Video Capture of Skin Motion using Calibrated Fabrics

Fabien DELLAS
MIRALab - University of Geneva
fabien.dellas@miralab.unige.ch

Lionel REVERET
Evasion - INRIA Rhône-Alpes
lionel.reveret@inria.fr

Abstract

We introduce a novel approach to capture linear and non linear motions of the skin surface due to muscle bugglings and other complex sub-surface interactions. Our method uses standard camera and consists of a video motion capture of the desired body surface covered by a stretchy and calibrated cloth. We develop a non model-based tracking method using regularity and topology of dedicated fabrics. Our tracking and 3D reconstruction methods support occlusions : we detect occlusions of 3D surfaces using temporal and spatial coherences, and are able to restart surface tracking once the occlusions have ended. The obtained 3D mesh can easily be integrated and adapted in a sequence of an animated virtual human. We demonstrate our method with the skinning of non-linear biceps deformations.

Keywords: Motion Capture, Skinning, Capturing techniques, Video tracking

1 Motivation

The production of believable 3D synthetic images of virtual humans remains exhaustive and tedious. For animating virtual characters, 3D animators work similarly as drawers for cartoons, that implies hundreds of hours for only few seconds of animation. The traditional 3D animation of an humanoid can be divided into three steps : first the character body shape is modeled in 3D, secondly the skeleton and its animation are implemented by key-frames techniques, finally the skin motion is simulated by relative influences of joints positions of the skeleton; this is the ”skinning” phase. In order to reduce animation work, a common procedure now consists of automatically animating the skeleton joints by motion capture [1]. However, this technique provides only rigid movements of the skeleton joints. The skin has to follow the skeleton motion, while preserving the human anatomy of muscles motion, although they are difficult to take into account by the spatio-temporal resolution of an optical motion capture process.

2 Contribution

Our work aims at measuring body surface deformations while moving and giving a practical 3D model to animators temporally coherent. 3D reconstruction and tracking are classical problems in computer vision, but in our case human skin gives poor texture information to track. Moreover skin deformations are complex and non-linear : muscles buggle, roll and slide. Their motion cannot be predicted with the only position and orientation of animation skeleton joints. We have manufactured a strecthy and textured suit in order to track and reconstruct the surface, inspired by recent reconstruction-oriented works for motion capture of clothes[2, 3]. Our approach allows 3D estimation and tracking even if the surface is partially occluded and recovers the occluded parts of the surface using temporal and spatial coherence. Furthermore our suit gives us at the

same time an estimation of the position and motion of animation skeleton joints.

3 State of the art

The classical method to simulate skin onto an animation skeleton is simple : each skeleton joint may influence the skin 3D shape, depending on the distance of the skin vertices to the joints location. The 3D position of each skin vertex is computed as a weighted linear combination of the 3D positions of this vertex transformed with respect to each nearby joint local coordinate system. It allows fast but approximative results [4]. Another flexible method consists of processing the skinning by Free Form Deformations (FFD) : 3D space surrounding the body shape is divided into volumes containing the pieces of the surface to skin. The deformable surface will be influenced only by its container[5].

Fine details such as muscle buggles are still difficult to model with such techniques and require additionnal manual work when animating the model. Recent works allows to reduce these problems by computing more pertinent interpolations under constraints. It consists of extracting rules from the skin, like volume conservation, and adding joints in order to simulate non rigid deformation like muscles bugglings[4, 6]. Other works are using simulation to predict the surface deformation due to muscles contractions [7, 8]. These methods divide the human body into three basis elements: the skeleton, represented by a hierarchy of segments; the muscles attached to the skeleton, modelised by ellipsoids and the tissues (skin and fat tissues covering the structure). When the skeleton moves, the principal axis of the ellipsoids are adjusted while conserving their volumes in order to simulate buggles. All these methods greatly increase the quality of the visual rendering but all share the same problem of not being automatically tuned. Empirical data have to be added. Our goal is to find a method to automatically estimate such data from video.

