouvements plus complexes, cette première étape de déformation pseudo-rigide nécessite des ajustements de manière à ce

que le maillage de référence soit aligné avec l'enveloppe visuelle. En utilisant le même processus d'échantillonnage décrit en section 4.1, on sélectionne un nouvel ensemble de points d'ancrage éloignés de la surface ciblée, selon les valeurs de l'EDT associée. Ces points sont projetés vers la surface selon la direction du gradient de l'EDT (figure 4). Les points d'ancrage déplacés lors de la première étape restent fixés. On effectue ensuite une nouvelle déformation ARAP comme décrit précédemment. Cette nouvelle étape permet d'aligner au mieux la pose du maillage avec les silhouettes capturées (*i.e.*, l'enveloppe visuelle). Etant donné que les déformations appliquées au maillage lors de cette étape d'ajustements sont de faible amplitude, la minimisation d'énergie converge plus rapidement que lors de l'étape de déformation initiale. Selon la complexité des geste de l'acteur, plusieurs étapes d'ajustement peuvent être appliquées. Dans notre cas, deux étapes d'ajustements sont généralement suffisantes.

5. Optimisation locale

L'étape précédente nous a permis de déformer le maillage initial pour lui faire adopter la pose décrite par la pose suivante. Cependant, après cette transformation globale, une étape d'optimisation locale est encore nécessaire pour que le maillage soit parfaitement adapté à la silhouette de la nouvelle pose. En effet certaines sections de la surface de l'objet sont animées de mouvements libres qui n'ont pas été détectés lors des étapes précédentes. Pour aligner le maillage avec l'enveloppe visuelle et pour supprimer les irrégularités qui peuvent survenir lors de la déformation, nous appliquons un algorithme de régularisation du maillage. Nous considérons le maillage comme un système mécanique dans lequel chaque sommet est soumis à deux contraintes :

Contrainte de régularisation : Cette contrainte correspond à un système masse-ressorts. On utilisera uniquement la composante tangentielle du vecteur obtenu. Chaque arête incidente applique sur le sommet une force répulsive ou attractive, selon sa longueur, de manière à uniformiser la distribution des sommets sur la surface :

$$f_r(p_i) = \alpha \sum_{j \in N(i)} (\|p_i - p_j\| - \bar{r}_i) \frac{p_i - p_j}{\|p_i - p_j\|}$$

α étant un coefficient de raideur, p_j un sommet du voisinage de dimension n de p_i et r étant la longueur à vide du ressort.

Contrainte de silhouette : On utilise la dérivée de l'EDT (voir section 3.1) afin de pousser chaque sommet vers la surface de l'enveloppe visuelle de l'objet (voir figure 5) :

$$f_s(p_i) = \sum_{j \in N(i)} (\vec{n}_{p_j} \cdot EDT(p_j))$$

avec \vec{n}_{p_i} et $EDT(p_i)$ respectivement la normale à la surface et la valeur de l'EDT en p_i . On utilise uniquement la composante normale de ce vecteur (figure 5).

Ces deux termes sont inclus dans l'expression suivante

du vecteur de déplacement exercé sur chaque sommet du maillage :

$$f(p_i) = \omega_1 f_r(p_i) + \omega_2 f_s(p_i)$$

où ω_1 et ω_2 sont des coefficients de pondération.

Afin d'appliquer ces déplacements sur le maillage, chaque sommet est considéré comme un solide caractérisé par une position et une vélocité à un temps t :

$$Y(t) = \begin{bmatrix} p(t) \\ v(t) \end{bmatrix}$$

Sa position à $t+1$ est :

$$Y(t+1) = F(t) = \begin{bmatrix} v(t) \\ \frac{1}{m} f(t) \end{bmatrix}$$

avec m la masse associée au point. Dans notre cas, nous utilisons un système non physique sans masses. Ce système mécanique est résolu de manière itérative en utilisant un intégrateur de Runge-Kutta.

