Studies in Computational Intelligence 509

Yann Savoye

Cage-based Performance Capture

Springer

Studies in Computational Intelligence

Volume 509

Series Editor

J. Kacprzyk, Warsaw, Poland

For further volumes:
http://www.springer.com/series/7092

Yann Savoye

Cage-based Performance Capture

Springer

Dr. Yann Savoye
Lyon
France

ISSN 1860-949X ISSN 1860-9503 (electronic)
ISBN 978-3-319-01537-8 ISBN 978-3-319-01538-5 (eBook)
DOI 10.1007/978-3-319-01538-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013949253

© Springer International Publishing Switzerland 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness that most frightens us. [...] Your playing small does not serve the world. There is nothing enlightened about shrinking so that other people won't feel insecure around you. We are all meant to shine, [...]. It's not just in some of us; it's in everyone. And as we let our own light shine, we unconsciously give other people permission to do the same. As we are liberated from our own fear, our presence automatically liberates others.

Marianne Williamson
A Return to Love: Reflections on the Principles of a Course in Miracles

Acknowledgments

The armadillo and horse models are courtesy of Stanford University.
The crane, handstand, and samba mesh sequences datasets are courtesy of MIT Adobe. Cited figures are copyrighted by corresponding authors and publishers (please see associated references as copyright notice).

Contents

Chapter 1
General Introduction

As an introduction to this book, we provide a general overview concerning the reusable digitalization of captured animation: from capturing the real-world to editing of video-based animation. In particular, we present and discuss a concise statement concerning the problem of non-rigid parametrization for video-based captured animation. To enforce the importance of this problem, we demonstrate the impact of this research area by the various domains of application. In addition, we clearly exhibit the motivations, objectives, and difficulties of the tackled class of problems. Hence, we yet uncover a research roadmap of our approach and original contributions to alleviate this problem. Finally, we draw the chapters' organization for this book.

1.1 Research Context

Forty Years. Over forty years have been required in the field of *Computer Graphics* and *Computer Vision* to be mature enough from the first use of computer-generated three dimensional graphical model (as seen in the movie *FutureWorld* in 1976), to the first accomplished photorealistic 3D movie (as seen in the movie *Avatar* in 2009). There is no doubt that *Computer-based Imagery Technologies* have significantly influenced and changed modern society, and have a significant impact on our everyday life over the last two decades. In the past, technology has made enormous leaps in term of power and availability. More precisely, tremendous technological progress led to affordable CPU computation power in recent years. Memory and graphic cards have boosted the generation of synthesis images. In the era of computers, the advent of technologies has enabled new possibilities in rendering, animation, modeling and real-time whole-body interaction (see contrast between Figs. 1.1 and 1.2).

Strong Duality. On the one hand, *Computer Graphics* techniques deal with the design of computer models or representation in order to produce high-quality images. On the other hand, *Computer Vision* tools are able to extract information from

Y. Savoye, *Cage-based Performance Capture*,
Studies in Computational Intelligence 509, DOI: 10.1007/978-3-319-01538-5_1,
© Springer International Publishing Switzerland 2014

Fig. 1.1 Futureworld: the first movie with a CG animation (Courtesy of MGM®)

Fig. 1.2 Avatar: performance capture movie (Courtesy of Weta Digital®)

real-world data captured by multimodal sensors, and to deduce intrinsic model. Both of them are naturally complementary and tend to reach convergence rapidly. In addition to the above, *Vision* is dual to *Graphics* and intersecting both fields can lead to more efficient, accurate, fast, automatic, robust and reliable tools with direct industrial applications. Clearly, the interdisciplinary between *Vision* and *Graphics* domains has demonstrated powerful benefits for production of massive 3D content.

Late-Breaking Convergence. Notwithstanding, *Computer Vision* and *Computer Graphics* are two different communities. Even if in-between convergence can be recently noticed, in practice both domains are slightly separated. This contrast illustrates how much video-based animation remains remarkably active and attractive for highly-competitive research. For instance, the film industry is potentially a showcase application area for video-based algorithms and progressively integrates late-breaking academic research results. Fundamentally, the ultimate goal of this convergence is to create the illusion of animation in its absence. Ultimately, the final validation will be achieve if there is categorically no manner for any viewer to distinguish if the scene is an optical illusion or if it is captured with a real camera. Overall, it is worth to mention that human perception is extremely accurate to detect artifacts in human-like motion.

1.1.1 From Real-World to Vision-based Animation

Images of the World in Motion. Reconstructing, synthesizing, analyzing and re-using dynamic shape or captured images of the world in motion is a recent outstanding challenge task. Furthermore, 3D video and vision-based animation have reached considerable attentions, and are now increasingly used for modeling and interacting with the physical world. While defining plausible deformation remains an extremely complex and time-consuming task, video-based capture tends to bring fine details of the reality representable, with respect to the natural properties of shapes in motion. After presenting the general context, we will precisely scrutinize key elements for obtaining suitable video-based animation.

Optical Illusion. Vision-based animation is a class of methods dealing with *Computer Animation* in a vision-oriented strategy, at the cross-fertilization of related key fields such as *Computer Vision*, *Computer Graphics* and *Numerical Optimization* (see Fig. 1.3). Let's offer concise definitions of those well-known fields:

- The *Computer Vision* field is dedicated to the acquisition of signal from real-world, the extraction of information from captured dynamic scenes (mainly 2D images, depth maps and videos) and to the analysis of the captured information in order to reconstruct unknown representations from various sensors. Nevertheless, animating people correctly requires a high-level of understanding about the human body moves and how to characterize human actions. For instance, four dimensional object reconstruction from a sequence of images is a classical problem in *Computer Vision* to infer such understanding.
- The *Computer Graphics* field is specialized in the synthesis of 3D representation entities (surface or underlying parametrized structures) and their processing. For instance, the class of problem related to shape deformation is a crucial aspect of deformable models. Then, the concern of structuring the input geometry in a meaningful fashion is of first importance in order to preserve desired properties, typically at the editing phase.
- The *Numerical Optimization* field is focused on developing effective mathematical tools in order to solve a numerical problem. The central aspect here is to find an optimal solution with respect to an objective function that might not have an exact

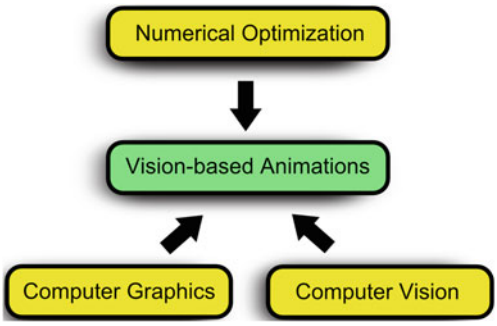

Fig. 1.3 Vision-based animation is at the crossroad of three technical pillars: computer vision, computer graphics and numerical optimization

solution in the strict sense. For instance, ill-posed or well-posed inverse problems in *Computer Graphics* and *Computer Vision* always require abundant use of linear or non-linear numerical optimization to recover unknown information.

1.1.2 High-Level Research Directions

More importantly, we are particularly interested in the complexity of representation for performance-based computer animation. More precisely, we intend to tackle the three following subproblems: *dimension reduction* from capture to re-use dense surface capture data in the perspective of whole body interaction, *animation reuse* of heterogeneous captured data toward *animation cartoonization*, and *dynamic reconstruction* of non-rigid performance motion from sensor-based data. Nevertheless, we would like to highlight some following pre-required notions that are transversal in these problems: *computer animation*, *performance-based animation*, *complexity of representation*, *capturing to reusing 3D videos*, and *video-based whole body interaction*.

Computer Animation. Animating an object has a direct link with everyday life and actions around us. Animation is a fundamental field that has grown immensely over the years to become a mainstream art. *Computer Animation* can be defined as the process employed for generating animated images by using *Computer Graphics* technologies. As far as *Three Dimensional Computer Animation* is focused on bringing believable behavior to a given digital representation, its principal interest is the analysis and synthesis of motion. Hence, the relation between *Computer Animation* with acquisition, capture, modeling, processing and reconstruction is then clearly established around the problematic of the motion. Moreover, since traditional animation is time-consuming, computer generated animation improves the creation of high quality film-size animation. Meanwhile, *Computer Animation* can be seen as the temporal evolution of a discrete structure representing the body surface in the case of character animation. The principle of animation is to enhance the realism by infusing life into characters while depicting their movements. Unfortunately, the task of creating animation remains laborious because it involves a lot of skills, and often many artistic iterations. It is worth to mention that substantial part of the creation of digital reconstructions, for large effects movies, may involve hundreds of person-years of manual artistic effort. For this reason, we can notice a growing demand for automatic content acquisition methods.

Performance-based Animation. Performance-driven approaches are motivated by the observation that high quality for life-like surfaces in motion cannot be reached only by an army of artists. Then, the main purpose of video-driven animation is the ability to produce photo-realistic animation of a digital replica of a moving object. Thus, generating highly realistic animated content requires non-invasive capturing technique to reproduce natural actor's performance in casual wardrobe and natural behavior. By bringing character to life, an animation engine designs smooth and

believable animations. Such engines must enable realistic animation with appealing effects in order to reproduce the visual realism of live performance. *Performance Capture* describes all emerging technologies used to measure motion of models that are nearly impossible for artist to recreate them with the same accuracy. Our world is full of complex objects in motion with complex physical properties allowing them to be deformed. Consequently, it makes sense to use a large choice of sensors for capturing real-world animation events and then to virtualize such effects. Reproducing effects from the real-world needs to solve difficult inverse problems.

As shown by Johansson in [1], the study of biological motion perception reveals that an animated creature can be exactly recreated using just a dozen of sphere craftily placed on the body surface. Nevertheless, one challenging direction is the life-like and non-rigid deformation of skins, that imitates natural movement of living creature comprising both humans in clothing and animals. The main objective is to overcome the long-established barrier known as the *uncanny valley*, referring to the perception that animation looks less realistic as it approaches human likeness. Achieving a photo-real digital actor and recreating decent life-like looking animation can enhance new levels of believability in films and computer games.

Complexity of Representation. Complex animation traditionally employs a multi-layered strategy to achieve multi-scale deformations. Such complex deformations reflect the combined influences of bones, skin, muscle and wrinkles in casual loose-fitting clothes (see Fig. 1.4). Contrary to motion capture, surface performance capture allows a realistic reconstruction of the shape, the motion, and the appearance of an actor performance rather than traditional pose parameters only. Even if geometry details are difficult to generate, they have a significant impact on the visual quality in order to mimic the natural complexity of the observed dynamic scene. For this reason, the explosion in the use of detailed and dense polygon meshes with hundreds of thousand to millions of polygons can be easily noticed in recent years. The inherent complexity of highly detailed shapes has a direct influence on the explosion in the amount of data to proceed. The richness of potential deformation is also a bottleneck for efficient storing and compression of 3D video data. Meanwhile, it is necessary

Fig. 1.4 Real world deformable and animated shapes: (from *left* to *right hand side*) a traditional Japanese indoor performance of Kyoto style dance with highly textured large apparel presenting clothing wrinkles, a tennis player with tight cloths and muscle bulging interacting in an outdoor sport scene with fast and large articulated motion, an Indian cobra snakes with high flexibility bending motion, a Bengal tiger's elongated body pose. (Images courtesy of Google®)

to preserve surface details in mesh animations because small-scale details have considerable impact on the visual quality. The need of intuitive techniques for massive and complex geometry motivates my work in the direction of animation reuse and controllable tools for 3D videos.

Dynamic Mesh Processing. Nowadays, the affordability of capturing device systems makes the problem of processing captured animated geometry attractive. A first consideration in *dynamic mesh processing* must be the setup of flexible geometric structures to encode dynamic mesh sequences, enabling a suitable representation for the evolving geometry. Such temporal data do not necessarily have one-to-one temporal association in term of vertex associations. In other words, the series of triangular meshes may not have an equal connectivity. In fact, capturing systems have a drawback in order to inject natural and inherent temporal correspondence between frame-to-frame without relying on incremental evolution of a given template or a given animatable underlying structure. Dealing with captured data turns the problem of dynamic mesh processing not just into a generalization of the static case, but introduces new problems such as the spatial and temporal coherence, the non-isometry of large deformation and topological changes. Another fundamental issue in mesh animation processing is to ensure shape preserving mesh deformation, while generating an animated mesh with fixed connectivity.

Capturing to Reusing 3D Videos. Affordable multiple-view acquisition platform and emerging low-cost sensors offer a solution for a quick and massive generation of 3D dynamic dense surface sequences that could fill 3D digital libraries for visual media productions. An increasing number of efficient techniques generate high quality real-world animation content with the aim to reuse the output models. Nevertheless, a main drawback is a considerable disparity between the towering 3D content. The growing convergence of computer games and films is a relatively new phenomenon that imposes to rethink current production cycles for massive 3D content. There will be no technological difference between what is played and what is watched. With massive amounts of 3D content, we notice the high demand of non-rigid structures able to improve the production process from the acquisition of 3D videos to their reuse in the context of film-game convergence. There is a significant need of a transversal and unique data representation from the acquisition to the reuse of 3D captured data that can be shared between animated feature films and console video games. Since modern game players expect high-tech cinematography in their games, such transversal data representation could help the high-quality cinematography to incorporate filmmaking techniques into their games. This film-game unified representation may radically improve the processing of massive amounts of 3D content. In addition, a compact set of reusable parameters offers the opportunity to deliver abundant 3D video content through the cloud.

Video-based Whole Body Interaction. In recent years, we notice an increasing interest for interaction with people in videos, from single actor performance to close multi-person interaction. *Mesh-based interaction* offered by performance capture can improve human-computer interaction. Thus, the human body becomes an interactive

control device and allows the possibility of human intervention during the animation playback. Moreover, the whole-body interaction is founded on the capture and processing of human signals from physical, physiological or cognitive aspect to generate feedback in a digital environment. The whole-body interaction explores richer ways of realistic interaction between humans in a virtual environment by a natural interaction with direct capture and analysis of the body language. The main idea is to use the human body as an interactive and non-invasive controller as an ubiquitous interface between the user and the software. Finally, human-computer interaction can leads to analysis of the motion and gesture recognition via a three dimensional reconstruction process.

1.1.3 Domains of Application

Once the general context of our work has been introduced, we will present potential domain of applications of targeted techniques in this section. In addition, some examples are evoked to illustrate the importance of resulting techniques in our everyday life, social human activities and directly used in production. By analyzing current real-world applications, we achieve a suitable point of view for identifying scientific locks and finally solving real problems. A large number of video-based graphics applications can be observed, from low-cost home entertainment room with real-time interaction to high-budget studio environment. Computer games and 3D animated films make heavy use of deforming geometric mesh sequences, and it is increasing day by day. In fact, the need of extracting manageable parametrization from real-world captured animation or directly the captured video stream is justified by a huge demand in obtaining large amount of high quality data with limit storage size and reusable efficiently. In order to demonstrate the importance of realistic animations, we propose an overview of various categories of applications as follows.

- **Cognitive Research:** Motion and shape analysis of human activity is at the center of interest since the very foundation of *Computer Vision*, robotics and visual perception of motion. Shape Motion Analysis can be described as a class of methods usually employed to record deformable model in a scene. First of all, this class of techniques propose correct answers to the basic but not simpler question: *what people are doing inside a video of them doing it?*
- **Edutainment:** Generating an animation has a significant impact to offer a new manner to the understanding and conceptualization toward an explanation of the world we live in. For instance, computer generated animation offers the possibility to generate contains for 3D book pictures, or documenting particular performance or other type of multimedia presentation.
- **Massive Entertainment Imaging:** The release of low-cost sensors such as webcams or Kinect® has changed the face of whole-body interaction in personal and social human activities. This recent technological advance allows the user to use

this technology in a non-controlled environment. Moreover, industries of motion films have a high demand of performance capture technologies to produce remarkable effects for feature films, television and advertising as well as for filmmaking and previsualization in preproduction. In addition, the fusion of performance capture and interactive technology is more and more present in high-end applications such as 3D cinematographic games, as seen in L.A. noire®.

- **Culture Heritage:** Encoding highly-detailed animation directly from cultural events could help to preserve and share the richness of world's cultural heritage, in a digital form and for future generations. By massively recording and encoding dancing master, traditional Japanese dance performances, dance styles, martial arts like samurais fighting battle with sword; researchers and engineers build invaluable 3D archive for visualization and animation of cultural heritage. More recently, capturing garment has an effective application in the textile industry.
- **Medical Applications:** Motion capture technologies are also used to analyze patient gait or pathological locomotion in the process of clinical rehabilitation. The analysis of dynamic shape is of paramount importance from biomechanics to 3D surgery monitoring of human and animal, resulting in the development of prosthesis.
- **Serious Game and Military Tele Surveillance:** Registration-to-animation from video stream arises in a number of applications such as the identification of a person in crisis management situation. Automatic activity recognition as seen in serious games or security surveillance infrastructure is an example worth mentioning. Not to mention that full-body registration and character animation have proven to benefit to the Defense, especially by enhancing the commander ability to fix a nearby enemy in potentially hostile military threats.
- **3D Interactive TV and Broadcasting:** 3DTV is a new emerging media to record an observed scene as high fidelity full 3D shapes. This novel technology allows production of free-viewpoint videos. In particular, the viewer has the possibility to select the desired viewpoint and viewing direction interactively on demand to observe the virtual environment thanks to a fly-thought camera. Such functionality is particularly attractive for navigating inside a sporting event. Existing application are commercialized already as of *LiberoVision*® or *Sport EyeVision* of Carnegie Mellon used for the SuperBowl. The strong emergence of 3D Television displays impulses a large demand in the massive generation of high quality content. The goal of such projects was to develop a free viewpoint system that allows the capture and interactive broadcasting of live events using multiple cameras.
- **Sport Science and Sport Motion Analysis:** Motion analysis and surface parametrization have a number of applications. Sport analysis performance observes how people behave when they are playing sports, and analyzes athletic performances to provide advices in learning golf. Representing athlete movements as a temporal series of three-dimensional shapes can permit a better understanding to improve sports training and to avoid injuries.

- **Immersive Tele-collaboration and Telepresence:** The need of efficient and unified representations from the mobile capture room to the stereoscopic display of single distant person can be seen in tele-immersion and collaboration application (for instance the Blue-c project [2] or the new BeingThere Project) where the coding and transmission phases are a strong bottleneck to enable natural interaction. Another example is the first TV hologram realized during the 2008 American Election, or the Tupac hologram at Coachella.

1.2 The Problem: Non-rigid Parametrization for Video-based Animation

Animation Parametrization. The dynamic capture and reuse of temporal sequences from actors' performance and is a valuable and hard problem for animation productions . In this section, we propose to introduce the new significant problem of *Video-Based Non-Rigid Animation Parametrization* (see Fig. 1.5). We also exhibit the advantage of an *animation parametrization* in comparison to parametrization-free techniques. Extracting all-in-one (compact, structured, skin-detached, non-rigid, reusable) non-articulated parametrizations from multi-view data or mesh sequences was relatively few studied before. In this book, we claim that the global pose for non-rigid captured surfaces can be simultaneously stored using cage-based strategies. Specifically, we prefer to avoid the issue of rigid and hierarchical angles representation and model-free representation for captured surfaces to allow more degrees of freedom. Such accurate and effective parametrization can be useful from the capture, registration, to the editing of 3D videos.

Cage-based Performance Capture. In this book, we mainly address the problem of extracting and reusing four dimensional (3D+t) , non-rigid, skin-detached and reduced parameters from captured animation data. We also proposed a research direction to extend our core framework directly toward the captured video stream. As the problem of *Cage-based Performance Capture* was never solved before, the main questions are: *how to setup an effective solution suitable from capture to reuse?* and *how to do it efficiently as-simple-as possible?* (see Fig. 1.6)

Purely Mesh-based Optimization. In particular, we aim to recover high-quality, reusable, abstracted representation parameters of a performer or various actors from already generated data, or unprocessed captured data that take into account the dynamics of skin and thigh clothing. The injection of a suitable parametrization

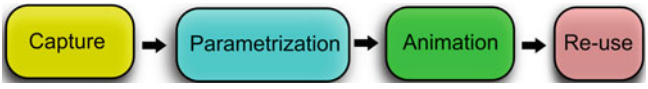

Fig. 1.5 Video-based non-rigid animation parametrization: captured data require a parametrization to allow efficient animation and further reuse

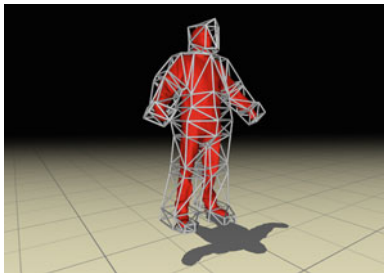

Fig. 1.6 Cage-based parametrization for 3D video: the dynamic geometry is depending on a set of space-time parameters, associated to the fixed connectivity of the bounding polytope, and attached to the enclosed model

into the pipeline is motivated by rendering high-resolution laser-scanned template animation at real-time rates. Consequently, we are interested in introducing an accessible but not simpler new linear least-squares optimization framework ables to encode body shape variations with controllable parameters, and without requiring an underlying rigid kinematics structure. Then, estimated parameters could be used directly by animator to re-edit animation, but also useful to compress large volume of temporal data, or to be transmitted over the network.

Keeping Things in Perspective. We keep in perspective the strong potential of our techniques to be apart of a template-based mesh registration process from image-based information and across time. First, we focus on preserving the quality of the reconstructed 3D video while encoding them with non-rigid and non-hierarchical underlying animatable structures. Multi-camera video recordings produce a huge amount of data for just few seconds of animation. Thus, defining a parametrization allowing high-quality playback for 3D video will be tremendously helpful to reduce unnecessary redundancies. In this work, we capitalize on the benefit of recent advances in cage-based systems to accurately reproduce realistic deformation. Nevertheless, injecting non-rigid and skin-detached global parameterization in a vision-oriented pipeline leads to additional and even harder problem to solve against traditional quasi-articulated parametrization.

1.3 Motivation, Objectives and Novelty

Motivation. In recent years, a lot of effort was invested in the acquisition of high-detailed dynamic surface from images, but relatively minor attention has been concentrated on establishing a suitable parametrization that is capable of preserving as-much-as possible the real-world captured properties. As definition, parametrization is the mathematical process that enables a complete set of effective coordinates or degrees of freedom necessary to characterize a geometric model. Beside, we observe a lack of global, flexible and reusable parametrization for editing mesh animation obtained by performance capture, and improving the registration process. We notice a lack of suitable, flexible animator-friendly parametrization for non-rigid surface without the need of additional skin corrections. More precisely, this work

is motivated by the emergence of two separate but well-explored domains: *cage-based deformation* using deformable mesh model and *video-based reconstruction* of dynamic 3D shape from multi-view images. Such motivation leads to investigate a novel deformable model where surface-based and space-based deformation are simultaneously used. A long-term vision is to use cage-based scheme for shape registration in the context of marker-less multiple-view acquisitions. To sum up, we are motivated to investigate promising research perspectives allowing animators to easily control animations that are captured from real-world.

Objectives. The overall objective is to introduce cage-based techniques for vision-based graphics context. The first objective of this book is to develop parametrization techniques in the context of video-based interaction, especially for multiple view setup. Fortunately, cage-based parametrization could simplify the workflow from capture to re-use of acquired data in the post-production process. Initially, we are interested in each step all along the production pipeline within a suitable parametrization can be useful. Hence, we are also interested in the use of a generic and common parametrization for the capture to the editing process. Moreover, we observe that creating computer-generated animations is time-consuming, and a lot of manual labor is quite often required, even if a commercial software suite is employed. The second objective of our work is to automatically extract a reduced number of controllable parameters. Making this process accessible and simpler for non-artist users is a challenging task. Another aim of our work is to identify bottlenecks of existing parametrization in order to propose solution having a better reproducibility. Finally, the ultimate objective is to provide a first solution to the problem of shape registration directly from silhouette images via the estimation of sparse controllable non-rigid parameters.

Novelty. In contrast to classical approaches, we prefer to avoid intrinsic or a kinematic parametrization for non-rigid surfaces. We therefore refer to a non-rigid type of lower dimensional subspaces to reproduce real-world and natural-looking motion. For the first time, we propose in this book to study and to introduce cage-based technique in the perspective of visual data acquired of by multiple cameras. This approach can uncover new possibilities in vision-based graphics. We also propose a *Graphics-to-Vision* methodology in order to extend well-adapted deformable techniques toward vision-oriented purposes.

1.4 Major Challenges

In this section, we exhibit the challenging side of the tackled problem and several difficulties in exploring a relatively few studied parametrization in order to facilitate all the workflow steps from acquisition to editing.

Difficulties in Capture and Re-use. Even if multiple view datasets of 3D video of people with clothing are increasingly available in a clean-up form, long-time

sequences are still difficult to store in a compact representation. Dealing with captured multi-view data is complicated because of their complexity and the amount of reconstructed output meshes to proceed. More specifically, we observe a key difficulty to encode or reuse multi-view captured meshes while preserving their natural-looking detailed deformation. Meanwhile, the human perception is exceedingly veracious, and eyes are highly sensitive to human motion error. The challenge is to estimate animation parameters able to reproduce visual pleasant mesh animation.

Technological Gaps. The gap between *Computer Vision* and *Computer Graphics* makes the convergence of new imaging technologies not trivial because of their different natures and specificities. We use advanced tools in five fundamental pillars: geometric processing, multi-view computer vision, mesh deformation, numerical optimization, and image processing. Mixing *Computer Graphics* structure with *Computer Vision* tools poses a number of problem. *Computer Vision* tools have a lack of accuracy and robustness by producing noisy, incomplete, inconsistent and sparse data that containing quite often false positive, reconstruction artifacts or outliers informations. In addition, generated information are quite often not robust under illumination change and uncontrollable environments. Meanwhile, *Computer Graphics* tools have complex structures even on clean synthetic data and have a lack of flexibility for reproducing and modeling real-world behavior realistically and temporally coherent in a simple manner. More importantly, the behavior of algorithms through the prism of black-box numerical optimization can be highly difficult to apprehend but crucial in the design of suitable solution for the tackled numerical problem. The interdisciplinarity emerging by fusing *Vision* and *Graphics* techniques turns the implementation into a time-consuming task.

Ill-posed Inverse Problems. Generally, this class of vision-oriented approaches needs to solve inverse and ill-posed problems where numerical optimization can be exceedingly difficult to set-up. Finding a suitable global parametrization is of utmost importance even if no perfect parametrization exists. Dynamic surface parametrization is a complex task because there are so many unknowns for which to solve. Another reason explaining this difficulty is that the estimation of desired parameters involves finding solutions that must not only fit some required mathematical criteria, but it must also naturally leads to pleasant looking to the viewer.

1.5 Our Original Contributions

After specifying the general context of the problem, we present the core contributions detailed further, and we emphasize their novelties and impacting aspect at the crossroad between *Computer Graphics* and *Computer Vision*. The research work presented in this book has been elaborated from scratch. We open a never-studied-before research direction called *Cage-based Performance Capture*. The piece of research we develop in this book mixed *modeling*, *animation*, *numerical optimization* and *vision*, and is mainly built around a core cage-based scheme.

Fig. 1.7 Our main contributions: in this book, we bring novel contributions in *inverse cage-based modeling*, *cage-based animation conversion* and multi-view data reuse, and *iterative cage-based registration* from multi-view silhouettes. Along this book, we establish a progressive graphics-to-vision strategy by relaxing prior knowledge from graphics tools toward vision-oriented applications while offering a comprehensive study

Contributions Workflow. To the best of our knowledge, since no previous work investigates cage structures as a potential non-rigid and animator-oriented parametrization in the context in video-based mesh registration, we propose a new deformable model methodology, well-adapted for 3D videos. We introduce a new class of approaches called *Cage-based deformation for video-based animation*. This book focuses on presenting novel collection of effective algorithms from *Graphics-to-Vision* relaxing prior knowledge on the data, where the ultimate goal is to reuse the performance capture data with better accessibility. As illustrated in Fig. 1.7 where the connexion between contributions are shown, the roadmap of this book can be summarized as follows:

- **Scalable Inverse Cage-based Modeling:** Firstly, inspired by traditional inverse kinematic techniques, we apply this paradigm on surfaces representation in an extrinsic fashion, thanks to cage inversion process. More precisely, we extend the *inverse kinematics* paradigm to cage-based modeling, in order to recover cage parameters from user-specified screen-space constraints (with the aim to be extended to image-based constraints). Hence, we develop a new *Inverse Cage-based Modeling* method for estimating cage parameters, relying on an inverse variational formulation. We show that the same variational principle can be formulated on skeletal structures.

- **Cage-based Animation Conversion and Multi-view Data Reuse:** Secondly, we extend the previous cage-based modeling approach to a data-driven retrieval process for converting quasi-articulated or fully non-rigid animated surfaces (assuming the consistent temporal connectivity is known) into a sequence of estimated optimal cage parameters via a process of *Cage-based Animation Conversion*. In addition, we also demonstrate an *Animation Cartoonization* application to multi-view data in term of cage-based surface exaggeration and video-based appearance stylization.

- **Iterative Cage-based Geometric Registration:** Thirdly, following previous work on global alignment of 3D video sequences, ICP-like Laplacian-driven mesh deformation and shape registration, we focus on leverage the problem of space-time shape registration by employing a cage-based registration technique. Thus, we propose a promising unsupervised approach to perform *Iterative Cage-based Registration* of reconstructed target point set from multi-view video recording. Then, this technique aims to introduce frame-to-frame cross-consistency into reconstructed

3D video data by non-rigidly aligning a laser-scanned template via frame-to-frame space deformation.

1.6 Thesis Organization

In this section, we present the organization of this book by detailing highlights and principle key ideas of all chapters' content, built around the cage-based strategy. In addition, we also emphasize the workflow of this book by showing the connection between all chapters from *Graphics* to *Vision*. At the heart of this book, we introduce the idea of *cage parametrization for video-based reconstructed surface* under motion and a *cage-based registration* for dynamic shape capture, but not limited to. Last but not least, we conclude this book by summarizing the benefit of our novel techniques and potential opportunity for future work. The contributions will be presented in 3 Chaps. 2, 3 and 4. The remainder of this book is organized as follows:

- **Chapter** 1 was dedicated to introduce a general overview concerning the reusablility of performance mesh animation and provides a summary of the general background of this book. First, we briefly described the vision-based animation research context. Then, we defined the problem of *video-based animation parametrization* and we describe a roadmap for addressing it in an improved manner.
- **Chapter** 2 isolates the problem of scalable geometry editing by the mean of inverse cage-based modeling in conjunction with a sparse collection of sparse screen-space constraints. We introduce the concept of *Cage-based Dual Laplacian* to achieve *Indirect Cage-Based Fitting*. This chapter presents our first contribution describing a simple linear optimization formulation to compute optimal cage parameters. We also review the rich literature concerning animatable surface parameterizations such as intrinsic, free-form and skeleton-based surface. More importantly, we will focus especially on exploring recent cage-based techniques and their meaningful properties comparing with previous deformation techniques.
- **Chapter** 3 covers the most impacting contribution by facing the problem of reparametrization non-rigid full-body captured mesh animation. In particular, we introduce a well-formed method to extract optimal low-dimensional cage-based parameters from non-rigid captured mesh sequences, already aligned in full-correspondence. A mathematical formulation of designed algorithms is proposed. A video-based evaluation of experimental results is also given. In contrast to previous work, our method facilitates the reusability of the captured mesh while preserving the global time-varying surface properties. We discuss the shortcomings of a non-hierarchic and non-rigid parametrization to represent animations of live-actor performance without articulated rigidity prior. In addition, we review related works concerning the acquisition and processing of dynamic captured surface in the context of 3D video editing. Finally, we offer a scheme to stylize shape motion and photometric multi-view data in a well-formed manner.

- **Chapter** 4 reflects upon the exceedingly hard problem of non-rigid alignment with cage-based parametrization to obtain temporally consistent dynamic outputs, without any assumption on temporal matching. In this chapter, we describe an effective framework to achieve target surface reconstruction and to establish robust temporal correspondence relying on geometric features only. Although it is still not the perfect solution, this first proposition toward *Cage-based Registration from Multi-View* is a considerable improvement for controllable dynamic shape capture and suitable output reusable by commercial animation engines. We initially provide a series of results using our *Iterative Cage-based Registration* technique in order to obtain a consistent dynamic mesh as output. In addition, we propose a review of related works helpful for registering frame-to-frame shapes in unsupervised manner.
- **Chapter** 5 concludes the research developed in this book with a summary about the contributions. In addition, we state and discuss conclusions and several limitations drawn from our experiments. Finally, a variety of potential research directions and open suggestions for further work are outlined.

References

1. G. Johansson, Visual perception of biological motion and a model for its analysis. Attention Percept. Psychophys. **14**, 201–211 (1973)
2. M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. Vande Moere, O. Staadt, Blue-c: a spatially immersive display and 3D video portal for telepresence. ACM Trans. Graph. **22**, 819–827 (2003)

Chapter 2
Sparse Constraints Over Animatable Subspaces

In this chapter, we present a cage-based inversion process for modeling tasks. This novel core methodology estimates coherent enclosed model deformation via sparse positional constraints. In particular, we investigate a new technique mixing surface-and-space deformation in order to fit input feature correspondences. Moreover, user-specified constraints are transferred into the cage domain with a well-adapted and detached Laplacian-type regularization. Thus, we suggest a novel algorithm to easily edit shape deformation using the *inverse kinematics* paradigm applied in geometric space. Finally, an intuitive user interface is proposed to allow users to easily specify desired screen-space constraints.

2.1 Introduction

Inverse Mesh Modeling. Nowadays, mesh editing and inverse modeling techniques play a decisive role in *Computer Graphics*. This research domain has been intensively studied over the years. Nevertheless, the relentless increase in demand of industry has motivated researchers to exhibit new coordinate systems as-well-as new optimization frameworks. Building accessible pipelines able to provide flexible output for animation re-use is a challenging issue. In nature, deformation techniques can be seen as an energy minimization process (defined locally or globally) that measures how much the object has been deformed from its initial pose given a support domain (for instance surface or volume representation). Approximating global shape characteristics of the surface aims to enforce specific surface resistance properties (such as rigidity, flexibility or elasticity). One significant challenge is to find a fast framework to achieve plausible and scalable boneless inverse kinematics that produces pleasing deformations and preserves the global properties of the surface. Avoiding the articulated skeleton parametrization to achieve large deformation allows us to considerate a large number of highly non-rigid deformations that occur in the physical world, also known as *free-form deformation*.

Y. Savoye, *Cage-based Performance Capture*,
Studies in Computational Intelligence 509, DOI: 10.1007/978-3-319-01538-5_2,
© Springer International Publishing Switzerland 2014

Constraint-based Mesh Optimality. With the explosion in polygons, mesh deformation tools definitively take advantage of parametrizations demonstrating independence toward the resolution of the surface-of-interest. Meanwhile, cage-based deformation techniques are increasingly used to control the deformation of an enclosed fine-detailed mesh in commercial software. Achieving deformation based on vertex constraints has been extensively studied for the case of pure meshes, but a few works specifically examine how such vertex constraints can be used to efficiently deform the models and to estimate the corresponding cage-pose. In this chapter, we show that mesh optimality can be efficiently reached within a framework employing *Cage-based Deformation* enhanced by *Laplacian Mesh Editing* principles. First, we provide a linear estimation framework for cage-vertex coordinates. Second, our deformation regularization is expressed exclusively on the cage connectivity rather than the enclosed mesh, yielding a computationally efficient solution which fully benefits from cage-based parameterization. We demonstrate a practical use of this scheme for dense mesh editing from sparse screen-space constraints reprojected into the world-space coordinates.

Contributions. In this chapter, we combine surface and space deformation techniques into an inverse variational modeling approach. Indeed, we focus on estimating an enclosed model that fits some sparse surfel displacement constraints, specified by the user (not limited to artists) over the enclosed mesh surface. We explore the least-squares cage geometry as an intermediate, transparent and off-the-shell tool, that is not directly editable by users. In other words, the model is embedded in a volumetric bounding cage using generalized barycentric coordinates with local properties. The optimal geometric solution is obtained by a single minimization within a linear black-box solver.

However, the key contribution of this chapter lies in the use of sparse linear surface constraints to estimate the global cage parameters reproducing the desired deformation of the enclosed model. Besides, such constraints are expressed on the enclosed model and transferred into the cage-based subspace (offered by the bounding cage). Consequently, we take advantage of optimal reduced parameters provided by the given coarse cage surrounding the surface. In addition, the cage structure is enhanced with Laplacian regularization leading to an effective over-determined system. The Laplacian cage maintains a spatial deformation of the mesh geometry, more powerful than applying *surface-based* or *cage-based* techniques separately. Finally, we show how Laplacian regularizations can be operated to non-manifold animatable subspaces such as traditional articulated kinematic skeletons.

