To appear in the 9th European Conference on Visual Media Production proceedings

Content-Aware Exaggerated Editing for Life-Like Captured Animations

Yann Savoye*

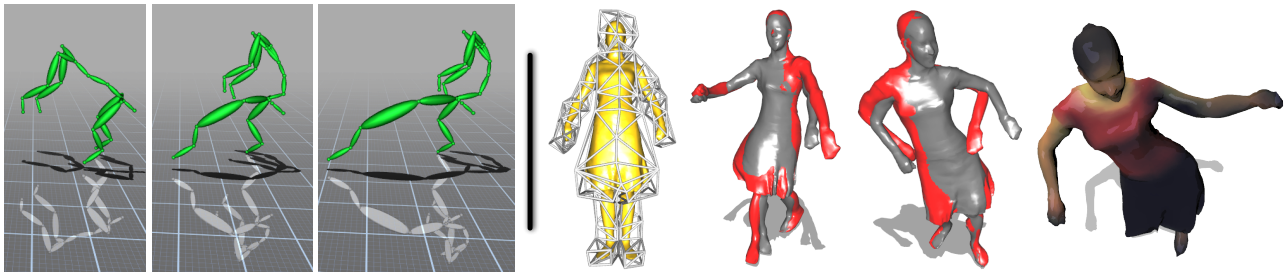

Figure 1: ***Context-Aware Cartoonization of Captured Animations.*** *An input skeletal Vicon® motion clip capturing a horse jump is re-edited into a kinematic-aware squashed-and-stretched figure, in the spirit of Tex Avery® TV cartoons, thanks to a controllable variational optimization (**left-hand side**). Dynamic captured surface of a real dancing performance (displayed in gray color) is approximated by cage-based encoding, to be exaggerated secondly using a squash-and-stretch non-rigid surface filter. For visual comparison, the stylized subspace-based shape is superimposed in red color. Finally, the resulting cartoon-style mesh animation is real-time rendered via video-infused toon appearance (**right-hand side**).*

Abstract

Cutting-edge efforts have been invested in the automatic production of breath-taking visual effects involving time-varying data captured from real-actor performances. However, one of the biggest challenges for computer-generated imagery is the puppetry of heterogeneous captured data, without the heavy use of trained artistic skills. We focus on achieving desired exaggerated animations coherently while preserving baked-in life-like visual cues. In this paper, we propose a new method to generate content-aware exaggerated captured animations by melting motion, shape and appearance properties. In particular, our suggested simple-yet-accurate approach explores two closely tools that serve the common theme of *Animation-Cartoonization*. The first one consists in realizing articulated-based stretchable cartoon editing from marker-based mocap clips. The second one generates video-based toon character from surface performance capture. Finally, we demonstrate the flexibility and stability of our approach on a variety of captured animations as input. Our well-formed scheme can be useful for low-budget productions of cinematographic games or movies.

1 Introduction

The analysis of shapes or stick figures in motion is the very foundation of structure-preserving editing from performance-based captured animations. Recent advances in dynamic surface capture and low-cost motion capture have made the creation of realistic animations a promising task for visual media production, such as movies or cinematographic video games. For instance, we observe the growing demand for re-synthesizing new expressive animation from already captured heterogeneous data of real actor performance. Unfortunately, most of interactive 3D software is ill-equipped to handle mixed sensor-based animations data such as 3D scans, multi-camera data and motion clips. In particular, the problem of processing information from the rich digitalization of living creatures to mesh puppetry animation is nearly unaddressed previously. Moreover, controllable creation of cartoon mesh animations, driven by real-life cues, is still a costly and time-consuming process and presents a number of technical challenges. Hence, generating as-photorealistic-as possible cartoon animation from markerless surface performance capture and extreme globally-coherent stretching from skeletal mocap area are key challenging problems. Synthesizing new puppetry animation, demonstrating fidelity to the spirit of TV cartoon with more exaggerated motion while preserving captured extreme cloth wrinkles of a real actor performance, is a fascinating research problem with direct applications in industry.

*ysavoye@siggraph.org

Problem and Terminology. Even if the capture of visual dynamics of living creatures in the form of 3D dense mesh surface or joint angles is popular, the coherent reediting of such animations is a difficult problem. Obviously, solutions must not only fit some basic mathematical criteria, but it must look and feel appropriate to animators and to the audience. In our terminology, *content-aware exaggeration* refers to the interactive perturbation of the input representation of live-action performance into a fully modular digital puppetry. We emphasize captured information such as articulated motion, shape evolution and related appearance with respect to the inherent content of the original input captured flow. Additionally, we define *life-like captured animations* as a class of time-varying consistent heterogeneous data grabbed by arbitrary sensors that encapsulate life-cues such as biological motion or subtle surface motion details like flesh elements and dynamic cloth wrinkles.

Motivations. Our main motivation is to explore opportunities offered by underlying subspaces to reuse heterogeneous time-varying captured data into animator-friendly animation system. Meanwhile, we focus on enhancing believability in cartoon animation by injecting life elements into fine-tuning reconstructed cartoon characters relying on disparate captured data. Our ultimate goal is to design reproducible techniques allowing the creation of visually rich animations from various captured data with simple tools. In this work, we are devoted to pushing the limits of what can be visually recreated from living-creatures physicality by generating *larger than life* squash-and-stretch effects. Resulting arbitrary underlying structures are reused secondly to drive stylized real-animation rendering with an inferred-to-depicted appearance that clearly communicates the performance's essence to the audience. In this paper, we try to bridge the gap between *Computer Animation*, *Computer Vision* and *Expressive Rendering* toward the reuse of captured anatomical shape, motion and multi-view video footage easily.

Contributions. The key contribution of this paper is an interactive process of cartoonization from mocap and multi-view data as shown on Figure 1. First, we propose a variational strategy for stretching skeletal captured animations that could drive shape deformation. Second, we propose a new approach to obtain video-based toon character from surface performance capture. In particular, our effective algorithm converts realistic spatiotemporal captured surface and multi-view data into exaggerated squash-and-stretch lifelike shape evolution coupled with video-aware cartoon-style expressive rendering. For the sake of clarity, it is worth to note that we propose a skeleton animation paradigm for sparse marker-based mocap but we must switch to cage-based animation for dense surface mocap data. Both techniques fit together toward the *Animation-Cartoonization* of heterogeneous captured data.

2 Background and Related Works

There is a great deal of recent research in a number of orthogonal areas that are of interest for our proposed method. In this section, we briefly review few recent works that are relevant to the problem of *content-aware exaggeration for captured animations*, covering the following main categories: captured body tracking, skeleton-based and cage-based encoding, motion and appearance stylization.