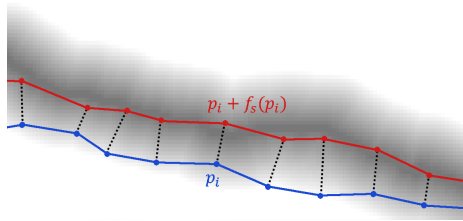


Figure 5: Contrainte de silhouette : les sommets sont déplacés le long de la normale, selon la valeur de l'EDT.

6. Résultats

6.1. Données

Nous avons testé notre méthode sur plusieurs jeux de données obtenus par reconstruction multi-vues de type *enveloppe visuelle*. La séquence *Cheerleader* est constituée de 200 volumes ayant une résolution de $182 \times 271 \times 171$ voxels (voir figure 11a). Le maillage de la pose de référence est constitué de 19234 sommets. La séquence *Astronaute* contient 25 volumes avec une résolution de $156 \times 115 \times 337$ voxels (voir figure 11b). Le maillage de la pose de référence est constitué de 8048 sommets. Ces deux séquences sont issues de prises de vues réelles en studio grâce à un ensemble de 24 caméras réparties autour de la zone de capture. La séquence *Simon* est constituée de 50 volumes d'une résolution de $114 \times 259 \times 95$ voxels. Ce jeu de données est une séquence de synthèse, réalisée à partir d'un modèle 3D dont la reconstruction multi-vues a été simulée par un logiciel. La séquence *Danseuse* a été générée en utilisant les images multi-vues de la plateforme GrImage[†], avec une résolution

moyenne de $150 \times 100 \times 300$ voxels. De la même manière, nous avons également généré la séquence *Capoeira*[‡] à partir des jeux de données décrits dans [dAST*08] avec une résolution de $200 \times 275 \times 200$ voxels. Etant donné le nombre réduit de points de vues (8 caméras) et la faible résolution des images, les enveloppes visuelles produites avec ces deux derniers jeux de données sont de moins bonne qualité.

6.2. Temps de calcul

Le calcul du champ de mouvements entre deux frames consécutives (section 3.2) s'effectue en moyenne en 40 secondes. L'étape initiale de déformation *as-rigid-as-possible* (section 4) nécessite environ 200 itérations pour converger avec la séquence *Cheerleader* et 320 itérations pour la séquence *Astronaute*. On échantillonne 10% des sommets du maillage de référence pour servir de points d'ancrages. On applique en générale deux étapes supplémentaires d'ajustement (section 4.3) qui nécessitent une centaine d'itérations en moyenne. L'étape d'optimisation locale (section 5) a été appliquée avec les coefficients suivants : $\omega_1 = 0.4$ et $\omega_2 = 0.6$. Les résultats visibles en figure 11 montrent que notre suivi de pose reste robuste. Le temps de calcul nécessaire à l'ensemble des traitements de maillage est d'environ 110 secondes pour chaque frame de la *Cheerleader* et de 80 secondes pour *Astronaute* et *Danseuse*.

6.3. Evaluation et comparaisons

L'utilisation d'une méthode de déformation pseudo-rigide nous permet de conserver, à chaque nouvelle pose, une structure de maillage aussi proche que possible que la pose précédente. La figure 6 montre que le maillage obtenu après cette déformation (figure 6e) conserve bien mieux les triangles qu'en appliquant les déplacements du flot de mouvements, comme c'est le cas dans [BLNL14] (voir figure 6c), sans chercher à conserver une rigidité locale. De cette manière, le maillage est temporellement cohérent entre les deux poses et l'étape finale d'optimisation locale (section 5) ne consiste qu'à adapter la surface aux mouvements non rigides qui peuvent survenir, notamment sur les vêtements amples. On peut observer avec la séquence *Cheerleader* que la surface des pompons, après la première étape d'alignement global (figure 2e), est correctement ajustée avec la pose ciblée lors de l'optimisation locale (figure 2f). Avec la séquence *Danseuse*, nous montrons que le maillage est correctement déformé de manière à suivre la forme de la robe en mouvement (figure 7). L'animation de ce type de mouvements est un avantage de notre approche par rapport aux modèles articulés qui ne sont pas adaptés aux déformations non rigides.