Chapter Reminder. The rest of the chapter is organized as follows. First, we briefly review most of the relevant work on interactive deformable surface and shape parametrization in Sect. 2.2. In particular, we propose a clear understanding concerning *surface-based*, *skeleton-based*, *space-based* and *cage-based* parametrization. Then, we give a conceptual and mathematical description of our inverse cage-based system in Sect. 2.3, and we show the effectiveness of our method on various datasets. Next, a generalization of the employed variational strategy to skeletal subspace structures is offered in Sect. 2.4. In this specific case, we found that Laplacian regularization

breaks the rigidity of skeletal subspaces by discretizing an elastic energy on their connected joints. Finally, this chapter is concluded in Sect. 2.5 with some discussions about benefits and limitations of our system.

2.2 Related Works on Interactive Deformable Surface

In this section, we present various relevant parametrizations in the area of *Interactive Deformable Surfaces* that are useful from digitized model to animatable surface. In particular, we focus on the representation of animatable discrete surface with reduced and reusable parameters. More importantly, we overview the large body of recent work on techniques addressing the problem of interactive mesh deformation using underlying deformable structure, with the aim to be used for video-based animation purpose. In particular, we are interested in three different levels: *surface-based*, *skeleton-based*, and *space-based* parameterization. Finally, our focus will be extensively oriented on *cage-based* paradigm.

2.2.1 Deformable Models

Theory of Deformable Models. Deformable models are efficient and popular tools in both *Computer Graphics* and *Computer Vision* where the main idea is to encode and manipulate a variety of flexible shape, at the heart of many applications. For instance, deformable models have been used in animation and simulation. Generally, deformable models can be described as a class of physics-based models that deform under the laws of Newton mechanics. Another advantage of attractive deformable models is to reduce the dependence to the object representation by providing information on its global behavior. In addition, deformable models provide promising techniques for registration. Simulating an animation can be seen as deforming a model across time. In case of dense meshes, a discrete deformable model is required

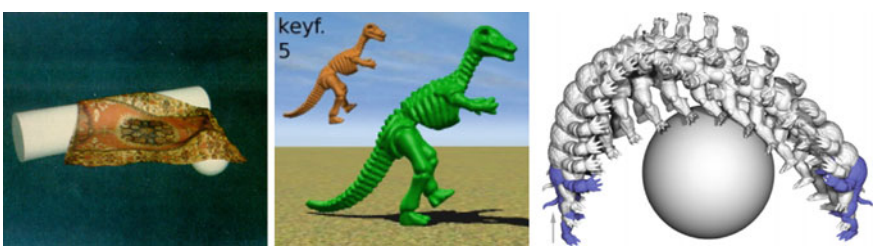

Fig. 2.1 Deformable models: (from *left* to *right*) an elastic deformable model (image courtesy of [2]), a reduced deformable model (image courtesy of [3]), a physically-based meshless deformable model (image courtesy of [4])

to efficiently performs a scalable deformation in interactive time. We formally define a deformable models as an active structure admitting a time-varying deformation map applied on its parameter components to form trajectories. For more details about deformable model for medical application, we refer the reader to the survey of Mcinerney et al. [1] demonstrating that deformable models are capable of accommodating variability of biological structures over time. Then, the problem of a discrete representation for deformable models make sense for computer scientists.

Discrete Deformable Models. One of the first pioneering technique for the class of *elastically deformable models* is presented by Terzopoulos et al. [2], which is based firmly in physics. Authors demonstrated that the natural dynamics of a deformable model can be described by constraints and equation governing the elastic deformation behavior as a potential energy of non-rigid curves. Moreover, they referred to the theory of elasticity to discretize the amount for resistance for stretching, bending and twisting in a closed-form equation:

$$\int_{\Omega} k_g \left\| G - G' \right\| + k_b \left\| B - B' \right\| dudv$$

where G is the fundamental form associated at the natural shape and B is the curvature imposed by the deforming constraints on a body Ω, with scalar factors k_g and k_b to balance the importance of both energy terms. Similar physically-based modeling techniques simulating elastic deformation with spring-force, and designed for animating deformable models are also proposed by Güdükbay et al. [5]. An interesting example of elastic deformation can be seen in the *snakes* active contour model as proposed by Kass et al. [6]. Nevertheless, the increasing polygonal complexity of scanned models invited researchers to introduce *reduced optimal control* for deformable model with the ultimate aim of improving modeling interactivity.

Reduced Optimal Control. An intriguing deformable object strategy is employed by Barbič et al. [3] to generate an animation that satisfies the goals of animators using reduced optimal control. These models have some shortcomings that need user intervention to achieve high performance. Furthermore, the use of a deformable model is mainly motivated by the need of realism with less efforts and manual work. Their techniques present a fast optimal control for deformable object to simulate complex physically-based mesh deformation and to reproduce dynamic effects. We notice that Adams et al. [4] design a meshless approach for deforming highly complex shapes in real-time that extends existing work on meshless finite elements. Finally, Sumner et al. [7] propose an impacting embedded deformation approach for shape manipulation that also benefits from reduced control via a deformation graph. High-quality shape deformation has gained a lot of attention in recent years. For instance, non-rigid mesh evolution can be performed using intrinsic surface deformation or extrinsic space deformation techniques (see Fig. 2.1).

2.2.2 Surface-based Parameterization

Differential Geometry. In recent years, differential geometry became an important topic of interest where manifold geometry is proceeded through linear algebra and differential calculus. Many efforts have been expanded on surface-based deformation. Consequently, *differential mesh editing* techniques are becoming a prominent suitable tool for preserving surface details during mesh operations. For instance, Laplacian coordinates provide an efficient intrinsic representation of the surface mesh to encode its local geometry as shown by Wardetzky et al. [8]. These editing systems represent local features using differential coordinates defined in a global coordinate system.

There are several types of highly impacting approaches exploiting a differential descriptor of the edited surface in terms of Laplacian and differential coordinates for mesh editing [9, 10]. Differential information as local intrinsic feature descriptors have been massively used for mesh editing in various frameworks over the decade. For instance, the proposed method of Lipman et al. [11] allows the reconstruction of the edited surface by solving a linear system that satisfies the reconstruction of the local details in a least-squares sense. In addition, the differential coordinates are not invariant with respect of rotation and scaling, and it is definitively a potential drawback for animated geometry. Nevertheless, these deformation techniques mimic elastic deformation via a Laplacian framework. Another interesting modeling technique using a functional optimization system is presented by Nealen et al. [12] where a set of curves is used to create sophisticated models. Another powerful editing framework also based on a descriptive set of wires is offered by Gal et al. [13].

Dual Laplacian. *Dual Laplacian* systems describe a collection of approaches where the Laplacian is not directly expressed on mesh vertices. Early approaches similar to the work of Au et al. [15] motivated the use of *Dual Laplacian* system by reducing distortion in parametrization and geometry. Another mesh editing method working in the dual domain for regions of interest can be found in the work of Qiong et al. [16]. Unfortunately, Laplacians cannot satisfy all natural properties. The observation of the local surface behavior has been proposed recently by Sorkine et al. [14], where *as-rigid-as-possible* surface modeling is performed by the minimization of the deformed surface under local rigidity transformation constraints (see Fig. 2.2).

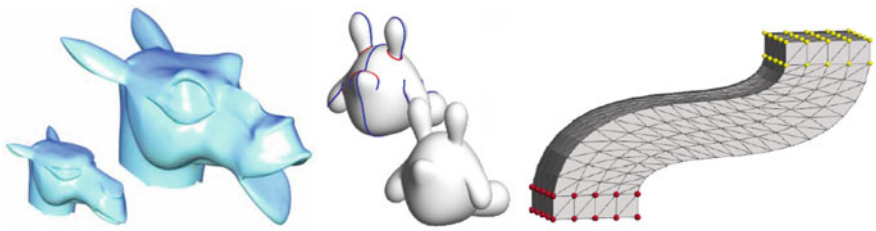

Fig. 2.2 Differential geometry: (from *left* to *right*) Laplacian mesh editing (image courtesy of [11]), fibermesh (image courtesy of [12]), and as-rigid-as possible surface modeling (image courtesy of [14])

2.2.3 Skeleton-based Parameterization

Articulation-Driven Skins. Introduced by the pioneering work of Magnenat-Thalmann et al. [17], *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 articulation to conduct a local parametrization for a variety of enveloping skinned mesh. This underlying structure allows user intervention to control large scale motion with fewer characteristic of the model, but does not provide realistic muscle movement and skin motion. In skeleton-based animation, the bone length is naturally preserved because the parametrization is purely kinematic and relies on non-independent parameters. Besides, articulated deformation ensures rigidity by blending affine per-joint transformation in hierarchical manner. We also notice that this type of parametrization is well-adapted for humanoid, quadruped and vertebral living animals that are naturally articulated, but it is difficult to maintain the illusion of an underlying bone structure in presence of additional non-rigid surrounding layers like clothing or musculature.

Rigging and Skinning. Generally, skeleton animation is often associated with a rigging and skinning function. Firstly, the rigging step attaches the skeleton to the mesh by computing a per-joint influence weight distribution for all mesh vertices. Secondly, at each frame the skinning step deforms the mesh from its default pose according to the blending of joint transformation. The first automatic rigging method for skeletal deformation is introduced by Baran et al. [18] with two principal steps: *skeleton embedding* and *skin attachment*. The skeleton embedding phase focuses on placing all skeleton joints in the optimal location according to the enclosed volume of the input character shape. The skeleton structure adjustment is then performed by estimating a refined embedding thanks to a complex multivariate objective function with respect to the mesh medial axis and a discrete penalty function. Next, the skin attachment process finds a bone weight distribution from a single mesh pose for attaching mesh vertices to skeleton. The per-vertex weight distributions generated from the joint proximity and a surface-based diffusion equilibrium equation use

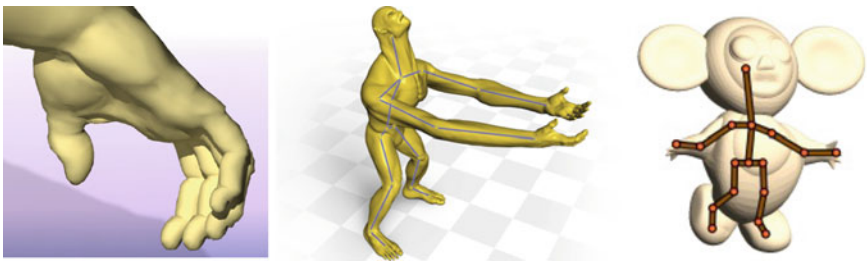

Fig. 2.3 Skeletal-based deformation: (from *left* to *right*) context-aware skeletal shape deformation (image courtesy of [20]), stretchable bone skinning (image courtesy of [21]), a skeleton is automatically adjusted and attached to a given skin using bone weight diffusion (image courtesy of [18])

Laplace's diffusion. Finally, this interesting technique seems limited to single component mesh and required the character to be input in the same orientation than the mesh model with approximately same scale. Notwithstanding, the problem of building character rigs from multiple sources was already attempted by Miller et al. [19].

Advanced Skinning. Skinning techniques can be seen as a formulation for mesh deformation resulting of control points displacement. Another well-known skinning algorithm to avoid the undesired candy-wrapper effect by employing a dual quaternion formulation is proposed in Kavan et al. [22]. The skin-skeleton paradigm is also tackled by Weber et al. [20] with a skeletal-based system producing naturally-looking skin deformation that runs according the context of input shape examples (see Fig. 2.3). More recently, jacobson et al. [21] extend the space of deformations for skinning for bending, stretching and twisting with only slight variant of existing skinning equations. More advanced physically-based skinning techniques have been also proposed recently for collision-free flesh-like character. For instance, a near-interactive simulation skinning technique is developed by McAdams et al. [23], to make physically-based elasticity practical for production of character rigging for skeleton-driven elasticity models with necessary control. The complexity of these techniques could be a potential drawback for representing complex surfaces deformation.

2.2.4 Space-based Parameterization

Free-Form Deformation. Introduced in 1986 by Sederberg and Parry [24], *Free-Form Deformation* (also known as FFD) is a pioneering class of classical solid modeling techniques where a rubber-like volumetric grid embeds an original shape of large-scale mesh or point-cloud to be altered. This first computer-assisted tool is based on space warping by blending the deformation at the control points. Conceptually explained by Coquillart et al. [25], the three dimensional lattice contains a rather small number of control points that encloses the model and the associated basis deforming functions should be smooth for aesthetics result. Manipulating the control points smoothly deforms the space enclosed in the lattice, and the embedded geometry deforms accordingly. Thus, the global-to-local deformation of the object is an intuitive appreciation with respect to the deformation. In addition, the original volume can be preserved using trivariate Bernstein polynomials. As seen in the work of Mac Cracken et al. [26] and Kobayashi et al. [27], the original FFD is generalized to various deformation approaches (arbitrary and cylindrical lattice), and various control point weighting factors.

Volumetric Preservation. Space-based parametrization offers improvements over surface-based approaches by allowing volumetric constraints. For instance, the *Volumetric Laplacian* studied by Zhou et al. [30] preserves volumetric properties for large deformations. Volume preservation is also addressed by Huang et al. [29] with a non-linear deformation energy projected onto the control mesh vertices.

Furthermore, there has been a terrific deal of work using an analogy to the traditional skeleton-based inverse kinematics but applied on boneless surfaces. Clearly, the important work of Der et al. [31] and Sumner et al. [28] introduce elegant mesh-based inverse kinematics techniques (see Fig. 2.4).

Fig. 2.4 Space-based deformation: (from *left* to *right*) mesh-based inverse kinematics (image courtesy of [28]), gradient domain mesh deformation (image courtesy of [29]), and volumetric Laplacian (image courtesy of [30])

2.2.5 Cage-based Parameterization

Cage-based Paradigm. In recent years, we notice an increasing interest for cages to manipulate 3D models efficiently. In this section, we review algorithms and properties of various cage-based deformation techniques. In order to deform a mesh, the cage is enhanced by a coordinate system, mainly based on the generalization of barycentric coordinates but also developed from more elaborated mathematics like Green's theory. By cage-based deformation techniques, we mean emerging deformation techniques where a bounding cage deforms an enclosed triangular mesh in a volumetric fashion, according to a particular coordinate system (such as *Mean Value Coordinates*, *Harmonic Coordinates* and *Green Coordinates*). Additionally, a coordinate system expresses the influence of the control cage on the deformable enclosed model. In the following, we focus on studying the peculiarity of cage-based paradigm to encode surfaces. Thus, we detail now the *automatic cage building* process and impacting coordinate systems (see Fig. 2.5).

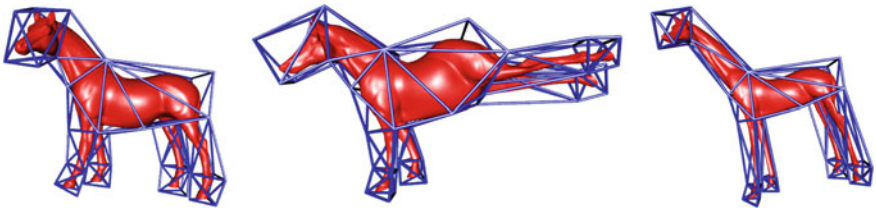

Fig. 2.5 Cage-based deformation: a low-dimensional deforming cage is used to drive natural deformations of a shape without relying on quasi-articulated priors (data courtesy of Ju et al. [32])

Automatic Cage Building. In cage-based approach, points controlling the deformations are the vertices of a specified polytope cage, bounding an enclosed template shape. A proper cage should envelop the original model without any self-intersection. The cage structure plays the role of a deformer. Thus, a cage is also considered better if it has significantly fewer vertices, while keeping global shapes similarities with the original model. Unfortunately, designing this coarse cage is the most tedious task for cage-based techniques. Solving the problem of automatic cage design can alleviate the user burden in setting up the cage. Relatively few cage-building techniques have investigated the complicated problem of generating a cage from a single model.

Even if few attention has been paid to the problem of automatic generation of coarse cage, promising solutions already exist. Importantly, a fully automatic method is presented by Xian et al. [33] to generate a coarse bounding cage that can keep the connectivity structure and significant features of the original mesh. In a first step, the bounding box is computed from the dense mesh by *Principal Analysis Component*. The box is slightly enlarged to ensure the cage to enclose the model. In a second step, the bounding box is discretized into voxels for finest shape enclosing, and the outer surface is triangulated. Finally, the output cage geometry is smoothed. They also proposed a formulation to compute the voxel resolution to have a desired sparse factor for the cage. Even if this technique provides a compelling direction to generate a cage, the resulting output geometry seems not a fairly abstracted coarse cage. In addition, Ben-Chen et al. [34] use an additional offsetting operation to enlarge the cage. Rather than a voxelization pass, the cage building technique of Deng et al. [35] propose an iterative simplification step using the well-known *Quadric Error Metric*. This procedure is followed by a self-intersection removal step. Topological intersections are detected using a marching squares method to avoid heavy topology repairing.

Most part of the time, the cage is built by hand or automatically obtained by adapting a bounding box with the condition that the cage should be coarse enough for easy manipulation. An automatic coarse cage-generation algorithm is also developed by Xian et al. [36], where the voxelization and principal analysis component are key ingredients to compute the oriented bounding box. Next, the OOB is sliced into multiple connected boxes. Each boxes are locally modified, reoriented, rescaled, and

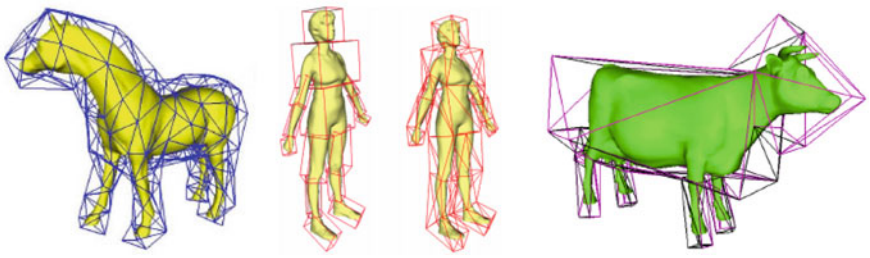

Fig. 2.6 Automatic cage generation: (from *left* to *right*) coarse bounding cage generation using voxelization (image courtesy of [33]), using oriented bounding box (images courtesy of [36])

merged together in order to generate the final cage mesh (see Fig. 2.6). This technique seems to provide meaningful cages as output for a given input model. In order to obtain a better abstraction of the envelop and a high degree of sparseness, Ju et al. [37] propose to reuse a library of pre-designed partial cage templates reassembled together with additional optimization to compose a complete cage. The resulting cage is well-adapted to the deformation behavior by ensuring tightness of the cage embedding. The technique of Kosinka et al. [38] relies on a injective shape deformation strategy using cube-like cages. Nevertheless, cube-like cages are not properly suitable for animated organic shape. Finally, the problem of rigidness for cage-based deformation is explored by Zhao et al. [39] with the incorporation of additional internal cage edges at the cage design step.

Coordinate Systems. There has been a great deal of work done in the past on developing techniques for deforming a mesh with generalized barycentric coordinates. Inspired from the pioneering work presented by Meyer et al. [40], cage-based methods are well-adapted for coherently deforming a surface by improving previous space-based deformation techniques. In particular, a cage parametrization allows model vertices to be expressed as a combination of cage-handles in conjunction with meaningful rigging functions. This family of methods tries to reach important properties: quasi-conformal mappings, shape preservation and smoothness. To animate the model, cage vertices are displaced and the vertices of the model move accordingly via a linear weighted combination of cage geometry parameters. By relying on a fixed cage connectivity as single geometric element, cage-based skinning brings more animator-friendly abstraction to manipulate a shape, other type of unstructured or degenerated geometries such as point clouds, polygon soup or particles. Recent years have seen an increasing interest for cage-based deformation: an emerging class of purely geometric space-based coordinates, widely used for controlling meshes enclosed in a flexible coarse bounding polytope. Then, we propose to review main coordinate systems (see Fig. 2.7).

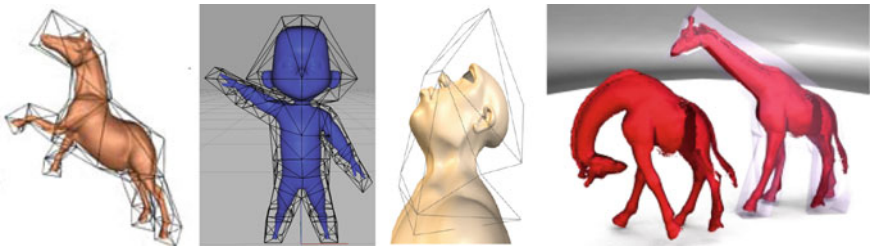

Fig. 2.7 Cage-based coordinates: (from *left* to *right*) mean value coordinates (image courtesy of [32]), harmonic coordinates (image courtesy of [41]), green coordinates (image courtesy of [42]), and variational harmonic mapping (image courtesy of [34])

Generalized Barycentricity. Various cage-vertex coordinate systems have been developed for space-based manipulations. Then, a rigging function expressed as generalized barycentric coordinates is associated with the cage-model paradigm to

describe realistic and plausible control behavior of the deformation. Motivated by the heavy use of interpolant function in surface parametrization, *Mean Value Coordinates* introduced by Ju et al. [32] is the first cage-vertex influence proposed as normalized generalization of homogeneous barycentric coordinates for cages. Unfortunately, negativity brought with *Mean Value Coordinates* leads to counter-intuitive deformation. To overcome this weakness, Lipman et al. [43] introduce the *Positive Mean Value Coordinates* by incorporating the visibility of the cage vertex from a point of the object in the coordinate system computation. More recently, *Poisson Mean Value Coordinates* developed by Li et al. [44] express *Mean Value Coordinates* with pseudo-harmonic properties. Finally, Langer et al. [45] also modify existing barycentric coordinates to *Higher-Order Barycentric Coordinates* by modifying *Mean Value Coordinates*. Generally, such coordinates have a lack of smoothness, and this drawback invited further exploration toward harmonicity.

Harmonicity. The importance of detail-preserving deformations and smoothness properties is the main issue for cage-based deformation. For this reason, harmonicity has attracted huge interest over the years to produce highly-detailed cage-based encoding. At first insight, a major research of DeRose et al. [46] explore *Harmonic Coordinates* for deforming characters for the movie Ratatouille at Pixar®. Then, the problem of designing and controlling volume deformations used to articulated characters are treated by Joshi et al. [41], with the introduction of *Harmonic Coordinates* significantly improves the deformation accuracy. In few words, *Harmonic Coordinates* are an elegant non-negative coordinates system for highly-detailed character animation, defined as the solution of Laplace's equation in the form of volumetric heat diffusion in the cage interior subject to appropriately chosen boundary condition. Meanwhile, Raif Rustamov [47] proposes a very interesting *Boundary Element Method* formulation to speed-up and to reduce the storage of *Harmonic Coordinates*. Moreover, a variant for *Harmonic Coordinates* is proposed by Lin et al. [48] with smooth indirect binding using segmented thin-layers. Nevertheless, Ben Chen et al. [34] prefer to estimate harmonic maps using a set of harmonic basis functions computed from a collection of example poses. These *Variational Harmonic Maps* are insensitive to cage geometry and do not require a voxelization procedure (see Fig. 2.8).

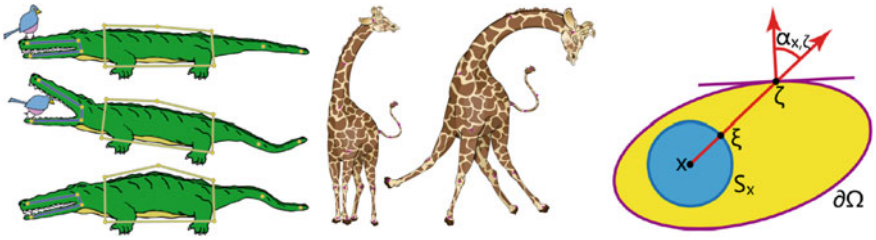

Fig. 2.8 Cage-based coordinates: (from *left* to *right*) bounded biharmonics weights (image courtesy of [49]), controllable conformal maps (image courtesy of [50]), and Poisson coordinates (image courtesy of [44])

Polyharmonicity. Furthermore, polyharmonicity has gained major attention recently and has demonstrated better shape-awareness and smoothness [51]. For instance, *Biharmonic Weights* described by Jacobson et al. [49] bring sparsity by restricting the weighting system to be *bounded biharmonic* with the use of a brush to define the region to be influenced. Nevertheless, automatic weight assignment method often generated weight matrices that required higher amount of storage space. Then, a Poisson-based weight reduction technique is proposed by Landerneau et al. [52] to produce geometric smooth weight reduction for cage-based deformation. Finally, late-breaking research of Jacobson et al. [53] demonstrate the usefulness of handle-based shape deformation using popular polyharmonic functions.

Quasi-conformality. Exact conformality is a very difficult property to obtain in conjunction with the cage-based strategy - and nearly impossible to formulate in 3D. Nevertheless, many improvements have been done toward better conformality with non-linear coordinates system. Outstandingly, Lipman et al. [42] introduce *Green Coordinates* leading to non-linear space deformations with shape-preserving properties. Then, *Green Coordinates* enforce quasi-conformal scale-preserving deformation and allow the use of partial cage. The use of partial cages are also investigated by Li et al. [54] with the introduction of cage-free local deformations using *Green Coordinates*. More importantly, we would like to point out that *Green Coordinates* ensure quasi-conformal mapping by integrating cage-face orientation, and overcome the lack of closed-form solution of *Harmonic Coordinates*. Even if the method of Weber et al. [50] produces pure conformal maps in 2D, unfortunately there is a scientific consensus that conformal maps in 3D nearly never exist.

Cage Reusability. The problem of reusing static cage templates from a small library is addressed by Ju et al. [37]. This technique proposes to rearrange pre-designed partial cage elements in order to form a desired final closed cage. Furthermore, learning deformation for a new shape from an existing pose is developed by Ben Chen et al. [55], demonstrating a cage-based method to transfer animation by employing spatial deformation transfer. Spatial deformation is also a useful tool for Chen et al. [56] for transferring deformation from one shape to another one using cage-based paradigm. Finally, the reuse of fairly coarse cages seems to be limited to deformation transfer only, and never associated before to the problem of capturing shapes.

2.3 Our Approach: Cage-based Shape Inverse kinematics

In this section, we propose the first formulation of *Inverse Cage-based Modeling* called *CageIK* where the problem of scalable mesh deformation is cast as an inverse variational problem ables to estimate cage parameters. In particular, we extend the *Inverse Kinematics* paradigm to classical cage-based deformation by recovering cage parameters from user-specified screen-space constraints. Simply put, we introduce the mathematical foundation of the core tool declined all along this book.

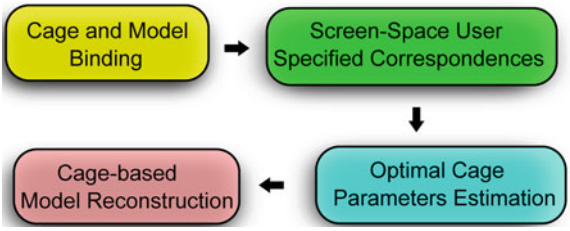

Fig. 2.9 Cage-based modeling: we introduce a novel hybrid method mixing surface-based and space-based deformation tools to perform cage-based inverse kinematics process

Boneless Mesh Deformation. *Boneless mesh deformation* techniques are well-known in the context of non-articulated surfaces deformation. This class of techniques is widely used to evolve a deformable triangulate mesh without any piecewise rigidity prior. Then, we conceive a novel *Dual-Laplacian* method for *Cage-Based Deformation Inversion* ables to achieve large mesh deformation. We examine the feasibility of an *inverse strategy* using a detached cage control, without explicit rigid parts. For this purpose, our method combines cages and Laplacian techniques to interactively fit the bounding cage expressing the desired and unknown deformed model. Then, we exhibit a new estimation technique, involving an optimal reduced number of parameters suitable for animation re-use (see Fig. 2.9 for the pipeline).

Novel Hybrid Mesh Deformation. The idea of combining space-based deformation techniques with surface-based techniques was firstly enunciated by Daniel Cohen-Or [57]. Meanwhile, the lack of reusable parameterization for non-rigid surface invited us to abandon underlying articulated skeletons and to propose a novel approach based on a dual formulation of the Laplacian expressed on the cage structure itself. In addition, the main challenge is to deal with the high number of degrees of freedom provided by the coarse cage. Then, we aim to estimate cage parameters expressing a desired mesh from an input configuration. To realize this *Cage-based Inverse Kinematics* process, we cast the problem as a minimization problem for cage retrieval. This optimization process retrieves the pose in a least-squares sense from arbitrary sparse features constraints expressed directly on the enclosed mesh. Thus, specified constraints are transferred to the cage domain using the rig. In our system, we employ a Laplacian regularization on the cage to surfacically constraint the spatial deformation.

Cage Optimality. In contrast to previous approaches, we propose an easy-to-implement approach but yet efficient to estimate optimal cage parameters through minimization process. In our system, we employ cage-Laplacian allowing us to perform a coherent cage estimation under sparse constraints expressed on the enclosed mesh. The main benefit of our method lies in the fact that the parameterization is fully geometry-independent to the enclosed mesh resolution. Even if our work is closed to the one of Borosan et al. [58] around the idea of producing an hybrid mesh deformation, our work is still novel because we rely on a linear optimization problem. The described optimization algorithm minimizing the deformation energy is also closely related to the work of Chen et al. [56] for integrating the cage into a

minimization framework. However, the key contribution is to solve a sparse linear system that estimates the optimal cage parameters reproducing the desired deformation of the enclosed model. Besides, such constraints features are defined over the model and directly transferred into the subspace domain using the indirection of a coarse control meshes.

2.3.1 Laplacian-based Mean-Value Cage

Mean-Value Subspace Deformation. A cage is described as a continuous bounding surface enveloping a given volume where an enclosed model is immerged. Then, model vertices are expressed as a weighted linear combination of cage vertices. The weights are given by a set of generalized barycentric coordinates stored in a $m \times n$ deformation weights matrix denoted by $\mathcal{H}$, where m is the number of vertices for the model and n for the cage. We also denote by $g_k\,(i) \in \mathbb{R}$ the normalized blend weights representing the deforming influence of the kth cage vertex on the ith model vertex. Furthermore, it is also possible to deform an arbitrary point on the enclosed mesh written as a linear combination of the coarse mesh vertex position via a constant weight deformation. Then, a point inside the cage is represented by an affine sum of cage-handles multiplied by a specific weight function called the *cage coordinates*. A deformed point $v_i' \in \mathbb{R}^3$ is then obtained by the following equation.

$$v_i' = \sum_{k=1}^{n} g_k\,(i) \cdot c_k'$$

where $v_i' \in \mathbb{R}^3$ is the deformed Cartesian coordinates according to a vector of cage geometry $\left\{c_1', \cdots, c_n'\right\}$. In order to produce as-local-as possible deformation mapping on the enclosed surface. In this chapter, both coordinate systems (*Harmonic Coordinates* and *Mean Value Coordinates*) are mathematically suitable for an effective inverse modeling system. *Mean Value Coordinates* are a simple and powerful closed-form method, with linear properties for creating functions that interpolate values assigned to the vertices of a closed mesh. For more technical details, we refer the reader to the paper of Ju et al. [32] offering a pseudo-code algorithm as well as the mathematical formulation for this coordinate system.

Laplacian Cage. In our work, we rely on a Laplacian-type regularization applied on the cage structure instead of its traditional expression on the enclosed mesh itself (as seen in *Laplacian Mesh Editing* techniques). Then, we call *Laplacian Cage*, a coarse triangular bounding polytope enhanced with a *Discrete Laplacian Operator*. There is a number of approximations for the Laplacian operator, and each has their own advantages and disadvantages. At first glance, the three-dimensional Laplace's equation of an input function $f\,(x, y, z)$ is a partial differential equation enforcing the neighborhood curvature coherence by setting the second derivative of equal to zero.

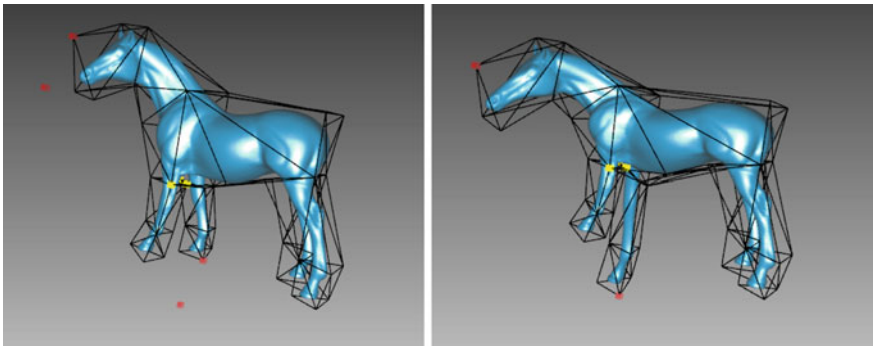

Fig. 2.10 Laplacian cage-based shape editing. The enclosed model can be directly edited by using traditional Laplacian mesh editing on the cage, followed by traditional cage-based deformation with the deformed cage (mean value coordinates are used in the example)

$$\Delta f = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0$$

where (x, y, z) is a triplet of Cartesian coordinates, the symbol ∇^2 denotes the *Laplace Operator* and ∂^2 is the second partial derivatives, and Δ is called the *Laplacian Operator* or simply the *Laplacian*. An example of *Laplacian Cage-based Shape Editing* is proposed in Fig. 2.10, where the enclosed model is edited via a Laplacian-based deformation of the cage.

Cotangent Weighting System. To ensure better shape-awareness, we inject a weighting system inside the Laplacian encoding, similarly to previous techniques. Efficient vertex-based Laplacian deformation techniques always recommend cotangent weighting system for a better approximation of the local mean curvature. To define an accurate discretization, various researchers proposed to weight the Laplacian with a scale-dependence weighting system. Then, the tangential component of the Laplacian is eliminated. For instance, Desbrun et al. [59] describe three discrete approximations. Each of them which boil down to a weighted sum of the one-ring neighbors of a vertex. An interesting discretization of per-edge cotangent weight w_{ij} associated to the oriented edge (i, j) linking the given cage-handle c_i to c_j is defined as follows (see Fig. 2.11):

$$w_{ij} = \frac{1}{2\text{Area}\,(c_i)} \left(\cot\left(\alpha_{ij}\right) + \cot\left(\beta_{ij}\right)\right)$$

where $Area\,(c_i)$ denotes the neighboring area of the adjacent faces of the cage-handle c_i. Moreover α_{ij} and β_{ij} denote the angles opposite to edge (i, j). This weighting system takes advantage of the curvature flow. Given the fact that a coarse cage connectivity has often irregular triangles with degenerated aspect ratios, the use of the cotangent weights compensates the irregular triangulation. More generally, this is

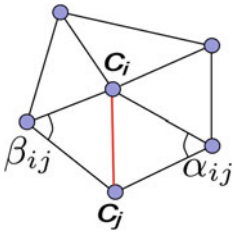

Fig. 2.11 Cotengent weight: We use a non-uniform weighting system for the Laplace Beltrami operator acting as cage-Laplacian

the better discretization of the *Laplace-Beltrami Operator* to functions on manifolds. Naturally, this discretization gives better results than the uniform weighting scheme by non-uniformly distributed mean curvature. The discrete version of the *Laplacian Operator* is also known as the discrete mean curvature normal operator because it evaluates the mean curvature normal.

Differential Cage Coordinates. In our work, we conservatively employ differential coordinates ($\hat{\delta}$-coordinates) to encode the local surface properties of the cage. Then, the differential coordinates $\hat{\delta}_i$ of c_i are obtained by linear transformation of the cage-handles. In other words, differential coordinates of a given cage-handle are computed as the default difference between its absolute Cartesian coordinates and the center of mass of its immediate neighbors in the cage mesh:

$$\hat{\delta}_i = \left(\hat{\delta}_i^x, \hat{\delta}_i^y, \hat{\delta}_i^z\right) = \sum_{j \in N(i)} w_{ij} \left(c_i - c_j\right)$$

where $\hat{\delta}_i$ is a 3-vector packing the differential coordinates for the ith cage handle and w_{ij} is the per-edge cotangent weight. Instead of using absolute coordinates, we describe the cage mesh geometry using a set of differential $\hat{\Delta} = \left\{\hat{\delta}_i\right\}$.