Captured-Body Tracking. A large variety of sensors roughly perform 4D acquisition of life-like performances. Consequently, heterogeneous outputs such as skeletal motion clips, multi-view data or silhouette-consistent mesh sequences are collected massively. As proposed by Kirk *et al.* [2005], early approaches estimate the subject skeletal location. More recently, Baak *et al.* [2011] demonstrated full-body motions tracking from a single time-of-flight camera. Moreover, modeling human body by sums of spatial Gaussian is attempted by Stoll *et al.* [2011] for fast marker-less motion capture. A promising breakthrough approach presented by Shiratori *et al.* [2011] overcomes classical outdoor acquisition problems by using outward-looking body-mounted cameras. Nowadays, non-intrusive techniques have been proposed to capture feature-rich output dynamic surfaces from multi-video stream without the need of wearable suit. For instance, the pioneering work on surface capture by Starck *et al.* [2007] is extended by Huang *et al.* [2011] to obtain temporally consistent non-rigid surfaces. In addition, De Aguiar *et al.* [2008] performs purely geometric mesh tracking using SIFT feature correspondences. However, only few techniques succeed in fitting simultaneously articulated skinned template to sparse multi-view silhouette with additional silhouette rim vertices correction as seen in [Vlasic et al. 2008; Ballan and Cortelazzo 2008; Gall et al. 2009]. Nevertheless, dynamic shape can be recovered using multi-view photometric stereo normal maps as proposed by Vlasic *et al.* [2009]. Not to mention that Liu *et al.* [2011] exhibit a segmentation-based and marker-less motion capture pipeline to estimate the pose-and-surface of two-people in the same video. Admitting our work is not devoted to the acquisition of animation, but inspired by the reconstruction of animatable human characters of Stoll *et al.* [2010], we focus firstly on reusing versatile data by exaggerating captured dynamic surface and performing variational stretchable skeletal editing.

Skeleton-based Animation. Biological motion perception reveals that an animated creature can be recreated globally just by a dozen of dots craftily placed on the body surface. Introduced by Magnenat-Thalmann *et al.* [1988], the skeleton-driven mesh deformation is a popular method thanks to its efficiency and simplicity. This class of deformation exploits the benefit of a generic hierarchy of kinematic articulations in order to conduct a local parametrization of a wide variety of enveloping skins. Traditional skeleton animation is often associated with a rigging and skinning function. An efficient technique to automatically rig a prior skeletal structure with a static two-manifold skin mesh is proposed by Baran *et al.* [2007] by employing two components: skeleton embedding and skin attachment. A similar example-based computation of joint influence weight distribution is proposed in [Weber et al. 2007]. While the dual quaternion blending detailed by Kavan *et al.* [2008] improves the geometric skinning, Larboulette *et al.* [2005] enhance animations of characters by adding a dynamic skin response to advocate unrealistic effects provoked by deforming a surface with underlying skeleton. More importantly, approximating non-rigid animation with skeleton-free skinning is an appealing idea, firstly addressed by James *et al.* [2005], but apparently unsuitable for captured dynamic surfaces. In contrast with methods [Kavan et al. 2010; Miller et al. 2010], our technique is the first to demonstrate silhouette-aware properties. On the top of that, we offer reusable and controllable parameters for re-skinning life-life captured surface of people in large clothing like a waving skirt.

Cage-based Encoding. Recent years have seen attractive interest for cage-based deformation: an emerging class of purely geometric space-based techniques, widely used for controlling meshes enclosed in a flexible coarse bounding polytope. Then, a rigging function expressed as generalized barycentric coordinates is associated with the *cage-model paradigm* to describe realistic and natural control behavior of the deformation, preserving the fine geometric details inherently. *Harmonic Coordinates* are introduced by Joshi *et al.* [2007] as the solution of volumetric heat diffusion in the cage interior with boundary conditions. Ben-Chen *et al.* [2009b] prefer to estimate harmonic maps using a set of harmonic basis functions computed from a collection of example poses. For the sake of sparsity, the deformation can be restricted to *Bounded Biharmonic Weights* as introduced by Jacobson *et al.* [2011]. Many improvements have been also proposed by Lipman *et al.* [2008] where *Green Coordinates* enforce quasi-conformal scale-preserving deformation and allow the use of partial cage. The problem of reusing static cage templates is addressed by Ju *et al.* [2008] where reassembly of pre-designed partial cage are proposed for compatible reuse. Furthermore, learning deformation for a new shape from an existing pose is developed Ben Chen *et al.* [2009a] to demonstrate a cage-based method for transferring animation. Unlike prior efforts, our work is the first to investigate skeleton-less cartoon filtering for dynamic captured surfaces relying on cage-based shape approximations.

Motion and Appearance Stylization. Non-photorealistic animation and rendering have recently become increasingly popular. Numerous approaches have investigated the problem of stylizing existing skeletal motion. Relatively small number of approaches emulate rubber-like exaggeration effects on skeletal motion data as shown in [Kwon and Lee 2007; Bregler et al. 2002]. Another approach of style-based inverse kinematics is proposed by Grochow *et al.* [2004]. Generating artist-inspired motion derived from 3D skeletal motion capture data is also studied by Bouvier *et al.* [2007]. Contrary to techniques of Kwon *et al.* [2011] that only address the stylization of articulated motions, our method demonstrates the generalization to high-quality 3D surface video relying on non-rigid underlying structure. To the best of our knowledge, motion stylization of cage-based shapes has never been addressed for non-rigid dynamic surfaces. Nonetheless, our method allows topological-coherent stretchable skeletal editing from mocap data with few manual constraints. Independently, intensive research in expressive rendering enables a wide variety of styles. Producing hand-drawn cartoon texture for 2D animation is presented in [de Juan and Bodenheimer 2004; Sýkora et al. 2011]. Traditional toon shading has been used by Kong *et al.* [2009] to depict morphological object features and extended by Barla *et al.* [2006] to support view-dependent effects. Another non-photorealistic shading is investigated by Rusinkiewicz *et al.* [2006] to exaggerate and depict shape. Meanwhile, in the field of vision-based projective rendering, multi-view texture mapping approach of Eisemann *et al.* [2008] only suggest preserving crisp of detailed texture appearance while avoiding artifacts. Finally, our work tries to rethink the capability of filling the gap between vision-based graphics and appearance stylization. To the best of our knowledge, content-aware stylization of captured life-like surface has never been studied before. Therefore, our work is the first attempt to exaggerate non-rigid deforming surface coherently in term of surface motion, mixed with the multi-view appearance.

The remainder of this paper is organized as follows. Our variational mocap-based stretchable editing procedure is explained in Section 3, with a novel comprehensive formulation. Next, our new techniques for exaggerating motion and depicting video-infused appearance of captured surfaces are described in Section 4. Then, experimental results are discussed in Section 5. Finally, the paper is concluded and an outlook to future work is given in Section 6.