[†]. <http://4drepository.inrialpes.fr/>

[‡]. <http://resources.mpi-inf.mpg.de/siggraph08/perfcap/>

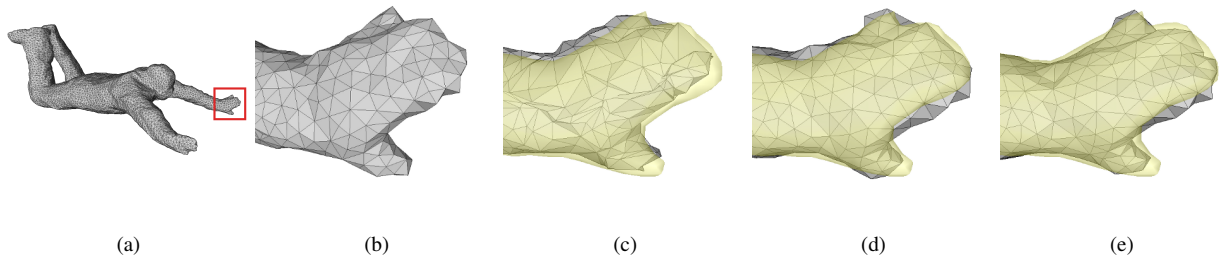


Figure 6: Différentes méthodes d'animation appliquées sur la séquence *Astronaute* (a). Maillage de la pose de référence (b). Déformation du maillage de référence (en gris) en appliquant des vecteurs du champ de mouvement [BLNL14] (c). Déformation du maillage en utilisant l'énergie ARAP [SA07] (d). Déformation du maillage en utilisant notre méthode (e). L'enveloppe visuelle de la nouvelle pose est représentée en jaune.

On peut également noter que notre système de déformation pseudo-rigide est plus adapté à notre cas d'utilisation que la méthode ARAP classique [SA07]. En effet, notre approche ne considère pas tous les points d'ancrage comme aussi fiables les uns que les autres puisque leur déplacement est guidé par un flot de mouvements qui manque de précision et peut contenir des vecteurs erronés. C'est pourquoi nous ajoutons à l'énergie globale les termes décrits dans l'expression (eq. 2). De plus, chaque point d'ancrage est associé à un poids w_a inclut dans le système final (voir section 4.2 - Résolution). Ce poids représente le *degré de confiance* que l'on accorde à chaque point d'ancrage, selon les critères décrits en section 4.1. Notre système s'avère ainsi plus souple et moins sensible aux erreurs du flot de mouvements que l'ARAP classique. La figure 6 montre qu'en utilisant l'énergie de l'ARAP seule, des erreurs dans le champ de mouvements peuvent entraîner des incohérences (voir la collision les extrémités de la main sur la figure 6d) là où notre méthode conserve un maillage de meilleure qualité (voir figure 6e).

Avec la séquence *Capoeira*, la qualité moindre des images et des silhouettes mène à des enveloppes visuelles très bruitées qui contiennent beaucoup d'imperfections (occlusions, trous ...). Notre méthode de suivi de forme, basée sur l'enveloppe visuelle, est de ce fait difficile à appliquer sur de longs intervalles de temps. On note cependant que le modèle animé produit par notre méthode présente une meilleure correspondance avec les silhouettes de l'acteur que le modèle généré par la méthode originale [dAST*08]. En effet la distance de Hausdorff moyenne entre le maillage animé et l'enveloppe visuelle (*i.e.*, intersection des silhouettes) est de 0.011933 avec la méthode originale contre 0.003291 avec notre approche (figure 8).

6.4. Applications

Grâce à notre méthode, les personnages reconstruits peuvent être aisément placés dans un environnement vir-

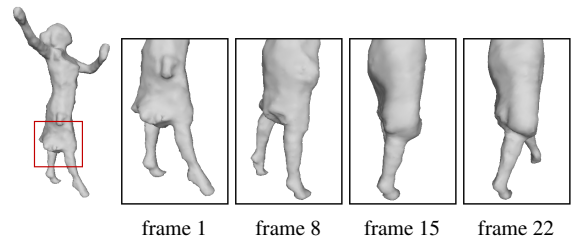


Figure 7: Exemple de mouvements non rigides sur la robe de la séquence *Danseuse*.