Discrete Cage Laplacian. A discrete Laplacian operator is defined at each cage-handle by the weighted sum of the difference vectors between the given cage-handle and its adjacent neighbors. A discrete Laplacian operator $\mathcal{L}_C(\cdot)$ is defined by its linear action on handle-based functions. The Laplacian operator can be linearly approximated at each vertex by the Laplace operator given a simple local parametrization. The properties of $\mathcal{L}_C$ is encoded by the coefficients matrix w_{ij}. When the weights in the sum are represented using the Voronoi area and cotangents, the Laplacian vector approximates the mean curvature at the cage-handle.

$$\mathcal{L}_C(c_i) = c_i - \frac{1}{d_i} \sum_{j \in N(i)} c_j = c_i - \frac{1}{d_i} \sum_{k=1}^{d_i} c_{i_k}$$

where d_i is the number of immediate neighbors $N(i)$ of ith cage handle (the degree or valence of i) and c_k is the kth direct cage-handle neighbor for a given cage-handle c_i.

Matrix-Form Cage-Laplacian. Thus, we propose a matrix-form expression discretizing this operator on the cage structure leading to the *cage-Laplacian* $\mathcal{L}_C$. First, we assume that i is the index of the inner cage-handle (also known as cage vertex) and j the index of a cage-handle belonging in the 1-neighborhood of i. Let A be the *Adjacency Matrix* of the cage connectivity, where its enteries A_{ij} are defined as follows:

$$A_{ij} = \begin{cases} 1 & (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $(i, j) \in E$ means that cage-handles i and j are adjacent. We denote by E the set of cage edges. Let $D = \text{diag}\,(d_1, \cdots, d_n)$ be the *Diagonal Degree Matrix*, where its enteries D_{ij} are defined as follows:

$$D_{ij} = \begin{cases} d_i & i = j, (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

Given its discretized form, the cage is easily processed by differential geometry algorithm. So, the *Laplacian* of the cage, measuring the deviation of a cage-handle from the centroid of its neighbors, is represented by a $n \times n$ matrix called the *Laplacian matrix* L transforming the absolute cage-handle coordinates to relative coordinates. The matrix is obtained by subtracting the *Adjacency Matrix* A to the *Diagonal Degree Matrix* D, then its enteries L_{ij} are defined as follows:

$$L_{ij} = D_{ij} - A_{ij} = \begin{cases} -1 & i = j \\ d_i & (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

The Laplace operator is also known as the divergence of the gradient at a scalar. Nonetheless, an important problem with the umbrella operator and a simple Laplacian discretization is they can provoke significant sliding and shape distortion when its applied to a coarse cage triangulation. Therefore, expressing the Laplacian on the cage can be seen as establishing a well-formed *Dual Laplacian* of the enclosed model.

2.3.2 Dual-Laplacian Surface Regularization

Regularization Energy Term. We propose to define a regularization term introducing additional information in order to solve an ill-posed problem or to prevent over-fitting. We determine the internal energy functional $E_{int}\left(\mathrm{c}'\right)$ that measures how smooth the cage is and how similar the deformed cage c' is to the original shape in term of local detail as follows:

$$E_{int}\left(\mathrm{c}'\right) = \left\| \mathcal{L}_C\left(c'\right) - \hat{\delta}' \right\|_2^2$$

This functional guarantees smoothness on large deformation in order to preserve the subspace boundary properties without rigidity assumption. Ensuring such a property leads to guarantee global characteristic of the model linearly.

2.3.3 Screen-space Surface Constraints

Barycentric Anchors. Contrary to existing frameworks where positional constraints enforce vertices to move to a specific target 3D position, we prefer to enforce surface features that are not limited to the set of enclosed mesh vertices. In other to deform the bounding cage, positional constraints are defined on the model using barycentric anchor points (Fig. 2.12). A barycentric anchor q_l on a piecewise linear surface can be evaluated and described using a linear combination of the barycentric coordinates $\{\gamma_1, \gamma_2, \gamma_3\}$ associated to three vertices $\{v_1, v_2, v_3\}$ of their surrounding triangle T_l that contains this anchor point as follows:

$$q_l = \sum_{v_i \in T_l} \gamma_i \cdot v_i$$

In the scenario where the user directly specifies source and target positions over the enclosed mesh surface in screen space, dragged-and-dropped barycentric anchors always offer suitable and precise sparse positional deformation constraints. To estimate the target point in world-space coordinates, we compute the intersection point between the ray passing through the target screen point and the parallel plane to the screen plane defined by the source point.

GPU Barycentric Map. Barycentric information are collected in a floating map. Then, this barycentric map is used to establish 2D-to-3D correspondences, directly

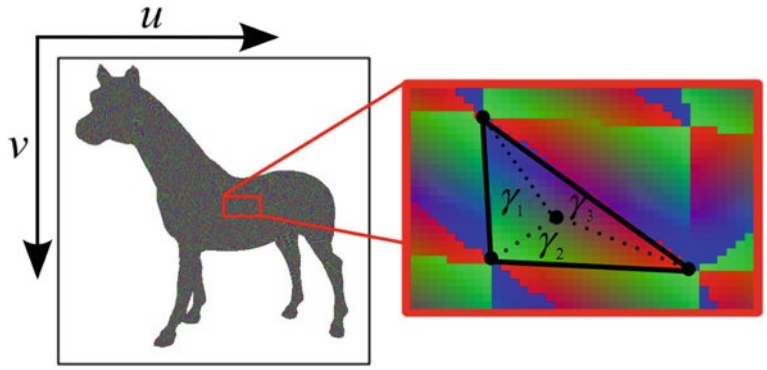

Fig. 2.12 GPU barycentric map: deformation handles are defined over the surface using barycentric coordinates expressed in image-space

by exploiting the GPU through a render-to-texture technique. More precisely, we store in a floating RGBA texture the index of the mesh face that reprojects into each screen-space pixel, and the associated barycentric coordinates corresponding to each pixel-based surface feature, according a given viewpoint (see Fig. 2.12). In other to deform the enclosed model via a space warping of the bounding cage, the user can drag-and-drop a collection of handles points over the enclosed mesh surface, directly by sketching a displacement vector on the screen for each handle. The origin point of this vector is called the source point and the final destination is called the target point. Finally, to estimate world space coordinates corresponding to the target point, we compute the intersection point between the ray passing through the target screen point and the parallel plane defined at the source point surfel. Thus, we can easily define a 2D-to-3D correspondence resulting in the association of a 3D target location coupled to a selected arbitrary point belonging to the surface space (see Fig. 2.13).

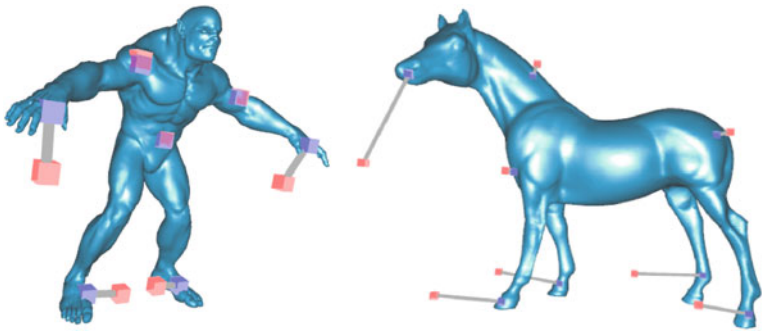

Fig. 2.13 Sparse deformation constraints: a small collection of user-specified geometric correspondences are defined in screen-space and backprojected over the surface in 3D. Such surface-based constraints serve to estimate a cage-based deformation field

Sparse Positional Constraints. Now, we provide a formulation expressing the Cartesian coordinates of a point q_l over the model only in term of the cage parameters:

$$q_l = \sum_{v_i \in T_l} \sum_{k=1}^{n} \gamma_i \cdot c_k \cdot g_k \, (i)$$

Then, we denote by q_l' the Cartesian coordinates position of the user-specified target point, associated to q_l to form a positional constraint. The last equation is a key component of the proposed method for the handling interaction. The transfer of surface constraints into the volumetric domain exploiting the cage indirection is expressed by this function. In other words, the last formulation permits to express surface constraints directly in terms of cage parameters linearly using an inverse near-conformal mapping, motivating the idea of boneless inverse kinematics.

Data Energy Term. We determine the external energy functional $E_{ext}\left(\mathrm{c}'\right)$ that measures the smoothness the cage under l positional constraints as follows:

$$E_{ext}\left(\mathrm{c}'\right) = \sum_{j=1}^{l} \left\| q'_j - q_j \right\|_2^2$$

2.3.4 Indirect Dual-Laplacian Cage-based Fitting

Least-Squares Cage. After the introduction of the key components of our method, in this section we propose to develop the core of our approach including the linear minimization process. During the minimization process, the cage is seen as a connectivity mesh and feature constraints are seen as external deformation features. Our surface-and-space based deformation technique preserves the mesh spatial coherence. The geometry of the cage can be efficiently reconstructed indirectly from their harmonic coordinates and relative coordinates by solving a system of linear equations. We cast the problem of deformation as least-squares Laplacian cage reconstruction process using a consistent minimization approach of an objective function requiring linear constraints such as the positional edited constraints. By transposing ideas presented by Sorkine et al. [60] to the cage domain, our cage parameters recover the sparse pose feature by minimizing an objective function in a least-squares sense in order to fit a continuous volume. Then, the geometry of the desired model is simply obtained by generating its position vertex according to the reconstructed cage parameters obtained on the concept of *Least-Squares Cage* (see Figs. 2.14, 2.15).

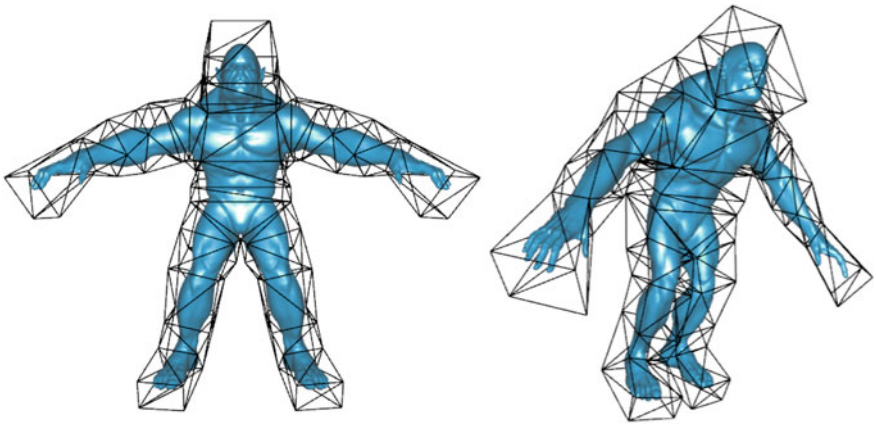

Fig. 2.14 Inverse cage-based modeling: subspace-regularized and space-based deformation under user-specified sparse constraints. A scanned model of an horse is enclosed in a coarse cage *(left)*, the user defined a set of 3D correspondences in screen-space. *(middle)*, then our cage fitting procedure retrieves the cage expressing the desired space-based deformation *(left)*. We use Harmonic Coordinates for this example

Given differential coordinates, the Laplacian operator of the default cage, the generalized barycentric weights $g_k\left(i\right)$ according the cage and the model at the default

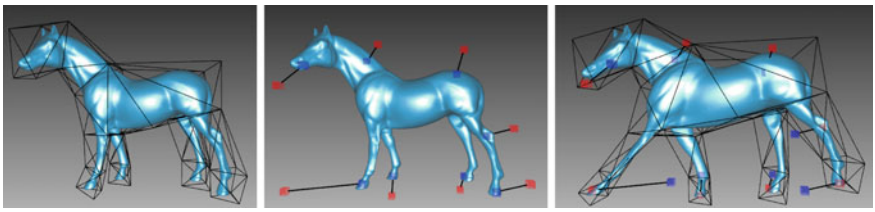

Fig. 2.15 Constraint-based scalable deformation: given a default cage-based shape (*left*), and user-specified constraints, our system retrieves a least-squares cage expressing the desired deformation in a scalable manner with shape resolution-independence (*right*)

pose, and a several 2D sparse surface constraints, the absolute coordinates of the model geometry can be reconstructed by estimating the absolute coordinates of the cage geometry. The combination of the differential coordinates and harmonic coordinates allows the reconstruction the edited surface by solving a linear system that satisfies the reconstruction of the local detail in least-squares sense.

Least-Squares Minimization. Since no exact solution generally exist, our linear least-squares system reconstructs the geometry of the coarse mesh, allowing to express the enclosed mesh using a linear cage-based deformation process. The key component of our inverse deformation algorithm is a least-squares minimization. We can reformulate the overall energy, leading to an overdetermined linear system to extract the cage parameters as follows:

$$\min_{\forall k,\ c'_k} \left(\alpha \sum_{k=1}^{n} \left\| \mathcal{L}_C \left(c'_k \right) - \hat{\delta}'_k \right\|_2^2 + \beta \sum_{j=1}^{l} \left\| q'_j - \sum_{v_i \in T^j} \sum_{k=1}^{n} \gamma_i \cdot c'_k \cdot g_k \left(i \right) \right\|_2^2 \right)$$

We denote by $\|\cdot\|_2$ the L_2 norm. Note that the first term of the energy (called the *smoothness term*) preserves global details of the cage and ensure a pleasant deformation under sparse constraints. The second term (called the *data term*) of the energy enforces the position of vertices to fit the desired model defined by positional constraints. Independent linear positional constraints are easily represented by separate equations for each three dimensions. The system can be weighted by both scalar $\alpha > 0$ and $\beta > 0$ to penalize or advantage each term with respect to the other one. In our experiments, we use $\alpha = \beta = 1$. Then, this linear least-squares system can be rewritten in the following matrix form:

$$\|\mathbf{A} \cdot \mathbf{X} - \mathbf{B}\|_2^2 = (\mathbf{A} \cdot \mathbf{X} - \mathbf{B})^T (\mathbf{A} \cdot \mathbf{X} - \mathbf{B}) = \mathbf{X}^T \mathbf{A}^T \mathbf{A} \cdot \mathbf{X} - 2\mathbf{B}^T \mathbf{A} \cdot \mathbf{X} + \mathbf{B}^T \mathbf{B}$$

where $\mathbf{A}$ stacks the cage-Laplacian, $\mathbf{B}$ stores the differential coordinates and target locations and $\mathbf{X}$ is the unknown location of cage vertices. To formulate a closed-form solution, we take partial derivatives with respect to the unknown vector and set them to zero. We then obtain the normal equation as follows:

$$2\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{X} - 2\mathbf{A}^T \cdot \mathbf{B} = 0$$

that we simplify in that manner:

$$\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{X} = \mathbf{A}^T \cdot \mathbf{B}$$

Our desired solution $\mathbf{X}$ can be written:

$$\mathbf{X} = \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T \cdot \mathbf{B}$$

Desired Model Reconstruction. To our best knowledge, the simple global optimization component of our framework does not already exist in the literature with such constraints to minimize. The overall energy-based warping performed by our technique recovers space deformation during the editing phase. After the cage retrieval process, the geometry of the desired enclosed model is reconstructed with linear combination function of cage geometry parameters related to the new estimation, preserving the fix connectivity of the cage. The system easily represented as a separate least squares system for each dimensions (x, y, z). The availability of iterative sparse linear solvers or direct method for solving the normal equation makes the computational kernel of our retrieval algorithm very efficient and simple to implement [61].

2.3.5 Experimental Results and Evaluation

Cage-based Inverse Modeling. This section describes our experiments using the proposed *Cage-based Inverse Modeling* system. Built upon the well-known least-squares formulation, our framework proposes a numerically stable and interactive mechanism to estimate new cage configuration. Our effective framework could serve for various geometric applications. Thus, we demonstrate the feasibility, stability and validity in practice with an experimental application as illustrated in Fig. 2.16. The cage is encoded with a winged-mesh structure to access faster to the one-ring-neighborhood of cage-handles and to efficiently compute the Laplacian coordinates. In the context of user-driven approach, we apply our algorithm by specifying sparse screen-space positional constraints over the enclosed surface. The indirect cage estimation improves the computation of the modeling thanks to the small system size evolving the cage indirection. The solving process of examples shown in Fig. 2.16 involving dense polygon meshes was generated under 100 μs. Alternatively, even if we think that better control of the object is possible if the cage is subdivided, cage coarsity must remain a requirement to pretend being *cage-based*. Finally, the main computational cost of our method is dedicated to the cage-model rigging and the solving process for the objective functional using a direct linear least-squares system. The running time is therefore most strongly dependent on the number of cage vertices.

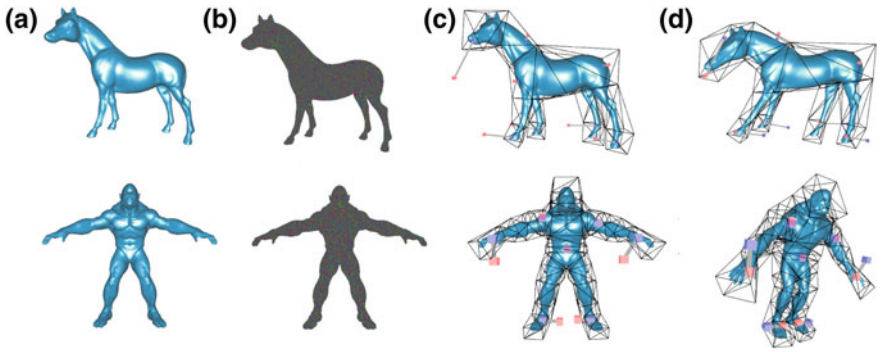

Fig. 2.16 Cage-based inverse modeling: the default enclosed model (**a**), its corresponding GPU barycentric map (**b**), user-specified constraints (**c**), the resulting cage-based deformation (**d**)

Sketching Positional Constraints. Additionally, we have developed an intuitive user interface that allows the user to model a new sophisticated model by sketching sparse positional constraints in the screen-space over the enclosed surface. This interaction technique provides a natural way to interact with basic input deformation features. Then, the indirect cage estimation improves the computation of the modeling because of the small system size involving the cage indirection. Our barycentric anchor system and surface-to-cage constraints transfer make our technique potentially interoperable with the emergent multi-touch surface interaction technology.

Benefits. Since global modeling operations naturally require high-level surface representations, our method improves the effectiveness of deformable surface by offering the possibility of encoding global shape deformation with respect of local influence. Then, our framework allows artists to re-use the estimated parameters during the modeling process with high flexibility. By solving the presented global linear least-squares system, this constraint-based deformation provides an efficient and robust computational framework. Moreover, it seems easy to integrate other type of linear deforming constraints. The linearity of the proposed objective functional makes the processing highly efficient. Nevertheless, our framework is not limited to *Mean Value Coordinates* or *Harmonic Coordinates* and could be adapted with other type of linear coordinate systems, such as *Biharmonic Coordinates*. Strong interest for cage-based deformation invites us to improve our framework in the future. We remark that the cage topology and geometry encapsulate a soft deformation resistance that allows us to produce a natural deformation for large deformation. In addition, the global explicit parameterization offers multiresolution editing.

Effectiveness. There are some limitations with our method that are theoretically derived from Laplacian and cage-based properties. The main drawback of our proposed technique arises from combining linear cage-based coordinates with Laplacian regularization. We remark that it is hard to obtain volume preserving with generalized barycentric coordinates. The major difficulty for these methods is to find efficient mappings which do not induce large distortion such as volume shrinking to the model

without enforcing additional constraints. Unfortunately, our system is not capable of ensuring isometry over the surface. Accordingly, volume constraints should be imposed to reduce such undesirable effects. Another notable limitation is the cage design. It is exceedingly tedious and time-consuming to extract automatically a reasonable default coarse bounding cage. An arbitrary cage may fail to express every desired pose correctly according to a given binding process.

2.3.6 A Mathematical Study

Cage-based Linear Algebra. In our technique, we do not apply the Laplacian directly on the enclosed mesh itself in order to reduce the computation cost and memory space, but more importantly to turn the shape modeling process into a scalable task. Notwithstanding, the deformation seems to preserve the local properties in the same visual fashion than if the *Laplacian Operator* will perform was defined over the enclosed model. Notwithstanding, we would like to develop a mathematical study to demonstrate that applying the Laplacian Operator on the cage structures is not mathematically equivalent (even with approximated derivation) to applying the *Laplacian Operator* on the enclosed model. In complement to this demonstration, we also propose some additional mathematical assertions concerning the cage-based linear algebra.

Subspace Decomposition. A vertex $p \in \mathbb{R}^3$ of the enclosed model can be decomposed into a cage-based subspace linear basis (only with near-conformal mapping properties), according to its associated to a set of rigging functions $h_k(p) : \Omega \mapsto \mathbb{R}^+$ (the rigging function can be defined using *Mean Value Coordinates* or *Harmonic Coordinates*) as follows:

$$p \approx \sum_{k=1}^{n} c_k \cdot h_k(p)$$

The well-known non-exact conformality of the cage-based rigging in three dimensional leads to an approximated projection of a given 3D point expressed in global euclidean into the cage-based subspace basis defined as the vector of n cage-handles $\{c_1, \cdots, c_n\} \in \mathbb{R}^{3\times n}$.

Subspace Extrapolation. As shown by Rustamov [47], it is possible to define the extrapolation of cage-based coordinates (only demonstrated for *Harmonic Coordinates*). For any default point p and its deformed corresponding point p', we can write:

$$p' - p = \sum_{k=1}^{n} h_k(p) \cdot c'_k - \sum_{k=1}^{n} h_k(p) \cdot c_k = \sum_{k=1}^{n} h_k(p) \cdot \left(c'_k - c_k\right)$$

Blending into the Subspace. Given $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}^+$ two factors and two points $v_a \in \mathbb{R}^3$ and $v_b \in \mathbb{R}^3$, we can write:

$$(\alpha \cdot v_a + \beta \cdot v_b) = \left(\alpha \cdot \sum_{k=1}^{n} c_k \cdot h_k(a) + \beta \cdot \sum_{k=1}^{n} c_k \cdot h_k(b) \right)$$

The last formulation leads to the following expressions:

$$\Leftrightarrow (\alpha \cdot v_a + \beta \cdot v_b) = \left(\sum_{k=1}^{n} \alpha \cdot c_k \cdot h_k(a) + \sum_{k=1}^{n} \beta \cdot c_k \cdot h_k(b) \right)$$

$$\Leftrightarrow (\alpha \cdot v_a + \beta \cdot v_b) = \sum_{k=1}^{n} c_k \cdot (\alpha \cdot h_k(a) + \beta \cdot h_k(b))$$

Enclosed Mesh Laplacian into the Subspace. Now, let's write the Laplacian of the enclosed mesh in term of the cage parameters. If we denote by $v_i \in \mathbb{R}^3$ and $v_j \in \mathbb{R}^3$ two different enclosed mesh vertices, $N(i)$ the one-ring neighborhood of v_i, and given the fact the *Laplacian Operator* of the enclosed mesh is written as follows:

$$\mathcal{L}(v_i) = v_i - \frac{1}{d_i} \sum_{j \in N(i)} v_j$$

where d_i is the number of direct vertices neighbors $N(i)$ for v_i. Then, we can rewrite this operator exclusively in term of the cage parameters, as follows:

$$\Leftrightarrow \mathcal{L}(v_i) = \left(\sum_{k=1}^{n} h_k(i) \cdot c_k \right) - \left(\frac{1}{d_i} \sum_{j \in N(i)} \left(\sum_{k=1}^{n} h_k(j) \cdot c_k \right) \right)$$

In addition, we can also rewrite this closed-form expression:

$$\Leftrightarrow \mathcal{L}(v_i) = \left(\sum_{k=1}^{n} h_k(i) \cdot c_k \right) - \left(\sum_{j \in N(i)} \left(\sum_{k=1}^{n} \frac{1}{d_i} \cdot h_k(j) \cdot c_k \right) \right)$$

Using previous assertions, we can simplify the last expression as follows:

$$\mathcal{L}(v_i) = v_i - \frac{1}{d_i} \sum_{j \in N(i)} v_j = \sum_{k=1}^{n} c_k \cdot \left(h_k(i) - \frac{1}{d_i} \sum_{j \in N(i)} h_k(j) \right)$$

Then, we can conclude that the cage-Laplacian has a multi-ring influence over the enclosed shape, and not exactly *model-Laplacian* equivalent because of its strong

dependence to cage-handles and associated rigging function. In the general case:

$$\mathcal{L}(.) \neq \mathcal{L}_C(.)$$

The resulting expression projects the mesh-Laplacian into the cage-based subspace with a minimal number of separable terms. This demonstration clearly shows that ensuring the Laplacian of the cage is not mathematically equivalent to ensure the Laplacian on the enclosed mesh.

Constraints Projected into the Subspace. Now, let's write the enclosed mesh-Laplacian term only in term of cage parameters.

$$\arg\min_{\forall i,\ v_i} \left(\sum_{i=0}^{m} \| \mathcal{L}(v_i) - \delta_i \|_2^2 \right) \Rightarrow$$

$$\arg\min_{\forall k,\ c_k} \left(\sum_{i=0}^{m} \left\| \sum_{k=1}^{n} c_k \cdot \left(h_k(i) - \frac{1}{d_i} \sum_{j \in N(i)} h_k(j) \right) - \delta_i \right\|_2^2 \right)$$

2.4 Beyond the Cage: A Generalization to Skeletal Structures

Generalized Laplacian Subspace. In this section, we show that our *Laplacian Subspace* technique can be generalized to other types of underlying animatable subspaces that are not purely geometric and two-manifold, such as a skeletal structure. By generalizing our spatial relationship principle on skeletal joint-handles, we show that the *Laplacian Regularization* canmodify the rigidness of an input rigid skeleton. We also demonstrate a useful interactive application for subspace reediting.

Skeletal Structures. Then, we describe a new and simple approach to edit traditional skeleton of animation with joint-based Laplacian-type regularization, in the context of exaggerated skeleton-based character animation. Despite decades of research, interactive character animations still have a lack of flexibility and editability in order to re-use real vertebral motion. A major problem for artists in production is to enhance the expressiveness of classical motion clips by direct manipulation of the underlying skeletal structure. Relatively small number of researchers present an approach for processing non-rigid effects on motion data in [62–64]. However existing techniques often avoid dealing with the potential of skeletal-based optimization, while preserving the joint coherence and connectivity. Besides, range of living creature in real-world have the flexibility to stretch to extreme positions and squash to astounding shapes. It can also be noticed that *squash-and-stretch* effects are easier to realize in traditional animation rather than mocap-based computer generated animation. For this reason, *Ratatouille* a Pixar® movie, did not use a rigid skeleton, while abandoning motion capture to reach such essential non-ultra realistic appeal.

2.4.1 Differential Skeletal Editing

Differential Joint Coordinates. The idea of the motion capture is to use sensor placed on the subject, and to collect data that describe their motion while they are performing motions. The pose of an articulated figure can be specified by its joint configuration in addition to the position and root orientation segment. Skeleton of animation $\mathcal{S} = (\mathcal{J}, \mathcal{B}, \mathcal{M})$ is composed of a hierarchy of joints enriched with motion data, organized in 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 ith joint at the frame t is denoted by $\mathbf{M}_i^t \in \mathbb{R}^{4\times4}$. The global location $\mathbf{p_i^t}$ of the ith joint at the frame t is then the homogeneous zero transformed by the sequence of transformation and can be written as: $\mathbf{p_i^t} = \mathbf{M_i^t} \cdot \left[\mathbf{0} \mid 1\right]^T$. We define the *Differential Joint Coordinates* δ_i^t of joint i at frame t as follows.

$$\delta_i^t = \mathbf{p_i^t} - \sum_{j \in N(i)} \frac{1}{d_i} \left(\left(\mathbf{M}_i^t - \mathbf{M}_j^t \right) \cdot \left[\mathbf{0} \mid 1\right]^T \right)$$

Skeletal Graph Laplacian. The degree of the ith joint denoted by d_i is equal to the number of joints linked to 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 as an open directed acyclic graph. So, we introduce the *Skeletal Graph Laplacian* L_S by $L_S = D - A$ where D is the diagonal matrix of joint degrees and A is the adjacency joint matrix. We denote by $\mathcal{L}_\mathcal{S}(.)$ the per-joint *Uniform Laplacian Operator* applied on the skeleton graph structure. The entries of the corresponding square symmetric $n \times n$ matrix L_S are setup 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 original 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.

2.4.2 Rigidity-Breaking Skeletal Optimization

Differential-Aware Stretching. Our algorithm takes an arbitrary input articulated motion signal and perturbates it in other to make it more stretch-like. In order to hack the rigidity, we prefer to deal with the skeletal structure as Euclidean joint coordinates.

Our technique has single key component: the global joint location optimization. At the beginning of pre-optimization, we reformulate the *Squash-and-Stretch* problem as *skeletal adaptation optimization*. Given the fact it is nearly impossible to get a plausible 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 location of a pre-animated pose satisfying stretching constraints. The core algorithm of our technique relies on the *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 driving idea is to employ fast and accurate skeleton fitting process, guided by kinematic-free constraints.

Least-Squares Skeleton. At each frame, the initial solution is the Euclidean parameters of the animated pose, generated by forward kinematics. Thus, *pose estimation* recovers the pose by minimizing an objective function that is a combination of penalty and data terms. Smoothness term is required for making the warp field regular. To control a desired pose, the user inputs the targeted global joint positions $\mathbf{q_l^t}$ for a collection $\mathcal{C}$ of edited joints. We conserve the Laplacian coordinates δ_i^t for each joint i in the skeleton hierarchy. The reconstructed positions $\widehat{\mathbf{p}}$ of the skeleton joints in world space coordinates are obtained by solving the following quadratic minimization problem separately from each Euclidean dimension:

$$\underset{\widehat{\mathbf{p}}^{\mathbf{t}}}{\operatorname{argmin}} \left(\sum_{l \in \mathcal{C}} \left\| \mathbf{q_l^t} - \widehat{\mathbf{p}}_{\mathbf{l}}^{\mathbf{t}} \right\|^2 + \sum_{i=1}^{n} \left\| \mathcal{L}_{\mathcal{S}} \left(\widehat{\mathbf{p}}_{\mathbf{i}}^{\mathbf{t}} \right) - \delta_i^t \right\|^2 \right)$$

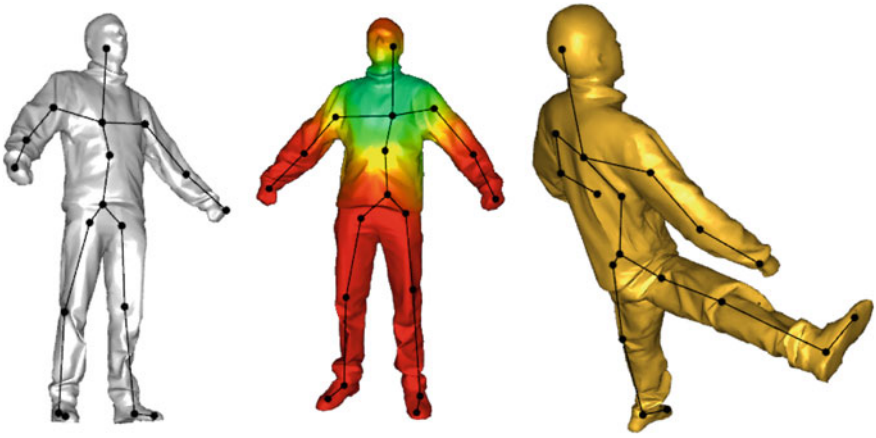

Fig. 2.17 Laplacian skeletal editing: optimizing traditional skeletal animated structure with differential-aware bone stretching effects allows animators to efficiently re-use previously captured MoCap data in the context of animation productions. We increase Squash-and-Stretch effects on existing realistic Basketball Dunk, Dancing, Horse Jump, Karate Kick-Off and motion clips (from *left* to *right* hand side). Original and intermediate edited poses are displayed in transparency and resulting edited poses in superposition

Automatic Stretching Constraints. Global joint location is estimated by minimizing the sum of squared difference between the data-driven pre-animated pose and input features cues. In order to avoid purely translation effects provoked by a single dragged-and-dropped joint, we fix by default the root joint, acting as skeleton gravity center. The root joint is generally located in a place where many bones come together. In our framework, bone elongation at frame t can also be automatically established along the bone direction for a selected joint l having the kth joint as parent with a time-dependent scaling factor α_t:

$$\mathbf{q}_\mathbf{l}^\mathbf{t} = \mathbf{p}_\mathbf{l}^\mathbf{t} \pm \alpha_t \cdot \frac{\left(\mathbf{p}_\mathbf{l}^\mathbf{t} - \mathbf{p}_\mathbf{k}^\mathbf{t}\right)}{\left\|\mathbf{p}_\mathbf{l}^\mathbf{t} - \mathbf{p}_\mathbf{k}^\mathbf{t}\right\|}$$

Poisson Skeletal Solver. Energy terms can be equilibrated by a weighting system in future work. Using such refined constraint formulation in conjunction with the Laplacian formulation on the bone structure is motivated by the fact that *Squash-and-Stretch* can be accomplished by differential scaling in Euclidean coordinates system. Hence, this minimization problem can be solved efficiently in least-squares sense based on the succeeding expression:

$$\mathbf{A} \cdot \mathbf{X}' = \mathbf{B}$$

where the matrix $\mathbf{A}$ stacks the skeletal-Laplacian and the matrix $\mathbf{B}$ stores the differential joint coordinates. This sparse editing technique performs spatio-temporal numerical optimization on MoCap data. Moreover, the global location of joints stored in the column matrix $\mathbf{X}'$ can be therefore found by solving at real-time rate the following small linear system that does not need to:

$$\mathbf{X}' = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{B}$$

2.4.3 Experimental Results

Laplacian Skeletal Subspace. To demonstrate the stability and the usefulness of our *Laplacian Skeletal Subspace* techniques, we have implemented an advanced and reproducible pipeline using only C++ and OpenGL. In all of our experiment, just a single editing feature is sufficient to adapt the whole skeletal pose in coherent manner, as shown in Fig. 2.18. Moreover, results have shown that the proposed technique is stable numerically enough to probably be used in low-budget productions, especially for processing sporty motion clips, as illustrated in Fig. 2.17. We have tested the robustness of our method on a corpus of motion clips with success. Few examples are shown in Fig. 2.18. In addition, we developed a simple interaction technique coupled with depth first search to handle the closest joint.

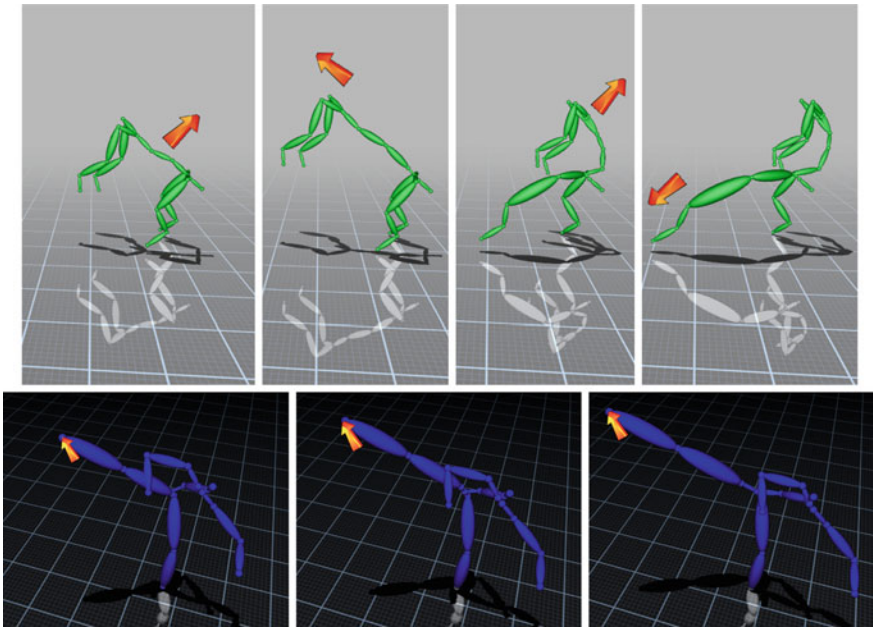

Fig. 2.18 Interactive skeletal editing: a single editing feature is sufficient to produce pleasant non-rigid gait style over the whole skeletal pose in coherent manner for a horse jump (*top row*) and a karate kick-off (*bottom row*)

Skeletal Shape Control. Even if human perception system tends to focus on kinematic parameters rather than on local features cues, real-life creatures actually elongate and stretch due to the elasticity of tendons and muscles. Consequently, breaking the rigidity of the underlying armature adds a pleasant realism with rubber-like effects. The resulting animated structure looks more pleasant while remains rigid in motion. Our technique could be potentially used in commercial 3D animation systems

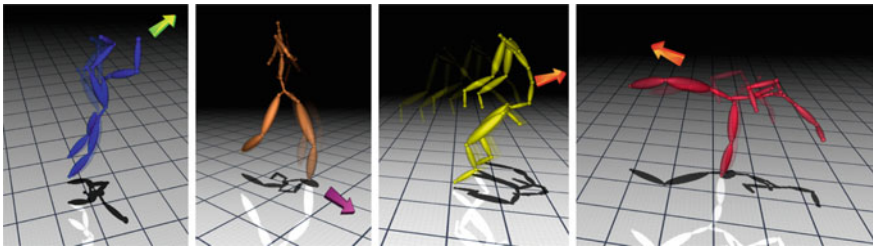

Fig. 2.19 Skeletal shape editing: our least-squares Skeleton can be used to drive the deformation of a static template mesh (*left*) having the same default topology characteristics. According to a given rigging function (*middle*), the deformed shape is obtained by a modified skeletal-based skinning operator (*right*)

to easily achieve non-rigid, rigged-and-skinned animation, skeletal shape control and elastic skin deformation from MoCap data. Our method yet offers the potential ability to control stretchable effect while allowing smooth behavior of skeletally-guided skin deformation, as far as the rigging function is smooth (see Fig. 2.19).