Other approaches study directly the surface body rather than simulating it. One of them captures directly the body surface[9]. The main idea is to take 3D scans of the body at different steps of a movement, for example the flexion of the arm, and interpolate between steps. Results are impressive but they rely on static poses of the body to capture the 3D scans. It is difficult to estimate if muscles have the same buggling during a real movement. Moreover 3D scan of body parts involves a substantial infrastructure. In [10], Plaenkers et al. process a model-based tracking with one view and a pseudo-anatomical model created with implicit surfaces. The surface refinement depends on many parameters due to the implicit surfaces and is not yet fully adapted to skinning methods for animators. Another experimental method consists of tracking interest points on a deformable surface with orthographic hypothesis which constraints the research space under the condition that a point is never occluded[11, 12]. This method implies the search of points that are never occluded during the sequence (inliers) and as such does not explicitly detect occlusion. A recent method allows to capture the surface of the body[13]. First a 3D “pin model” is generated and fitted with a human model manually. This model is based on the skeleton with normals that represents the volume of the body. Thanks to silhouette extraction combined with learning, normals are refined and provide the surface of the body which deforms during movements. This last work presents impressive results but implies fitting of the pin model to the video of the subject: it is currently made thanks to an accurate motion capture system. Our goal is to use standard video equipement to study muscles deformations.

4 Overview of our non-linear motion capture system

We want to obtain real surface data rather than simulate them. Instead of predicting surfaces given the animation skeleton, we have taken the reverse approach : directly capture data with calibrated cameras and deduce the position and orientation of each skeleton joint to give a reference to animators. Our problem can be divided into two parts: firstly, the 2D tracking problem, and secondly, the 3D reconstruction problem. The main difficulty is to guarantee temporal coherence while reconstructing surfaces even

if tracked points are occluded.

5 Tracking features

5.1 Video protocol and stretchy suit

Natural skin does not give enough textured information to be easily and reliably tracked, so the first idea is to texture it in order to simplify the tracking. Our solution is to wear a stretchy and textured suit which follows accurately skin deformations. We concentrate our work on the arm. We define a detailed protocol based on anatomy [14] in order to better examinate each possible movement. We define a rest position. It is important because our method take this position as a reference : the desired tracked surface must be builded in order to analyse the structure of the pattern and to compute its statistics. Our suit is textured with a checker. Each square is initially 2 cm large when the tissue is unstretched (cf fig 1).

5.2 Extraction and Tracking

We use Lucas Kanade method on the two views [15, 16] and get interest points of every visible corners of the checker. We track them at each time step during the sequence. Because our suit is well-textured, interest points are easily detetected and tracked between two steps in real-time (25 Hz interleaved video frame rate, cf fig 1). Then we have to match tracked features between two views. The matching process is made by evaluating the closest distance between the epipolar line of the first view and potential corresponding points of the second view. Our texture is largely auto-similar and thus provides more than one candidate. The matching problem is ultimately solved using the full 3D triangulation and the statistical criterion described in section 6.1.

6 3D Reconstruction and occlusion management

We search the 3D point Q which minimizes distances between the projection rays given by the image points of N views [**?**]. We have to solve

Figure 1: Interest points in green in the two views on the rest pose

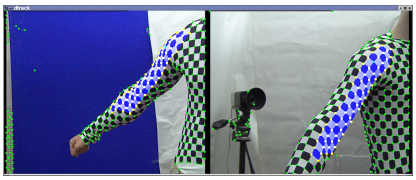

Figure 2: 3D points reprojected in the two views in yellow

a least-squares problem. We compute it by performing a SVD decomposition (cf fig 2).

6.1 Statistical learning of the edges's dimensions

In order to deal with occlusions and to increase reconstruction stability we have to know the position of each point relatively to each other, i.e. the mesh position and regularity. We process once the triangulation on the initial posture. It gives us the surface topology which should be respected during all the sequence. The mesh we obtain is the exact copy of our tissue texture, and we can calculate 4-Neighbours of each point, which give us informations about edges. Thus we dispose of an implicit 3D model without such techniques like model-based tracking. Then we compute at the initialization statistics about edges : their average length, variance and standard deviation.

$$D_i = \frac{1}{n_v}\sum_{i=1}^{n_v} d_i$$

$$\mu = \frac{1}{n_p}\sum_{i=1}^{n_p} D_i \quad \nu = \frac{1}{n_p}\sum_{i=1}^{n}(D_i - \mu)^2 \quad \sigma = \sqrt{\nu}$$

with d_i the distance between the point and its neighbours i, n_p, n_v the number of points and neighbours.
Since the mesh is the exact replic of our tissu we

can expect statistics that match with real lenghts of edges: we obtain $\mu = 25.2$ mm and $\sigma = 2.5$ mm, that is conceivable due to the stretching.

