Video-based Toon Character from Surface Performance Capture

Yann Savoye*
INRIA Bordeaux University, France.

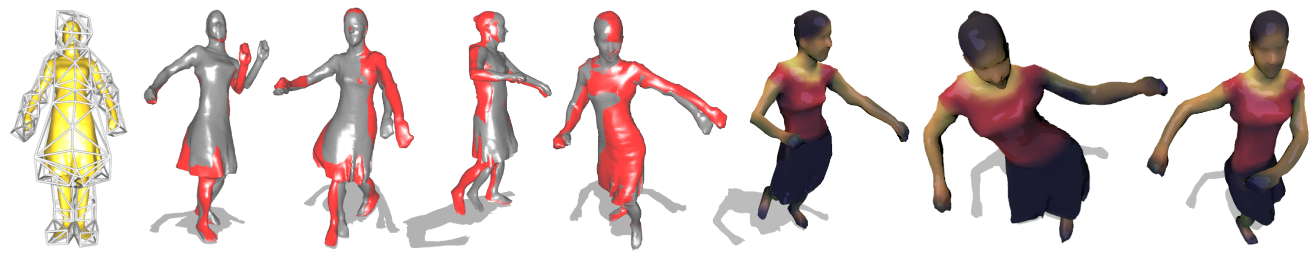

Figure 1: ***Real-Life Cartoon 3D Video:*** *Sparse-and-detached surface trajectories are first estimated over the whole mesh sequence of a captured dance performance, to be exaggerated secondly thanks to a non-rigid squash-and-stretch motion filter. For visual comparison, the reconstructed cage-based shape is displayed with grey color, superposed with the corresponding new filtered mesh with red color. Finally, the resulting cartoon-style mesh animation is real-time rendered using stylized subspace deformation coupled with video-based toon appearance.*

1 Introduction

Recent advances in dynamic surface capture have made the creation of a realistic animation a promising task for modern visual media production such as computer generated movies or 3D cinematographic video games. Nonetheless, automatic creation of cartoon mesh animations, driven by real-life cues, is still a costly and time-consuming process and presents a number of hard technical challenges. To the best of our knowledge, we propose the first attempt at generating as-photorealistic-as possible cartoon animation from markerless surface performance capture. The purpose of synthesizing new puppetry animation demonstrating fidelity to the spirit of comic book style with more exaggerated motion while preserving extreme captured cloth wrinkles is difficult to achieve. Consequently, the key contribution of our work focuses on a novel cartoon stylization approach for 3D video that efficiently reuse temporally consistent dynamic surface sequence captured from real-world actor performance. In particular, our simple and effective algorithm converts realistic spatiotemporal captured surfaces and multi-view data into an exaggerated life-like squash-and-stretch shape evolution coupled with context-aware cartoon-style expressive rendering.

2 Our Approach

Inspired by the philosophy adopted for the production of *The Adventures of Tintin: Secret of the Unicorn*, an upcoming screen adaptation including performance capture produced by Weta Digital®, we propose an original technique for generating quality life-like non-photorealistic animation from highly detail animation and photometric cues captured from real actor performance. Contrary to the techniques presented in [Kwon and Lee 2011], that only address the stylization of articulated rigid motion transferred to skinned meshes, our method generates high-quality cartoon-style 3D video from real-world dynamic scenes relying on non-rigid underlying animation structure. Since full-body performance capture is limited by the physicality of real actors actions, we add more non-natural squash-and-stretch effects on the global characteristic of photo-real dynamic surface motions. In addition, original video-driven texture is not convincing for comic adaptation and invites us to mimic hand-drawn appearance without the intervention of an artist during the whole process. To the best of our knowledge, our proposed method is the first attempt to coherently stylize highly non-rigid captured surface deformation while preserving photorealistic cues using non-rigid animation subspace temporal filtering combined with video-based toon-style shading. As a result, the core algorithm of our approach is decoupled in two major steps: *exaggerating motion of life-like surface* and *depicting video-infused appearance*.