2.5 Discussion and Conclusions

Low-dimensional Parametrizations. In this chapter, we have presented a concise overview of various low-dimensional parametrizations useful to process sparse deformation constraints on a given static surface. Next, we have introduced an inverse modeling framework based on the estimation of a surfacically-constrained space-deformation mapping through the prism of numerical optimization. To the best of our knowledge, our well-designed technique does not previously exist in the literature. Then, we exhibit an external parameterization involving a reduced number of parameters suitable for maintaining one-to-one correspondence during the deformation phase. The dimension reduction offered by the cage connectivity is a strength for surface encoding and for further network transmission of deformation parameters.

CageIK. In the first contribution, a mix of generalized barycentric coordinates and Laplacian coordinates are injected inside a linear minimization framework to reconstruct a desired global pose from a default enclosed mesh. This *Dual Laplacian Cage-based Mesh Editing* technique allows users to produce visually pleasing deformations. Furthermore, the linearity of the core objective functional makes the processing highly efficient and improves the effectiveness of deformable surface computation. Obviously, cage connectivity and its location encode non-trivial soft kinematic constraints, because the cage topology drastically restricts the deformation under constraints. This could be particularly attractive for producing pseudo-kinematic space deformation over non-rigid surfaces without explicit skeletal constraints to be solved numerically.

Our method offers the possibility to encode global deformation of the shape with respect of local shape influences. Furthermore, we demonstrate the strength of near-conformal mapping and differential geometry to reconstruct an enclosed mesh by estimating its corresponding and desired cage locations. In addition, this cage-based approach is a turn-key tool for artist in order to generate, reuse, modify the resulting set of parameters, interact with the resulting data easily or potential deformation transfer. That is worth mentioning that our framework is not mathematically restricted to a specific coordinate system as-far-as the selected cage-based system implies linear constraints. The algorithm is effective, conceptually simple and comprehensible enough to be widely implemented and tested with previous mesh-based frameworks.

Laplacian Skeletal Editing. We demonstrate the strength of Laplacian and differential editing on underlying animatable structures not limited to bounding cages. So, we have also introduced a novel, accurate, robust, easy to use, skeletal editing and interactive techniques allowing unskilled animators to manipulate arbitrary skeletal

poses with real-time stretchable control. Breaking the physics of articulated models allows us to obtain more realistic to stretchable-like effect, independently of the skin layer. Spatial coherence of joints is preserved by differential-aware scaling represented by a quadratic energy functional leading to a new *Skeletal Poisson Solver*. The usefulness and flexibility of our Laplacian approach is fully demonstrated to emphasize effective high-quality skeletal-based poses. In the context of visual media productions, our pipeline will probably reduce the amount of time needed to obtain controllable realistic deformations for expressive characters. More importantly, this contribution makes complex interactions with fully connected skeletons more accessible to non-professionals by relying on interactive editing.

Limitations, Benefits and Future Work. Nevertheless, a strong limitation appears in the cage design in itself. In fact, it is exceedingly tedious to define a default cage connectivity that could lead to a perfect inversion process. The main benefit of our method is that the minimization framework is fully independent to the model resolution. Clearly, the influence of the input cage connectivity is not easy to predict for the rest of the editing phase. Intuitively, the default cage encodes non-trivial and implicit soft kinematic constraints. This means that the cage-handles configuration at the default pose already affects the reconstructed enclosed. Contrary to techniques that only solve differential deformation over the original mesh domain, our proposed technique realizes scalable mesh editing on large-scale models or laser-scanned model by leveraging the memory bottleneck. According to the high-demand for methods based on the generalization of barycentric and Laplacian coordinates, we believe these novel approaches will offer promising futures directions. This is also supported by the strong interest in hybrid mesh deformation, boneless inverse kinematics and also geometric skeletons. Even if our current system is yet effective, there are also numerous ways in which it could be improved. For example, a potential future work will be the incorporation of other suitable linearized constraints.

References

1. T. Mcinerney, D. Terzopoulos, Deformable models in medical image analysis: a survey. Med. Image Anal. **1**, 91–108 (1996)
2. D. Terzopoulos, J. Platt, A. Barr, K. Fleischer, Elastically deformable models, in SIGGRAPH '87, 1987, pp. 205–214.
3. J. Barbič, M. Silva, J. Popović, Deformable object animation using reduced optimal control, in SIGGRAPH '09, 2009.
4. B. Adams, M. Ovsjanikov, M. Wand, H-P. Seidel, L. J. Guibas, Meshless modeling of deformable shapes and their motion, in SCA '08, 2008, pp. 77–86.
5. U. Güdükbay, B. Ozguc, Animating deformable models: different approaches, in Proceedings of the Computer, Animation, 1995.
6. M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models. Int. J. Comput. Vision **1**(4), 321–331 (1988)
7. R. W. Sumner, J. Schmid, and M. Pauly. Embedded deformation for shape manipulation. ACM Trans. Graph. 26(3), 80:1–80:7 (2007).

8. M. Wardetzky, S. Mathur, F. Kälberer, E. Grinspun, Discrete laplace operators: no free lunch, in SGP'07: Proceedings of the fifth Eurographics Symposium on Geometry Processing, 2007, pp. 33–37.
9. Y. Lipman, O. Sorkine, D.C. Or, D. Levin, C. Rössl, H-P. Seidel, Differential coordinates for interactive mesh editing, in Proceedings of International Shape Modeling **2004**, 181–190 (2004)
10. O. Sorkine, Differential representations for mesh processing. Comput. Graph. Forum 25, 789–807 (2006).
11. Y. Lipman, O. Sorkine, M. Alexa, D.C. Or, D. Levin, C. Rössl, H-P. Seidel, Laplacian framework for interactive mesh editing. Int. J. Shape Model. textbf11(1), 43–61 (2005).
12. A. Nealen, T. Igarashi, O. Sorkine, M. Alexa, *FiberMesh: designing freeform surfaces with 3D curves, in Proceedings of ACM SIGGRAPH* (ACM Trans, Graph, 2007)
13. R. Gal, O. Sorkine, N.J. Mitra, D.C. Or, iwires: an analyze-and-edit approach to shape manipulation. ACM Trans. Graph. Siggraph **28**(3), 1–10 (2009)
14. O. Sorkine and M. Alexa. As-rigid-as-possible surface modeling, in Symposium on Geometry Processing, 2007, pp. 109–116.
15. O.K-C. Au, C-L. Tai, L. Liu, H. Fu, Dual laplacian editing for meshes. IEEE Trans. Visual Comput. Graphics textbf12(3), 386–395 (2006).
16. Q. Luo, B. Liu, Z. Ma, H. Zhang, Mesh editing in roi with dual laplacian, in CGIV '07: Proceedings of the Computer Graphics, Imaging and Visualisation, 2007, pp. 195–199.
17. N. Thalmann, R. Laperrière, D. Thalmann, Joint-dependent local deformations for hand animation and object grasping, in Proceedings on Graphics, Interface '88, 1988, pp. 26–33.
18. I. Baran, J. Popović, Automatic rigging and animation of 3d characters, in ACM SIGGRAPH 2007 papers, 2007.
19. C. Miller, O. Arikan, D. Fussell, Frankenrigs: building character rigs from multiple sources, in Proceedings of I3D'10, 2010, pp. 31–38.
20. O. Weber, O.Sorkine, Y. Lipman, C. Gotsman, Context-aware skeletal shape deformation, in Proceedings of Eurographics, Compute. Graph. Forum (2007).
21. A. Jacobson, O. Sorkine, Stretchable and twistable bones for skeletal shape deformation, in Proceedings of ACM SIGGRAPH ASIA, ACM Trans. Graph. 30(6), (2011).
22. L. Kavan, S. Collins, Jiří Žára, C. O'Sullivan, Geometric skinning with approximate dual quaternion blending. ACM Trans. Graph. **27**(4), 105 (2008)
23. A.M. Adams, Y. Zhu, A. Selle, M. Empey, R. Tamstorf, J. Teran, E. Sifakis, Efficient elasticity for character skinning with contact and collisions. ACM Trans. Graph. 30(4), (2011).
24. T.W. Sederberg, S.R. Parry, Free-form deformation of solid geometric models, in SIGGRAPH '86, 1986, pp. 151–160.
25. S. Coquillart, P. Jancéne, *Animated free-form deformation: an interactive animation technique, in SIGGRAPH, 1991* (Comput, Graph, 1991)
26. R. MacCracken K.I. Joy, Free-form deformations with lattices of arbitrary topology, in SIGGRAPH '96, 1996, pp. 181–188.
27. K.G. Kobayashi, K. Ootsubo, t-ffd: free-form deformation by using triangular mesh, in Proceedings of the eighth ACM Symposium on Solid Modeling and Applications, 2003, pp. 226–234.
28. R.W. Sumner, M. Zwicker, C. Gotsman, J. Popović, Mesh-based inverse kinematics. ACM Trans. Graph. **24**(3), 488–495 (2005)
29. J. Huang, X. Shi, X. Liu, K. Zhou, L-Y. Wei, S-H. Teng, H. Bao, B. Guo, H-Y. Shum, Subspace gradient domain mesh deformation. ACM Trans. Graph. **25**(3), 1126–1134 (2006)
30. K. Zhou, J. Huang, J. Snyder, X. Liu, H. Bao, B. Guo, H-Y. Shum, Large mesh deformation using the volumetric graph laplacian. ACM Trans. Graph. **24**, 496–503 (2005)
31. K.G. Der, R.W. Sumner, J. Popović, Inverse kinematics for reduced deformable models, in SIGGRAPH '06: ACM SIGGRAPH. Papers **2006**, 1174–1179 (2006)
32. T. Ju, S. Schaefer, J. Warren, Mean value coordinates for closed triangular meshes. ACM Trans. Graph. **24**(3), 561–566 (2005)
33. C. Xian, H. Lin, S. Gao, Automatic generation of coarse bounding cages from dense meshes, in Shape Modeling, International, 2009, 2009.

34. M. B-Chen, O. Weber, C. Gotsman, *Variational harmonic maps for space deformation* (ACM Trans, Graph, 2009)
35. Z.-J. Deng, X.-N. Luo, X-P. Miao. Automatic cage building with quadric error metrics. J. Comput. Sci. Technol. **26**(3), 538–547 (2011)
36. C. Xian, H. Lin, S. Gao, Automatic cage generation by improved obbs for mesh deformation. Vis. Comput. **28**(1), 21–33 (2012)
37. T. Ju, Q.-Y. Zhou, M. Panne, D.C. Or, U. Neumann, Reusable skinning templates using cage-based deformations. ACM Trans. Graph. **27**(4), 1–23 (2008)
38. J. Kosinka, M. Barton, Injective shape deformations using cube-like cages. Comput. Aided. Des. Appl. **7**(3), 309–318 (2010)
39. Y. Zhao, X-G. Liu, Q-S. Peng, H-J. Bao, Rigidity constraints for large mesh deformation. J. Comput. Sci. Technol. 24, (2009).
40. M. Meyer, H. Lee, A. Barr, M. Desbrun, Generalized barycentric coordinates on irregular polygons. J. Graph. Tools. **7**(1), 13–22 (2002)
41. P. Joshi, M. Meyer, T.D. Rose, B. Green, T. Sanocki, Harmonic coordinates for character articulation. ACM Trans. Graph. 26(3) (2007).
42. Y. Lipman, D. Levin, D.C. Or, Green coordinates. ACM Trans. Graph. 27(3) (2008).
43. Y. Lipman, J. Kopf, D. C. Or, D. Levin, Gpu-assisted positive mean value coordinates for mesh deformations, in SGP, 2007, pp. 117–123.
44. X.-Y. Li, S.-M. Hu, Poisson coordinates. IEEE Trans. Visual Comput. Graphics **19**(2), 344–352 (2012)
45. T. Langer, H.-P. Seidel, Higher order barycentric coordinates. Comput. Graph. Forum **27**(2), 459–466 (2008)
46. T.D. Rose, M. Meyer, *Harmonic coordinates* (Pixar Technical Memo, Pixar Animation Studios, 2006)
47. M.R. Rustamov, Boundary element formulation of harmonic coordinates, Technical report, (2008).
48. C-A. Lin, E. Miller, G.S. Lee, ibind: smooth indirect binding using segmented thin-layers, in SIGGRAPH Talks, 2009.
49. A. Jacobson, I. Baran, J. Popović, O. Sorkine, Bounded biharmonic weights for real-time deformation. ACM Trans. Graph. 30 (2011).
50. O. Weber, C. Gotsman, Controllable conformal maps for shape deformation and interpolation, in ACM SIGGRAPH 2010 papers, SIGGRAPH '10 vol 29 (2010).
51. O. Weber, R. Poranne, C. Gotsman, Biharmonic coordinates. Comput. Graph. Forum. **31**(8), 2409–2422 (2012)
52. E. Landreneau, S. Schaefer, Poisson-based weight reduction of animated meshes. Comput. Graph. Forum. **29**, 1945–1954 (2010)
53. A. Jacobson, T. Weinkauf, O. Sorkine, Smooth shape-aware functions with controlled extrema. Proceedings of Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. Comput. Graph. Forum. **31**(5), 1577–1586 (2012)
54. Z. Li, D. Levin, Z. Deng, D. Liu, X. Luo, Cage-free local deformations using green coordinates. Vis. Comput. **26**, 1027–1036 (2010)
55. M.B. Chen, O. Weber, C. Gotsman. Spatial deformation transfer, in Proceedings of the SCA'09, 2009 pp. 67–74.
56. L. Chen, J. Huang, H. Sun, H. Bao, Cage-based deformation transfer. Comput. Graph. **34**, 107–118 (2010)
57. D.C. Or, Space deformations, surface deformations and the opportunities in-between. J. Comput. Sci. Technol. **24**(1), 2–5 (2009)
58. S. Zhang, P. Borosan, R. Howard, A. Nealen. Hybrid mesh editing, in Proceedings of Eurographics 2010 (short papers), Sweden, (2010).
59. M. Desbrun, M. Meyer, P. Schröder, A.H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow, in SIGGRAPH, 1999 pp. 317–324.
60. O. Sorkine, D.C. Or, Least-squares meshes, in Shape Modeling. International **2004**, 191–199 (2004)

61. M. Botsch, D. Bommes, L. Kobbelt, Efficient linear system solvers for mesh processing, in IMA Conference on the Mathematics of Surfaces, 2005 pp. 62–83.
62. J. Kwon, I.-K. Lee, Rubber-like exaggeration for character animation, in Proceedings of the 15th Pacific Conference on Computer Graphics and Applications, 2007 pp. 18–26.
63. J.W. Davis, V.S. Kannappan, Expressive features for movement exaggeration, in SIGGRAPH '02, 2002 pp. 182–182.
64. C. Bregler, L. Loeb, E. Chuang, H. Deshpande, Turning to the masters: motion capturing cartoons, in SIGGRAPH '02, 2002 pp 399–407.

Chapter 3
Reusing Performance Capture Data

This chapter describes a novel approach, where *Laplacian Cages* enhanced by harmonic weights are used for reparametrization of non-rigid dynamic surface, already acquired by multiple view techniques. We finally provide a series of results with a video-based evaluation to measure the accuracy our *Cage-based Animation Reparametrization* technique applied on captured mesh sequences. Finally, we propose an *Animation Cartoonization* algorithm dealing with the exaggeration of dynamic surface in the cage-space with additional video-based appearance stylization.

3.1 Introduction

Dense Surface Capture. Modeling dynamic 3D scene across time from multiple calibrated views is a challenging problem that has gained full attention of the *Computer Vision* and *Computer Graphics* communities in recent years. Nevertheless, the relentless increase in demand of 3DTV industry has inspired researchers to exhibit new approaches able to produce reusable contents from massive data. Current pipelines achieve temporal mesh deformation using one or many linear or non-linear numerical optimizations. Video-based approaches are proven to be convenient to acquire human performances. The field of targeted applications is very large, ranging from content generation for the 3D entertainment and motion picture industries, video game to sport analysis. Even if the human motion can be abstracted by the kinematic of rigid parts, the observed surface to track is purely non-rigid because of small detail variations induced by clothes, hair motion and certain degrees of flexibility by virtue of organic motions. Full-body performance capture is a promising emerging technology that has been intensively studied in *Computer Graphics* and *Computer Vision* over the last decade. Highly-detailed performance animations are easier to obtain using existing multiple views platforms, markerless capture and 3D laser scanner.

Y. Savoye, *Cage-based Performance Capture*,
Studies in Computational Intelligence 509, DOI: 10.1007/978-3-319-01538-5_3,
© Springer International Publishing Switzerland 2014

Boneless Dynamic Surface Re-Skinning. Even if skeleton-based parametrization is widely use to represent the characteristic of the human motion analysis as seen in Poppe et al. [1], there is an avenue for investigating new type of animator-oriented full-body parametrizations. We reach quasi-conformal mapping to restore highly-deforming body information by offering more degrees of freedom. A major challenge is to exhibit an efficient framework to achieve plausible boneless reskinning that produces pleasing deformations and preserves the global appearance of the surface and offers flexible reusable outputs for animation. For this reason, we propose a new fully linear framework to reskin highly non-rigid captured dynamic mesh models with full correspondence and without the need of secondary deformation (like simulation, or additional surface-based deformation).

Contributions. In this chapter, we investigate the feasibility of extracting optimal reduced animation parameters without requiring an underlying rigid kinematic structure, but offered by a given coarse cage surrounding the surface. Then, we explore the potential of introducing harmonic cage-based linear estimation to reskin captured surface in motion. To the best of our knowledge, we propose the first algorithm for encoding performance capture meshes using a cage-based strategy. The main advantages of our novel approach are its linear single pass estimation of the desired surface, easy-to-reuse output cage sequences, and reduction in storage size of animations. In a first time, we introduce the cage-based parametrization for highly non-rigid surfaces that are already aligned. The surface model, initially acquired by a laser or initial dense reconstruction method, is smoothly and volumetrically embedded in a coarse but topologically similar mesh, called *the cage*, whose vertices serve as control points. The resulting reduction in control parameters and space embedding yields a compelling new trade-off to tackle the animation reuse problem.

Our results show that estimated parameters allow a sufficient silhouette-consistent generation of the enclosed mesh under sparse frame-to-frame animation constraints and large deformation. Once again, the cage layer is enhanced with Laplacian regularization. In a second time, we proposed a technique for *Animation Cartoonization* using cage-based subspace exaggeration with the extracted parameters and stylization of multi-view data.

Chapter Reminder. The rest of the chapter is organized as follows. Relevant works concerning markerless full-body performance capture and non-rigid shape capture using multi-view system are briefly reviewed and discussed in Sect. 3.2. Next, both problems of *Cage-based Animation Fitting* and *Animation Cartoonization* are attempted respectively in Sects. 3.3 and 3.4. We show the effectiveness of our methods with several examples. Finally, this chapter is discussed in Sect. 3.5 and concluded in Sect. 3.6.

3.2 Related Works from Capture to Re-Use of Dynamic Surfaces

There is an important deal of recent research that is of interest from capture to re-use of animatable surface. In this section, we briefly review some of relevant works to the problem of *Acquiring Surface in Motion* and *Performance Capture Reuse*.

3.2.1 Acquiring Surface in Motion

In this section, we describe recent works in the field of 3D Video and Vision-based Graphics. In fact, a lot of research has focused on the problem of acquiring and processing performance-based surface in motion, full-body performance capture from multiple views and interactive mesh deformation. First, markerless performance capture can be defined as the generation of spatio-temporally coherent and connectivity-preserving geometry. Techniques for video-based performance capture make often use of video-driven laser-scanned template, template-free deformation, as-well-as skeleton-based or boneless dynamic surface reconstruction. More precisely, we will detail effective pipelines to produce 3D video from multi-view setup and dedicated to the digitalization of real-world animation.

Marker-based Skin Capture. Early approaches focused on video-based capture of human skin rather than large waving apparel. As an illustration, one of the first outstanding high-quality template-based approach technique for capturing and animating dynamic skin motion is proposed by Park et al. citePark2006 where a significant large number of reflective markers are distributed over the actor body and segmented into set of rigid parts. Because of markers proximity, the mesh deformation procedure tries to fit surface landmarks to time-varying markers position using a radial basis functions and allows to animate natural muscle deformation (see Fig. 3.1). This data-driven methods is extended by Park et al. [3] in order to collect database of dynamic

Fig. 3.1 Human skin deformation: a scan body mesh is animated using radial basis functions driven by a dense collection of marker trajectories (image courtesy of [2])

skin deformations by recording the skin surface motion of various performances. From this complete collection, a statistical model of the deformations is extracted to identify static and dynamic parts.

Skeleton-Free Surface Capture. Nevertheless, one of the first markerless technique concerning acquisition of deformable human geometry from silhouette is presented by Sand et al. [4]. Their approach retrieves the actor mesh pose by adapting a collection of elongated deformable primitives parametrized with the skeleton motion without relying on carved space, but only on extracted silhouette observation. Skeleton-free parametrization of whole-body scan was firstly proposed by Allen et al. [5]. In this work, a consistent and common parametrization is realized on a large database of diverse scanned body shape using principal analysis component and mesh fitting to a high-resolution template surface. Using dense collection of marker-based placed on the body, this technique is able to capture surface details as well (see Fig. 3.2). Unfortunately, the representation employed for the fitting process do not exhibit a reduction of parameters, and makes the reuse of this parametrization nearly impossible for animation purpose. Furthermore, the non-linear optimization scheme is demonstrated to capture naked skin deformation without large clothes. In addition, Anguelov et al. [6] improved this technique by distinguishing the pose and body information in order to perform shape-completion of people by registration. *Principal Component Analysis* is also used to parameterize the shape transformation. Finally, a follow-up method proposed by Anguelov [7] learns a model of human body surface variations from shape examples.

Fig. 3.2 Space of human body: a consistent parametrization is established from a large database of diverse scanned body shape in order to fit a high-resolution template surface (image courtesy of [5])

Template-Less Surface Capture. In contrast to previous methods using marker-based skin capture, a variety of non-intrusive optical human capture techniques have been proposed from multiple viewpoints. One of the first pioneering work in markerless surface capture for performance-based animation, in order to increasing visual realism in animation, is proposed by Starck et al. [8, 9] with a fully automated system to capture shape and appearance. In this approach, highly realistic animated content is generated through capturing motion of real human performance from multi-video, without requiring specialist suits and invasive markers. The volumetric surface geometry is reconstructed by a mix of visual hull and volumetric stereo scene reconstruction with feature matching. Moreover a

spherical parametrization is employed to achieve remeshing of the reconstructed surface and control over the output animation is established using a move graph structure (see Fig. 3.3).

Fig. 3.3 Surface performance capture: surface capture of a hiphop dancer is recorded in a dedicated chroma-key blue-screen studio with 8 HD Thomson Viper cameras. At each frame, a temporal independent close polygonal model with small-scale surface details is generated as output of the system (image courtesy of [8])

High-Resolution Shape Capture. Following the pioneering work of Starck et al. [8], numerous of techniques have been developed in order to fit an articulated skinned template to silhouette or visual hull information. In recent work, a fast pose reconstruction method followed by non-rigid deformation of the laser scanned template to fit the silhouettes video stream is proposed by Vlasic et al. [10] in order to capture global pose and surface details from multi-view video recordings. At every frame, a pose estimation starts to geometrically fit the skeleton to the visual hull. Following the pose estimation, the rigged template mesh of the performer is deformed using linear blend skinning driven by the estimated articulated pose. Finally, the non-rigid shape estimation is adjusted to satisfy silhouettes contour constraints by non-rigid Laplacian-type deformation. Additionally, temporal filtering eliminates temporal model sliding (see Fig. 3.4). Even if this technique may require human intervention, its outstanding advantage is the full correspondence of output meshes, suitable for editing or high-quality animation reusing. Another intriguing method by Vlasic et al. [11] dynamically performs high-resolution shape capture using multi-view photometric stereo normal maps.

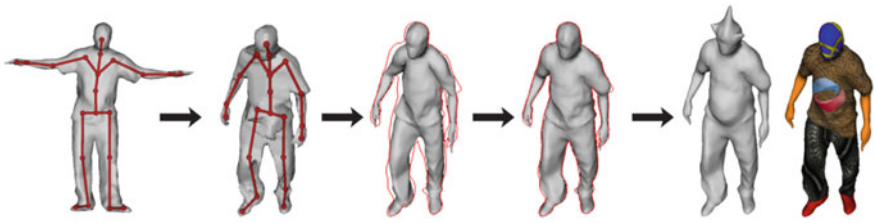

Fig. 3.4 Skeleton and shape from silhouettes: the global pose of the mesh is first obtained by a geometric fitting of the skeleton to visual hull. Then, surface details are added from multi-view video recordings (image courtesy of [10])

Template-based Performance Capture. De Aguiar et al. [12] propose an important markerless approach to capture human performances from multi-view video and to produce a dense and feature-rich output. In this work, authors rely on a purely geometric-based approach for video-based capturing human performances without the need of wearable body suit. This technique fits a deformable laser-scan template mesh to sparse multi-view silhouettes by explicitly abandoning traditional skeletal shape parametrization. At the first pose, an iterative closest point technique is utilized as an initial alignment after scanning. The reconstruction takes benefit of surface-and-volume shape deformation techniques. In the first step, global model pose is captured using volumetric deformation with tetrahedral Laplacian to match established temporal image SIFT feature correspondence. In the second step, the tetrahedral pose is improved using silhouette rims. Lastly, refined surface details are captured by adaptation along silhouette contours and model-guided multi-view stereo. Nevertheless, additional user intervention seems to be required to enforce key handles. Generally, silhouette overlap technique may fail if the source template is to far away from the target silhouettes (see Fig. 3.5).

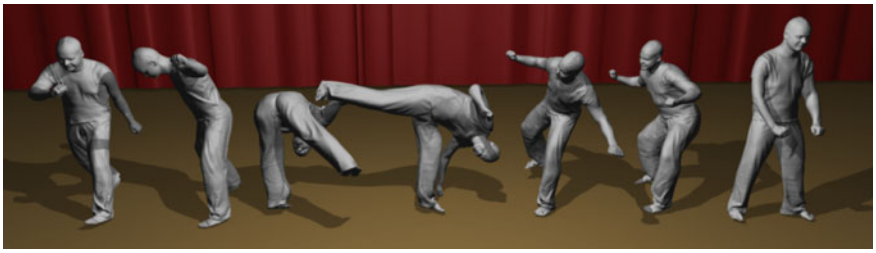

Fig. 3.5 Sparse performance capture: markerless mesh capture can also be achieved using a skeleton-free tetrahedral deformable model. Then, the deformation is driven by temporal correspondences established from extracted image features along multi-video sequences (image courtesy of [12])

Animatable Performance Capture. The previous approach of De Aguiar et al. [12] is enhanced with a novel optimization scheme proposed by Gall et al. [13] for capturing skeleton-based pose estimation and surface local estimation of a human or an animal subject using articulated template model and multi-view silhouette information. This approach operates in two steps: skeleton-based pose estimation and surface estimation. The pose estimation achieves a particle-based local and lower dimensional optimization of kinematic chain represented by a set of twist based on skinning and SIFT features. Then, the surface estimation incorporates silhouette alignment constraints into a Laplacian deformation framework. This method is also further extended by Stoll et al. [14] where physically-based cloth simulation is employed in clothes region of the animatable template mesh thanks to a cloth segmentation and statistical body model. The idea is to turn performance capture systems

into a pipeline ables to provide a complete animatable output with a high-level of understanding on what is observed (see Fig. 3.6).

Fig. 3.6 Animatable character: after the capture and the surface refinement steps, the captured surface is decorrelated into rigid and non-rigid components. The clothing-part mesh is then simulated using cloth simulation (image courtesy of [14])

Low-Cost Performance Capture. The problem of non-rigid deformation from multi-video stream is also attempted by Ballan et al. [15] with a conceptually simple and reproducible approach that does not require expensive laser scanner. The proposed contribution is a markerless motion recovery technique using a textured articulated deformable mesh model driven by silhouette information and optical flow features in a four cameras setup. In their motion estimation approach, a hierarchical organization of bones composed of 46 degree of freedom is adopted to drive a pre-acquired skin using a passive body scanner. Hence, a simple rigging process attaches the reconstructed skin to the input skeleton and lead to a linearly skinned deformable model. As pre-process, each video stream is first treated independently in order to extract a set of good feature to track and to establish a collection of valid pruned correspondences by unifying optical flow and 2D silhouette contour matching. The reconstruction problem is formulated as the constrained minimization of a feature-based reprojection error, by reconstructing the current pose using the pose

Fig. 3.7 Markerless motion capture of skinned models: a textured deformable mesh with skeleton is driven by silhouette information and optical flow features in a four cameras setup (image courtesy of [15])

information at the previous frame. To obtain a very good fit between estimated and observed data, the objective function is minimized in a non-linear manner in 2D image space to perform the optimal estimation for the skeleton joint parameters. A weighting constraints system is developed to enforce the contribution of optical correspondence compared to the silhouette correspondence.

The input mesh is deformed according to the underlying skeleton configuration and linear skinning operator, allowing to keep the pixel discrepancy error around an mean value. This error is the result of the imbrication of background subtraction error, mismatches between the actual 3D shape and artifacts coming from motion blur under face movements. The main advantage of this technique is the very limited number of cameras required to produce an accurate and robust reconstruction over different type of complex motion such as pirouette and break-dancing sequences. Moreover, this dynamic mesh deformation technique does not assume the need of laser scanned trackable mesh template. In addition, the employed algorithms are widely used in *Computer Graphics* and *Computer Vision* and simple enough yet efficient to extract information from captured images and the underlying deformation structure. This technique seems relatively easy to reproduce compared to similar pipelines. However, the non-linear formulation could lead to potential time-consuming optimization (approximately 30 min of computation per-frame). Similarly to other pipeline, the fitting process requires characters with Skin-tight clothing. Consequently, no finest non-rigid surface components like cloth wrinkles can be incorporated to the quasi-articulated surface deformation (see Fig. 3.7). Low-cost performance capture using cheaper devices like Kinects®bring mesh-based capture more accessible and popular as shown by Ye et al. [16].

Multi-persons Performance Capture. Liu et al. [17] propose the first segmentation-based markerless motion capture method to track two persons in the same video footage. To achieve this goal, a multi-image segmentation is employed in order to perform a feature-to-person assignment that drives the skeleton-based mesh deformation. Besides, a probabilistic scheme improves the mesh fitting by resolving intersection, especially in close contacts. Even if this technique is demonstrated for two people, its generalization to more crowded scene seems doable (see Fig. 3.8).

Heterogeneous Capturing Systems. We conclude that there is a few of heterogeneous pipelines that produce at least a collection of output temporal full-body mesh sequence, not often aligned. Moreover, skeletal representation or other types of reduced parameterization is not always offered as output by such systems. More recently, the massive content of output captured data (images, calibrated silhouettes, mesh sequences, skeletons) invites researchers to investigate the *Performance Capture Reuse* problem for authoring such massive content.

Fig. 3.8 Interacting characters: multi-view image segmentation improves the pose fitting and surface estimation of two-people in observed sequences of close physical contacts by realizing a feature-to-person assignment (image courtesy of [17])

3.2.2 Performance Capture Reuse

3D Videos. With the emergence of 3D commercial displays thanks in part to the 3DTV research project [18], there is a high demand for reusing collected data beyond capturing, representing and transporting 3D video. In this section, we present most of outstanding 3D video editing techniques. 3D video technologies are able to capture high-fidelity full 3D shape, motion and texture to offer free-and-rotate special effects. The full 3D video pipeline is presented by Takai et al. [19].

Reusable Performance Capture. This section gives an overview of various techniques involved in the reuse of performance capture animation and multi-view data. 3D video data consist in a series of time-varying information with various natures (calibration data, meshes, textures, silhouettes) that suggest potential post-production reuse to establish more information or synthesize complementary information, as seen in the Minimal 3D video of Tung et al. [20]. Nonetheless, the reuse of captured mesh-based animation data is more challenging problem than the reuse of conventional motion capture data in order to resynthesis new dense animation for two reasons: the high amount of data produced and the faithful adaptation of a mesh animation. A large number of applications of reusable performance capture can be noticed: space-time editing, stylization of the shape motion and appearance, texture or deformation transfer, mesh segmentation, summarization of redundancy, and compression, and skin reparametrization.

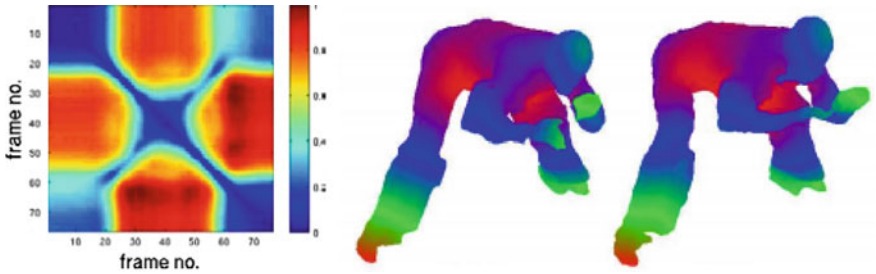

Fig. 3.9 Shape similarity and segmentation (from *left* to *right*): a similarity descriptor between surface capture sequences (image courtesy of [21]), and 3D video segmentation (image courtesy of [22])

Shape Segmentation, Similarity and Summarization. PerfCap mesh sequences have complex deformation characteristics that could be challenging to analyze toward segmentation and summarization. Then, 3D video performance segmentation is proposed by Tung et al. [22] where Reeb graph-based descriptors are used to characterize the topology. Robust body-part segmentation labels are transferred frame-by-frame using motion flows. Huang et al. [23] propose to examine the shape similarity inside a sequence. This technique reuses performance capture data by applying histogram and shape-flow descriptor to identify frames with similar shape and motion for people with loose clothing and complex motion. A comprehensive performance evaluation of skeleton and non-skeleton shape descriptors is presented by Huang et al. [24]. Such techniques can also be used for temporal shape matching. Furthermore, automatic summarization of 3D video is proposed by Huang et al. [21] (see Fig. 3.9). Volume-sampling and Shape Histogram is adopted to identify the self-similarity of 3D video sequences. A graph is constructed from the self-similarity and the shortest path is found as the set of keyframes. The global optimization provides a concise representation.

Graph-based Dynamic Surface Synthesis. Relatively few techniques have been developed to provide a higher controllable structure over several animation sequences. Recent work generalize the concept of motion graph to dynamic surfaces. A video-based representation has been introduced by Starck et al. [25] for free-viewpoint control of realistic mesh animation from multiple view video sequences captured in a ten cameras studio. The system combines image-based reconstruction and video-based animation. Surfaces are reparameterized on the spherical domain to blend motion sequences to provide seamless motion cycles and smooth transitions between different motions. A multiple resolution with coarse-to-fine spherical matching algorithm has been developed to provide correspondences between spherical geometry images.