6.2 Validity criterion

We can deduce a local criterion for each point that decides if a point is occluded or not : we define a validity interval for the distance between a point i and each neighbour j : $d_{i,j}$ with $j = 1..4$:

$$\mu - c\sigma \leq d_{i,j} \leq \mu + c\sigma \quad (1)$$

where c caracterizes the elasticity of the tissue. We choose $c = 2.5$, which corresponds to 95% of a Gaussian distribution. The reconstruced points that do not satisfy the criterion are recomputed with a pose estimation method.

6.3 Estimation of one-view occluded points

If a point does not satisfy the criterion, it may remain visible in one of the views. Thus we can see this point reconstruction problem as a classic pose estimation problem with one view. Certain recent approaches use homography between two squares for estimating the points position, but in our case the four points of a square are not in the same plane so the reconstruction may be invalid [2]. We choose to use triangle pose estimation. We make one assumption for estimating these occluded points: distances between a point and its neighbours are very similar between two time steps. This is true if we assume very small local deformations between two steps and according to the pattern resolution. Each point of our mesh gets two neighbours or more, so with each occluded point we get always a triangle and three distances that allows us to constraint the point's depth. We are looking for the three depht coefficients λ : we can find them by an optimisation process. Temporal coherence helps us to make it converging to the good minimum : we initialize the optimisation problem with the lasts known depth factors λ :

$$Min(\sum_{j,l} \left(\|Q_j(\lambda_j)Q_l(\lambda_l)\|_2 - d_{jl}\right)^2)$$

considering

$$P_i Q_j = \lambda_j q_j \quad \forall j, l = \{1, 2, 3\}$$

We solve it by Newton's method and it converges in few iterations.

6.4 Filling holes

If a tracked point is totally occluded in all views, i.e. the local criterion is still above to the statistical threshold, the last solution is to interpolate it. A relaxation method can be used to fill holes. We define three ways to interpolate : using a simple interpolation with neighbours (SI); using the average neighbours deplacement between two time steps (ATD); using interpolation between last known tangents assuming local invariance between two steps (TTI). We compute interpolation error comparing with non-occluded reconstructed points and choose to conserve the last one :

Method	SI	ATD	TTI
Variance error (mm)	2-5	1-8	0-3

The last statistics confirm our hypothesis: the distances are localy invariant between two steps.

6.5 Tracking guidance and adaptative geometry constraints

When points could not be tracked, that is to say may be occluded, we replace the tracking with an estimation of where they might be: we reproject the result of the pose estimation in all views and replace positions of tracked points. This prediction aims at providing a guidance to the tracker, without loosing temporal coherence, when later the point will not be occulded any more (for example points on the arms when the elbow is rotated back and forth). This process works for small occlusions but we can better constraint the problem thanks to an analysis of topology combined with an innovation tracking : we capture a smooth surface, thus we can deduce "continuity constraints" of lines composing it. As a consequence we can constraint non plausible lines directions. In parallel of Lucas Kanade tracking we use a simple interest lines detection initialized with the first reconstructed mesh. If points are occluded we try to detect again lines at each step until occlusion has stopped. The 2D innovation points are conserved in all views using a new 3D reconstruction if they pass the validity criterion on edge

statistics. The whole process is summerized in figure 3.

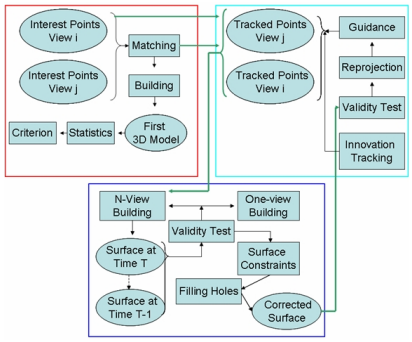

Figure 3: The whole process : in red initialization, in blue reconstruction, in cyan tracking guidance

6.6 Estimation of the rigid motion

With a goal of reusing deforming surfaces in a standard animation framework, we had to find a practical way to control deformations. Most of the time animators use an animation skeleton, thus it is required to match the deformation with the orientation and position of the skeleton joints. Thanks to the temporal coherence guaranted by our tracking, We just have to identify certain tracked points which correspond to joints of the skeleton, and then calculate their positions and orientations. For example in our case we identify 5 points as it is done in a traditionnal motion capture sequence (cf fig 4).