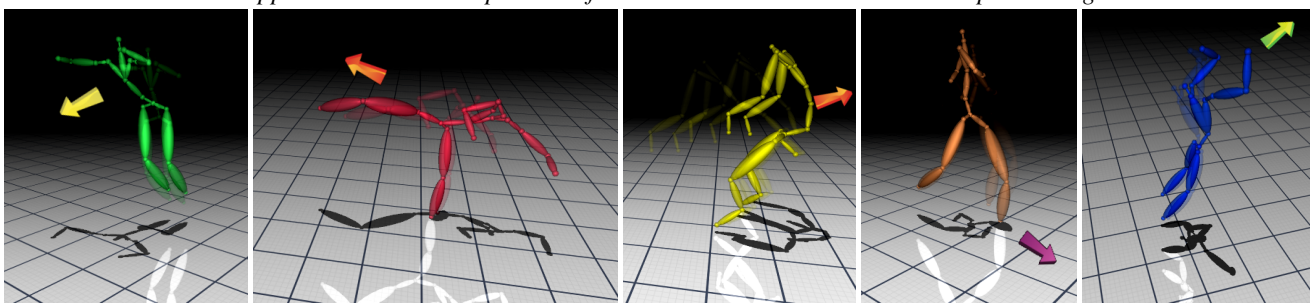

Figure 2: ***Variational Motion-Aware Skeleton Editing.*** *Optimizing traditional skeletal animated structure with differential-aware bone stretching effects allows animators to re-use previously captured mocap data efficiently in the context of cartoon production. We increase Squash-and-Stretch effect on existing realistic* Jumping, Karate Kick-Off, Horse Ruade, Dancing *and* Basketball Dunk *motion clips (**from left to right hand side**). Original and intermediate edited poses are displayed in transparency, and resulting edited poses in superposition.*

3 Skeletal-based Exaggerated Editing

In this section, we describe a painless variational skeleton editing approach to re-use skeleton of animation with joint-based Laplacian-type regularization, in the context of exaggerated skeleton-based character animation. Living creatures stretch because of the elasticity of tendons and muscles. Consequently, breaking the rigidity of the underlying armature adds pleasant realistic to real vertebral motion. A major subproblem for artists in production is to enhance the expressiveness of classical motion clips. It can be observed that most of characters in TV cartoons have the flexibility to stretch to extreme positions and squash to astounding shapes. In addition, we notice this effect is easier to realize in traditional animation than in computer-generated animation. For this reason, *Ratatouille* a Pixar® movie, did not use a rigid skeleton and abandons motion capture to reach such essential ultra non-realistic appeal. Whereas, we build upon the *Theory of Plasticity* to deal with the potential of skeletal-based optimization, while preserving the joint coherence and connectivity.

3.1 Skeletal Graph Laplacian

The idea of the motion capture is to use sensors for collecting data that describe performing motion of observed articulated subjects. The pose of an articulated figure is specified by its joint configuration with respect to the position and orientation of the root joint. A skeleton of animation $\mathcal{S} = (\mathcal{J}, \mathcal{B}, \mathcal{M})$ is composed of a hierarchy of joints decorated with motion data, organized in a series of frames. $\mathcal{S}$ is made of a set $\mathcal{J}$ of n joints, a set $\mathcal{B}$ of bones connecting joints, and a set $\mathcal{M}$ of k motion frames. Two joints i and j are connected into a unique bone only if $(i, j) \in \mathcal{B}$. Motion data consist of bundle of signal defined as a continuous function $\mathbf{f}(t)$. Compiled global rigid transformation matrix from captured motion data associated to the i^{th} joint at the t^{th} frame is denoted by $\mathbf{M}_i^t \in \mathbb{R}^{4\times4}$. The global location $\mathbf{p}_i^t \in \mathbb{R}^4$ of the i^{th} joint at the t^{th} frame is the homogeneous zero transformed by the sequence of transformation and can be written as: $\mathbf{p}_i^t = \mathbf{M}_i^t \cdot \left[\vec{0}\,|1\right]^T$. The neutral element $\left[\vec{0}\,|1\right]^T \in \mathbb{R}^4$ projects the origin of local frame from Eulerian-to-Cartesian space. We define the discrete *Differential Joint Coordinates* $\boldsymbol{\delta}_i^t$ as the differential encoding of the joints deformation matrices in relationship with joint i at t^{th} frame by a geometrization process as follows:

$$\boldsymbol{\delta}_i^t = \left(\mathbf{M}_i^t - \sum_{j\in N(i)} \frac{1}{d_i}\left(\mathbf{M}_i^t - \mathbf{M}_j^t\right)\right) \cdot \left[\vec{0}\,|1\right]^T$$

The degree of the joint i denoted by d_i is equal to the number of joints linked to the given joint i. The set of immediate adjacent joints to i is denoted by $N(i) = \{j \mid (i, j) \in \mathcal{B}\}$. In addition, the skeleton topology is represented by an open directed acyclic graph. In order to provide resistance while offering elastic properties to the rigid body deformation, we generalize the stiffness of Hooke's law to non-manifold skeletal structure describing the connectivity graph between degrees of freedom through a Laplacian formulation. This type of coordinates can be computed in the same spirit for geometric skeleton by directly encoding joints location in a differential manner. Then, we introduce the *Skeletal Laplacian Matrix* $L_S \in \mathbb{R}^{n\times n}$ by $L_S = D - A$ where D is the diagonal matrix of joint degrees and A is the adjacency joint matrix. We denote by $\mathcal{L}_S(\cdot)$ the per-joint *Uniform Laplacian Operator* applied on the skeleton graph structure. The entries of the corresponding square symmetric matrix L_S are set up as follows:

$$L_S[i, j] = \begin{cases} 1 & \text{if } i = j \\ -1/d_i & \text{if } (i, j) \in \mathcal{B} \\ 0 & \text{otherwise} \end{cases}$$

We focus on adding non-rigid effects to an existing captured skeletal structure, while preserving its consistency and native connectivity. In our approach, we are referring to the well-known Laplacian shrinking effects (*i.e.,* shearing and stretching distortion) in order to apply non-rigid warps over the rigid skeleton topology.

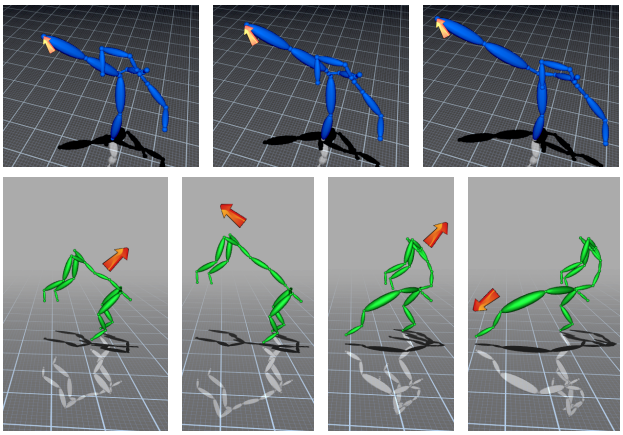

Figure 3: ***Edited Skeletal Structures in Motion.*** *A single editing feature is sufficient to produce pleasant cartoon gait style over the whole skeletal pose with spacetime coherence for a horse jump (**bottom row**) and a karate kick-off (**top row**).*