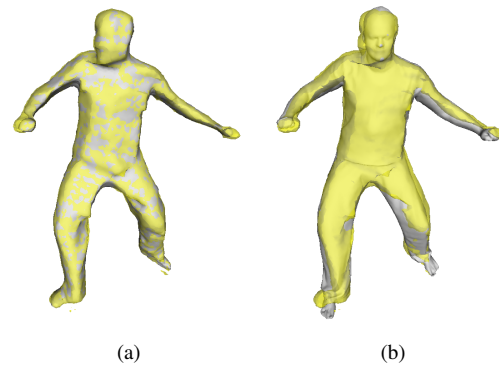


Figure 8: **Séquence Capoeira**. Correspondance entre le maillage de référence animé (en gris) et l'enveloppe visuelle de la frame suivante, qui représente la pose ciblée (en jaune). Comparaison entre notre méthode (a) et [dAST*08] (b).

tuel comme de simples maillages animés. Le rendu d'une scène virtuelle contenant des personnages reconstruits est ainsi facilité par rapport à l'utilisation de séquences de maillages où chaque *frame* de rendu doit être effectuée avec la pose correspondante. Nos maillages temporellement co-

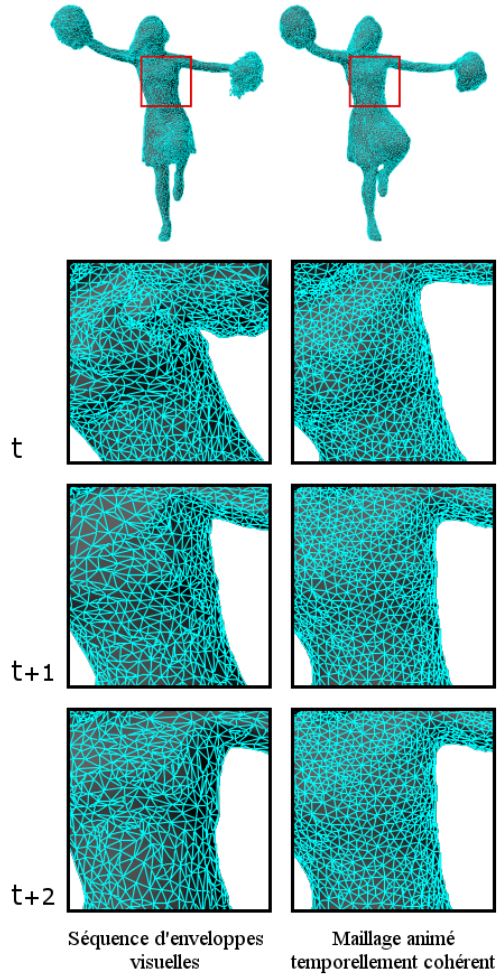


Figure 9: Comparaison entre une séquence de maillages obtenus par enveloppe visuelle (gauche) et notre maillage animé temporellement cohérent (droite).

hérents peuvent être directement exportés vers des logiciels d'animation 3D. La continuité structurelle de nos maillages au cours de la séquence (chaque sommet peut être suivi tout au long de l'animation, voir figure 9) permet de mettre en place des interactions entre le personnage animé et son environnement virtuel comme des collisions par exemple. Une *UV map* unique peut également être conservée tout au long de l'animation au lieu d'être recalculée à chaque *frame*.

6.5. Limitations et perspectives

Lors du traitement de la séquence *Simon*, il s'avère que certains détails fins sont difficilement gérés par notre méthode, notamment lors de la phase d'optimisation locale du maillage. En effet la reconstruction des mains du modèle fait apparaître les doigts du personnage, mais la finesse du vo-

lume à cet endroit entraîne une collision des sommets du maillage lors de l'application des contraintes de silhouette et de régularisation (section 5) comme le montre la figure 10.

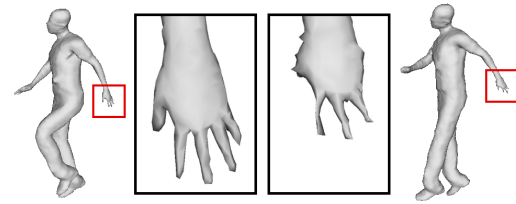


Figure 10: Dégradation du maillage au cours de la séquence *Simon*.