*ysavoye@siggraph.org

3 Exaggerating Motion of Life-Like Surface

The continuous observed dynamic scene is represented by a regular sampled sequence of non-rigid temporally consistent triangular meshes $\mathcal{A} = \{\mathcal{M}_0(F, V_0), \cdots, \mathcal{M}_t(F, V_t)\}$ with consistent global connectivity F given by the trackable laser scanned template shape. We denote by $\mathcal{M}(V, F)$ a triangular mesh with V the set of n vertices and $\mathbf{v}_i^t \in \mathbb{R}^3$ the location of the i^{th} vertex at frame t. Let $\Omega \subset \mathbb{R}^3$ designates the volumetric domain included by m cage control handles. The bounding cage is represented by a piecewise linear surface defined by these handles. Given the original input mesh sequence $\mathcal{A}$ and the specified bounding polyhedra $\mathcal{B}$ associated with given improved rigging function, a low-dimensional non-rigid surface motion signal is a function $s(t) = (\mathbf{c}_1^t, \cdots, \mathbf{c}_m^t)$ describing the shape in motion into the harmonic projective subspace basis where $\mathbf{c}_j^t \in \Omega$ is a 3-vector indicating the location of the j^{th} cage handle in global coordinates system at the time step t. Since it is impractical to achieve global perturbation directly on the surface itself, we establish a high-level shape analysis process to temporally register the surface motion signal. At each frame, this signal represented by a vector $\mathbf{c}^t$ of compact and non-hierarchical surface-free motion parameters is learnt by solving the following objective functional:

$$\underset{\mathbf{c}^t}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \left\| \mathbf{v}_i^t - \sum_{j=1}^{m} w_{ij} \cdot \mathbf{c}_j^t \right\|_2^2 \right)$$

where $w_{ij} : \Omega \to \mathbb{R}$ is the influence weight given by the function w precomputed, once at bind time, for a cage handle j associated to an enclosed mesh vertex i. We choose w to restrict the degree of freedom to harmonic (easily extendable to biharmonic) deformation of a specified flexible cage as proposed in [Joshi et al. 2007], retaining good surface characteristic. The extracted signal is rearranged to form a set of per-handle trajectories defining a value for a given cage handle as a function of the time. We apply the cartoon animation filter suggested in [Wang et al. 2006], as signal enhancement filter to add follow-through exaggeration and anticipation effects of large-scale deformation on the motion signal without losing small-scale details already encapsulated in the acquired surface as follows:

$$\widetilde{s}(t) = s(t) - s(t) \otimes LoG$$

where $s(t)$ is the input signal to be altered, LoG is the Laplacian of Gaussian, $\widetilde{s}(t)$ the output filtered signal and $\otimes$ the convolution operator. This approach convolves each handle trajectory with an inverted Laplacian of a Gaussian filter to create a cartoon-style subspace thanks to the negative lobes of LoG filter. The desired squash-and-stretch surface is then synthesized by transferring the filtered signal to the surface via the same space deformation technique previously used during the extraction procedure.

4 Depicting Video-Infused Appearance

The second phase of the algorithm consists in establishing video-driven appearance that looks like a hand-drawn picture in respect to the sequence of captured images, as shown in Figure 2. On account of tincture and skin pigmentation properties are intrinsically space-time invariant, we opt for reconstructing an omnidirectional multi-view color mapping by globally aggregating per-vertex reprojected color through time. We estimate a pseudo-material intensity distribution over the surface by blending calibrated multi-view real matted image over the whole temporal sequence. To be noted that exploring the complete multi-view sequence allows us to collect an average overtime color information representing the texture more globally, especially in occluded zones of the surface for an arbitrary frame or view. The weighted blending function takes into account the angle between the vertex normal, the camera viewing vector and the vertex visibility. To eliminate artifact or noisy texel and to reduce over-photorealistic local details while ensuring global seamless intensity, we iteratively apply a per-vertex smoothing filter followed by a sharpening operator on the color components.