Huang et al. [26] present a framework for concatenative human motion synthesis from 3D video sequences according to user-defined space-time warping constraints, which preserves the detailed dynamics. Transitions between 3D video sequences are identified without the requirement for temporal correspondence using 3D shape similarity over an adaptive temporal window. Huang et al. [28] focus on editing dynamic surface geometry by synthesizing new 3D character animation from a public captured performance database with three interpolation methods for space-time propagation. The parametric control of captured mesh sequences provided by Casas et al. [29, 30] is based on the assumption that the input 3D video sequences are already aligned and contain similar actions. Casas et al. [27] describe a parametric model for real-time interactive character animation from 4D actor performance capture. This approach is also based on a useful database of temporally aligned mesh sequence reconstructions of an actor performing multiple motions. Finally, parametric motion graphs are introduced to enable movement control and transition between different motions (see Fig. 3.10).

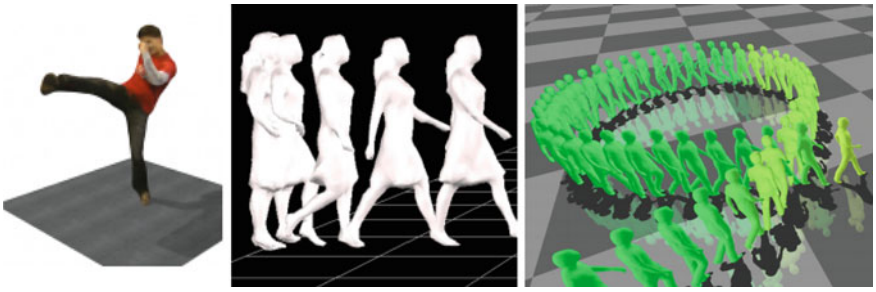

Fig. 3.10 Graph-based Dynamic Surface Synthesis: video-based character animation (image courtesy of [25]), motion graph on surfaces (image courtesy of [26] and [27])

Space-Time Reediting and Stylization. The problem of space-time reediting and stylization of performance capture data has received almost none attention in the past. Tejera et al. [31] propose a space-time editing technique of 3D sequences with a learnt motion model gives a flexible interactive approach to mesh sequence editing with a similar level of control for conventional skeletal animation. This space-time editing technique is based on a Laplacian-type deformation approach. This allows constraints such as foot or hand position to constrain the captured movement and to interact with objects of the scene (see Fig. 3.11).

Re-Parametrization, Compression and Transfer. The lack of controllable parameters for non-rigid world and massive database of raw data invite researchers to re-parametrize, compress and transfer attributes of 3D videos. Even if the reusing traditional motion capture was extensively studied over the years [32], the reuse of captured dynamic surface is a novel problem where adequate solutions are not just the simple transposition of solution from mocap paradigm. For instance, one of the first techniques aiming at reparametrizing an existing arbitrarily deforming

Fig. 3.11 Space-time reediting and stylization (from *left* to *right*): space-time control of captured mesh sequences (image courtesy of [30]), and motion stylization of dynamic surface (image courtesy of [31])

animation sequence is proposed by Kircher et al. [33] to allow the user to change the geometry of the surface in a temporally coherent manner. More recently, de Aguiar et al. [34] propose a fully-automatic method to convert skeleton-less mesh animation into a compact skeleton-based animation. This extraction effectively bridges the gap between the purely mesh-based and skeleton-based animation representation. Moreover, James et al. [35] present an automatic method for constructing skinned mesh approximations to parametrically coherent mesh sequences using mean shift clustering. Building on this idea, Kavan et al. [36] present an algorithm to construct high quality skinned approximations of complex highly deformable animations using approximate dense dimensionality reduction approach.

Fig. 3.12 Re-parametrization, compression and transfer: (from *left* to *right*): 3D geometry videos (image courtesy of [37]), performance capture transfer (image courtesy of [38]), texture transfer (image courtesy of [22])

The reuse of performance capture also includes two central notions: the *compression* and *transfer*. In the one hand, Quynh et al. [37] present *Conformal Geometry Videos* (CGV), as a novel extension of the traditional geometry videos by taking into the consideration of the isometric nature of 3D articulated motions. *Geometry videos* are a compact representation encoding each frame with a 2D domain where

3D articulated motion can be uniquely determined. On the other hand, Baran et al. [38] transfer existing performance capture mesh deformation from one character to another, by determining the minimum amount of information necessary to specify this transfer with patch-based LRI Coordinates. Meanwhile, Tung et al. [22] perform dense matching of 3D dynamic surfaces captured by a set of multi-view video camera. This technique allows texture transfer between surface models extracted from various sequences and motion transfer from a 3D video to an unanimated model (see Fig. 3.12). Feng et al. [39] present a new data-driven approach for image-based synthesis of realistic video animations containing user-defined human motions. One key component of their approach is the generation of target textures from a multi-view video.

3.3 Our Approach: Cage-based Animation Fitting

Cage-based Performance Animation. Even if a lot of methods reconstruct continuous surfaces from several viewpoints, we notice a lack of global, flexible and reusable parametrization. However, finding a suitable non-rigid animation model with reduced control parameters is a key problem for video-based animation. Unlike methods based on an underlying articulated skeleton, we aim to estimate non-rigid subspace deformations reproducing the time-varying dynamic surfaces without secondary deformations. Since the *reconstruction* step is not claimed as part of our contribution in this chapter, we deal with surfaces having the specificities to encapsulate life-like features from video footage. Thus, we propose a new approach estimating an optimal set of cage vertices position allowing to reskin template while preserving its fitness to the multi-view silhouettes. The proposed method uses a single linear pass ensuring cage smoothness under sparse subspace constraints. Then, we exhibit an external parametrization dealing with a significant number of reduced parameters comparing to the number vertices of the enclosed mesh. However, the key contribution is to solve a sparse linear system to estimate the optimal cage parameters reproducing the desired deformation of the enclosed model, given data-driven constraints. This chapter shows that cage parametrization are useful to reuse output of video-based acquisition. To the best of our knowledge, this is the first approach using cage-based animation for performance capture.

3.3.1 Non-rigid Dynamic Shape Analysis

Characterize Captured Animation. The analysis of dynamic shapes with realistic motions is the very foundation of structure-preserving editing from performance-based captured animations. Then, we propose to characterize captured animation across time using a deviation descriptor. To do this, we propose to use the *edge length deviation* to detect rigidities, rigidity boundaries and non-rigidities of cap-

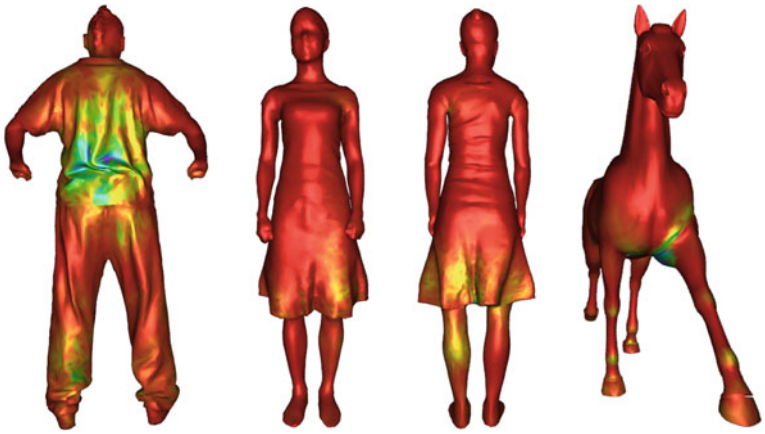

Fig. 3.13 Dynamic surface motion analysis: the color-coded edge-length deviation on the whole mesh animation reveals that non-rigid human animations in clothing cannot be cast as piecewise-rigid model, contrary to the synthetic horse animation. Highest non-rigid areas (with respect to the full surface variations) are displayed in *green-to-yellow colors*. No boundaries are clearly identifiable on performance capture data

tured dynamic surface with constant connectivity as proposed by Tierny et al. [40]. The *edge length deviation* provides a good descriptor of the deformation gradient, by identifying surfels that moves rigidly over time. Figure 3.13 shows some shape examples describing the visual characteristic of the model deformation with its corresponding *edge length deviation*. In contrary, the quasi-articulated animation provides by the synthetic dataset.

Non-rigidities Detection. Consequently, it seems infeasible to decompose highly non-rigid surfaces into rigid parts or to extract kinematic skeleton directly from the surface deformation. More importantly, the use of a single skeleton as shape-preserving encoding for captured animated surfaces seems impractical because the cloth layer provokes unpredictable additional elastic deformations that perturbates the identification of boundaries corresponding to joint areas though the surface. This effect makes the automatic decomposition into invariant rigid surface region nearly impossible for surfaces that cannot be categorically classified as quasi-articulated meshes. Because cloth deformation are correlated to organic articulated motion, it seems hard to distinguish the articulated deformation component without explicitly separating both deformation behaviors. In addition, we think that video-based meshes cannot be re-edited in a near-conformal manner from an extracted rigid skeleton parameterization without secondary deformation. Finally, assuming no skeleton prior will help a better shape understanding for acquired dynamic surfaces, that exhibit arbitrary elastic deformation instead of strong piecewise rigidities and isomorphism. Finally, we believe that *edge length deviation* is an interesting global shape descriptor to separate real-world and synthetic datasets.

Reskinning Non-rigid Surfaces. Since skeletal representation appears impractical to preserve non-rigid surface features of an input dense captured surfaces, we propose to use a detached cage with a given rigging function. This encoding restricts deformations on a latent space that is low-dimensional, with more precise surface control for non-rigid tight cloth dynamics. Then, cage-handles offer a suitable flexibility to enforce local rigidities over the enclosed surface. Hence, cage-based deformation offers an elegant free-form abstraction to capture low-frequency shape deformations correctly with reusable output parameters.

3.3.2 Cage-based Conversion from Performance Mesh Animation

Animation Conversion. Markerless highly-detailed performance capture is one of the most emerging technology that has been recently studied in vision-based graphics and 3D video with appreciation to robust multi-camera studio and 3D laser scanner. Achieving *inverse animation* by approximating dynamic meshes using rigid skinning has inspired researchers to convert video-based reconstructed mesh sequences into rigid kinematic parameters as introduced by de Aguiar et al. [34]. In contrast with previous techniques using skeleton-based animation paradigms, we describe an efficient linear estimation framework to convert non-rigid performance animation into cage-based animation. Our approach retrieves a small collection of animation parameters through a single-pass minimization process without the need of an underlying rigid kinematic structure. This novel *cage-based animation conversion* process has the main advantage of retaining sufficient surface details of arbitrary deformations aswell-as offering flexible reusable outputs for animation control and potential transfer. Our method allows to build an animation model of vastly reduced size and improves editability, while preserving life-like motion details up to small bounded error.

Cage Sequence Extraction. Before introducing technical details, we briefly describe the procedure of our method in this overview. Our method retrieves the space warping of the observed surface across time. The main challenge is to deal with the high number of degrees of freedom provided by the coarse cage. Our pipeline takes as input a video-based reconstructed mesh sequence in conjunction with a default cage. As output, the system produces a sequence of cage-handles location parameters that fits for each frame, and shares same mesh connectivity. Then, we estimate a sequence of cage parameters expressing the mesh at each animation frame using *cage-based inversion*. This optimization procedure retrieves the cage pose in a least-squares sense from sparse data-driven motion constraints expressed directly over the enclosed surface and transferred into the subspace domain using its indirection. This new technique produces suitable outputs for animation compression and reuse. In addition, differential geometry is well-known for reconstructing an edited surface [41]. That is why a straightforward and effective way to retrieve animation parameters is to combine these techniques. We exploit a bounding polytope topologically-aware in respect of the deformable surface to encode the animation. Thus, we embed models

$\mathcal{M}$ in a coarse cage $\mathcal{C}$ using *Harmonic Coordinates* having local smooth properties. The set of n cage vertices is denoted by $\mathcal{V}_{\mathcal{C}} = \{c_1, \ldots, c_n\}$ where c_k is the location of the kth cage vertex. The set of m model vertices with $\mathcal{V}_{\mathcal{M}} = \{v_1, \ldots, v_m\}$ where v_i is the location of the ith model vertex. We refer respectively by v_i^t and c_k^t the geometry of the input model and estimated cage parameters at the frame t.

Laplacian-based Harmonic Subspace. The subspace domain is the volume enclosed in the cage interior, according to few control points. The spatial relationship between the cage and the model is pre-computed at the default pose using harmonic rigging, enabling the illusion of realistic deformation of a model accurately [42]. An Harmonic coordinate $h_k(i)$ is a generalized barycentric weight representing the deforming influence of the kth cage vertex on the ith model vertex. *Harmonic Coordinates* are obtained by volumetric heat diffusion process in the cage interior, as the precomputed solution of Laplace's equation with Dirichlet boundary condition. For each control point c_k, we define a harmonic function $h_k(\cdot)$ by enforcing the interpolation property $h_k(j) = 1$, if $k = j$, and 0 if not, and asking that h_k be linear on the edges of the cage. Our rigging process follows implementation details of Joshi et al. [42], and are derived from the well-known *heat diffusion equation* described as follows:

$$\frac{\partial u}{\partial t} - \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = 0$$

where $u(x, y, z, t)$ is a time-dependent rigging function, representing the vector of estimated weights, reaching the desired steady-state. Then, we propose to define the *Laplacian Operator* on the cage instead of the laser-scanned model to improve the scalability of the fitting process and to keep model detail properties good enough. We therefore refer to our notion of *Laplacian Cage* defining a cage structure enhanced with a Laplacian regularization thanks to a *Dual Laplacian Operator* $\mathcal{L}_C(\cdot)$. Applied on the cage shape, this operator is defined by the difference vectors between each cage vertex and its adjacent neighbors with cotangent weighing system. The cage differential coordinates $\hat{\delta}$ encode each control vertex relatively to its neighbourhood

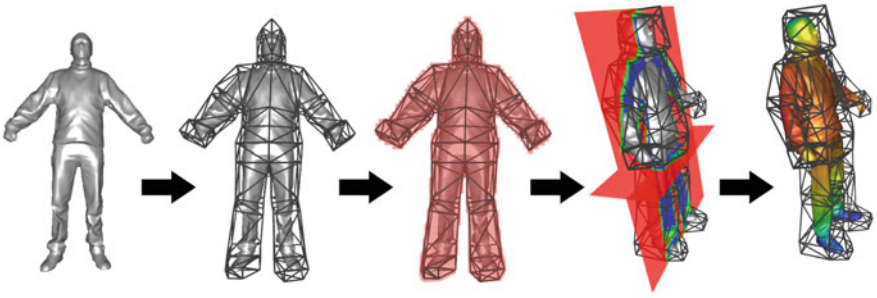

Fig. 3.14 Harmonic rigging process (from *left* to *right* hand side): laser-scanned template, cage-based model embedding, cage voxelization, voxel tagging and harmonic weights computation

preserving the local geometry. Besides, *Laplacian Cage* has interesting properties to convert large deformation over laser scanned surfaces using sparse constraints, as well as to ensure their local features independently of the high-resolution (see Fig. 3.14).

3.3.3 Estimation of Space-Time Cages for Non-rigid Surfaces

Recovering Cage Parameters. In our application, we employ a full-body template mesh, already captured using a laser scanner. This static template is then approximated using a linear piecewise surface. The subspace connectivity is preserved across time because the default cage is only seen as a connectivity mesh and animation-driven feature constraints are seen as an external deformation stress. This surface-and-space based deformation technique preserves the spatial coherence of the warping. Then, the cage geometry can be reconstructed efficiently by indirectly exploiting the rigging and by solving a system of sparse linear equations. We cast the problem of deformation as a *Least-Squares Laplacian Cage* reconstruction process using a consistent minimization approach of an objective function. The subspace solver recovers desired cage poses by minimizing an objective function in a least-squares sense in order to fit a continuous volume. Then, the geometry of the desired model is simply obtained by generating its vertex position according to the expressed cage parameters obtained on the concept of *Least-Squares Cage* (see Fig. 3.15).

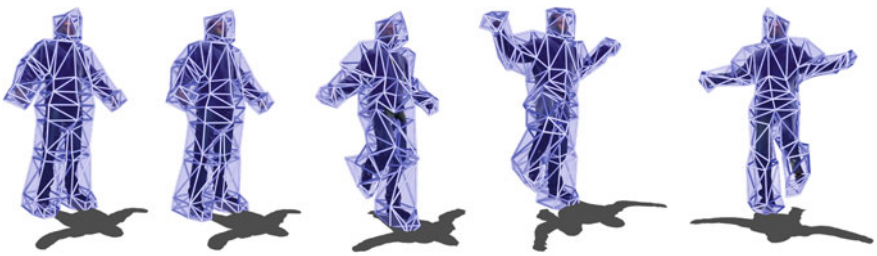

Fig. 3.15 Cage conversion: estimated cage fitting for a sequence of performance capture data with highly non-rigid surface deformation with small-scale and large global deformation. The enclosed mesh is resulting of our cage-based animation conversion

Space-Time Cage Optimization. We reformulate the *Animation-Conversion* problem as a minimization problem for cage recovery. Our optimal estimation of cage parameters is formulated as a least-squares consistent minimization of an objective function, expressed exclusively in term of cage geometry (for frame-to-frame animation). To retrieve cage parameters, we formulate and solve an overdetermined linear system to minimize the overall energy as follows:

$$\min_{\forall k, c_k^{t+1}} \left(\underbrace{\sum_{k=1}^{n} \left\| \mathcal{L}_C^t(k) - \hat{\delta}_k^t \right\|_2^2}_{\text{smoothness term}} + \underbrace{\sum_{i \in \mathcal{S}} \left\| v_i^{t+1} - \sum_{k=1}^{n} c_k^{t+1} \cdot h_k(i) \right\|_2^2}_{\text{data term}} \right)$$

The subset $\mathcal{S}$ is composed of sparse deformation constraints, driven by the data-driven surface evolution. Such constraints are automatically selected irregularly inside regions of interest over the enclosed surface. The optimization process takes advantage of the cage-based mesh parameterization to recover the optimal cage geometry in a least-squares sense. The smoothness term contains the Laplacian-based cage regularization, preserving the cage global details under sparse linear subspace constraints and allowing a temporally-coherent estimation. Meanwhile, the data energy term transfers a collection of surface deformation constraints into the subspace domain by the quasi-conformal harmonic mapping. That means the enforced deformation constraints are directly expressed in terms of cage parameters. Both energy terms can be weighted to favor one comparing to the other.

Accordingly, Laplacian-based harmonic subspaces with sparse linear constraints permit the estimation of coherent parameterization. This estimation ensures a sufficient preservation of silhouette rims vertices that are already encoded in the input data. Harmonic coordinates reproduces as-local-as possible deformation over the laser-scanned deformable surface without space-consuming mesh-based approach. Finally, the geometry of performance animation is simply restored and rendered real-time using our 3D free-viewpoint video player. Expressed as a linear combination of estimated cage geometry with harmonic weighting, cage-based deformation is used to regenerate absolute coordinates of the enclosed model across time.

Energy Terms. Note that the first term of the energy preserves the cage smoothness and ensures a pleasant deformation under a reduced number of constraints in comparison to the enclosed mesh density. Hence, the space-based distortion energy is measured by the Laplacian on the cage. The total local distortion measure for a deformation process is given by a quadratic energy term. Furthermore, the second term of the energy enforces the position of vertices to fit the desired model defined by positional constraints. To our best knowledge, the simple global optimization component of our framework with such formulated constraints to minimize does not already exist in the literature.

Underlying Surface Fitting. The second term of the objective function measures how much the cage enforces sparse positional constraints. The transfer of surface-based constraints into the subspace domain exploiting the cage indirection is expressed by this energy term. In other words, the last formulation enables to express surface constraints directly in terms of cage parameters linearly using an inverse near-conformal harmonic mapping. Given the control mesh at the previous frame state in a deformation mesh sequence, we need to exploit frame-to-frame deformation of the finer mesh to automatically constructed an altered control mesh for every frames in the sequence. As shown on the results, the cage retrieval process requires a reduced

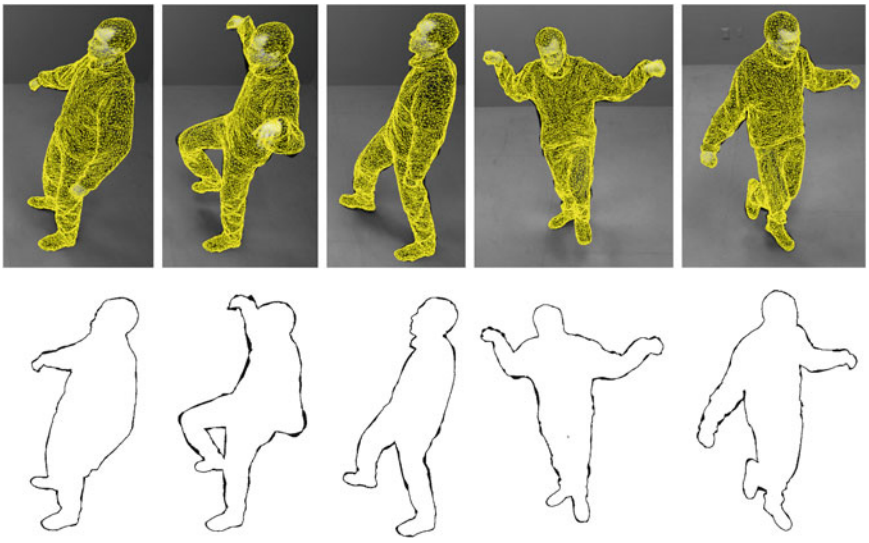

Fig. 3.16 Cage-based shape reprojection error: projection of the wireframe mesh generated by the estimated cage with 5 % of selected constraints (*top*), silhouette overlap between the rendered silhouette and the extracted silhouette (*bottom*)

number of corresponding input vertices and their displacement to form sparse linear subspace constraints to be able to estimate a cage expressing a surface fitting to the silhouette good enough.

3.3.4 Experimental Results and Video-based Evaluation

Effective Cage-Fitting Framework Implemented with OpenGL and C/C++, our prototype framework proposes a stable mechanism to fit a given cage for various potential applications. This section describes our experiments using the described system where various evaluations of the clean-up algorithm are conducted. In practice, to demonstrate the feasibility of our method for life-like surface, we choose to use as input captured dynamic surface having strong local variations driven by the clothing. We can only test our algorithm on few challenging real-world dataset because the setup stage is really fastidious and delicate to setup for non-artist users. The entire process takes less than two seconds per frame without any code optimization, and solvers only running on CPU. The algorithm performance is validated by both qualitative and quantitative evaluations. We show that the cage reproduces the 3D motion of life-like character accurately without requiring a dense or conformal mapping (see Figs. 3.16 and 3.17).

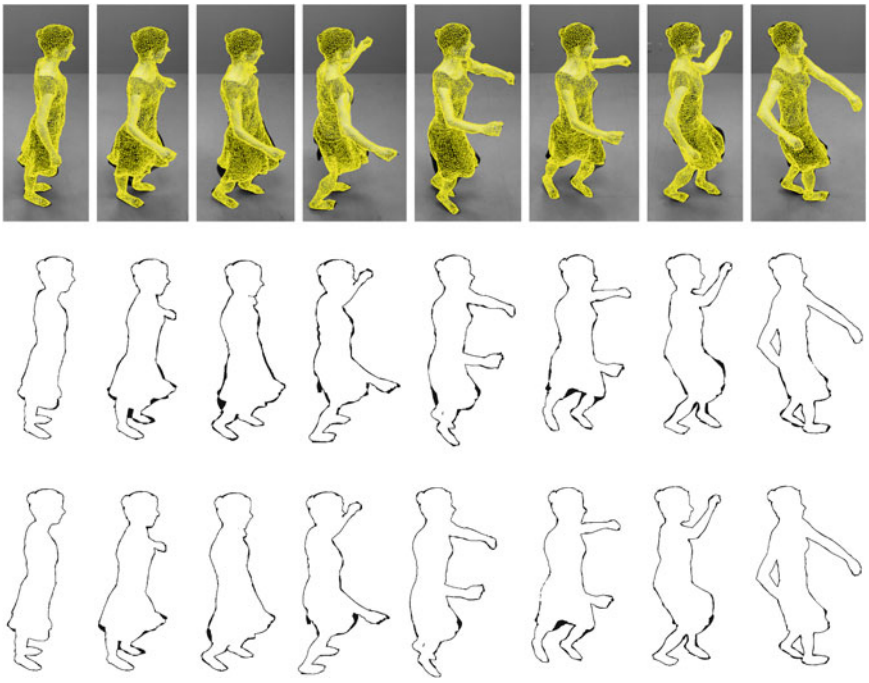

Fig. 3.17 Cage-based shape reprojection error: projection of the wireframe mesh generated by the estimated cage with 5 % of selected constraints selected (*first row*), silhouette overlap between the cage-based silhouette and the extracted image-based silhouette (*second row*), silhouette overlap between the original mesh silhouette and the extracted image-based silhouette (*third row*)

Silhouette Overlap Error. We propose to use various error metrics to evaluate the accuracy of the estimated shape in respect with a default template and the observed multi-view video streams. Conveniently, the exclusive-or (XOR) operation between the rendered silhouettes of the cage-based shape and the segmented video-image silhouette, yielding those pixels that are not overlapping. This *Silhouette Overlap Error* [43] can be seen as an energy function, defined at a given viewpoint as follows:

$$E_{XOR}(\mathcal{C}) = \sum_{i=0}^{N} \sum_{x=0}^{X} \sum_{y=0}^{Y} \left(P_s(x, y) \wedge \overline{P_m(x, y)} \right) \vee \left(\overline{P_s(x, y)} \wedge P_m(x, y) \right)$$

where $\mathcal{C}$ is the mode parameters (here the vector of cage location parameters), X and Y are the dimensions of the image, N is the number of camera. $P_s(x, y)$ is the 0 or 1 value of the pixel (x, y) in the captured image silhouette, while $P_m(x, y)$ is the equivalent in the reprojected model image given that the cage parameter is $\mathcal{C}$ (Figs. 3.18, 3.19).

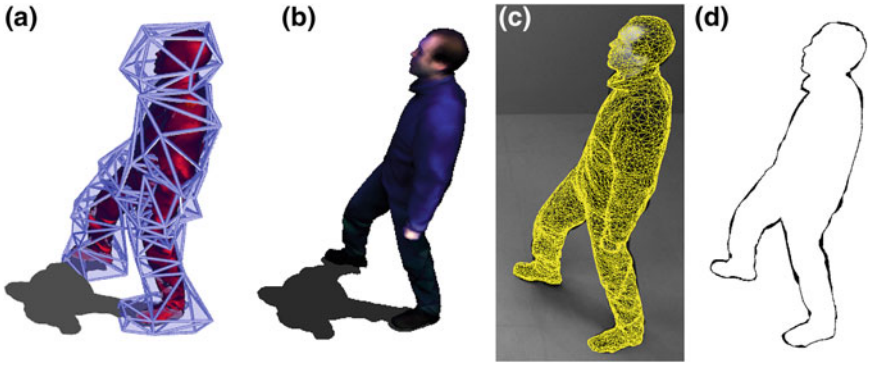

Fig. 3.18 Qualitative evaluation. Evaluation of the cage-based reconstructed surface: Estimated cage (**a**), textured enclosed surface (**b**), cage-based model reprojected (**c**), silhouette overlap error (**d**)

Silhouette Consistency. The performance of our system was tested on multi-view video sequences recorded with eight cameras at a resolution of 1600×1200 pixels. The template is composed of 10002 vertices and the cage is composed of 141 vertices (80 % of parameter reduction of the enclosed model). To validate our method, some experimental results are shown on real datasets (see Fig. 3.21). Qualitative evaluations are based upon visual comparisons (silhouette consistency) for each reconstructed frame with respect to the original shape, and according to various percentage of selected vertex constraints (see Fig. 3.19). We also provide rendering feedback to allow qualitative evaluation on two different sequences with a total of 348 frames (see Fig. 3.17). Then, we run our cage-based reskinning method to measure how much the estimated cage-based deformation of the template can fit the observed silhouettes without applying an additional silhouette regularization on the enclosed surface. For our evaluation, we measure the fidelity of our output with several error metrics such as edge length variation, root mean square volume variation comparing to the input

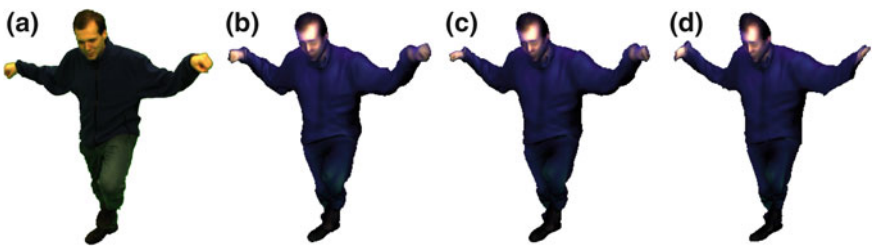

Fig. 3.19 Constraints density influence: influence of the number of positional constraints on the quality of enclosed mesh expressed by the estimate cage (the number of positional constraints are expressed in percentage of enclosed mesh vertices). Real image (**a**), 100 % constraints (**b**), 5 % constraints (**c**), 1 % constraints (**d**)

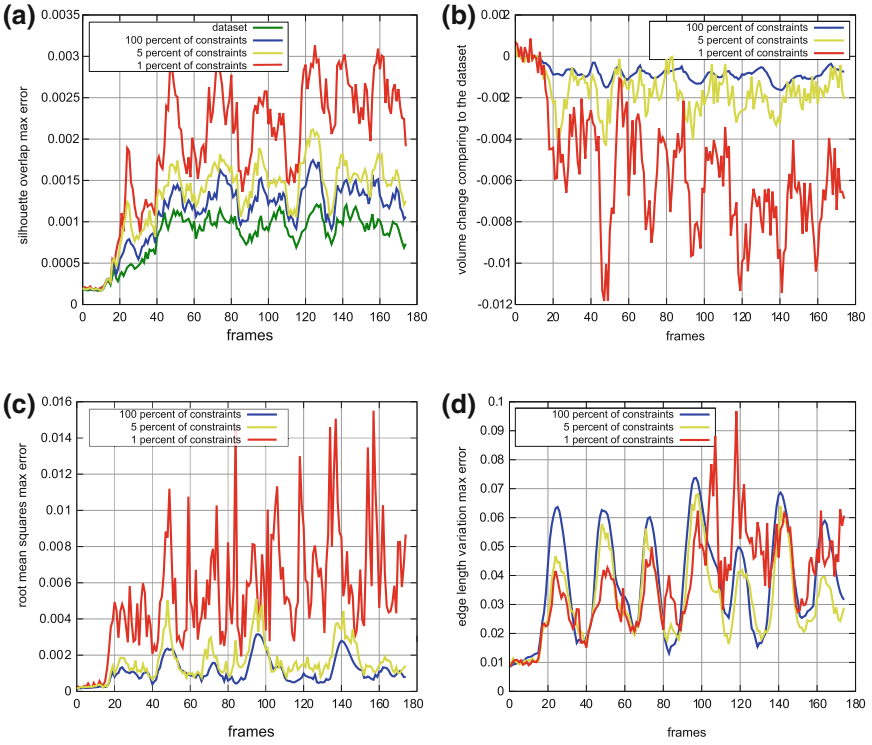

Fig. 3.20 Quantitative evaluation: quantitative evaluation on MIT Crane dataset. Silhouette overlap max error (**a**), volume change (**b**), root mean square max error (**c**), edge length variation (**d**)

dataset models as shown on Fig. 3.20. For visual evaluation (see Fig. 3.21), we simply back-project cage-based meshes into the video stream according to camera calibration information. For qualitative evaluation, the fidelity of our output is measured with the average silhouette overlap error comparing to the input dataset models. The multi-view silhouette overlap error between the reprojected model and the silhouette image over time reflects the estimation error. The difference between the overlap of our generated animation is similar to the one obtained with the input animation with a sufficient number of constraints. Furthermore, the cage retrieval process estimates silhouette-consistent cage parameters interactively under sparse linear subspace constraints as shown in Fig. 3.21. Obviously, the quality of the reskin decreases with less constraints with our basic strategy.

Validated Claim. For the first time, we validate the hard claim concerning the feasibility of using cage-based scheme to reskin highly-detailed mesh animation representing realistic human in motion. In addition, the deformation driven by the cage offers an affordable silhouette-consistency with respects to all images recorded by all cameras. our technique is well-suited for encoding and compression stage because the fixed connectivity of the cage is preserved across the sequence . The most sur-

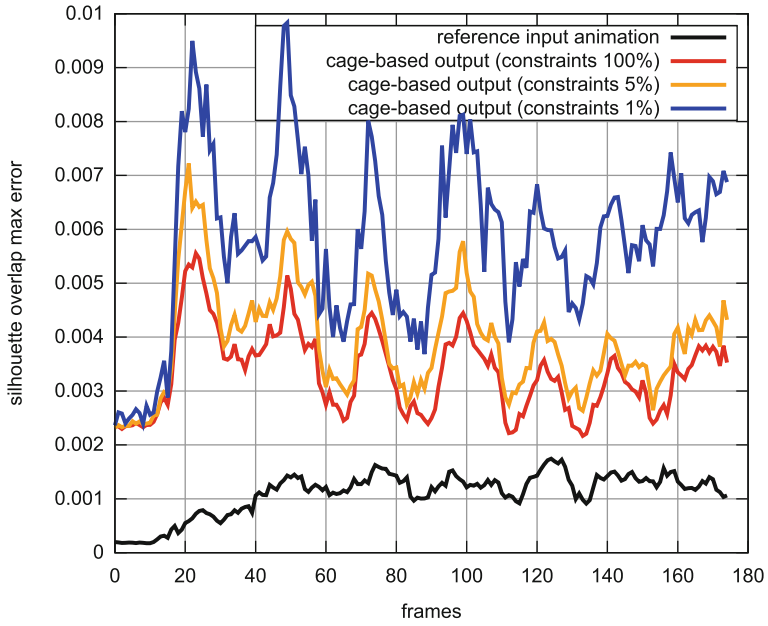

Fig. 3.21 Deformation constraints: impact of the quantity of deformation constraints (in percentage of enclosed mesh vertices) on the quality of the estimated cage-based mesh. The silhouette overlap error is expressed in percentage of foreground pixels that are similar to the reference silhouette image

prising aspect that we would like to point out is that the harmonic mapping combined with a detached set of cage-handles is able to reproduce animation with plausible silhouetteness. To show the accuracy of our cage fitting, we have developed a 3D video player that displays the cage-based performance animation in real-time. Consequently, cage-based deformation allows 3D video players to produce a smooth and accurate frame-to-frame playback of time-varying meshes.

Reskinning Feasibility. We have shown the feasibility of involving cage-based parametrization to reskin video-based mesh sequences. The main advantage of our framework is its linear form, as-well-as the reduction of mesh parameters, which is independent of the surface resolution and allows potential further reuse. Our harmonic cage-based encoding allows mesh rim vertices to fit the silhouette more precisely than articulated skeletons because of better local controls. Our minimization framework can efficiently extract fitted subspace for mesh animations by retrieving the cage pose associated to all animation frames, while preserving the input cage connectivity. In addition, our techniques drastically decrease the size of dataset without a strong lost of visual quality (see Fig. 3.22).