3.2 Stretchable Skeletal Optimization

Our algorithm takes an arbitrary articulated motion signal as input. Hence, we perturbate its Euler and Euclidean representations in such a way that the output motion looks more cartoon-like. Our technique has two key components: spatiotemporal motion filtering and global joint location optimization. In order to hack the inherent rigidity, we prefer to deal with the skeletal structure in its euclidean form. At the beginning of pre-optimization, we apply the cartoon animation filter on joints angles as suggested by Wang *et al.* [2006] in order to add the follow-through, exaggeration and anticipation effects on the motion signal. The filtered motion signal $\tilde{\mathbf{f}}(t)$ has intriguing properties, especially on kinematic subchains that are not explicitly edited automatically or hand-drawn specified. This filter involving a Laplacian of Gaussian *LoG* is defined as follows:

$$\tilde{\mathbf{f}}(t) = \mathbf{f}(t) - \mathbf{f}(t) \otimes LoG$$

To continue, we reformulate the *Squash-and-Stretch* problem as a skeletal adaptation optimization. Given the fact it is nearly impossible to get an utmost squash-and-stretch by working exclusively in Euler space with inverse kinematics, we prefer to optimize the whole skeletal structure in term of global joint locations from a pre-animated pose satisfying stretching constraints. The core algorithm of our technique relies on *Skeletal Graph Laplacian*, with the aim to ensure spatial relationship of joints under sparse differential-aware stretching features over the whole joint hierarchy. As a result, the key idea is to employ a fast and accurate skeleton fitting algorithm, guided by kinematic-free constraints. At each frame, the initial solution is the Euclidean parameters of the current pose, generated by forward kinematics. Thus, our system recovers the pose estimation by minimizing an objective function composed of a combination of penalty and data terms. The smoothness term is required to make the warp field regular. To control a cartoon desired pose, the user inputs the targeted global joint positions $\left\{\mathbf{q}_l^t\right\}$ for a collection C of few edited joints. We conserve the Laplacian coordinates $\boldsymbol{\delta}_i^t$ for each joint i in the skeleton hierarchy. By discretizing the potential energy of plasticity, the reconstructed positions $\widehat{\mathbf{p}}^{\,\mathbf{t}}$ of the skeleton joints coordinates are obtained in the world space by solving the following linear minimization problem:

$$\underset{\forall i,\ \widehat{\mathbf{p}}_i^{\,t}}{\operatorname{argmin}} \left(\sum_{l \in C} \left\| \mathbf{q}_l^t - \widehat{\mathbf{p}}_l^t \right\|_2^2 + \sum_{i=1}^{n} \left\| \mathcal{L}_S\left(\widehat{\mathbf{p}}_i^t\right) - \boldsymbol{\delta}_i^t \right\|_2^2 \right)$$

Global joint locations are estimated by minimizing the sum of squared difference between the data-driven pre-animated pose and input feature cues. This objective function sustains a heavy penalty if neighboring joint i and j are mapped far apart. In order to avoid purely translation effect provoked by a single dragged-and-dropped joint, we fix the root joint by default, acting as skeleton gravity center. The root joint is located generally in a region where many bones come together. In our framework, bone elongation can also be automatically established along the bone direction. Our minimization formulation takes advantage of local anisotropy by stretching bones in the minimum curvature direction. The use of an advanced constraint formulation, united with Laplacian on the bone structure, is motivated by the fact that the *Squash-and-Stretch* can be accomplished by differential scaling in Euclidean coordinate system. Hence, this minimization problem can be solved efficiently in least-squares sense based on the succeeding classical expression: $\mathbf{A} \cdot \mathbf{X}' = \mathbf{B}$. Moreover, the global location of joints can be found by solving in real-time a very small sparse linear system using this closed-form expression: $\mathbf{X}' = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{B}$. This sparse editing technique performs spatio-temporal numerical optimization on mocap data as illustrated in Figure 2 and 3.

As post-optimization step, it will be useful to enforce temporal smoothness by penalizing inter-keyframe deviation. For example, the edited motion signal can be simply filtered using iterative smoothing convolution. The resulting animated structure looks yet very pleasant while remains rigid in motion. In the context of processing video-based tracking outputs where surface and skeleton are captured simultaneously, our skeletal-based editing technique could offer the ability to control stretchable rigged shape, coupled with the skinning scheme proposed by Jacobson *et al.* [2011]. Assuming that a template mesh is attached to the skeleton at the rest pose with respect to a smooth rigging function similar to Baran *et al.* [2007], non-rigid edits are transferred to a skeletally-guided template skin as illustrated in Figure 4.

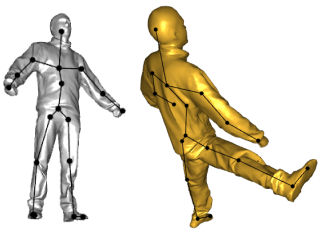

Figure 4: ***Skinned Edited Skeleton***

4 Surface-based Exaggerated Editing

Acquiring the surface behavior with markerless optical capture, rather than sparse maker-based technique, allows us to considerate the appearance editing. Meanwhile, we are forced to reconsiderate the underlying structure because the skeleton appears impractical to preserve small scale surface features of input dense surface mocap under exaggerated global surface motion. As illustrated in Figure 5, real-world captured animations need more than the natural behavior of skeletons. Hence, cage-based deformation offers an elegant free-form abstraction to capture the low-frequency shape deformations correctly with reusable output parameters as shown in Figure 6.

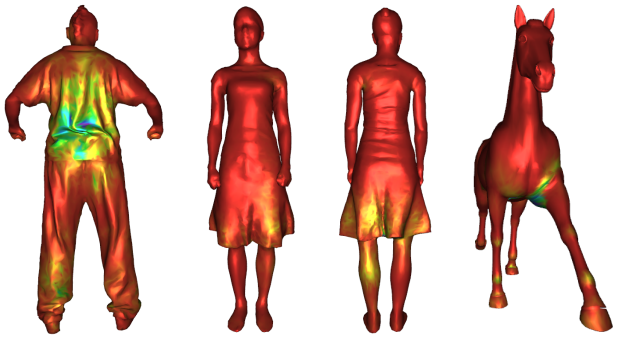

Figure 5: ***Surface Motion Analysis:*** *The color-coded edge-length deviation (computed with the techniques of [Tierny et al. 2008] on the whole mesh animation) reveals that non-rigid human animations in clothing cannot be cast as piecewise-rigid model contrary to the synthetic horse animation. Highest non-rigid areas (with respect to the full surface variations) are displayed in green-to-yellow colors.*

Inspired by the philosophy used for the production of *The Adventures of Tintin: Secret of the Unicorn*, a noteworthy screen adaptation including performance capture produced by Weta Digital®, we propose an original technique for generating quality life-like non-photorealistic animation from highly detailed animation and photometric cues captured from real actor performance. 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 motion. In addition, original video-driven texture is not convincing for comic adaptation and invite us to mimic non-ultra toon 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 stylize highly non-rigid captured surface deformation coherently while preserving photorealistic cues. In particular, we employ 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 distinct steps: *exaggerating motion of life-like surface* and *depicting video-infused appearance*.