(a) **(b)** **(c)** **(d)**

Figure 2: ***Appearance Generation*** *is performed in various steps:* ***(a)*** *captured surface reprojection,* ***(b)*** *multi-view texture mapping,* ***(c)*** *per-vertex filtering and* ***(d)*** *per-pixel cartoon style enhancement.*

As illustrated in Figure 3, the final lighting equation is then obtained by a controllable blending of the treated video-based photometric cues with the traditional toon shading. The filtered space-based shape animation is rendered using variable quantization and fixed-function outlining. Variable quantization involves coloring object surfaces in a step-wise colorization manner and outlining effect enhances the body shape by drawing its suggestive contour with thick and black line segments, such as in [Barla et al. 2006]. For the sake of simplicity, directional lighting is used in our rendering engine.

Figure 3: ***Video-based Toon Texture.*** *Our body and clothing appearance model is described as the fusion of video-based photometric information with controllable quantization and outlining.*

5 Results and Discussion

As proof of concept, we use the technique of [Vlasic et al. 2008] to generate the sequence of mesh animation from eight cameras regularly spaced in a chroma-key green room. We bound the laser scanned template to a cage containing approximately hundred vertices, distributed in a shape-aware manner. Our technique is generalizable to any temporal consistent mesh sequence, such as [Huang et al. 2011]. Our process is done with interactive rate and produces visually aesthetic effects. The resulting animation is real-time rendered using cage-based skinning with harmonic coordinates. The behavior of our techniques is illustrated in the accompanying video as well as in Figure 1 and 4. We demonstrate the efficiency of our method on various challenging dataset particularly on sequence of the samba dance, where the highly non-rigid natural-looking wrinkles on loose apparel like a skirt and large deformation are emphasized by style subspace while ensuring its motion semantic.

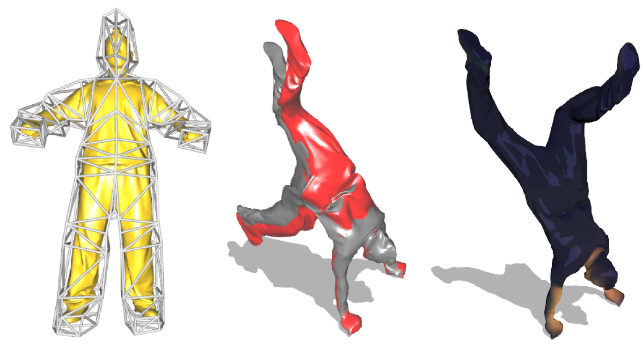

Figure 4: ***Video-based Toon Character:*** *Cage-based handles leading to temporal trajectories (**left**). Superposition of original space-based deformation in gray color with cartoon filtered surface motion in red color (**center**). Video-injected toon appearance (**right**).*

6 Conclusions and Future Work

In this work, we revisited the *illusion of life* principle by applying a visual metaphor to emphasize natural expressivity of the motion and appearance of performance capture meshes with non-natural appealing effect. Accordingly, our novel method intentionally does not fit neatly in either the animation category or live action and leads to the introduction of a new application for multi-view performance capture. The proposed semi-automatic approach brings real-life animation to a controllable cartoon adaptation in the spirit of the captured visual information. To reach best perceptual impact, we decoupled the performance capture stylization on underlying non-rigid surface motion parameters as well as the reconstructed appearance. Finally, injecting cartoon exaggeration in video-based life-like non-rigid surface brings a heightened believability of animated toon characters. In the future, we plan to explore more opportunities in the field of computer-generated dynamic surface motion while preserving the life-like visual appealing.

References

Barla, P., Thollot, J., and Markosian, L. 2006. X-toon: An extended toon shader. *Symposium on Non-Photorealistic Animation and Rendering (NPAR)*.

Huang, P., Budd, C., and Hilton, A. 2011. Global temporal registration of multiple non-rigid surface sequences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Joshi, P., Meyer, M., DeRose, T., Green, B., and Sanocki, T. 2007. Harmonic coordinates for character articulation. *ACM Trans. Graph. 26* (July).

Kwon, J.-Y., and Lee, I.-K. 2011. The squash-and-stretch stylization for character motions. *IEEE Transactions on Visualization and Computer Graphics*.

Vlasic, D., Baran, I., Matusik, W., and Popović, J. 2008. Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph. 27* (August), 97:1–97:9.

Wang, J., Drucker, S. M., Agrawala, M., and Cohen, M. F. 2006. The cartoon animation filter. *ACM Trans. Graph. 25* (July), 1169–1173.