Observations. Our method is able to deal with non-rigid deformation already encapsulated in the input data, and that are thoroughly impossible to recover precisely by

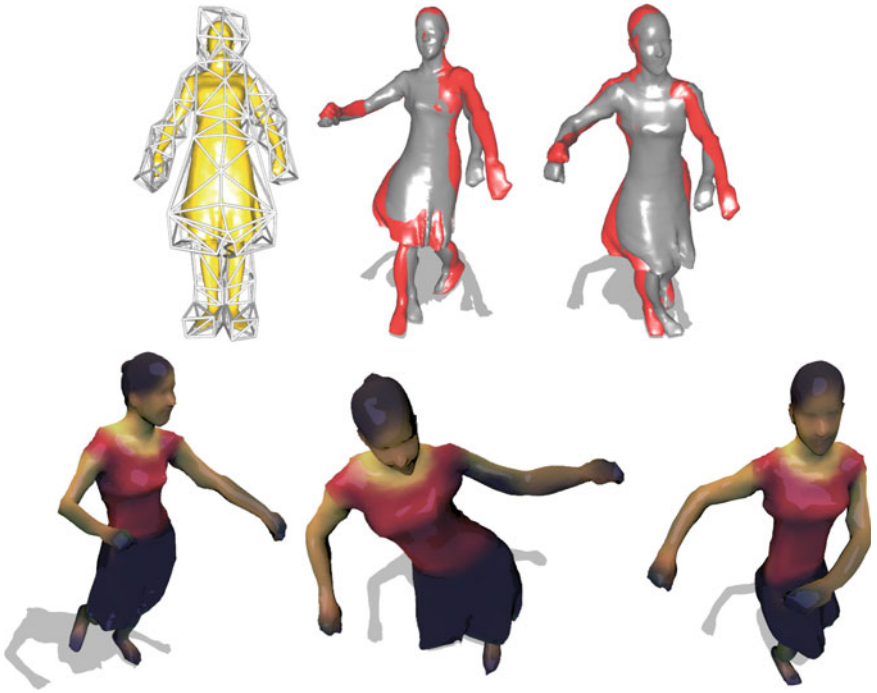

Fig. 3.22 Cartoon 3D video: sparse-and-detached surface trajectories are first estimated over the whole mesh sequence of a captured dance performance, to be exaggerated secondly thanks to a non-rigid squash-and-stretch motion filter. For visual comparison, the reconstructed cage-based shape is displayed with *gray color*, superposed with the corresponding new filtered mesh with *red color* (*top row*). Finally, the resulting cartoon-style mesh animation is real-time rendered using stylized subspace deformation coupled with video-based toon appearance (*bottom row*)

a rigid skeleton and linear skinning weights. As a result of more degrees of freedom, our harmonic inverse cage-based estimation allows rim vertices to preserve silhouette consistency better than inverse skeleton-based estimation. Dealing with a non-rigid underlying structure helps the estimation to obtain physically plausible results even if the observed surface is enhanced with deforming piece of clothes. Easily implementable in existing pipelines, our accurate and simple method can be improved by employing sparse solver processing on GPU. Unfortunately, our cage conversion method suffers from drawbacks directly derived from Laplacian and space deformation properties. For example, the volume shrinking can provoke interior folding, undesired connectivity distortion and potential cage fold-over under inconsistent subset constraints. Another major limitation of our method is that the deformation result depends on the enclosed shape and the cage tessellation. Moreover, the automatic generation of abstracted cages is also an open and hard problem for performance capture applications. Clearly, the resulting framework is suitable for compressing and transmitting high-quality video-based performance capture mesh sequence in

minimal way with quite less amount of data than it is commonly done nowadays. To summarize, boneless cage-based approach is more convenient than kinematic hierarchy-based approach to guarantee the deformation fidelity of non-rigid surface to multi-video stream.

3.4 Beyond the Cage: Animation Cartoonization

3.4.1 Foundation of Animation Cartoonization

Always Exaggerate. Recent advances in dynamic surface capture have made the creation of realistic animations 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.

Performance Capture Puppetry. 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-life visual cues. Thus, 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 opportunities to serve the theme of *Animation-Cartoonization*. We 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.

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. In contrast with the work of Kwon et al. [44] that addresses the stylization of articulated motions, our method demonstrates the generalization to high-quality 3D surface video relying on non-rigid underlying structures. To the best of our knowledge, motion stylization of cage-based shapes has never been addressed for non-rigid dynamic surfaces. Independently, intensive research in expressive rendering enables a wide variety of styles. Producing hand-drawn cartoon texture for 2D animation is presented in [45, 46].

Traditional toon shading has been used by Kang et al. [47] to depict morphological object features and extended by Barla et al. [48] to support view-dependent effects. Non-photorealistic shading is also investigated by Rusinkiewicz et al. [49] to exaggerate and depict shapes. Meanwhile, in the field of vision-based projective rendering, multi-view texture mapping approaches such as the one of Eisemann et al. [50] 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.

Acquiring dynamic surface behaviors with markerless optical capture, rather than with 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 variations of input dense surface mocap under global exaggeration. As illustrated in Fig. 3.13, real-world captured animations need more than the natural behavior of skeletons to express local shape variations. Hence, cage-based deformation offers an elegant free-form abstraction to capture the low-frequency shape deformations as-much-as possible with reusable output parameters as shown in Fig. 3.23.

3.4.2 Cartoonization of Multi-view Data

Life-Like Non-photorealistic Animation. Inspired by the philosophy adopted for the production of *The Adventures of Tintin: Secret of the Unicorn*, a recent 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. Since full-body performance capture is limited by the physicality of real actors actions, we add 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*.

Structure-Preserving Editing. The analysis of shapes in motion is the very foundation of structure-preserving editing from performance-based captured animations. Recent advances in dynamic surface capture have made the creation of realis-

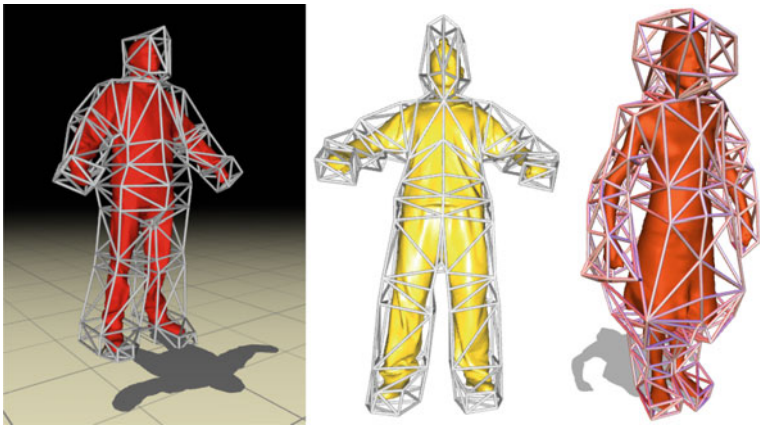

Fig. 3.23 Generic cage-setup: a default humanoid-type cage connectivity is used to perform scalable editing of whole-body models. This abstraction of skin behavior can be adjusted to a large configuration of human body topology such as girls in skirt or men in trouser. We bound the space of laser-scanned body template with a cage containing a compact set of controllable handles distributed in a shape-aware manner

tic 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 and multi-camera data. 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 marker-less surface performance capture and extreme globally-coherent stretching are key challenging problems. Synthesizing new puppetry animation is a fascinating research problem with direct applications in industry, espacially to demonstrate fidelity to the spirit of TV cartoon with more exaggerated motion while preserving captured extreme cloth wrinkles of a real actor performance.

Content-Aware Exaggeration. 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 non-rigid mesh 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.

Expressive Life-Like Animation. Our main motivation is to explore opportunities offered by underlying subspaces to reuse time-varying captured data into an 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 section, we are devoted to push 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. Here, 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.

Animation-Cartoonization. Our novel key contribution is an interactive process of *Animation-Cartoonization* from multi-view data. We propose a new approach to obtain video-based toon character from surface performance capture. 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 content-aware cartoon-style expressive rendering. Consequently, the key contribution of our work focuses on a novel cartoon stylization approach for 3D video that efficiently reuse temporally consistent dynamic surface sequences, captured from real-world actor performance.

3.4.3 Space-Time Exaggerating of Life-Like Surfaces

Animation Cage-Space. The continuous dynamic surface representing the actor performance is denoted by:

$$\mathbb{S}^t = \left\{(x, y, z) \in \mathbb{R}^3,\ f\ (x, y, z, t) = 0\right\}$$

and stored by a discrete sequence of q non-rigid temporally consistent meshes denoted by $\mathcal{A} = \left\{\mathcal{M}_1(F, V_1), \ldots, \mathcal{M}_q(F, V_q)\right\}$ 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 ith 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

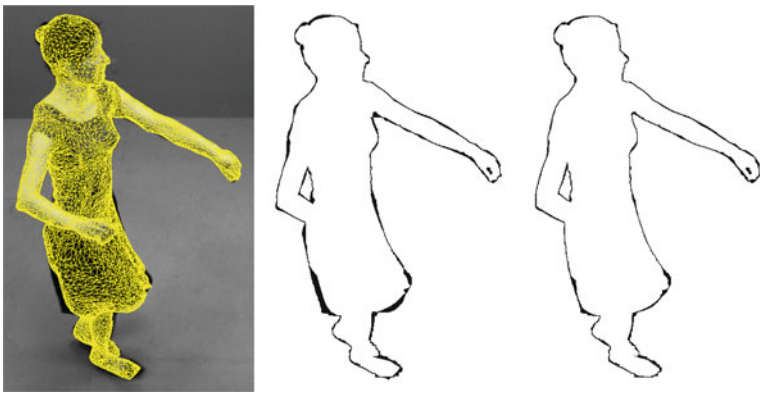

Fig. 3.24 Silhouette-aware cage-based encoding: backprojected cage-based mesh congruence with the corresponding captured image (*left-hand side*), silhouette overlap error of the cage-based mesh with the extracted real silhouette (*middle*), silhouette overlap error of the original model with the extracted real silhouette (*right-hand side*)

by these handles. Given the original input dense mesh sequence $\mathcal{A}$ and the specified bounding polyhedron $\mathcal{B}$ associated with given rigging function w, a low-dimensional non-rigid surface motion signal is a function $s(t) = (\mathbf{c}_1^t, \ldots, \mathbf{c}_m^t)$ 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 jth cage-handle in global coordinate system at a given time t. Figure 3.23 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 Fig. 3.24.

Given the fact that achieving global perturbation directly on the surface itself seems 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.

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

where $w_{ij} : \Omega \to \mathbb{R}$ 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 by Joshi et al. [42], retaining substantial global-to-local surface characteristics.

The cage tessellation with a given rigging function restricts deformations to a space of natural warping because human activities are often located on a latent space

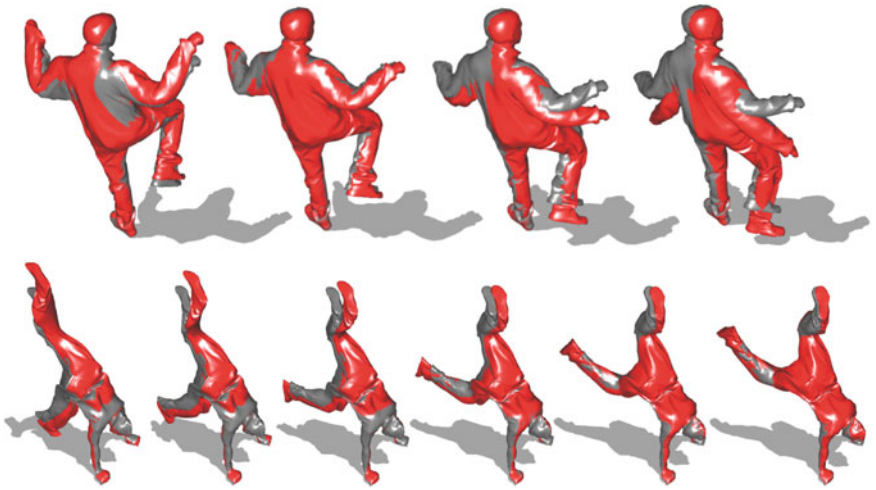

Fig. 3.25 Handle-and-motion aware surface filtering: cage parametrization is a meaningful low-dimensional subspace enabling analyze-to-synthesis approach on motion variation in body shapes that evolve non-rigidly over time. Thus, we boost the captured surface dynamics (in *red color*) by applying the *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

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 reskin the enclosed surface, without losing too much quality as seen in Fig. 3.24. This allows 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 drastically.

Cage-based Subspace Filtering. 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 [51], 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 Fig. 3.25. To achieve this, we obtain the filtered cage-based shape geometry directly, using the following linear cage-based filter operator:

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

The extracted signal is rearranged to form a set of per-handle trajectories defining 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. The formulation of the Laplacian of Gaussian with a standard deviation $\sigma = 1.4$ is provided as follows.

$$LoG(x, y, z) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2 + z^2}{2\sigma^2}\right] exp[-\frac{x^2 + y^2 + z^2}{2\sigma^2}]$$

Cage-based Interpolation. 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 $\widetilde{c}$.

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

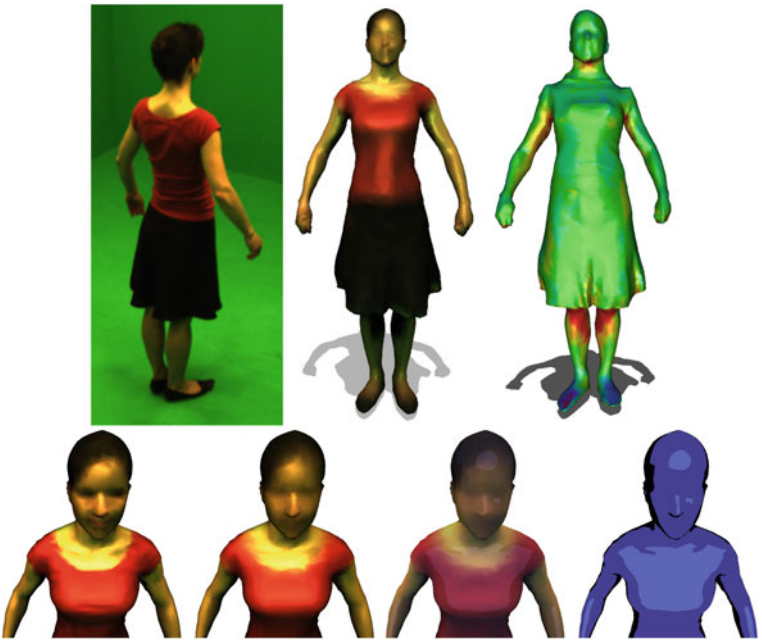

Fig. 3.26 Appearance editing: (*Top*) Captured images are projected on the model, bypassing partial hidden cameras visibility. (*Bottom* from *left* to *right*) The appearance model is performed in various steps: multi-view projective texturing, per-vertex filtering, and per-pixel cartoon style enhancement. The traditional toon shading is also displayed for comparison with our final result

3.4.4 Depicting Video-Infused Appearance

Life-likeness Appearance. The second step consists in establishing video-driven expressive appearance with respect to the sequence of captured images, as shown in Fig. 3.26. We conduct an inferred-to-depicted approach to obtain a richness and life-likeness expressive appearance. We assume 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 ith vertex is obtained by a per-vertex multi-view projective function described as follows.

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

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

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

where for each projected texture $(\mathbf{n}_i \cdot \mathbf{e}_l)$ is the angle between the outer-pointing surface normal at the ith vertex denoted by $\mathbf{n}_i$ and the viewing vector pointing toward the direction of the considered lth camera denoted by $\mathbf{e}_l$. 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 lth camera. In addition, the projection operator $\Pi_l\,(\cdot)$ associated to the projection matrix of the lth 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 lth camera at time t for the given image coordinates. The resulting weight determines the importance of the input camera for ensuring photo-consistent blending.

Per-Vertex Color Filtering. Merging the multi-view footage into an average over-time color distribution encodes the texture 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 reflectance. 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–Concentration* process over textural elements that also ensures global seamless intensity. To accomplish this,

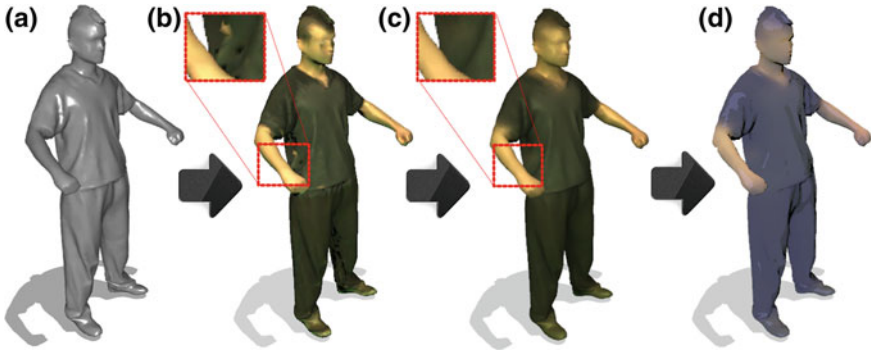

Fig. 3.27 Appearance generation is performed in various steps: captured surface reprojection (**a**), multi-view texture mapping (**b**), per-vertex filtering (**c**) and per-pixel cartoon style enhancement (**d**)

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 ith vertex, as described as follows.

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

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

Stylized Rendering Equation. As illustrated in Fig. 3.27, 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. [48]. 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:

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

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 to amplify more cartoony effects. 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 in the cartoon world. 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.

3.4.5 Experimental Results

To demonstrate the effectiveness of our techniques, we have implemented a straightforward prototype software. In this section, we briefly describe our encouraging results, and we discuss related observations and limitations. Then, the effectiveness of our system is evaluated on performance mesh animations, reconstructed from multi-video sequences by Vlasic et al. [10].

Toon Performance Capture. Moreover, we validate the proof of concept of our video-based toon generation from surface performance capture, by applying our *Animation-Cartoonization* algorithms on the output of Vlasic et al. [10] composed of mesh animations and captured image sequences. Hence, we demonstrate the efficiency of our methods on various challenging datasets, particularly on the samba dance sequence, where highly non-rigid deformation appears on the wrinkled skirt. Large deformation are also emphasized by employing a stylized subspace ables to ensuring the near-conformality of original small-scale details as-much-as possible under global shape exaggeration. The *cage-based surface filtering* process is performed at interactive rate, assuming that *Harmonic Coordinates* are computed as pre-process. Furthermore, as seen in Fig. 3.28, the video-infused appearance produces visually aesthetic effects at real-time rate. Finally, we illustrated the capabilities of our techniques to produce mixed animation-oriented editing such as elongated adjustment of the boneless structure in motion, automatic handle-aware filtered captured surface and inferred-to-depicted life-like appearance. Figure 3.22 illustrates the overall behavior of our techniques. We believe our proposed techniques are vivable enough to be extended for the low-budget productions of games and cartoon featured movies.

Fig. 3.28 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 on the Tilke model (*left-hand side*) and the mohawk hairstyle of Daniel (*right-hand side*)

3.5 Discussion

Silhouette-Aware Cage-based Fitting. Our harmonic cage-based estimation allows rim vertices to preserve silhouette consistency. We think that our techniques offer potentially better reskinning than the skeleton-based approach. This cage-based optimization could be a useful tool to improve the incremental reconstruction across time, because this formulation can provide a better control over the surface to allow rim vertices to fit the silhouette without prior knowledge of rigid parts. In this chapter, we demonstrate the strength of *Harmonic Coordinates* used inside a linear minimization framework to reskin an enclosed animation. We show that our method is well-adapted in the context of a multiple-views setting with a proper experimental validation. Even if cage-based parametrization implies a restriction on degree of freedom in term of deformation, we demonstrate that our method is well-adapted for highly non-rigid deformation. The cage plays the role of a mathematical subspace basis, not rendered at the final stage of the rendering process. Then, the disturbing visual distortion of the cage due to the small-scale variation sensitivity is not really problematic because the deformed model keep pleasant aspect and the cage is not dedicated to be rendered. Even if the input dataset of mesh sequence is globally articulated but highly non-rigid locally because of cloth wrinkles, our technique seems perfectly well-suited for non-articulated deformable shapes. Our method is efficient to handle plausible deformation for video-based animations.

Animation Cartoonization. Since capturing-to-reusing animation is a challenging technical task, *Animation-Cartoonization* must preserve the life-likeness of captured data from performers, whatever animators do over animatable structures. We propose conceptually well-formed editing techniques reducing the amount of skilled artistry to create compelling exaggerated editing. In this chapter, we study an abstraction tailored to ensure the baked-in likelifeness of shape, motion and appearance under perturbations. We push captured animation forward its natural physicality thanks to the content-aware editing. To the best of our knowledge, we propose the first filtering technique for cage-based dynamic surfaces in conjunction with an inferred-to-depicted free-viewpoint appearance.

Performance Capture Stylization. In this chapter, we also revisited the *illusion of life* principle by emphasizing the motion and appearance expressivity of performance capture meshes thanks to 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. 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 surface motion parameters as-well-as the reconstructed appearance. Then, we take advantage of the stylized shading to hide serious texel artifacts caused by the video-based texture mapping. Finally, injecting non-rigid cartoon exaggeration in video-based life-like surface brings a heightened believability to exaggerated realistic-looking toon characters.

Compression and Transmission. Finally, our novel approach is also intriguing for 3D video compression and animation reuse according to the reduction parameters and the compact representation offered by the coarse cage geometry. Our approaches could greatly reduce the cost of storing film full-length of deformed models constituting the captured animation sequence. In the context of 3D immersive and collaborative telepresence applications such as techniques of Gross et al. [52] where data must be transmitted massively over the network, our techniques will probably provide suitable output for fast and efficient animation transmission, without requiring a progressive transmission as proposed. Comparison with other compression techniques is left for future work.

Benefits. A cage structure is a skin-detached low-dimensional subspace, well-suited to drive real-time high-resolution animation. To summarize, boneless cage-based skinning seems to be intuitively more convenient than kinematic hierarchy-based approach to guarantee the fidelity to multi-video data. The main advantage of our algorithm is the simplicity of our energy functional, evolutive enough to be widely implemented and tested with previous framework. Moreover, our least-squares formulation with linear constraints has a closed-form global minimum. The proposed method can speed up animation workflows in current post-productions using sensor-based animations. We exploit the benefit of small controllable subspaces to provide interactive and accessible editing process, to non-expert animators. A sequence of

fitted cage can be reused to produce a large variety of mesh sequences by rebinding the default cage to the same enclosed model at the various resolution. 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. This effect is particularly interesting to study more advanced stylization tools for performance capture. Nevertheless, we are carefully conscientious that our techniques can be improved to generate more elaborated animation puppetry from captured live-action.

Limitations. 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. Building a proper bounding cage is very tricky for non-artist user, especially if the cage is fairly coarse. Even though, the purely geometric tessellation of cage allows low-rank approximation and re-skinning of highly deformable animations, this detached parametrization may fail in case of topological changes. Commonly with all subspace-based techniques, the major limitation of our method is that the deformation result depends on the specified cage tessellation. 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 large deformation. The resulting material shading can obscure important features and can be blurry because of the texture filter employed for removing the noise. Our methods open promising opportunities in the field of non-photorealistic rendering from sensor-based animation.

3.6 Conclusions

Cage-based Optimization. Even if there is a strong interest for *template-based approaches* and *multi-view performance capture*, to the best of our knowledge no previous work tried to use cage-based parametrization for reskinning performance meshes, before this work. In this chapter, we have investigated opportunities in-between *cage-based reskinning* and *cage-based stylization* with numerical optimization. We have developed a space-based deformation framework incorporating cage-based optimization in the context of multi-view setup. To deal with a generic shape abstraction and the local deformation behavior of the input captured data, we uniquely rely on manually designed default cage.

Future Work. In the future, we plan to investigate the possibility of achieving incremental 4D reconstruction, without relying on pre-aligned shape dense sequences. Our cage-based fitting method could be improved easily by integrating several image-based reconstruction cues such as sparse features like surface texture, silhouette and motion features observed in multiple viewpoint images. In addition, we would like to investigate more accurate coordinate systems in order to obtain better quasi-conformality, local detail preserving and low-distortion parameterization with respect

to the captured meshes. Concerning the stylization of performance capture data, we plan to explore more opportunities to preserve the life-like visual appealing. Moreover, it may be of interest to run user-study, and to test different type of motion and appearance filters, tightly coupled with subspace-based surface skinning. As the first study exploring the potentiality of combining cage-based stylization or reskinning for performance capture mesh animation, this work opens up a lot of potential extensions for future research.

References

1. R. Poppe, Vision-based human motion analysis: an overview. Comput. Vis. Image Underst. **108**, 4–18 (2007)
2. S.I. Park, J.K. Hodgins, Capturing and animating skin deformation in human motion, in *SIGGRAPH '06*, pp. 881–889 (2006)
3. S.I. Park, J.K. Hodgins, Data-driven modeling of skin and muscle deformation. ACM Trans. Graph. **25**, 881–889 (2008)
4. P. Sand, L. McMillan, J. Popović, Continuous capture of skin deformation. ACM Trans. Graph. **22**, 578–586 (2003)
5. B. Allen, B. Curless, Z. Popović, The space of human body shapes: reconstruction and parameterization from range scans. ACM SIGGRAPH **2003**, 587–594 (2003)
6. D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: shape completion and animation of people. ACM Trans. Graph. **24**, 408–416 (2005)
7. D. Anguelov, Learning models of shape from three-dimensional range data. PhD thesis (2006)
8. J. Starck, A. Hilton, Surface capture for performance-based, animation. IEEE CGAA **27**, 21–31 (2007)
9. J. Starck, G. Miller, A. Hilton, Volumetric stereo with silhouette and feature constraints. Br. Mach. Vis. Conf. (BMVC) **3**, 1189–1198 (2006)
10. D. Vlasic, I. Baran, W. Matusik, J. Popović, Articulated mesh animation from multi-view silhouettes. ACM Trans. Graph., 27, 1–9 (2008)
11. D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, W. Matusik, Dynamic shape capture using multi-view photometric stereo, in *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* (2009)
12. E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun, Performance capture from sparse multi-view video, in *SIGGRAPH'08*, pp. 1–10 (2008)
13. J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn, H.-P. Seidel, Motion capture using joint skeleton tracking and surface estimation, in *CVPR* (2009)
14. C. Stoll, J. Gall, E. de Aguiar, S. Thrun, C. Theobalt, Video-based reconstruction of animatable human characters, in *ACM SIGGRAPH Asia 2010* (2010)
15. L. Ballan, G. Maria Cortelazzo, Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes, in *3DPVT* (2008)
16. G. Ye, Y. Liu, N. Hasler, X. Ji, Q. Dai, C. Theobalt, Performance capture of interacting characters with handheld kinetics, in *ECCV* (2012)
17. Y. Liu, C. Stoll, J. Gall, H.-P. Seidel, C. Theobalt, Markerless motion capture of interacting characters using multi-view image segmentation, in *CVPR* (2011)
18. L. Onural, T. Sikora, J. Ostermann, A. Smolic, M.R. Civanlar, J. Watson, An assessment of 3dtv technologies, in *NAB Broadcast Engineering Conference Proceedings* **2006**, 456–467 (2006)
19. T. Takai, S.Nobuhara, H.Yoshimoto, T. Matsuyama, 3d video technologies: capturing high fidelity full 3d shape, motion, and texture, in *ISMAR* (2006)
20. T. Tung, T. Matsuyama, Minimal 3d video, in *ACM SIGGRAPH ASIA 2009 Sketches* (2009)

21. P. Huang, A. Hilton, J. Starck, Automatic 3d video summarization: key frame extraction from self-similarity, in *3DPVT* (2008)
22. T. Tung, T. Matsuyama, Dynamic surface matching by geodesic mapping for 3d animation transfer, in *CVPR* (2010)
23. P. Huang, A. Hilton, J. Starck, Shape similarity for 3d video sequences of people. Int. J. Comput. Vis. **89**, 362–381 (2010)
24. S. Nobuhara A. Hilton P. Huang, T. Tung, T. Matsuyama, Comparison of skeleton and non-skeleton shape descriptors for 3d video, in *3DPVT* (2010)
25. J. Starck, G. Miller, A. Hilton, Video-based character animation, in *SCA'05*, pp. 49–58 (2005)
26. P. Huang, A. Hilton, J. Starck, Human motion synthesis from 3d video, in *CVPR'09*, pp. 1478–1485 (2009)
27. J.-Y. Guillemaut, D. Casas, M. Tejera, A. Hilton, 4d parametric motion graph for interactive animation, in *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games 2012 (I3D'12)* (2011)
28. P. Huang, A. Hilton, Surface motion graphs for character animation from 3d video, in *SIGGRAPH 2009: Talks*, pp. 56:1–56:1 (2009)
29. J.-Y. Guillemaut, D. Casas, A. Hilton, Interactive parametric control of mesh sequences, in *SCA'10: poster* (2010)
30. J.-Y. Guillemaut, D. Casas, M. Tejera, A. Hilton, Parametric control of captured mesh sequences for real-time animation, in *Proceedings of the 4th International Conference on Motion in Games (MIG)*, pp. 242–253 (2011)
31. M. Tejera, A. Hilton, Space-time editing of 3d video sequences. *Conference on Visual Media Production*, pp. 148–157 (2011)
32. W. Geng, G. Yu, Reuse of motion capture data in animation: a review, in *ICCSA'03*, pp. 620–629 (2003)
33. S. Kircher, M. Garland, Editing arbitrarily deforming surface animations, in *SIGGRAPH '06*, pp. 1098–1107 (2006)
34. E. de Aguiar, C. Theobalt, S. Thrun, H.-P. Seidel, Automatic conversion of mesh animations into skeleton-based animations. *Comput. Graph. Forum Computer Graphics Forum (Proc. Eurographics EG'08)*, pp. 389–397 (2008)
35. D.L. James, C.D. Twigg, Skinning mesh animations. ACM Trans. Graph. **24**, 399–407 (2005)
36. L. Kavan, P.-P. Sloan, C. O'Sullivan, Fast and efficient skinning of animated meshes. *Computer Graphics Forum (Eurographics 2010)* (2010)
37. D.T.P. Quynh, Y. He, X. Chen, J. Xia, Q. Sun, S.C.H. Hoi, Modeling 3d articulated motions with conformal geometry videos (cgvs), in *MM '11*, pp. 383–392 (2011)
38. I. Baran, D. Vlasic, E. Grinspun, J. Popović, Semantic deformation transfer. ACM Trans. Graph. **28**, 1–36 (2009)
39. F. Xu, Y. Liu, C. Stoll, J. Tompkin, G. Bharaj, Q. Dai, H.-P. Seidel, J. Kautz, C. Theobalt, Video-based characters: creating new human performances from a multi-view video database, in *SIGGRAPH* (2011)
40. J. Tierny, J.-P. Vandeborre, M. Daoudi, Fast and precise kinematic skeleton extraction of 3d dynamic meshes, in *ICPR* (2008)
41. O. Sorkine, D. Cohen-Or, Least-squares meshes, in *Shape Modeling International* **2004**, 191–199 (2004)
42. P. Joshi, M. Meyer, T. DeRose, B. Green, T. Sanocki, Harmonic coordinates for character articulation. ACM Trans. Graph. **26**, 1–71, (2007)
43. E. de Aguiar, C. Theobalt, M. Magnor, H.-P. Seidel, Reconstructing human shape and motion from multi-view video, in *CVMP*, pp. 42–49 (2005)
44. J. Kwon, I.-K. Lee, The squash-and-stretch stylization for character motions. IEEE Trans. Vis. Comput. Graph. **18**, 488–500 (2011)
45. C. de Juan, B. Bodenheimer, Cartoon textures, in *Proceedings of SCA'04*, pp. 267–276 (2004)
46. D. Sýkora, M. Ben-Chen, M. Čadík, B. Whited, M. Simmons, Textoons: practical texture mapping for hand-drawn cartoon animations, in *Proceedings of NPAR'11* (2011)

47. D. Kang, J.-M. Chung, S.-H. Seo, J.-S. Choi, K.-H. Yoon, Detail-adaptive toon shading using saliency, in *VIZ '09*, pp. 16–20 (2009)
48. P. Barla, J. Thollot, L. Markosian, X-toon: an extended toon shader, in *Proceedings of NPAR'06*, pp. 127–132 (2006)
49. S. Rusinkiewicz, M. Burns, D. DeCarlo, Exaggerated shading for depicting shape and detail. *ACM Transactions on Graphics (Proc. SIGGRAPH)* (2006)
50. M. Eisemann, B. De Decker, M. Magnor, P. Bekaert, E. de Aguiar, N. Ahmed, C. Theobalt, A. Sellent, Floating textures. *Computer Graphics Forum (Proc. of Eurographics)* (2008)
51. J. Wang, S. Drucker, M. Agrawala, M.F. Cohen, The cartoon animation filter, in *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, pp. 1169–1173 (2006)
52. M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. Vande Moere, O. Staadt, blue-c: a spatially immersive display and 3d video portal for telepresence. ACM Trans. Graph. **22**, 819–827 (2003)

Chapter 4
Toward Non-rigid Dynamic Cage Capture

This chapter explores the problem of non-rigid alignment using cage-based parametrization to obtain consistent dynamic meshes without any assumptions on temporal matching. Then, we describe a reconstruction framework to retrieve 3D geometric information from multi-view silhouette cues at each time step. To achieve this goal, we establish robust temporal correspondences relying only on geometric features. This investigation toward cage-based registration from multi-view silhouettes is a first insight in the context of dynamic mesh capture with controllable, reusable and low-dimensional outputs designed for commercial animation engines. We finally provide a series of results using our iterative cage-based registration technique in order to obtain a consistent dynamic mesh structure.

4.1 Introduction

Cage-based Registration. Over the decade, spatial and temporal reconstruction from multiple views image sequences is a serious problem in *Computer Vision* as well as in *Computer Animation*. Methods dealing with the silhouettes fusion of an object are a well-explored field in *Computer Vision*. Unfortunately, reconstructing the visual hull at each frame is inconsistent and invites us to investigate spatio-temporal coherency of an incremental shape evolution across time. Unlike methods based on an underlying skeleton or a frame-to-frame shape matching, we use the visual hull in conjunction with a closed bounding cage to provide an output dynamic shape. In this chapter, we propose a new *cage-based registration* technique to recover non-rigid observed shapes using a meaningful animator-friendly and non-rigid reduced embedding. This subspace offers natural silhouette-aware complexity encapsulation and possible reuse of time-varying estimated parameters associated the underlying flexible structure.

Y. Savoye, *Cage-based Performance Capture*,
Studies in Computational Intelligence 509, DOI: 10.1007/978-3-319-01538-5_4,
© Springer International Publishing Switzerland 2014

Markerless Reconstruction Using Deformable Cage. This chapter develops the concept of *Cage-based Spatio-Temporal Reconstruction* where the problem is formulated as the video-based incremental reconstruction of a mesh sequence, using spatio-temporal deformation without any correspondence prior knowledge. The main idea of our method is to build a coarse control mesh to enforce the template local rigidity and to drive the template deformation across time through this underlying parametrization. In this chapter, we also introduce a mathematical framework to estimate cage parameters at each frame using calibrated multi-view image sequences and cage-based deformation. Thus, the proposed method can be seen as a novel constrained least-squares variational approach. More specifically, consistent dynamic shape capture from casual videos is a fundamental task at the cross-fertilization of *Computer Vision* and *Computer Graphics*. Notwithstanding, recent advances in low-cost dynamic scanning turn the cross-parametrization of non-rigid animatable surface into an ill-posed vision-oriented problem. Non-rigid structures in motion are ubiquitous in the world we live in, and their complex behaviors respond to the nature's law. Thereupon, acquiring dynamic whole-body live-action via shape registration is a challenging research task that plays a decisive role in various vision-based animation applications, such as 3D videos.

Consistent Shape Registration. Removing inconsistency in temporal series of unstructured geometric data could demand a heavy computational cost. A common source of strong motion inconsistencies in capture are artifacts such as holes and topological inconsistencies, resulting of the temporally independent generation of two consecutive inconsistent meshes. Shape registration is a fundamental tool in *Computer Vision* that could help to inject some consistent shape parametrization in the captured data, by resolving the problem of shape adjustment. Integrating a low-dimensional parametrization into the captured mesh sequence may ensure temporal consistent globally parameterization through an animatable structure while online capture and may directly offer reusable meaningful parameters to the artist during the capture process. We present a framework for shape alignment for closed genus-zero surfaces. The cage topology associated to the captured shape remains consistent. Deformable models reduce the dimensionality of an inverse problem, allowing us to look for solutions only spanned by the deformation. Shape registration computes correspondences across frames. As output, every mesh in the series share the same connectivity. Then, we inject a one-to-one mapping in term of vertices correspondence between successive generated frames of sequence.

Motivation. Numerical geometry of non-rigid shapes has been extensively studied over the years. Nevertheless, registering highly non-rigid dynamic captured surface is a hard and fundamental problem of crucial importance that attracts an ever-increasing interest in recent years. A potential application of non-rigid registration technology featured in the video-clip G-Force of Flo Rida where non-rigid facial multi-view scans leads to an ear-to-ear head alignment. Separately, cage-based methods have already shown their efficiency to drive expressive non-rigid animation of feature-length animated film. In this chapter, we propose a new understanding of life-like dynamic captured surface of people in presence of clothing. Then, we cast the anima-

tion reconstruction as cage-based registration of a given laser-scanner template mesh, guided by reconstructed geometry information from input multi-view silhouettes.