Concernant la génération du maillage de référence, nous considérons que la pose initiale de l'acteur permet une reconstruction cohérente et que la topologie de la surface n'est pas modifiée au cours de la séquence (on conserve un maillage fermé de genre 0). Cependant les changements de topologie qui peuvent survenir au cours de l'animation, notamment à cause des artefacts et des imprécisions de la reconstruction multi-vues, ne sont pas gérés actuellement et peuvent donc mener à des erreurs. A l'avenir, une méthode de reconstruction mélangeant stéréo et silhouette doit permettre de générer un modèle plus précis que la méthode actuelle uniquement basée sur l'enveloppe visuelle. Notre modèle de maillage dynamique pourrait également intégrer les changements de topologie, à la manière de l'approche proposée par Letouzey et Boyer [LB12] par exemple.

7. Conclusion

La méthode que nous avons présentée permet de générer des maillages animés à partir de séquences d'enveloppes visuelles volumiques issues de la reconstruction d'un acteur au cours du temps. Un flot de mouvements, estimé par appariement de voxels, est utilisé pour guider le déplacement d'un ensemble de points d'ancrages échantillonnés automatiquement. Ces points d'ancrages guident une déformation globale pseudo-rigide de l'ensemble du maillage pour adopter la pose décrite par le volume au pas de temps suivant. Cette méthode d'animation ne pose pas de restriction sur la nature des données et fonctionne sur des surfaces génériques, quels que soient les costumes ou accessoires portés par l'acteur. Une étape d'ajustement local permet d'aligner au mieux la surface avec l'enveloppe visuelle de la nouvelle pose. Les seules restrictions actuelles concernent les changements de topologie qui peuvent survenir au cours de la séquence et qui pourront faire l'objet de travaux futurs.

Remerciements

Nous souhaitons remercier nos partenaires d'XD Productions pour leur accompagnement dans la conduite du projet RECOVER3D. Ce projet est soutenu par le Fonds national