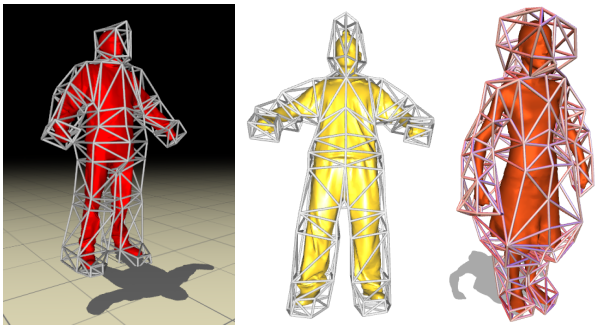

Figure 6: ***Generic Cage-Setup.*** *A default humanoid-type cage connectivity is used to perform scalable editing of whole-body models. This abstraction of skin behavior can be adjusted to a large configuration of human body topology such as girls in skirt or men in trouser. We bound the space of laser-scanned body template with a cage containing a compact set of controllable handles distributed in a shape-aware manner.*

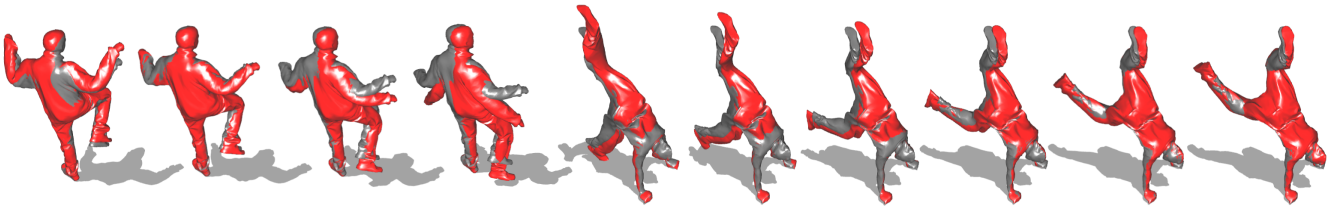

Figure 7: ***Handle-and-Motion Aware Surface Filtering.*** *Cage parametrization is a meaningful low-dimensional subspace enabling analyze-to-synthesis approach on motion variation in body shapes that evolve non-rigidly over time. Thus, we boost the captured surface dynamics (in red color) by applying the LoG filter on high-level time-varying parameters in order to preserve low-frequency information during global exaggerated surface editing. During filtering the cage-based structure allows details to be stable over time.*

4.1 Exaggerating Motion of Life-Like Surface

The continuous dynamic surface acquired by multi-view camera tracking $\mathbb{S}^t = \left\{(x,y,z) \in \mathbb{R}^3, f(x,y,z,t) = 0\right\}$ is represented by a sequence of q non-rigid temporally consistent triangular meshes $\mathcal{A} = \left\{\mathcal{M}_1(F,V_1), \cdots, \mathcal{M}_q(F,V_q)\right\}$ with consistent global connectivity F given by the tracked laser-scanned template shape. We denote by $\mathcal{M}(V,F)$ a triangular dense mesh with V the set of n vertices and $\mathbf{v}_i \in \mathbb{R}^3$ the location of the i^{th} vertex. Let $\Omega \subset \mathbb{R}^3$ designates the volumetric domain enclosed by m cage-handles of control. The bounding cage is represented by a piecewise linear surface defined by these handles. Given the original input dense mesh sequence $\mathcal{A}$ and the specified bounding polyhedra $\mathcal{B}$ associated with given rigging function w, 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 coordinate system at a given time t. The Figure 6 illustrates the genericity of our technique by adjusting roughly a common full-body cage template connectivity to a wide variety of topology for people in clothing. The accuracy of this parametrization to preserve the observed local silhouette consistency is demonstrated in Figure 9.

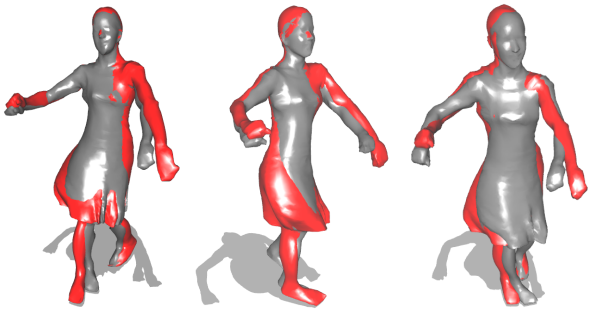

Figure 8: ***Cage-based Squash-and-Stretch.*** *The dense dynamic surface is converted into non-rigid subspace-based animation with low-dimensional parameters. Then, the motion-aware stylization of the life-like surface is performed by applying the cartoon animation filter in the reduced cage-space. As a result, we generate a sequence of stylized mesh with near-conformal mapping but smooth stretching, perserving small-scale details as-much-as possible with respect to the extracted cage handles.*

Given the fact that achieving global perturbation directly on the surface itself is impractical, we establish a high-level shape analysis process to register the surface motion signal temporally. At each frame, this best-fitting signal represented by non-hierarchical compact surface-free motion parameters is a low-rank approximation faithfully estimated by solving the following objective functional:

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

where $w_{ij} : \Omega \rightarrow \mathbb{R}$ is the influence weight given by a weighing function, 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 deformation of a flexible cage defined in the reduced domain Ω as proposed in [Joshi et al. 2007], retaining substantial global-to-local surface characteristics.

The cage tessellation with a given rigging function is meaningful to restrict deformations to a space of natural warping because human activities are often located on a latent space that is low-dimensional. Even though non-rigid cloth dynamics cannot be defined in term of sparse local rigid transformation, harmonic handles offer a suitable fashion to enforce local rigidities over the enclosed surface. In particular, handle-aware harmonic scalar fields define local soft rigidities properties over the enclosed surface, as seen in Figure 8. This restricts the surface to be deformed coherently with respect to the cage tessellation under cage-based shape cartoon filtering without decorrelating rigid and non-rigid surface components. The deformation is forced along the geodesics which impose a smooth and locally rigid deformation. Modeling the rigidness of the ambient space rather than the shape itself allows us to filter coarse-scale surface features without destroying fine-scale details.

Animators may use *Laplacian Mesh Editing* on the cage structure to stretch the model with user control. Alternatively, 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 dynamic surface without losing small-scale details already incorporated in the acquired surface, as illustrated in Figure 7. To achieve this, we obtain the filtered cage-based shape geometry directly, using the following linear cage-based filter operator:

$$\widetilde{\mathbf{v}}_i^{\,t} = \sum_{j=1}^{m} w_{ij} \cdot \left(\mathbf{c}_j^t - \mathbf{c}_j^t \otimes \ LoG \right)$$

The extracted signal is rearranged to form a set of per-handle trajectories defining a value for each cage-handles as a function of time. This approach convolves each handle trajectory with an inverted *Laplacian of Gaussian* filter to create a cartoon-style subspace thanks to the negative lobes of the *LoG* filter. For the sake of explanation, we define $\otimes$ as the convolution operator applied on a given cage-handle. Thus, the desired squash-and-stretch surface is synthesized by transferring the filtered signal to the surface via space deformation. Additionally, new in-between frames can be generated using a smoother *Cage-based Cosine Interpolation* function, avoiding discontinuities directly from an inter-frame factor α varying from 0 to $\pi/2$ and the filtered cage parameters $\widetilde{\mathbf{c}}$, as follows:

$$\widetilde{\mathbf{v}}_i^{\,t^+} = \sum_{j=1}^{m} w_{ij} \cdot \left(\cos(\alpha) \cdot \widetilde{\mathbf{c}}_j^{\,t-1} + (1 - \cos(\alpha)) \cdot \widetilde{\mathbf{c}}_j^{\,t} \right)$$

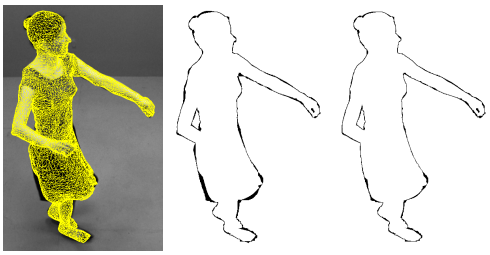

Figure 9: ***Silhouette-Aware Cage-based Encoding:*** *Backprojected cage-based mesh congruence with the corresponding captured image (**left-hand side**), silhouette overlap error of the cage-based mesh with the extracted real silhouette (**middle**), silhouette overlap error of the original model with the extracted real silhouette (**right-hand side**).*