Goals. Little attention has been paid to the possibility of registering via a meaningful reduced set of parameters, abstracting observed deformations rather than the shape evolution itself. Thus, we aspire to accomplish dynamic shape capture from sparse multi-view videos, by the mean of markerless registration of a static cage-based template and visual hull prior. Our goal is to estimate optimal time-varying target cage location expressing the unknown deformation of the template under unsupervised correspondences. Correspondences are established between the source enclosed model at its current state, and the target visual hull reconstructed from multi-view calibrated silhouettes. Injecting global cross-connectivity in temporally inconsistent representation is the ultimate goal of our investigation. In this chapter, we try to estimate optimal cage parameters expressing a plausible deformation with respect to the unknown physical motion observed in the multi-view video stream. To highlight key characteristics of the proposed solution, we catch the reader's attention that we do not employ any user-specified marker or landmark on the template to force point-to-point correspondence. We assume any collection of target cage, any user-specified correspondence or an example-based learning space of deformation as input. User intervention is only required at the binding step to roughly adjust the geometry of the given generic cage connectivity in order to correctly envelop the template mesh.

Contributions. We address the particularly hard and ill-posed inverse problem of non-rigid registration in the presence of observed surface wrinkles. Consequently, we propose an incremental *cage-based animation reconstruction* strategy. First, we cast the animation-reconstruction of static laser-scanner mesh template from temporally-inconsistent point clouds as an unsupervised cage-based registration (see Fig. 4.1 for the pipeline overview). We non-rigidly register time-varying unaligned dense point clouds with unknown correspondences. Thus, we perform scalable handle-aware biharmonic shape registration, relying on the higher-level of shape abstraction offered by the space-based paradigm. In order to recover a space-time surface as output, we propose an elasto-plastic deformation model allowing free-form shape registration iteratively. Our non-rigid skin-detached subspace preserves the template local properties under noisy variations encapsulated in the captured point clouds.

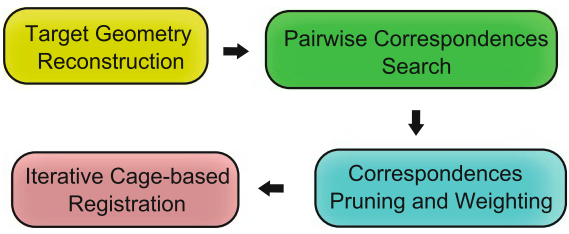

Fig. 4.1 Cage-based registration: Our non-rigid registration scheme has many key ingredients: Target geometry reconstruction, pairwise correspondences search, correspondences pruning and weighting, and iterative cage-based registration

In particular, we try to leverage the problem of highly non-rigid space-time registration by employing an elasto-plastic coarse cage as reduced and deformable geometric model. Elastic-rigidities are expressed through high-level Laplacian-of -Biharmonics control, defined in term of rigged cage-handles. At the heart of our data-driven system, a weighted least-squares formulation allows the production of aligned and highly-detailed boneless meshes. Finally, we successfully demonstrated a first step toward plausible deformation using our algorithm on a collection of challenging real-world datasets composed of anatomical shape in clothing. To the best of our knowledge, our technique is the first to investigate cage-based control for markerless dynamic shapes registration of observed full-body deformation.

Our pipeline enables *scalable curve-based surface registration* of video-driven whole-body performance in a low-dimensional animation-space, loosely decoupled for the shape resolution. More importantly, the output skin-detached curves expresses highly-detailed template from inconsistent temporal scans with a low dimensional consistent space-time subspace. Consequently, non-rigid surfaces can be registered frame-to-frame via a linear combination of biharmonic basis functions, decoupling the state size from the controlled geometry and dynamic complexity of the mesh. To the best of our knowledge, our technique is the first to investigate *handle-aware elastic overlapping-rigidities* for registering life-like full-body dynamic shapes under noisy clothing variations.

Chapter Reminder. The reminder of this chapter is organized as follows. A large overview of related works from reconstruction to registration is offered in Sect. 4.2. Our handle-aware detached registration from multi-view videos is explained in Sect. 4.3, with a novel comprehensive formulation. Then, experimental results with real datasets are detailed in Sect. 4.3.6. A potential extension toward captured casual images is sketched in Sect. 4.4. Next, the overall procedure is discussed in Sect.4.5. Finally, this chapter is concluded, and an outlook to future work is given in Sect. 4.6.

4.2 Related Works from Reconstruction to Registration

There is a terrific deal of recent research that is of interest toward *Cage-based Performance Capture*. In this section, we briefly review some of recent works that are relevant to the problem of image-based shape reconstruction and non-rigid dynamic shape registration for captured videos.

4.2.1 Image-based Shape Reconstruction

Silhouette Foreground Extraction and Calibration. Silhouettes images are reusable information that detach 2D region corresponding to the actor from the background. The problem of foreground separation from background modeling

was addressed in a numerous of work, but relatively few techniques offer robust result under uncontrollable lighting. For instance, Gallego et al. [1] present a foreground segmentation technique for multi-view acquisition system. This segmentation technique takes benefit of the spatial redundancy in the multi-view data to perform the region-based foreground segmentation process by combining segmentation and the 3D reconstruction. The foreground segmentation is improved by iteratively enhancing the shape-from-silhouette volume. Goldlucke et al. [2] prefer to deals with simultaneous depth map estimation and background separation in a multi-view setup with a collection of fixed calibrated cameras. Therefore, the background separation takes benefit from graph-cut in order to gain robustness against noise. Finally, Gong et al. [3] incorporate a pixel-based learning method to modulate the temporal background changes promptly, in conjunction with a *graph-cut* method to propagate per-pixel evaluation results over nearby pixels. Background subtraction techniques are used to obtain actor silhouettes as observed from the different camera viewpoints to be reuse in shape-from-silhouette reconstruction approaches.

Silhouette-based Shape Reconstruction. The reconstruction of a 3D real world scene from a sparse set of silhouettes taken from several viewpoints is a classical problem in *Computer Vision*. Thus, image-based visual hulls became very popular with applications including the human body pose. We notice three classes of shape-from-silhouette approaches: voxel-octree carving, polyhedral reconstruction and probabilistic visual hull. Paving the way, Laurentini [4] obtains a 3D shape for the first time with a volume intersection approach based on the 2D silhouettes of a given object (see Fig. 4.2).

Szeliski [8] and Müller et al. [9] prefer to use an octree as a tree of recursively subdivided cubes for computing octree bounding volume from multiple silhouettes. To optimize memory needed for a volumetric representation, Erol et al. [10] rely on an octree-based visual hull specific adaptive sampling algorithm. Aganj et al. [5] also compute a spatio-temporal visual hull for dynamic scenes. More recently, Knoblauch et al. [11] prefer an approach for volumetric visual hull reconstruction using a voxel grid that focuses on the moving target object. Furthermore, Knoblauch et al. [12] also introduce a region-of-interest visual hull refinement for tele-immersion environments. Finally, Matusik et al. [13] describe an efficient and

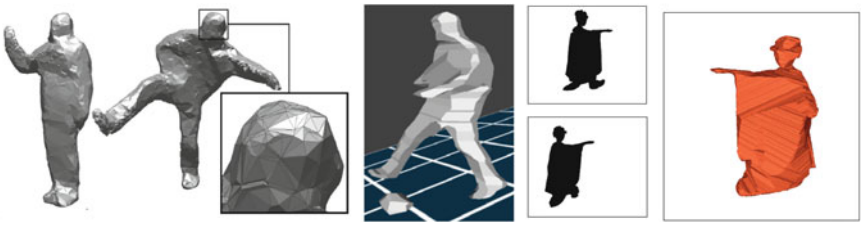

Fig. 4.2 Silhouette-based shape reconstruction: (from *left* to *right*) Spatio-temporal shape from silhouette (images courtesy of [5]), polyhedral visual hulls (images courtesy of [6]), and GPU surface-based shape reconstruction (images courtesy of [7])

important image-based approach to take advantage of epipolar geometry and incremental computation.

Nevertheless, various polyhedral techniques abandoned the need of a volumetric structure. For instance, the pioneering polyhedral representation for the visual hull directly from the silhouettes is offered by Matusik et al. [6]. An advantage is that polyhedral techniques do not requires a voxelization procedure. Münch et al. [14] also exploit the special structure of generalized cone polyhedra. Nonetheless, a large body of existing techniques incorporate a probabilistic model to refine the volumetric reconstruction process. Isidoro [15] proposes a stochastic mesh-based reconstruction with a novel probabilistic framework to estimate surface coloring and object shape. A Bayesian approach to image-based visual hull reconstruction is introduced by Grauman et al. [16] using a probabilistic principal components analysis-based technique to estimate a maximum a posteriori reconstruction of multi-view contours. In contrast, Snow et al. [17] prefer to employ an energy minimization formulation of the voxel occupancy problem.

In addition, the GPU could be used to optimize the computation at high speed. For instance, Ladikos et al. [18] present two efficient GPU-based visual hull computation algorithms to reach real-time performance. In addition, Yous et al. [7] also propose a GPU surface-based shape reconstruction from a set of silhouette images that demonstrates efficiency. Finally, contrary to previous methods limited to indoor reconstruction, only a few techniques similar to the one of Kim et al. [19] preform multiple view reconstruction of dynamic outdoor scenes adapting studio technology to uncontrolled environments.

Surface Polygonization. A mesh-based polygonization procedure is then required to obtain a triangular surface from volumetric carved representation. The so-called *marching cubes* algorithm has been widely adopted for extracting a polygonal surface mesh from a volumetric visual hull. For instance, Mercier et al. [20] use image pixels with marching cubes for constructing a triangular mesh. Milne et al. [21] also employ a marching cubes algorithm to assign polygonal patches to surface voxels. The lack of interpolation information along each voxel edge, which is required by the marching cubes algorithm is overcome in the work of Liang et al. [22] with exact intersections between voxel edges and the visual hull boundary. More recently, a fast meshing algorithm providing a topologically correct interpolation of the surface points is proposed by Salvador et al. [23].

More advanced techniques for polygonization are presented by Hornung et al. [24] to realize robust reconstruction of watertight 3D models from non-uniformly sampled point cloud without normal information. In their approach, the surface maximizes the global confidence and efficiently produce manifold triangle mesh with a minimal number of vertices. Another method to extract iso-surfaces from distance volumes is offered by Wood et al. [27] with a semi-regular surface as output meshes using geometrically adaptive sampling. Moreover, an interesting algorithm for generating a watertight level-set from an octree is presented by Kazhdan et al. [25] where a provably watertight mesh can be extracted from the octree without necessitating the refinement of nodes. Meanwhile, the final optimized surface is obtained as a

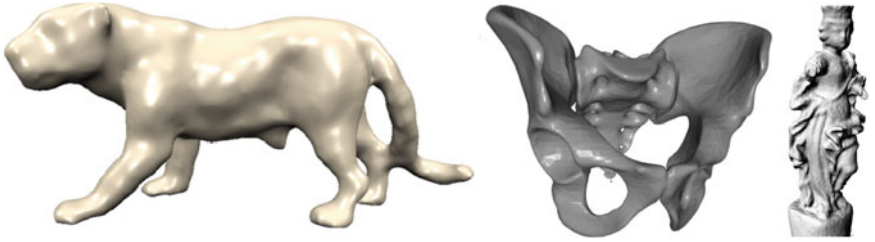

Fig. 4.3 Surface polygonization: (from *left* to *right*) Robust reconstruction (image courtesy of [24]), watertight level-set (image courtesy of [25]), and optimized min-cut/maxflow surface (image courtesy of [26])

min-cut/maxflow solution of the weighted graph by Sormann et al. [26]. Another surface extraction approach using isosurface is used by Goldlucke et al. [28] for polygonization of space-carving input. Sreevalsan-Nair et al. [29] prefer a dual contouring approach as an alternative to standard marching cubes to approximate an isosurface from trivariate data given on a volumetric mesh. Another variant of *dual contouring* polygonization of octree-based reconstructions is achieved by Montenegro et al. [30]. Finally, Hoppe et al. [31] present a general method for automatic reconstruction of accurate, concise, piecewise smooth surface models from scattered range data (see Fig. 4.3).

Multi-view Stereo and Volumetric Graph-Cut. Multi-view stereo has been widely used to refine the visual hull reconstruction. Various researchers investigated silhouettes and stereo fusion for online reconstruction to refine the visual hull [35–38]. More recently, Wu et al. [39] obtain a complete and accurate 3D model from multi-view images by fusing multi-view and photometric stereo for 3D reconstruction under calibrated illumination. Volumetric graph-cut is a key optimization tool in multi-view stereo for Liu et al. [40]. For instance, motivated by approaches where the graph cut algorithm serves as discrete optimization, Tran et al. [41] describe a graph cut algorithm to recover the 3D object surface using both silhouette and foreground color information. Esteban et al. [42] also present an intriguing approach to high-quality 3D object reconstruction by using multiresolution gradient vector flow diffusion approach for the stereo-based energy term. A multi-stereo correlation voting approach and a gradient vector flow diffusion is also used by Esteban et al. [43] to recover the contour generators of the model. A global graph-cut algorithm operating on a 3D embedded graph is also employed by Paris et al. [44].

3D video is obtained by Tung et al. [34] from multi-view reconstruction techniques where dynamic 3D shape reconstruction is operated with unique Markov Random Field using graph-cuts. The problem of obtaining complete, detailed reconstructions of shiny texture-less objects is addressed by Vogiatzis et al. [45] with a formulation of photometric stereo which combines multiple viewpoints to allow closed surface reconstructions. Even if Vogiatzis et al. [46] present a formulation that benefits from a computationally tractable global optimization using graph-cuts, Wu et al. [39] obtain a complete and accurate 3D model from multi-view images

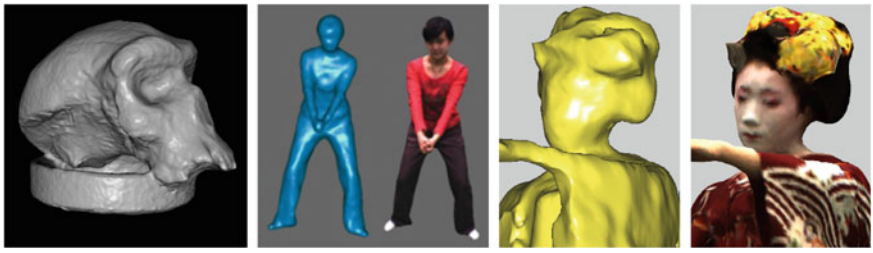

Fig. 4.4 Multi-view stereo and volumetric graph-cut: (from *left* to *right*) Multi-view stereo revisited (image courtesy of [32]), multiview stereo (image courtesy of [33]), and super-resolution using graph-cuts (Image courtesy of [34])

captured using a multi-illumination multi-view stereo algorithm. Ladikos et al. [47] apply a volumetric graph-cut on a narrow band around the current surface estimation to determine the optimal surface inside this band.

Paalanen et al. [48] present a simple multi-view stereo algorithm implemented for GPUs. Even if Sinha et al. [49] also formulate multi-view 3D shape reconstruction as the computation of a minimum cut on the dual graph of a semi-regular tetrahedral mesh, Hornung et al. [50] prefer a volumetric stereo algorithm based on finding the minimum cut in an octahedral graph structure embedded into the volumetric grid. More importantly, Goesele et al. [32] present an extremely simple yet robust multi-view stereo algorithm to merge depth maps into a single mesh using a straightforward volumetric approach. Finally, a point cloud based multi-view stereo algorithm for free-viewpoint video is proposed in Liu et al. [33] to produce the most competitive performance among current algorithms under sparse viewpoint setup according to both static and motion data sets (see Fig. 4.4).

Human Markerless Mocap from Silhouettes. Many existing systems for human reconstruction use model-based fitting driven by silhouette sequences. For instance, Cuzzolin et al. [51] propose an action recognition algorithm in which the image sequences capture a moving human body produced by a significant number of cameras are first used to generate a volumetric representation of the body by means of volumetric intersection. This shape representation is used for training a support vector machine allowing the characterization of human body postures from the computed visual hull. Furthermore, Cheng et al. [52] describe voxel-based hand posture estimation using only voxels to estimate model parameter of a given hand model consists. A non-intrusive approach is presented by de Aguiar et al. [53] to estimate a kinematic model and its parameters of motion from a sequence of voxel-volumes. De Aguiar et al. [54] reconstruct time-varying geometry model of a human actor from a generic human body model and a silhouette-based analysis-by-synthesis method. In addition, Kun et al. [55] propose a new markerless shape and motion capture approach from multi-view video sequences.

Various researchers also introduce *human pose estimation* techniques based on image processing as seen in [56, 57]. In addition, Starck et al. [58] reconstruct a scene captured in multiple camera views based on a prior humanoid model of the

scene geometry. Moreover, Curio et al. [59] present a pose recovery and reconstruction system that combines mapping of image features into a posture space with reduced dimensionality. Next, Caillette et al. [60] propose a new evaluation scheme based on volumetric blobs-fitting reconstruction, where appearance and image features are represented by Gaussian mixtures. Michoud et al. [61] and Yang et al. [62] present a marker-free 3D human motion capture for real-time 3D human visual hull reconstruction with skeleton extraction from voxel.

4.2.2 Non-rigid Shape Registration

Rigid Shape Registration. Shape registration plays an important role in numerous geometric applications both in *Computer Vision* and *Computer Graphics*. Firstly, the *iterative closest point* paradigm was proposed for rigid shape registration. Besl and Mc Kay [63] describe a general approach for computing efficient rigid shape registration by dealing with six degrees of freedom. The iterative closest point algorithm converges monotonically to the nearest local minimum. In addition, Rusinkiewicz et al. [64] propose efficient variants of the original ICP algorithm in order to improve convergence for model-based fitting by identifying six stages for the ICP algorithm: selection, matching, weighting, rejecting and error metric minimizing.

Even if Segal et al. [65] introduce a generalization of the ICP algorithm which takes into account the locally planar structure of both scans in a probabilistic model, the main drawback of this class of approach is its limitation to rigid transformation and its time complexity. However, Jost et al. [66] review various existing registration methods for geometric data and propose an accelerated algorithm using a heuristic approach to find the closest points. Since the first *Iterative Closest Point* (ICP) algorithm employed to minimize the difference between two clouds of points by Besl and Mc Kay [63] a lot of non-rigid ICP techniques are nowadays ubiquitous for fine registration of surfaces (see Fig. 4.5).

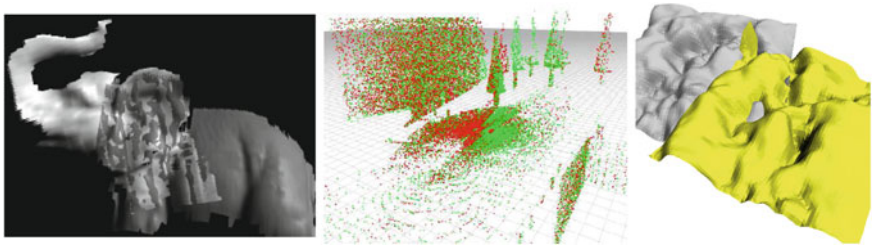

Fig. 4.5 Rigid shape registration: (from *left* to *right*) Variants of the original ICP algorithm (images courtesy of [64]), generalized-ICP (images courtesy of [65]), and registration methods for geometric data (images courtesy of [66])

Fig. 4.6 Articulated-based ICP: (from *left* to *right*) Motion sampling and clustering motion (images courtesy of [72]), dynamic piecewise-rigid object (images courtesy of [73]), and pairwise skeleton correspondence without assume a template (images courtesy of [74])

Articulated-based ICP. The articulated generalization of the ICP for non-rigid spacetime registration has been relatively well-studied over the last decade as seen in [67–69] to recover an articulated model using the skeleton-guided fitting to point cloud. For instance, Anguelov et al. [70] adopt an unsupervised non-rigid ICP technique based on the correlated correspondence algorithm. Furthermore, Anguelov et al. [71] automatically recover an articulated model with a learning technique introducing contiguity constraints and EM algorithm. The unsupervised registration problem is formulated as an automatic label assignment problem via motion sampling and clustering motion by Chang et al. [72]. An algorithm for acquiring the 3D surface geometry and motion of a dynamic piecewise-rigid object is proposed by Pekelny et al. [73] using a single depth video camera.

In addition, an articulated global registration algorithm is performed to reconstructed articulated models from dynamic range scan sequence by Chang et al. [75]. Zheng et al. [74] register captured point clouds using a curve-skeleton and pairwise skeleton correspondence without assuming a template. Meanwhile, Sigal et al. [76] automatically recover a detailed parametric model using a combined discriminative and generative articulated pose. Finally, Chu et al. [77] propose a hierarchical human motion estimation technique using particle filtering and soft-joint constrained ICP with voxel-based reconstruction. In essence, registering dynamic geometric via the skeletal paradigm is well-adapted when the observed data assume articulated deformation, nevertheless this parametrization is not well-suited for non-articulated geometric object or object that does not present sufficient strong rigidness. Contrary to the class of *Hierarchical Articulated ICP* algorithm, we do not assume dynamic piecewise-rigid motion and stick figure deformable model (see Fig. 4.6).

Nearest Neighbor and Correspondence Computation. Approximating nearest neighbor or establishing matches is a key component for ICP algorithms. Arya et al. [78] focus on an optimal algorithm for approximate *nearest neighbor* searching in fixed dimensions by constructing efficient data structures to partition the space. Moreover, a well-adapted formalization to the k-nearest neighbor problem for obtaining the nearest point of a point cloud to a given query is proposed in

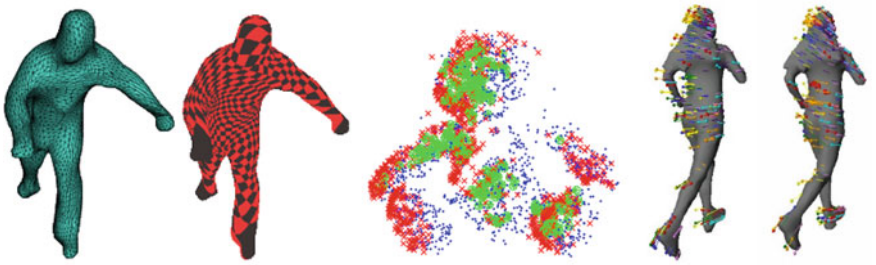

Fig. 4.7 Nearest neighbor and correspondence computation: (from *left* to *right*) Dense 3D correspondence finding (images courtesy of [84]), optimal correspondences from pairwise constraints (images courtesy of [87]), and sparse photometric features (images courtesy of [85])

[79, 80]. Establishing a pair-correspondences between the source surface and the nearest point on the destination surface is proposed by Kok-Lim Low [81].

Starck et al. [82] achieve non-rigid surface correspondence without prior model of human shape or kinematic using a consistent spherical surface mapping, spherical remeshing and matching to estimate dense temporal correspondence. Another coarse-to-fine strategy to perform robust geodesic matching of 3D dynamic shapes is proposed by Tung et al. [83]. In addition, Naveed et al. [84] present a dense 3D correspondence finding using prefiltered SIFT between successive frames and alignment-by-deformation. Doshi et al. [85, 86] propose an empirical evaluation of sparse photometric features useful for non-rigid surface deformation reconstructed from video sequences. Enqvist et al. [87] study the problem of establishing optimal correspondences from pairwise constraints using SIFT and graph method. More recently, Zhang et al. [88] introduce biunique correspondence is introduced to enhance the performance of *Iterative Closest Point* algorithms by searching multiple closest points (see Fig. 4.7).

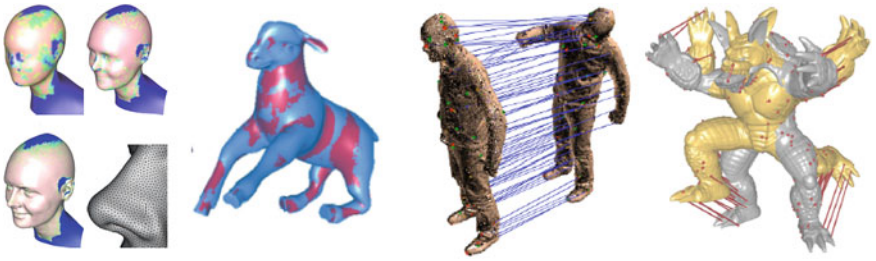

Fig. 4.8 Template-based ICP registration: (from *left* to *right*) Implicit ICP with local prior probability (Images courtesy of [93]), generalized surface flows (Images courtesy of [96]), locally convergent deformable ICP algorithm (images courtesy of [95]), and rigid cluster for surface correspondence (images courtesy of [97])

Geometry Registration and Template-based ICP. Non-rigid Registration was firstly addressed with geometry data using template-based ICP. For instance, Stoll et al. [89] propose a template deformation technique for point cloud fitting guided by user-specified correspondences. Iterative template-based fitting method with Laplacian-based mesh deformation to point cloud is proposed by Yeh et al. [90] with Laplacian-based mesh deformation to point cloud with dual-domain relaxation. Moreover, Hasler et al. [91] also use ICP-based Laplacian mesh deformation approach driven toward a given point cloud of interest for registration. In addition, Schneider et al. [92] also employ an ICP scheme to non-rigidly fit a template mesh to an unstructured point cloud, controlled by a PCA-based morphable model. Finally, various techniques propose non-rigid ICP variant toward geometric registration using an implicit ICP with local prior probability [93], an iterative model assembly allowing hole filling for animation reconstruction [94], and a locally convergent deformable ICP algorithm that works under topological noise [95] (see Fig. 4.8).

More recently, Huang et al. [97] try to estimate rigid clusters over the surface by grouping correspondences that generate same transformation. Amberg et al. [98] use different regularization schemes and collection of landmarks inside an ICP algorithm for surface registration and locally affine deformation. Finally, Eckstein et al. [96] generalize *surface flows* for non-rigid shape matching for near-isometric and quasi-rigid deformation prior.

Elastic Non-rigid Shape Registration. A collection of variant of ICP describes key ideas about *elastic non-rigid* method for registering dense deformable model with coarse to fine strategy [99], incremental deformation [100] and elastic orientation-preserving matching using linear programming to minimize physically motivated energies [101]. The key idea is to detect meaningful correspondences over the surface using variant of ICP and *Integer Linear Program* Formulation. Finally, Sagawa et al. [99] proposed an elastic convolved ICP estimating rigid local transformation according to a weighting scheme for registering a range scan sequence (see Fig. 4.9).

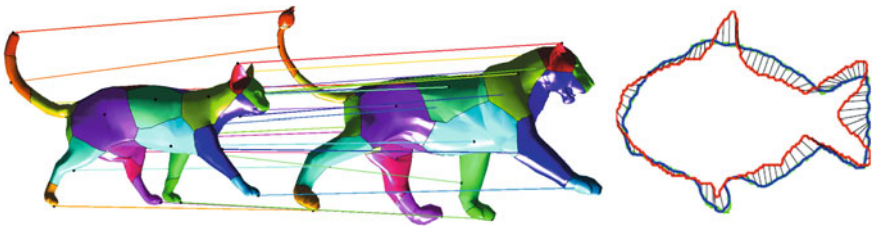

Fig. 4.9 Elastic non-rigid shape registration: Elastic orientation-preserving matching using linear programming (images courtesy of [101]) in 3D, and elastic free-form shape registration (images courtesy of [100] in 2D)

Non-rigid ICP-like Shape Registration. The non-rigid ICP-like shape registration is an important problem for dynamic scanning and shape completion [102].

For instance, Allen et al. [103] create a whole-body morphable model using a non-linear optimization formulation to characterize the space of human body variation. Anguelov et al. [104] propose a data-fitting approach to fit a skinned model to unstructured range scan meshes. In addition, many approaches develop a local-to-global or a coarse to fine hierarchical procedure for the registration process [105, 106], or try to register a set of affine transformation [107].

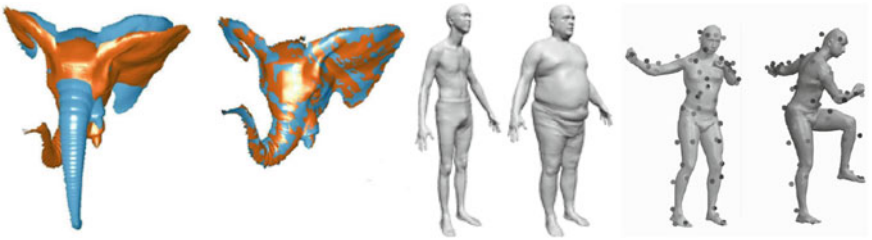

Fig. 4.10 Non-rigid ICP-like shape registration: (from *left* to *right*) Non-linear implicit registration optimization (images courtesy of [112]), data-fitting approach (images courtesy of [103]), and dynamic scanning and shape completion (images courtesy of [102])

More recently, Brown et al. [108] rely on a locally weighted ICP method and Thorstensen et al. [109] describe a variational framework for non-rigid surfaces using matching techniques. Furthermore, various techniques combine normal flow constraints in addition to point coordinates to derive a non-rigid ICP-like scheme [110, 111]. Finally, Cheng et al. [112] establish dense correspondences using non-linear implicit registration optimization and Papazov et al. [113] focus on modeling as-rigid-as-possible shape registration using a closest-point search and a deformation graph for a range scan. Note that establishing dense correspondences can be a complex task in presence of noisy data or data with local variation resulting of cloth deformation (see Fig. 4.10).

Space-Time Non-rigid Surface ICP. We notice a well-known class of non-rigid ICP-like registration approaches addressing the problem of space-time reconstruction from video capture [114]. For instance, Li et al. [115] realize *spatio-temporal registration* with pairwise correspondence and deformation Model. In our work, we also seek globally consistent space-time geometry and we share the same goal with the work of Wand et al. [94] toward reconstructing a consistent frame sequence from with same connectivity a given space-time point cloud. Space-time surface reconstruction is performed using registration flow in [96, 116] but these approaches are limited to near-isometric deformations. Moreover, Mitra et al. [117] describe an algorithm for intra-frame motion between level-set scans. Our approach treats non-rigid registration as an optimization problem similarly to Huang et al. [97], but the video-based nature of our method invite us not to assume approximately isometric deformations. More recently, Münch et al. [110] propose interesting ideas concerning fuzzy point correspondences (see Fig. 4.11).

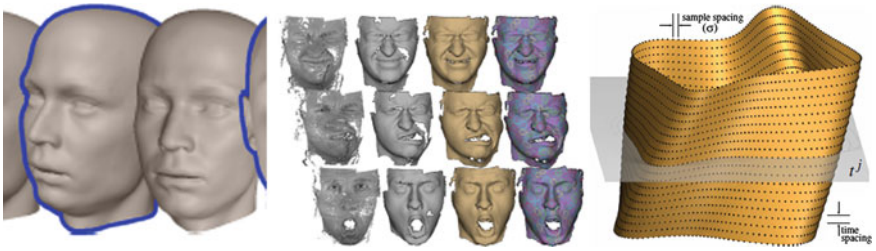

Fig. 4.11 Space-time non-rigid surface ICP: (from *left* to *right*) Surface registration with landmark prior knowledge (images courtesy of [98]), reconstruction of deforming geometry (images courtesy of [94]), and space-time surface (images courtesy of [117])

Vision-Oriented ICP Registration. More importantly, recent years have seen a wide collection of ICP-like techniques achieving impressive registration stage from sensors. For instance, Liao et al. [118] propose a matching technique to reconstruct complete 3D surface deformation over time by a single camera with initial alignment with global warping between two consecutive frames. Cheng et al. [119] also propose a robust non-rigid registration of large difference deformed models using an efficient RANSAC algorithm. Moreover, non-linear optimization for registering partial depth scan is introduced by Li et al. [120] by estimating affine transformation. Once again, Li et al. [121] also propose follow-up approach concerning video-based and template-based surface registration using a non-rigid ICP formulation.

Patch-based approaches are a recent class of vision-oriented ICP-like registration requiring iterative surface segmentation. More importantly, Budd et al. [123] propose an incremental patch-based registration with modified ICP driven by hard-and-soft constraints on volumetric deformation, SIFT matching and geometric correspondences. This approach is extended by Klaudiny et al. [124] using cooperative patch-based method for 3D surface fitting. The term *cooperative* means that all patches are minimized independently to reach a suboptimal solution iteratively. Global temporal registration of multiple non-rigid surface sequences using patch-based approach

Fig. 4.12 Vision-based ICP registration: (from *left* to *right*) Non-linear optimization for registering partial depth scan (images courtesy of [120]), animation cartography (images courtesy of [122]), and patch-based registration (images courtesy of [123])

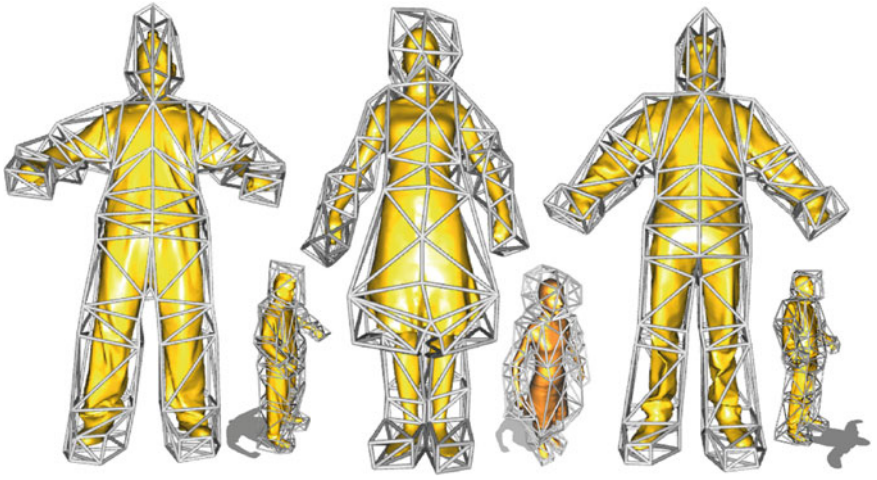

Fig. 4.13 Reconstruction process: Our visual shape registration scheme has many key ingredients: Volumetric spare carving, watertight surface extraction, surface denoising and smoothing, and target point set

is proposed by Huang et al. [125], based on the surface similarity between graphs. Moreover, Budd et al. [126] propose a hierarchical shape matching approach with pairwise non-rigid alignment. The deformation is composed of iterative surface fitting of geodesic patches with appearance and geometric feature in a coarse-to-fine alignment.

Finally, recent complex techniques provide elegant pipelines with efficient vision-based registration. Similarly to temporally coherent completing of dynamic shapes of Li et al. [127] using a local ICP based system, our system relies on the visual hull prior and surface fairing, but with a better level of abstraction concerning the underlying subspace. *Animation transplantation* is proposed by Sussmuth et al. [128] to transplant animation to a different static mesh using pairwise non-rigid deformable retargeting via affine registration. *Animation Cartography* is proposed by Tevs et al. [122] for intrinsic reconstruction of shape and motion with global optimization from dynamic point cloud sequence. A multiple kinect-based scanning system performing a non-rigid full-body human registration using ICP closest point is proposed by Tong et al. [129] (see Fig. 4.12).

4.3 Our Approach: Handle-Aware Detached Registration

Skin-detached Registration. In this section, we describe our *skin-detached registration* technique build upon silhouette-consistent reconstruction of unrelated series of target point clouds. As the key insight, an iterative *elasto-plastic optimizer* solves the cage-handles registration by alternating between normal-guided pairwise correspondences computation from target-to-template and template-based warping. Our non-rigid ICP-like technique is piloted by weight-control to balance between elasticity and plasticity energies. Finally, the registered dynamic surface is resituated according to temporally-registered cage-handle curves and associated space-based deformation (see Fig. 4.14).

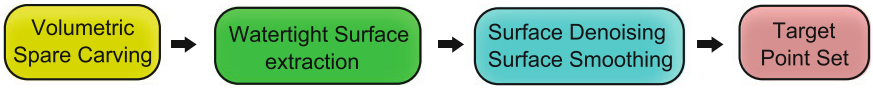

Fig. 4.14 Reusable skin-detached embedding. A generic humanoid-type cage connectivity performs as-conformal-as possible encoding of whole-body models for scalable reediting. This abstraction of skin behavior can be adjusted to a large configuration of the organic topology such as girls in skirt or men in trouser. The same user-specified coarse polytope can bound the space of different laser-scanned body template using a compact set of controllable handles distributed in a shape-aware manner

ICP-like Cage. We propose to achieve the registration process of a deformable object through non-rigid iterative closest point algorithm and Laplacian-based cage-based deformation where constraints over the enclosed surface are projected into the biharmonic cage domain. This ICP-like technique is a straightforward technique inspired from the standard algorithm for registration developed by Besl and McKay [63]. Our ICP-like algorithm is based on a similar iterative procedure converging to local minima, and the iterative loop is broken when the surface change falls below a preset threshold. In our application, the target information is composed of point clouds (with estimated normal information) generated from multi-view reconstruction algorithms.