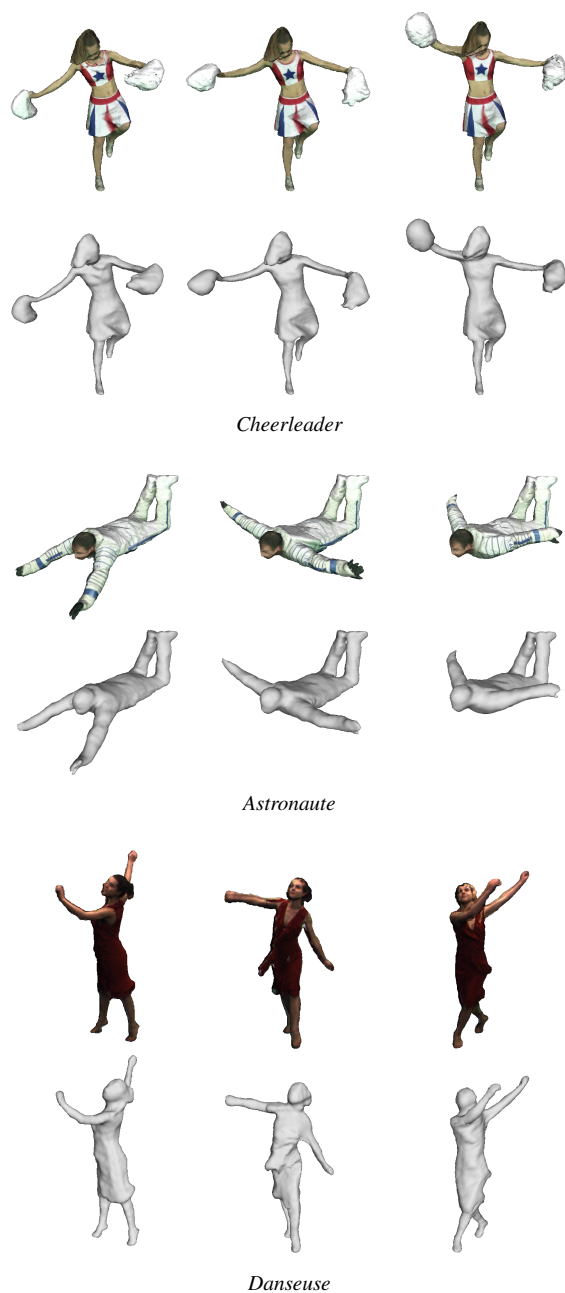


Figure 11: Résultats obtenus avec les séquences *Cheerleader*, *Astronaute* et *Danseuse*. Comparaison entre la reconstruction par enveloppe visuelle (en haut) et le maillage animé (en bas).

pour la société numérique (FSN) dans le cadre des *Investissements d'Avenir*. Nous remercions également le groupe de recherche *3D Video and Vision-based Graphics* du *Max-Planck-Center for Visual Computing and Communication* (MPI Informatik / Stanford) et l'équipe *Morpheo* de l'INRIA Grenoble pour les jeux de données fournis.

Références

- [AFM*06] ALLARD J., FRANCO J.-S., MÉNIER C., BOYER E., RAFFIN B. : The grimage platform : A mixed reality environment for interactions. In *IEEE International Conference on Computer Vision Systems, ICVS* (janvier 2006), pp. 46–46.
- [AG04] ANUAR N., GUSKOV I. : Extracting animated meshes with adaptive motion estimation. In *International Workshop on Vision, Modeling and Visualization (VMV)* (novembre 2004), pp. 63–71.
- [BBPL06] BITTAR E., BENASSAROU A., PLOTON D., LUCAS L. : Hierarchical tracking of intra-cell structures in 4d images. In *3rd International Conference on BioMedical Visualization* (juillet 2006), pp. 82–90.
- [BLNL14] BLACHE L., LOSCOS C., NOCENT O., LUCAS L. : 3d volume matching for mesh animation of moving actors. In *Eurographics Workshop on 3D Object Retrieval, 3DOR* (avril 2014), pp. 69–76.
- [BTP14] BOUAZIZ S., TAGLIASACCHI A., PAULY M. : Dynamic 2D/3D Registration. In *Eurographics 2014 - Tutorials* (2014), Holzschuch N., Myszkowski K., (Eds.).
- [CBI10] CAGNIART C., BOYER E., ILIC S. : Probabilistic deformable surface tracking from multiple videos. In *11th European conference on Computer vision (ECCV)* (2010), vol. 6314 de *Lecture Notes in Computer Science*, pp. 326–339.
- [dAST*08] DE AGUIAR E., STOLL C., THEOBALT C., AHMED N., SEIDEL H.-P., THRUN S. : Performance capture from sparse multi-view video. *ACM Transactions on Graphics. Vol. 27*, Num. 3 (août 2008), 98 :1–98 :10.
- [GSdA*09] GALL J., STOLL C., DE AGUIAR E., THEOBALT C., ROSENHAHN B., SEIDEL H.-P. : Motion capture using joint skeleton tracking and surface estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (juin 2009), pp. 1746–1753.
- [JSW05] JU T., SCHAEFER S., WARREN J. : Mean value coordinates for closed triangular meshes. *ACM Transactions on Graphics. Vol. 24*, Num. 3 (juillet 2005), 561–566.
- [kin15] KINOVIS. <http://kinovis.inrialpes.fr/>, 2015.
- [Lau94] LAURENTINI A. : Visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 16*, Num. 2 (1994), 150–162.