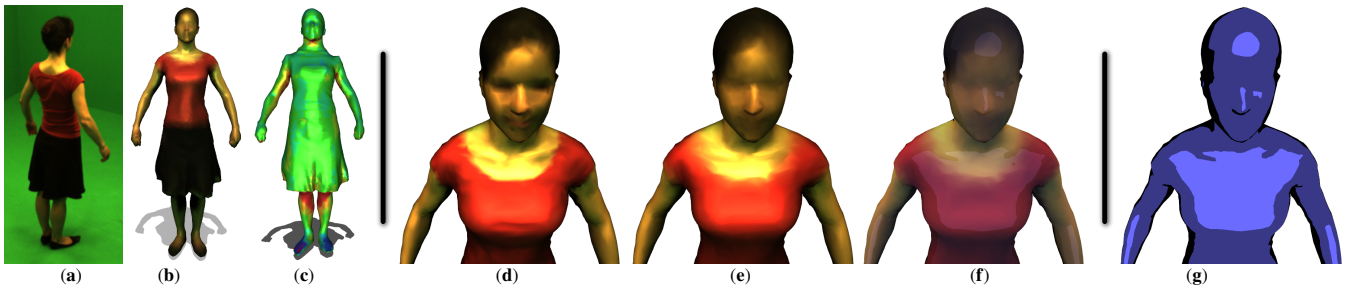

Figure 10: ***Appearance Editing:*** *Captured images (**a**) are projected to the model (**b**), bypassing partial hidden cameras visibility (**c**) in various steps: multi-view projective texturing (**d**), per-vertex filtering (**e**) and per-pixel cartoon style enhancement (**f**). The traditional toon shading is also displayed (**g**) for comparison with our final result (**f**).*

4.2 Depicting Video-Infused Appearance

The second step consists in establishing video-driven expressive appearance with respect to the sequence of captured images, as shown in Figure 10. We conduct an inferred-to-depicted approach to obtain a richness and life-likeness expressive appearance. We assume that the non-rigid surface is observed by a sparse network of k calibrated pinhole cameras during q frames. On account of the tincture and the skin pigmentation remain space-time unchanged, we reconstruct a seamless multi-view color distribution over the surface by averaging overall frames of the image pixel colors corresponding to vertices reprojection into calibrated images. This video-infused color component $\rho(i)$ associated to the i^{th} vertex is obtained by a per-vertex multi-view projective function described as follows:

$$\rho(i) = \frac{1}{q}\sum_{t=1}^{q}\frac{1}{k}\sum_{l=1}^{k}\omega_{li}^{t}\cdot\mathcal{I}_{l}^{t}(\Pi_{l}(\mathbf{v}_{i}^{t}))$$

subject to the normalization $\sum_{l}\omega_{li}^{t}=1$ with the weighted blending function:

$$\omega_{li}^{t}=\chi_{l}(\mathbf{v}_{i}^{t})\cdot\frac{1}{D_{l}\left(\mathbf{v}_{i}^{t}\right)^{2}}\cdot(\vec{\boldsymbol{n}}_{i}\cdot\vec{\boldsymbol{e}}_{l})$$

where for each projected texture $(\vec{\boldsymbol{n}}_i\cdot\vec{\boldsymbol{e}}_l)$ is the angle between the outer-pointing surface normal at the i^{th} vertex denoted by $\vec{\boldsymbol{n}}_i$ and the viewing vector pointing toward the direction of the considered l^{th} camera denoted by $\vec{\boldsymbol{e}}_l$. We incorporate a view-dependent rescaling factor to penalize the photometric contribution of distant vertices from cameras by injecting a normalized depth operator $D_l(\cdot)$ applied for a given vertex in respect to the l^{th} camera. In addition, the projection operator $\Pi_l(\cdot)$ associated to the projection matrix of the l^{th} camera provides the reprojected image coordinates for a given vertex. The local visibility function $\chi_l(\cdot)$ is set to 1 if the vertex visible or 0 otherwise. The vertex visibility is decided by relying on rendered depth maps and surface orientation. Finally, the color operator $\mathcal{I}_l^t(\cdot)$ returns the color component in the matted image captured by the l^{th} camera at time t for the given image coordinates. The resulting weight determines the importance of the input camera for ensuring photo-consistent blending.

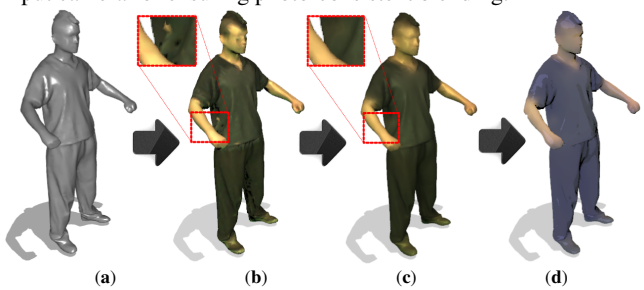

Figure 11: ***Appearance Generation*** *is performed in various steps: (**a**) captured surface reprojection, (**b**) multi-view texture mapping, (**c**) per-vertex filtering and (**d**) per-pixel cartoon style enhancement.*

Merging the multi-view footage into an average overtime color distribution encodes the texture more globally by preventing partially occluded region. Unfortunately, the resulting appearance contains ghosting or blurring effect due to reprojection errors, lighting conditions and non-lambertian surface reflectances. We exploit the lower sensitivity of the human visual system to generalize over these kinds of *painted shoebox* inconsistency. Thus, to eliminate noisy texels and reduce non-characteristic hyper-real saliency, we employ a *Diffusion-Contraction* process over textural elements that also ensures global seamless intensity. To accomplish this, we apply a per-vertex smoothing and sharpening filter iteratively on the color components where $\mu\in[0,1]$ is a dumping factor and $N(i)$ is the set of direct neighbors for the i^{th} vertex, as described as follows:

$$\forall i;\quad \rho(i)\longleftarrow(1-\mu)\cdot\rho(i)+\mu\cdot\tfrac{1}{|N(i)|}\sum_{j\in N(i)}(\rho(j)-\rho(i))$$

$$\forall i;\quad \rho(i)\longleftarrow(1-\mu)\cdot\rho(i)+\mu\cdot\tfrac{1}{|N(i)|}\sum_{j\in N(i)}(\rho(i)-\rho(j))$$

As illustrated in Figure 11, the final shading equation is described as a controllable mixture of the filtered photo-aware cues with the traditional toon varnish. The toon shading consists in 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 [Barla et al. 2006]. The continuous video-infused color component σ is obtained by interpolating the discrete distribution ρ estimated over the surface. Finally, the resulting GPU shading equation for pixel-wise intensity distribution τ is generated by a controllable blending of life-like and cartoon color components as follows:

$$\tau=\int_{\mathbb{S}^t}((1-\beta)\cdot toonmap\,[\mathbf{n}\cdot\boldsymbol{\ell}]+(\beta\cdot\sigma))$$

where $\mathbf{n}$ is the outer surface normal distribution, $\boldsymbol{\ell}$ is the ray from the single light source and $\beta\in[0,1]$ is a control interpolating factor balancing the impact of both components. The dot product between the surface normal and the light direction is used as an index to select the desired threshold shading in the input 1D toon map denoted *toonmap*. The colormap can be chosen to govern the color tone and amplify more cartoony effect. In our experiments, we use a blue-to-black stepwise map. For the sake of simplicity, directional lighting is used with a typical rendering pass invoked in real-time. Working with laser-scanned meshes allows us to estimate directly a color per-vertex that is a sufficient support for containing texture information. So the construction of a parametrized texture atlas is bypassed. Since everything is in the eyes of the beholders, our photo-realism undercoat is paramount in the world of make-believe and adds fuel to the fire of efforts to bring characters to life. The cartoon varnish helps photo-real characters to adventuring in a stylized world. We observe that the over-real appearance makes the perception of shape tridimensional-less. Thus, we emphasis 3D perception of image-based textural cues via toon shading.