7 Results

7.1 Occlusions detection and Pose estimation

Without any occlusion our method works very well (cf fig 5). Occlusions are detected in red.

Figure 4: Joints identification of the arm

Figure 5: Example of tracked reconstructed surface without occlusion

Figure 6: Invalid tracked surfaces in red

We can see the validity criterion is pertinent, and even before a point disappear totally, it introduces spatial incoherences (cf fig 6). After this step the fist occludes the tracked points in the second view. Thus we estimate the 3D thanks to the last known distances of the reconstructed model and the other view (cf fig 7). We combine 3D reconstruction and 2D line tracking to guide the tracking of points where they will not be occluded any more.

7.2 Integration in a sequence

We have used our reconstruction as the control polygon of a subdivision surface integrated in an arm sequence to show how to use our results in an animation software. The points we defined previously are the same used in a standard motion capture sequence. As a consequence, we only have to match them with the skeleton of the animated 3D model under the software and scale the surface using the joints, than can be done automatically. In our case, only the first

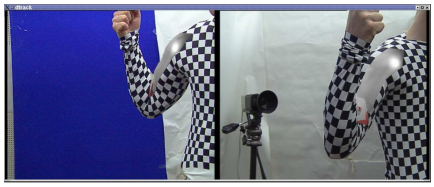

Figure 7: Reconstruction with large occlusions

3 points are used to match the surface and the second and the third one (counting from up to down) defines the scale factor.

8 Discussion and futur works

Our method for reconstructing muscles in motion gives a good estimation of non-linear surface movements and can be integrated easily in an animation software. It can be fully automatised and improved by adding cameras. Methods we use to solve reconstruction and pose estimation problem are sub-optimal and could be improved significantly by using other recent techniques [2, 3]. A very interesting issue will be to study which points give the surface its non-linear deformation and to find how to amplify them in order to increase or decrease muscles buggling in a realistic way. Then, we could analyse and reconstruct all the possible movements in order to create a statistical model which could generate new movements we did not capture. Indeed we captured all basis movement, thus a new movement will necessary be a combination of them. With a optimisation process we could find the right one and deduce innovant postures and associated deformations.

References

[1] L. Herda, P. Fua, R. Plänkers, D., R. Boulic, and D. Thalmann. Using skeleton-based tracking to increase the reliability of optical motion capture. *Human Movement Science Journal*, 2001. In Press.

[2] D. Pritchard and W. Heidrich. Cloth motion capture. In *Eurographics*, 2003.

[3] Igor Guskov Sergei Klibanov Benjamin Bryant. Trackable surface. In *Eurographics*, 2003.

[4] Xiaohuan Corina Wang and Cary Phillips. Multi-weight enveloping : Least-squares approximation techniques for skin animation. In *SCA*, 2002.

[5] Singh et al. Skinning characters using surface-oriented ffd. In *Graphic Interface*, 2000.

[6] Alex Mohr and Michael Gleicher. Building efficient , accurate character skins from examples. In *to appear in ACM SIGGRAPH*, 2003.

[7] Ferdi Scheepers, Richard E. Parent, Wayne E. Carlson, and Stephen F. May. Anatomy-based modeling of the human musculature. In *ACM SIGGRAPH*, number Annual Conference Series, pages 163–172, 1997.

[8] Luciana Porcher Nedel. Animation of virtual human bodies using motion capture devices. In *Proc. 2nd Brazilian Workshop on Virtual Reality - WRV'99*, 1999.

[9] Brett Allen, Brian Curless, and Zoran Popovic. Articulated body deformation from range scan data. In *ACM SIGGRAPH*, 2002.

[10] R. Plänkers and P. Fua. Articulated soft objects for video-based body modeling. In *ICCV*, Vancouver, Canada, July 2001.

[11] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *IEEE Conf.Computer Vision and Pattern Recognition*, 16, 2000.

[12] L. Torresani and Chris Bregler. Space-time tracking. In *ECCV*, pages 801–812, 2002.

[13] Jovan Popovic and Peter Sand. Continuous capture of skin deformation. In *ACM SIGGRAPH*, 2003.

[14] Henry Rouvire et Andr Delmas. *Anatomie descriptive topographique et fonctionnelle Tome 1*. Masson, 1960.

[15] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.

[16] Jianbo Shi and Carlo Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593 – 600, 1994.