- [LB12] LETOUZEY A., BOYER E. : Progressive shape models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (juin 2012), pp. 190–197.
- [LCF00] LEWIS J. P., CORDNER M., FONG N. : Pose space deformation : A unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques* (2000), SIGGRAPH '00, pp. 165–172.
- [LLCO08] LIPMAN Y., LEVIN D., COHEN-OR D. : Green coordinates. *ACM Transactions on Graphics*. Vol. 27, Num. 3 (août 2008).
- [LLV*12] LI H., LUO L., VLASIC D., PEERS P., POPOVIĆ J., PAULY M., RUSINKIEWICZ S. : Temporally coherent completion of dynamic shapes. *ACM Transactions on Graphics*. Vol. 31, Num. 1 (février 2012), 2 :1–2 :11.
- [LM99] LACHAUD J.-O., MONTANVERT A. : Deformable meshes with automated topology changes for coarse-to-fine 3D surface extraction. *Medical Image Analysis*. Vol. 3, Num. 2 (1999), 187–207.
- [LSG*11] LIU Y., STOLL C., GALL J., SEIDEL H.-P., THEOBALT C. : Markerless motion capture of interacting characters using multi-view image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (juin 2011), pp. 1249–1256.
- [LSI*13] LUCAS L., SOUCHET P., ISMAËL M., NOCENT O., NIQUIN C., LOSCOS C., BLACHE L., PRÉVOST S., REMION Y. : Recover3d : A hybrid multi-view system for 4d reconstruction of moving actors. In *4th International Conference on 3D Body Scanning Technologies* (novembre 2013), pp. 219–230.
- [NM04] NOBUHARA S., MATSUYAMA T. : Heterogeneous deformation model for 3D shape and motion recovery from multi-viewpoint images. In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)* (septembre 2004), pp. 566–573.
- [PLBF11] PETIT B., LETOUZEY A., BOYER E., FRANCO J.-S. : Surface flow from visual cues. In *International Workshop on Vision, Modeling and Visualization (VMV)* (octobre 2011), pp. 1–8.
- [PSDB*10] POPA T., SOUTH-DICKINSON I., BRADLEY D., SHEFFER A., HEIDRICH W. : Globally consistent space-time reconstruction. *Computer Graphics Forum - Proceedings of Eurographics Symposium on Geometry Processing*. Vol. 29, Num. 5 (2010).
- [SA07] SORKINE O., ALEXA M. : As-rigid-as-possible surface modeling. In *Proceedings of Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP* (2007), pp. 109–116.
- [SH07] STARCK J., HILTON A. : Correspondence labelling for wide-timeframe free-form surface matching. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (octobre 2007), pp. 1–8.
- [ST94] SAITO T., TORIWAKI J.-I. : New algorithms for euclidean distance transformation of an n-dimensional digitized picture with applications. *Pattern Recognition*. Vol. 27, Num. 11 (1994), 1551–1565.
- [TM10] TUNG T., MATSUYAMA T. : Dynamic surface matching by geodesic mapping for 3D animation transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (juin 2010), pp. 1402–1409.
- [TTB12] THIERY J.-M., TIERNY J., BOUBEKEUR T. : Cager : Cage-based reverse engineering of animated 3d shapes. *Computer Graphics Forum*. Vol. 31, Num. 8 (2012), 2303–2316.
- [VBG*13] VAILLANT R., BARTHE L., GUENNEBAUD G., CANI M.-P., ROHMER D., WYVILL B., GOURMEL O., PAULIN M. : Implicit skinning : Real-time skin deformation with contact modeling. *ACM Transactions on Graphics*. Vol. 32, Num. 4 (juillet 2013), 125 :1–125 :12.
- [VBMP08] VLASIC D., BARAN I., MATUSIK W., POPOVIĆ J. : Articulated mesh animation from multi-view silhouettes. *ACM Transactions on Graphics*. Vol. 27, Num. 3 (août 2008), 97 :1–97 :09.
- [VBR*99] VEDULA S., BAKER S., RANDER P., COLLINS R., KANADE T. : Three-dimensional scene flow. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (1999), vol. 2, pp. 722–729.
- [VZBH08] VARANASI K., ZAHARESCU A., BOYER E., HORAUD R. : Temporal surface tracking using mesh evolution. In *10th European Conference on Computer Vision (ECCV)* (2008), vol. 5303 de *Lecture Notes in Computer Science*, pp. 30–43.
- [WP02] WANG X. C., PHILLIPS C. : Multi-weight enveloping : Least-squares approximation techniques for skin animation. In *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Computer Animation, SCA* (2002), pp. 129–138.