5 Results and Discussion

To demonstrate the robustness and usefulness of our techniques, we have implemented a straightforward prototype software. In this section, we briefly describe our encouraging results, and we discussed related observations and limitations.

Experimental Results. We tested our real-time stretchable skeletal editing technique successfully on a corpus composed of more than forty motion clips, captured by a Vicon® system or motion capture suits. Editing seems to be more pleasant for sporty motions, with the root joint locked, rather than walk-cycling. Small amount of user interaction is required to obtain promising results yet, with stability and coherence. In addition, we developed a simple user interaction technique to handle the closest joint to edit, coupled with depth first search. Moreover, to validate the proof of concept of our video-based toon generation from surface performance capture, we reuse the outputs of [Vlasic et al. 2008] composed of mesh animations and image sequences captured by eight cameras regularly spaced all around a chroma-key rectangle room. We demonstrate the efficiency of our methods on various challenging datasets, particularly on a samba dance sequence, where highly non-rigid natural-looking wrinkles on the skirt and large deformation are emphasized by stylized subspace while ensuring the near-comformality of original small-scale details. The cage-based surface filtering process is performed at interactive rate assuming that *Harmonic Coordinates* are computed as pre-process. Furthermore, as seen in Figure 12, the video-infused appearance produces visually aesthetic effects in real-time.

Figure 12: ***Video-Infused Toon Appearance.*** *Our body and clothing appearance model is described as the fusion of video-based photometric information with controllable quantization and outlining underlining Tilke cloth (**left-hand side**) and the mohawk hairstyle of Daniel (**right-hand side**).*

In other to assess the final results, the accompanying video illustrates the capabilities of the techniques to produce mixed animation-oriented editing such as real-time elongated adjustment of the bone structure in motion, automatic handle-aware filtered captured surface and inferred-to-depicted life-like appearance. Figure 1 illustrates the overall behavior of our techniques. Results show that the proposed techniques are promising enough to be used probably in the low-budget productions for games and movies.

Limitations. Even if our skeletal-based technique is capable of producing coherent squash-and-stretch effets over the bone structure, the *Skeletal Laplacian* does not preserve the natural kinematic curvature under manipulations. Putting scales into the bone matrices can induce unwanted shears. The inherent limitation of cage-based deformation is the well-known tedious process to arrange cage-handles in a shape-aware manner and the time-and-memory consuming process for rigging computation. Even though, the purely geometric tessellation of cage allows low-rank data reduction and re-skinning of highly deformable animations, this detached parametrization may fail for topological changes. Our cage-based signal perturbation will benefit of weight reduction to avoid shape sensitivity. We note that it is difficult to go further away from as-photorealistic-as possible surface exaggeration with the current cage-based cartoon filter for dynamic surface having extremely large deformation. The resulting material shading can obscure important features and can be blurry because of the texture filter employed for removing the noise. As a first step, currents limitations of our *Cartoon-Reality* methods open promising opportunities in the field of non-photorealistic animation.

Benefits. Skeleton and cage structures are skin-detached low-dimensional subspace, well-suited to drive real-time high-resolution animation. Firstly, we introduced a skeleton in motion editing technique that allows animators to manipulate arbitrary animation with real-time control. Breaking physics of rigid motion allows us to obtain cartoon-like effects, independently of the skin layer. Differential-aware scaling, represented by an energy functional leading to a *Poisson Skeletal Solver*, preserves the spatial coherence of joints. Secondly, we also 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 effects. 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. We decoupled the performance capture stylization on underlying non-rigid surface motion parameters as well as the reconstructed appearance. We take advantage of the stylized shading to hide issues inherent in video-based surface texturing. Finally, injecting non-rigid cartoon rent in video-based surface texturing. Finally, injecting non-rigid cartoon exaggeration in video-based life-like surface brings a heightened believability of non over exaggerated animated toon characters.

6 Conclusions and Future Work

Since capturing-to-reusing animation is a challenging technical task, *Animation-Cartoonization* must remain ultimately a unique piece of art, preserving the life-likeness of captured data from performers, whatever the animator's manipulations over animatable structures. We propose conceptually well-formed editing techniques reducing tremendous amount of skilled artistry to create compelling exaggerated editing. In this paper, we study endo-and-exo underlying abstraction behavior tailored for ensuring the baked-in likelifeness of shape, motion and appearance under structural perturbation. Demonstrated by two complementary techniques, we push captured animation forward its natural physicality thanks to content-aware editing from reality to cartoon state. To the best of our knowledge, we propose the first controllable variational techniques for breaking rigidities of skeletal captured motion clips, the first filtering technique for cage-based captured dynamic surface and free-viewpoint inferred-to-depicted appearance.

The proposed method can speed up animation workflows in current post-productions of sensor-based animations. We exploit the benefit of small controllable subspaces to provide interactive and accessible editing process, to non-expert animators. In term of human visual perception, seeing people in different motion and appearance could be an enthralling scenario for understanding the *uncanny valley* in perceiving subtle changes in between a barely human and fully human animation. Nevertheless, we are carefully conscientious that our technique can be improved in order to the generate more elaborated animation puppetry from captured live-action. For this reason, we plan to explore more opportunities to run user-study, to test different type of motion and appearance filters, tightly coupled with subspace-based surface skinning. Finally, we have developed fast and accurate techniques in order to ensure coherence and control of the heterogeneous captured data and life-like appearance under exaggerated editing process.

References

BAAK, A., MÜLLER, M., BHARAJ, G., SEIDEL, H.-P., AND THEOBALT, C. 2011. A data-driven approach for real-time full body pose reconstruction from a depth camera. In *IEEE 13th International Conference on Computer Vision (CVPR)*.

BALLAN, L., AND CORTELAZZO, G. M. 2008. Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. In *3DPVT*.

BARAN, I., AND POPOVIĆ, J. 2007. Automatic rigging and animation of 3d characters. In *ACM SIGGRAPH 2007 papers*, ACM, New York, NY, USA, SIGGRAPH '07.

BARLA, P., THOLLOT, J., AND MARKOSIAN, L. 2006. X-toon: an extended toon shader. In *Proceedings of NPAR'06*, ACM, New York, NY, USA, 127–132.

BEN-CHEN, M., WEBER, O., AND GOTSMAN, C. 2009. Spatial deformation transfer. In *Proceedings of the SCA'09*, ACM, New York, NY, USA, 67–74.

BEN-CHEN, M., WEBER, O., AND GOTSMAN, C. 2009. Variational harmonic maps for space deformation. *ACM Trans. Graph. 28* (July), 34:1–34:11.

BOUVIER-ZAPPA, S., OSTROMOUKHOV, V., AND POULIN, P. 2007. Motion cues for illustration of skeletal motion capture data. In *Proceedings of NPAR'07*, ACM, New York, NY, USA, 133–140.

BREGLER, C., LOEB, L., CHUANG, E., AND DESHPANDE, H. 2002. Turning to the masters: motion capturing cartoons. In *Proceedings of SIGGRAPH '02*.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from sparse multi-view video. In *ACM SIGGRAPH 2008 papers*, ACM, New York, NY, USA, SIGGRAPH '08, 98:1–98:10.

DE JUAN, C., AND BODENHEIMER, B. 2004. Cartoon textures. In *Proceedings of SCA'04*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 267–276.

EISEMANN, M., DE DECKER, B., MAGNOR, M., BEKAERT, P., DE AGUIAR, E., AHMED, N., THEOBALT, C., AND SELLENT, A. 2008. Floating textures. *Computer Graphics Forum (Proc. of Eurographics)*.

GALL, J., STOLL, C., AGUIAR, E. D., THEOBALT, C., ROSENHAHN, B., AND PETER SEIDEL, H. 2009. Motion capture using joint skeleton tracking and surface estimation. In *In IEEE Conf. on Computer Vision and Pattern Recognition*.

GROCHOW, K., MARTIN, S. L., HERTZMANN, A., AND POPOVIĆ, Z. 2004. Style-based inverse kinematics. In *ACM SIGGRAPH 2004 Papers*, ACM, New York, NY, USA, SIGGRAPH '04, 522–531.

HUANG, P., BUDD, C., AND HILTON, A. 2011. Global temporal registration of multiple non-rigid surface sequences. In *CVPR 2011, Colorado Springs, CO, USA, 20-25*, 3473–3480.

JACOBSON, A., AND SORKINE, O. 2011. Stretchable and twistable bones for skeletal shape deformation. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH ASIA) 30*, 6.

JACOBSON, A., BARAN, I., POPOVIĆ, J., AND SORKINE, O. 2011. Bounded biharmonic weights for real-time deformation. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*.

JAMES, D. L., AND TWIGG, C. D. 2005. Skinning mesh animations. In *ACM SIGGRAPH 2005 Papers*, ACM, New York, NY, USA, SIGGRAPH '05, 399–407.

JOSHI, P., MEYER, M., DEROSE, T., GREEN, B., AND SANOCKI, T. 2007. Harmonic coordinates for character articulation. *ACM Trans. Graph. 26* (July).

JU, T., ZHOU, Q.-Y., VAN DE PANNE, M., COHEN-OR, D., AND NEUMANN, U. 2008. Reusable skinning templates using cage-based deformations. *ACM Trans. Graph. 27* (December), 122:1–122:10.

KANG, D., CHUNG, J.-M., SEO, S.-H., CHOI, J.-S., AND YOON, K.-H. 2009. Detail-adaptive toon shading using saliency. In *Proceedings of VIZ'09*, VIZ '09, 16–20.

KAVAN, L., COLLINS, S., ŽÁRA, J., AND O'SULLIVAN, C. 2008. Geometric skinning with approximate dual quaternion blending. *ACM Trans. Graph. 27* (November), 105:1–105:23.

KAVAN, L., SLOAN, P.-P., AND O'SULLIVAN, C. 2010. Fast and efficient skinning of animated meshes. *Computer Graphics Forum 29*, 2.

KIRK, A. G., O'BRIEN, J. F., AND FORSYTH, D. A. 2005. Skeletal parameter estimation from optical motion capture data. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

KWON, J.-Y., AND LEE, I.-K. 2007. Rubber-like exaggeration for character animation. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, 18–26.

KWON, J., AND LEE, I.-K. 2011. The squash-and-stretch stylization for character motions. *IEEE Transactions on Visualization and Computer Graphics*.

LARBOULETTE, C., PAULE CANI, M., AND ARNALDI, B. 2005. Dynamic skinning: adding real-time dynamic effects to an existing character animation. In *Spring Conference on Computer Graphics*.

LIPMAN, Y., LEVIN, D., AND COHEN-OR, D. 2008. Green coordinates. *ACM Trans. Graph. 27* (August), 78:1–78:10.

LIU, Y., STOLL, C., GALL, J., SEIDEL, H.-P., AND THEOBALT, C. 2011. Markerless motion capture of interacting characters using multi-view image segmentation. In *CVPR'11*.

MAGNENAT-THALMANN, N., LAPERRIÈRE, R., AND THALMANN, D. 1988. Joint-dependent local deformations for hand animation and object grasping. In *Proceedings on Graphics interface'88*, 26–33.

MILLER, C., ARIKAN, O., AND FUSSELL, D. 2010. Frankenrigs: building character rigs from multiple sources. In *Proceedings of I3D'10*, 31–38.

RUSINKIEWICZ, S., BURNS, M., AND DECARLO, D. 2006. Exaggerated shading for depicting shape and detail. *ACM Transactions on Graphics (Proc. SIGGRAPH)*.

SHIRATORI, T., PARK, H. S., SIGAL, L., SHEIKH, Y., AND HODGINS, J. K. 2011. Motion capture from body-mounted cameras. *ACM Transactions on Graphics 30*, 4.

STARCK, J., AND HILTON, A. 2007. Surface capture for performance-based animation. *IEEE Comput. Graph. Appl. 27* (May), 21–31.

STOLL, C., GALL, J., DE AGUIAR, E., THRUN, S., AND THEOBALT, C. 2010. Video-based reconstruction of animatable human characters. *ACM Trans. Graph. 29* (December), 139:1–139:10.

STOLL, C., HASLER, N., GALL, J., SEIDEL, H.-P., AND THEOBALT, C. 2011. Fast articulated motion tracking using a sums of gaussians body model. In *ICCV'11*.

SÝKORA, D., BEN-CHEN, M., ČADÍK, M., WHITED, B., AND SIMMONS, M. 2011. Textoons: Practical texture mapping for hand-drawn cartoon animations. In *Proceedings of NPAR'11*.

TIERNY, J., VANDEBORRE, J.-P., AND DAOUDI, M. 2008. Fast and precise kinematic skeleton extraction of 3d dynamic meshes. In *ICPR*.

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.

VLASIC, D., PEERS, P., BARAN, I., DEBEVEC, P., POPOVIĆ, J., RUSINKIEWICZ, S., AND MATUSIK, W. 2009. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graph.*.

WANG, J., DRUCKER, S., AGRAWALA, M., AND COHEN, M. F. 2006. The cartoon animation filter. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006) 23*, 3 (July), 1169–1173.

WEBER, O., SORKINE, O., LIPMAN, Y., AND GOTSMAN, C. 2007. Context-aware skeletal shape deformation. *Computer Graphics Forum (Proceedings of Eurographics) 26*, 3.