4.3.1 Non-rigid Registration Setup

Multi-view Full Body Recording. The full-body performance of a real-actor is filmed by a synchronized network of only eight fixed and pre-calibrated high-res pinhole cameras, regularly spaced in a chroma-key green room. We assume the continuous observed dynamic scene is represented by a regular sampled sequence $\mathcal{S}_l = \{\mathcal{I}_l^1, \cdots, \mathcal{I}_l^b\}$ of b foreground-segmented binary silhouette images corresponding to each camera l. We denote by $\mathcal{I}_l^b$ the binary silhouette image associated

to the lth camera at time b. Built upon the estimated 4×3 projection matrix describing the configuration of the lth camera, the projection operator $\Pi_l(\cdot) : \mathbb{R}^3 \mapsto \mathbb{N}^2$, mapping a silhouette pixel coordinates to a given 3D world point included in the camera frustum. Assuming static prior knowledge on the specific shape of interest, we propose to evolve the fixed single-component connectivity $\mathcal{F}$ offered by the given dense, laser-scanned, static template mesh $\mathbb{M} = (\mathcal{V}, \mathcal{F}, \mathcal{G})$ of the corresponding actor to fit the deformation observed in the extracted multi-view silhouette images sequences. The geometry of the closed genus-zero template mesh, composed of n model vertices, is written by $\mathcal{V} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ where $\mathbf{v}_i \in \mathbb{R}^3$ is the location of the ith model vertex. The discrete normal vector field to the template surface is denoted by $\mathcal{G} = \{\mathbf{n}_1, \cdots, \mathbf{n}_n\}$ where $\mathbf{n}_i \in \mathbb{R}^3$ is the normalized surface gradient of the ith model vertex.

Laplacian Cage-based Shape. In the following, we will refer to $\Omega \subset \mathbb{R}^3$ as the bounded polyhedral domain included by m control cage-handles enveloping the motion-less template mesh (see Fig. 4.14). The set of cage-handles is denoted by $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_m\}$ where $\mathbf{c}_k$ is the current location of the kth cage-handle in global coordinates system. Our input cage polytope structure is augmented by the Laplace-Beltrami Discrete Operator $\mathcal{L}(\cdot)$ with non-uniform cotangent weights to enforce as-low-as possible distortion of the irregular cage connectivity [130]. In addition, associated cage differential $\boldsymbol{\delta}$-coordinates encode each cage-handle relatively to its neighborhood in the enclosing cage connectivity [131]. The geodesic-aware volumetric relationship between the cage subspace and the static template is encapsulated by bi-harmonic rigging, performed at the default pose. All of these components help to achieve scalable template mesh registration.

4.3.2 Target Point-Clouds Reconstruction

Carving-based Reconstruction. Shape-from-silhouettes techniques are more adapted since we do not want any assumption on the captured images (that is not the case of other techniques such as shape-from-shading). We present a robust algorithm to reconstruct a watertight triangle mesh from multi-view silhouettes. Assuming the assumption that the foreground can be separated from the background and the surface is non-Lambertian, we choose the non-invasive voxel-based method to reconstruct the markerless visual hull at low volumetric resolutions. The visual hull is the consistent maximum volume, carved with respect to a set of foreground silhouette.

We conserve all voxels that back-projection inside the silhouette images onto the optic ray. We start to compute the voxel carving from a fully occupied voxel grid with low volumetric resolution (we use 128^3 voxels, or 256^3 to better carve out fine details), and then we eliminate voxels that do not reproject inside the silhouette to obtain a visual hull with maximal volume compatible with silhouette from multi-view camera images (see Figs. 4.15 and 4.16). Even if an octree is a more accurate hierarchical structure for sculpturing object digital, we just rely on

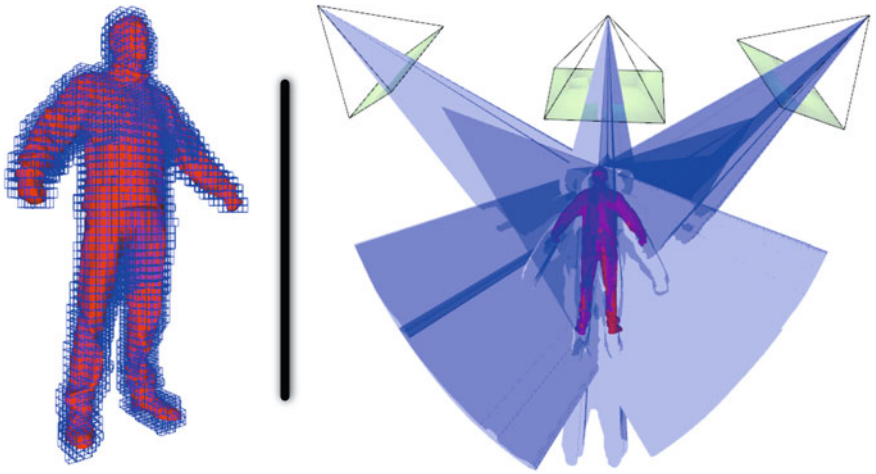

Fig. 4.15 Voxel carving and visual cone: Space-carving driven by fusing silhouettes is compared to the static input laser scanned on the first frame (*left hand side*). According to given calibrated cameras and their foreground-segmented silhouette, the visual cone is traced through the silhouette bins contour. The silhouette defines a back-projected generalized cone that contains the actual object (*right hand side*)

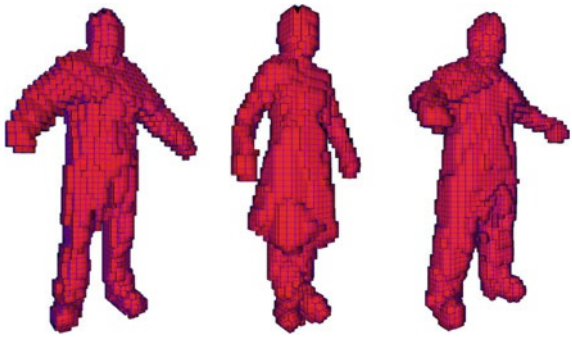

Fig. 4.16 Voxel carving. A 3D volumetric representation is obtained by voxel carving process from multi-view calibrated silhouettes

a voxel grid for now. We define the cone $\mathcal{H}_i$ generated by the silhouette $\mathcal{I}_l^t$ as the set of 3D points that reproject inside the given silhouette. The reconstructed visual hull $\mathbb{V}^t$ defined by the silhouette set can be written as the following cone intersection: $\mathbb{V}^t = \bigcap_{i=1}^{n} \mathcal{H}_i = \left\{x \in \mathbb{R}^3 : \forall l, \Pi_l(x) \in \mathcal{I}_l^t\right\}$. A spacetime isosurface $\mathbb{S}^t = \left\{(x) \in \mathbb{R}^3, f(x,t) = 0\right\}$ can be extracted using a polygonization process of the boundary voxel crust (Fig. 4.17).

Outer Watertight Polygonization. As a first step for the surface extraction, we polygonize the boundary of connex boolean voxels that facing toward the exterior to generate the desired winged-edge mesh connectivity and reconstructed topology. We polygonize a low-genus isosurface by placing a single vertex mesh at the outer-pointing corner of boundary voxels. the *polygon-based crust* formed by boundary

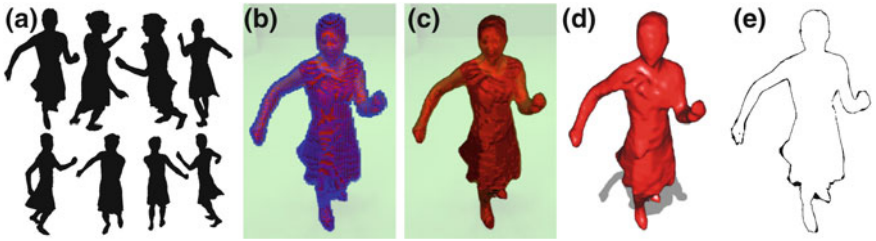

Fig. 4.17 Video-based reconstruction of geometric targets. Extraction of multi-view silhouettes as binary maps (**a**), Voxel-based shape obtained from sparse of 8 converging multi-view silhouettes (**b**), Watertight surface reconstruction (**c**), and Optimized surface (**d**), reprojected in casual images. We compare the silhouette-awareness of the reconstruction by apply the XOR operator between the recorded silhouette map and the silhouette image rendered from the visual hull. This silhouette overlap metric demonstrates an acceptable pixel discrepancy (**e**)

voxels for a wrapper surface. Our meshing algorithm is quite sensitive to dirtiness noise. To generate an unknown oriented watertight shell that maximizes the silhouette consistency, we extracting a topologically noisy watertight meshes connecting outer-pointing faces of the volumetric boundary crust. Consequently, the final optimized output surface is then crack-free watertight (see Fig. 4.18).

Polygonized Surface Smoothing. We apply an iterative surface smoothing operator to improve the surface confidence of the extracted two-manifold, that always bound the ideal surface. The outer-hull vertices are contracted in such way they are lying close to the observed silhouette contours. The results lead to a better geometric thickness and visually silhouette consistent for input viewpoints. This post-reconstruction processing improves the surface fairness by removing high-frequency noise. Then, in a second step we optimize the surface fairness to obtain a pleasant smooth silhouette-aware geometry by attenuating the grid artifact. The extracted surface is denoised by applying an iterative Laplacian smoothing where each reconstructed surface vertex $\mathbf{p}_j$ are moved using the following smoothing operator $\mathbf{p}_j \leftarrow \mathbf{p}_j - \Delta^2(\mathbf{p}_j)/d$ where d is the valence of $\mathbf{p}_j$. The smoothing pass tries to clean the noise induced by the voxel volume discretization. Our dense mesh generation will benefit of further refinement such multi-view volumetric stereo techniques.

Vertex Normal Estimation. Since we obtain an extracted connectivity, we can efficiently and effortlessly compute a normal distribution field. The outward orientation field of reconstructed vertices is obtained by computing the gradient of the watertight surface mesh. Then, the normals associated to each point in the cloud is easily estimated using its associated the surface gradient.

Derivation of Silhouette-Aware Point-Cloud. An unknown watertight surface is roughly reconstructed from multi-view silhouettes, without heavy processing. As far as the connectivity of all reconstructed targets are temporally connectivity-inconsistent, the resulting connectivity is without of interest for the rest of our algorithm. Then, we derive each oriented target point cloud from each optimized

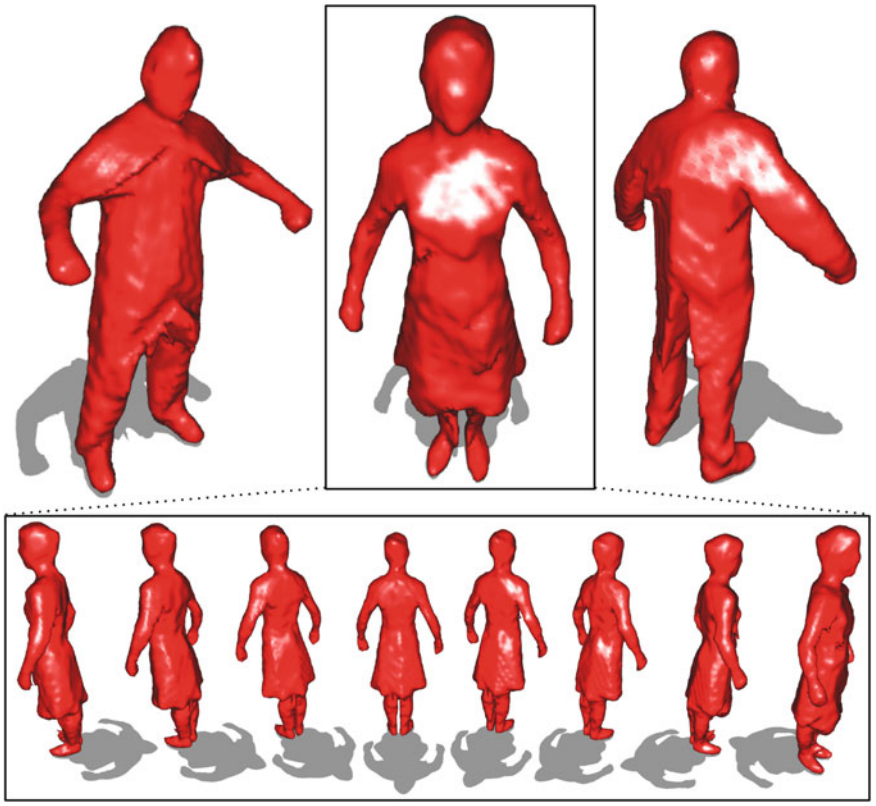

Fig. 4.18 Carved Visual Hulls. A smoothed watertight surface is reconstructed as the isosurface offered by the fusion of silhouette maps (*top row*) to rely on a geometric target prior. Free-viewpoint rendering of the reconstructed shape for a single frame (*bottom row*)

reconstructed surface. Consequently, a collection of temporally independent, unorganized, and uniformly sampled point cloud is naturally derived from the reconstructed manifold (see Fig. 4.17).

Our system does not exploit the target connectivity or any already-known surface flow assumption. As time-varying geometric output, we obtain a sampled sequence of b temporally-inconsistent but oriented point clouds $\mathcal{P} = \{\mathcal{P}_0, \cdots, \mathcal{P}_b\}$ that are partially overlapping. Each point cloud $\mathcal{P}_i = \{(\mathbf{p}_j, \mathbf{m}_j)\}$ defines a set of 3D points $\mathbf{p}_j \in \mathbb{R}^3$ with corresponding normals $\mathbf{m}_j \in \mathbb{R}^3$. If more accurate surface is required, it also possible to refine the surface using multi-view stereo. Because our technique needs an intermediate geometric entity that is reconstructed, our following registration process is potentially effective for low-cost sensors.

4.3.3 Normal-Guided Pairwise Correspondences

Temporally Inconsistent Reconstruction. The independent reconstruction of the visual hull at each frame leads to temporally incoherent surface (no one-to-one vertex correspondence available). The surface extraction process does not necessarily conduct to same topology because of reconstruction artifact, implicit object deformation and surface details (no temporal correspondence can be established inherently). Then, we develop a *Normal-Guided Pairwise Correspondence* algorithm to establish mapping between the template mesh and geometric targets.

Pairwise Correspondences. Since the correspondence information is unknown, we need to add an alignment ingredient in our algorithm. Then, we leverage the problem of temporal matching by the mean of the *Iterative Closest Points* paradigm. To determine reliable data-driven temporal correspondences with no prior knowledge, candidate geometric feature correspondences have to be established automatically by approximating k-nearest neighbors queries for all given targets points [78]. We use a similar fuzzy-yet-robust geometric strategy than Budd et al. [123] to infer a minimal set of compatible feature correspondences updated at each intra-frame iteration. We determine a suitable k-nearest neighbor to establish unsupervised correspondences between the source template and target point cloud. Our strategies significantly improve the data-driven fitting.

We preprocess the template in its current pose into a Bounding Volume Hierarchy structure [132] that optimizes the nearest neighbor search via a space partitioning. Then, the nearest-neighbor queries descend into the tree faster than in kd-tree by minimizing root-to-leaf traversal time. More importantly, we prefer to use the BVH for approximating k-nearest neighbor since this structure can be rebuilt every frame faster than classical kd-tree to handling changing template geometry. We opt for a top-down construction of the BVH following ideas of Wald et al. [133]. While we try to construct soft geometric constraints in order to be injected into an iterative subspace solver, approximating non-exact nearest neighbor search is well-suited for the optimization strategy.

Normal-Guided Pruning. In a first *target-to-source* pass, unsupervised geometric correspondences are established from target-to-source locations because the target is invariant during intra-frame iterations. Consequently, for each $\mathbf{p}_j \in \mathcal{P}$ of the point cloud we compute a limited set of nearest neighbor candidates. For the selection of pair, we use a nearest-neighbor approximation using BVH structure.

$$\mathcal{J} = \left\{ \left(\mathbf{v}_i, \mathbf{p}_j\right) \ : \ \forall \mathbf{p}_j \in \mathcal{P} \ ; \ \exists \mathbf{v}_i \in Knn\left(\mathbf{p}_j\right) \right\}$$

where $Knn\left(\cdot\right) \subset \mathcal{V}$ is the k-nearest neighbors subset of the source template for a given target point query. The resulting set $\mathcal{J}$ is composed of collection of compatible pair $\left(\mathbf{v}_i, \mathbf{p}_j\right)$ verifying the following normal divergence criterion. Conflicting outliers correspondences are eliminated by a normal-guided pruning routine (see Fig. 4.19).

Similarly to Budd et al. [123], the compatible correspondences must satisfy a cost measuring how much the smoothed normal divergence lies under a given threshold:

$$\mathbf{n}_i \cdot \mathbf{m}_j < \cos(45°)$$

where $\mathbf{m}_j$ is the smooth normal direction of a given target point, and $\mathbf{n}_i$ the smooth normal direction of a given source template vertex. Each pair $(\mathbf{v}_i, \mathbf{p}_j)$ are locally weighted according to the following weighting function measuring the divergence in terms of distance and normals:

$$\chi_j(i) = \frac{\max(\mathbf{n}_i \cdot \mathbf{m}_j, 0)}{d}(\mathbf{v}_i, \mathbf{p}_j)^2$$

where $d(\cdot)$ is the Euclidean distance. The numerator is the *clamped dot product* of source and target vertex normals. The denominator is the squared euclidean distance of the source and target vertices, acting as the inverse distance weighting.

In a second *source-to-target* pass, we rearrange the correspondences to construct the ultimate compact set $\mathcal{S}$. To avoid constraints conflicts, a unique target location $\mathbf{q}_k \in \mathbb{R}^3$, for the kth current template vertex, is obtained by averaging compatible target points in the current point cloud of $\mathcal{P}$ to proceed, and the resulting pairwise correspondence $s_k \in \mathcal{S}$ is constructed. In more details, if the kth template mesh vertices is softly constrained by a set of compatible target points (at least one), its final target location $\mathbf{q}_k$ for a given constrained template mesh vertices is defined as the weight centroid of compatible targets. The unknown desired target location is formulated as the centroid of weighted k-nearest multivariate neighbors for a given query. We weight the importance of all neighbors using a common *Euclidean Inverse Distance* function, multiplied by a clamped dot product of pairwise normal.

For each constrained vertex $\mathbf{v}_i$ having at least one pair in J, the corresponding target location $\mathbf{q}_i \in \mathbb{R}^3$ (and associated normal $\mathbf{r}_i \in \mathbb{R}^3$) is the weighted average of congruent target point, obtained by the standard Shepard interpolation procedure:

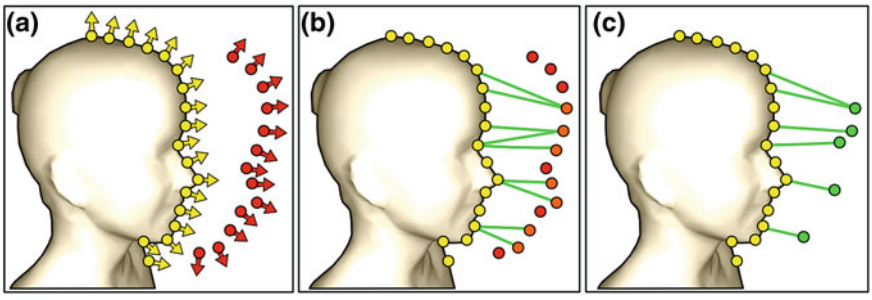

Fig. 4.19 ICP-like Correspondences. Unsupervised geometric correspondences are established from target to source locations, exploiting normal and geometric information (**a**), Reliable data-driven geometric temporal correspondences are established automatically with no prior knowledge by approximating k-nearest neighbors queries for all given targets points (example shown only for correspondences for orange colored target) (**b**), Conflicting outliers correspondences are rejected by a normal-guided pruning mechanism to infer a minimal set of compatible feature correspondences (**c**)

$$\mathbf{q}_i = \frac{\sum_{j \in J} \left(\chi_j(i) \cdot \mathbf{p}_j\right)}{\sum_{j \in J} \chi_j(i)} \quad \text{and} \quad \mathbf{r}_i = \frac{\sum_{j \in J} \left(\chi_j(i) \cdot \mathbf{m}_j\right)}{\sum_{j \in J} \chi_j(i)}$$

After this outliers pruning and aggregation stage, each as-collinear-as possible correspondence is locally weighted defined by $\gamma_k = \max(\mathbf{n}_k \cdot \mathbf{r}_k, 0)$ as the clamped dot product of pairwise smoothed normals of input source and generated target location. The final set of robust locally-weighted correspondences $\mathcal{S}$ is denoted by:

$$\mathcal{S} = \{s_k : (\mathbf{v}_k, \mathbf{q}_k, \gamma_k)\}$$

4.3.4 Iterative Elasto-Plastic Optimization

Cage-based Template Fitting. The frame-to-frame passive shape reconstruction generates a temporal scene flow with the lack of voxel-to-voxel correspondences. Consequently, a constant connectivity is not maintained over time. In addition, reconstruction from sparse views make the ambiguous synthesis of concave area impossible and provoke a *creamy cake* effect on the watertight meshes. These drawbacks motivate us to develop a data-driven registration via cage-based template mesh fitting to obtain a highly-wrinkled dynamic meshes with full correspondence. Ambiguities due to a sparse network of camera and incomplete visibility can be overcome by the mean of a template-based mesh processing. A mesh-based template has a rich shape prior useful to resolve ambiguous space-time reconstruction inconsistency, by restricting the reconstruction to its topology and details. The visual hull-based registration of an inanimate template mesh provides an incremental video-based connectivity-preserving time-varying evolution of input inanimate template.

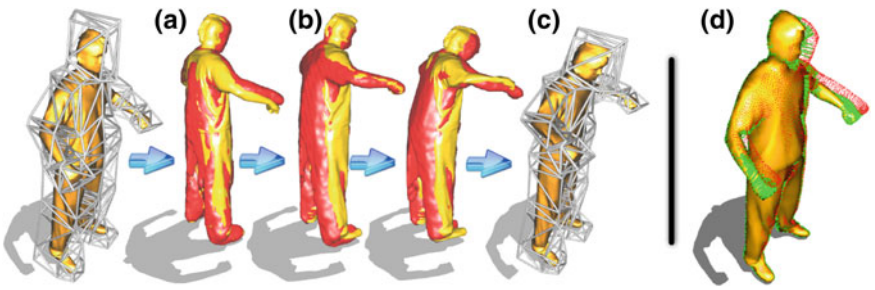

Fig. 4.20 Dynamic registration of life-like organic shapes. A humanoid-type cage is roughly bound to a laser-scanned template mesh, to perform scalable registration of an observed whole-body motion (**a**), We aim at deforming the source template (in yellow color) toward a noisy target point cloud (in *red color*) (**b**), reconstructed from sparse captured silhouettes. Our new skin-detached registration procedure pulls cage-based geometry toward estimated target locations (**c**), Normal-guided pairwise correspondences (in *green color*) are automatically established (**d**)

Registration Process and Solver. Laplace's equation on the cage structure is a partial differential equation enforcing the neighborhood cage curvature coherence by setting the second derivative of equal to zero. Biharmonic fields convey a sense of thin-plate deformation-restriction energy. Global cage-handles locations is estimated by minimizing the sum of squared difference between the current template pose and data-driven matching cues. Then, we solve a single gauss-newton optimization problem expressed in a large sparse matrix-form. Theoretically, the perfect convergence criteria will be reached when the cage-based surface is completely pulled inside the smooth outer voxel hull. Then, we employ a subspace solver to pull the cage-based shape toward the visual hull (see Fig. 4.20).

Iterative Subspace Solver. For each noisy point cloud, our cage-handle curves registration process alternates between *deformation optimization* and *correspondences* successively, in order to approximate surface motion subtle by body deformation and clothing. At each intra-frame iteration u^+, we initialize the cage geometry with location $\mathbf{c}^u$ resulting of the previous iteration u. Then, we update $\mathcal{L}(\cdot)$, the corresponding $\boldsymbol{\delta}$-coordinates and we infer $\mathcal{S}$. Finally, driven by the correspondences propagated in the subspace, a new per-iteration configuration $\mathbf{c}^{u^+}$ is estimated by solving the following the global quadratic variational objective function. The employed factor α and β will be discussed in the next section.

$$\underset{\left\{\mathbf{c}_1^{u^+},\cdots,\mathbf{c}_m^{u^+}\right\}}{\operatorname{argmin}} \left(\underbrace{\alpha \cdot \sum_{j=1}^{m} \left\| \mathcal{L}^u\left(\mathbf{c}_j^{u^+}\right) - \boldsymbol{\delta}_j^u \right\|_2^2}_{\text{Higher-level Laplacian}} + \underbrace{\beta \cdot \sum_{s_k \in \mathcal{S}^u} \gamma_k \cdot \left\| \mathbf{q}_k - \sum_{j=1}^{m} w_{kj} \cdot \mathbf{c}_j^{u^+} \right\|_2^2}_{\text{ICP-like Target Constraints}} \right)$$

where $w_{kj} : \Omega \mapsto \mathbb{R}^+$ is the biharmonic weight for a given cage-handle j with respect to the kth template vertex. Our biharmonic coordinates computation follows core technical ideas proposed in [134]. Our rigging weights are the unique solution, obtained by finite-element approximation to the fourth-order elliptic biharmonic equation with Dirichlet boundary conditions. For the sake of simplicity, we express the diffusing bi-Laplacian kernel $\nabla^4 \equiv \nabla^2\nabla^2$ onto the volumetric cells discretization of the cage interior domain, at the binding step only:

$$E(w_{kj}) = \int_{\Omega} \left\| \nabla^4 w_{kj} \right\|^2 dx$$

The cage-Laplacian regularization ensures the system to be overdetermined and enforces temporally smoothness prior by injecting weak penalty assumption. At each iteration, the vertex positions of enclosed shape are updated by a dense surface deformation field resulting of cage-based deformation with the new cage parameters. Besides, we restituate the registered cage-based shape geometry $\widetilde{\mathcal{V}} = \left\{ \widetilde{\mathbf{v}}_1^{\,u^+}, \cdots, \widetilde{\mathbf{v}}_n^{\,u^+} \right\}$ directly by linear blending of estimated cage parameters, via

the following linear non-rigid skinning operator:

$$\forall i \in [1, n]\,; \widetilde{\mathbf{v}}_i^{\,u^+} = \sum_{j=1}^{m} w_{ij} \cdot \mathbf{c}_j^{u^+}$$

Even if the convergence is reached generally after less than one hundred of iterations, we fix a maximum number of iteration to stop the iterative process just in case of the evolution of the surface does not lie under a specified root mean square tolerance. The general procedure for a complete frame can be seen as solving a non-linear system with an inexact Gauss-Newton method. During deformation, the cage connectivity remains unchanged.

Sparse Matrix Form. This well-conditioned minimization problem can be mathematically solved efficiently and simply using a global weighted least-squares formulation. So, we rewrite this system into its succeeding over-determined matrix-form expression: $\mathbf{WA} \cdot \mathbf{X}' = \mathbf{WB}$, where $\mathbf{X}'$ is a $3\,m$-dimensional unknown objective vector. The left-hand side matrix $\mathbf{A}$ is constructed by stacking the Laplacian elements of the cage and involved biharmonic weights. The $3\,m$-dimensional right-hand side $\mathbf{B}$ is formed by stacking target location of all correspondences and the differential coordinates of the cage. The diagonal matrix $\mathbf{W}$ compiles the weights α, β and γ. Then, the normal equation can also be expressed by:

$$(\mathbf{WA})^T (\mathbf{WA}) \cdot \mathbf{X}' = (\mathbf{WA})^T (\mathbf{WB})$$

Thus, the global location of cage-handles can be therefore found by solving the following classical closed-form expression in interactive time:

$$\mathbf{X}' = ((\mathbf{WA})^T (\mathbf{WA}))^{-1} (\mathbf{WA})^T (\mathbf{WB})$$

Note that the most time-consuming part of the solving process is in the computation of the square matrix $\left((\mathbf{WA})^T (\mathbf{WA})\right)^{-1}$. In practice, we estimate the optimal least-squares cage location $\mathbf{X}' = \left\{\mathbf{c}_1^{u^+}, \cdots, \mathbf{c}_m^{u^+}\right\}$ with off-the-shell Conjugate Gradients algorithm to avoid forming the normal equations and to exploit sparsity of the left-hand side. Then, the optimal solution is attained by a steepest-descent black-box, through walking in the direction of the gradient.

Biharmonic Overlapping Fields. Overlapping fields developed by Au et al. [135] conduct us to uncover that the biharmonic coordinates defined a set of per-handle fields that overlap over the enclosed mesh with meaningful properties. In particular, the so-expected multi-rings local rigidity constraint in non-rigid ICP problems is here implicitly leverage by the combined rigidity field that could be written as $\Psi = \sqrt{\sum_i \|\nabla \varphi_i\|^2}$ for free, as seen in [135]. This formula corresponds to the square root sum of the squared gradient magnitude of biharmonic fields. We denote by φ_i the biharmonic fields associated to each i cage-

Fig. 4.21 Handle-aware local rigidities. Cage-handles with biharmonic mapping convey a sense of handle-aware soft-rigidities influence over the surface. Thus, biharmonic fields produce a local overlapping control. Each colored sub-mesh corresponds to a region of interest where the influence is highest for a given handle. The proximity-rigidity influence smoothly decays non-linearly from the center of implicit patch to the farthest vertex on the whole mesh, if responding to the unbounded heat-like diffusion

handle. This rigidity field can be reused further for shape understanding and shape analysis. Assigning heavy rigidity influences to region that are spatially closed to the handles but decays geodesically far from the handle. These properties are fundamental in our approach to maintain local rigid motion of the surface during registration. Directly related to biharmonic overlapping fields, the cage-Laplacian energy term allows the actual combined rigidity fields to stay smooth during inverse modeling. But most important: it brings non-uniform flexibility over the enclosed mesh, necessary for registration large surface stretching, and that is impossible to realize with other underlying-based ICP techniques previously (see Fig. 4.21).

Output Space-Time Surface. The output of our algorithm is a collection of space-time cage-handle curves, allowing the real-time reconstruction of time-varying mesh sharing the same fixed connectivity derived from the original input genus-zero template. In other words, the overall algorithm output is a low-dimensional non-rigid surface motion signal $s(t) = (\mathbf{c}_1^t, \cdots, \mathbf{c}_m^t)$ describing the resulting shape in motion into the biharmonic projective subspace basis where $\mathbf{c}_j^t \in \Omega$ is a 3-vector indicating the location of the jth cage-handle in global Cartesian coordinates system at the time step t. Storing the resulting sparse-and-detached space-time curves with

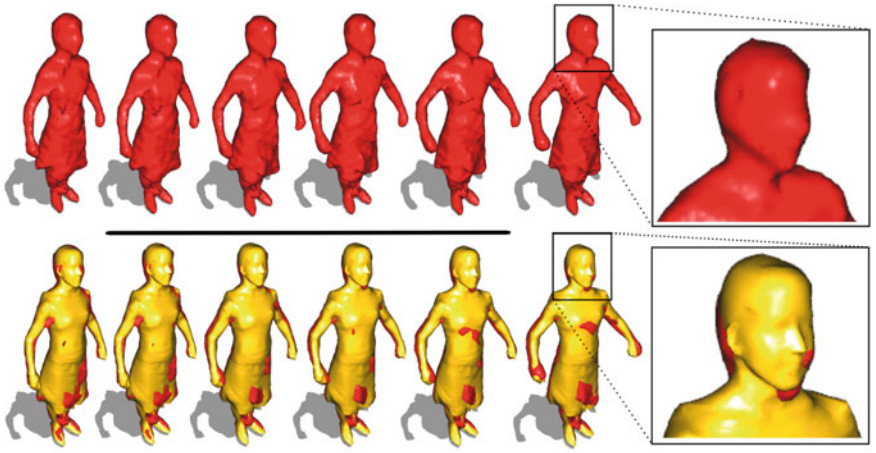

Fig. 4.22 Small-scale detail-preserving. The temporal sequence of unregistered reconstructed meshes from sparse silhouettes is shown in *red color* (*top row*). Superposed in *yellow*, the cage-based surface fitting is driven by unsupervised temporal correspondences automatically established with the reconstructed targets (*bottom row*). Our technique enforces sharp shape details, corresponding to the nose and eyes of the input template, over topology inconsistency of targets as shown in the close-up frames

the pre-computed biharmonic coordinates are sufficient to generate a regular sampled sequence of non-rigid temporally consistent 4D space-time surface meshes $\mathcal{A} = \{\mathcal{M}_0(F, V_0), \cdots, \mathcal{M}_t(F, V_t)\}$ with consistent global connectivity F offered by the input laser-scanned template and V the estimated per-frame cage-based geometry (see Figs. 4.22 and 4.23).

4.3.5 Weight-Control Update Rules

Adaptive Weighting System. We integrate an adaptive weighting system (temporally updated) for the linear constraints in order to have a better deformation quality. Incorporating such system is definitively a useful tool in Laplacian-type editing technique to produce as-rigid-as possible deformation. The problem is: how to defined that weights correctly and what is the strategy to choose for the temporal update of such weights across iterations. The developed technique must ensure the rigidity of the enclosed surface, and the registration must be robust under noisy target point cloud. Preservation of such consistency leads to an iterative optimization problem. The core component of the registration is minimizing the sum of squared distance between nearest pairwise points. The registration is completed when the error change between iterations falls below a selected threshold. The surface flow resulting of the registration can be seen as a displacement field computed by cage-based skinning.

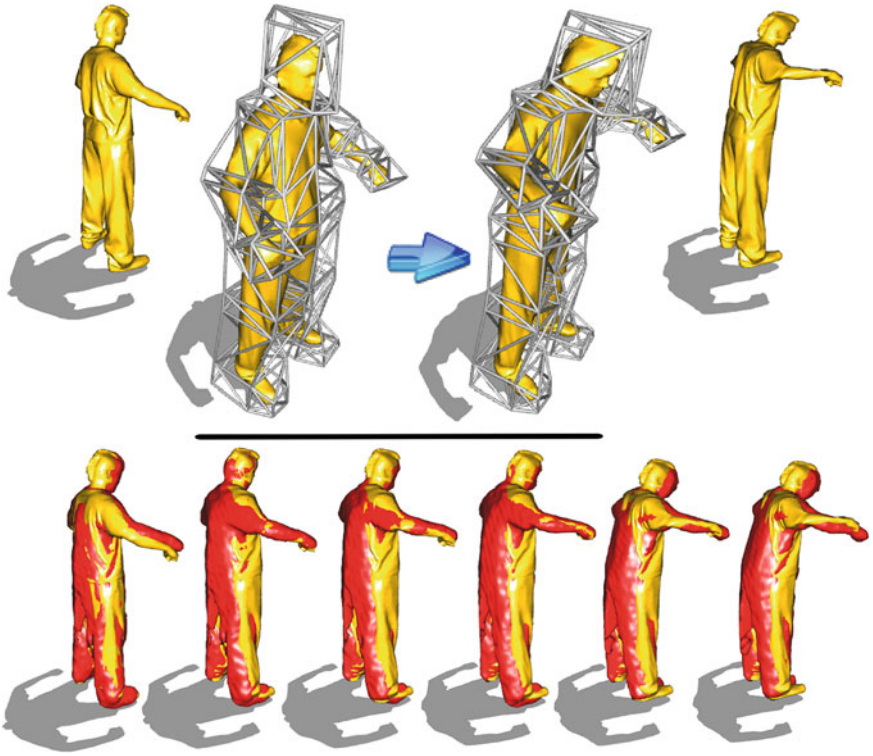

Fig. 4.23 Detail-preserving registration. Our registration process co-registers enjailed surface and cage-handles trajectories (*top row*) over reconstructed noisy point clouds while retaining baked-in template surface details (*bottom row*). The template-based prior is fundamental to overcome ghost-phantom geometry features

Data-Term Enforcement and Stiffness Relaxation. Decomposing the non-linear registration into successive and infinitesimal linear deformation steps yields a careful balance of the objective function terms using an automatic weighting system. Our heuristic rules for adjusting weight controls are motivated by the need of stronger relaxation of the shape prior across iteration for higher flexibility and monotone convergence to a local minimum. This relaxation is plausible because the confidence in correspondences are improved and convergence is closer along iteration. The closed-form formulation of our rules are obtained by experiments and provided as follows. Thus, the data-term weight-control is initialized at $\beta = 0.01$ at each new frame and increased along intra-frame iterations by following an exponential geometric growth as parametric update-rule to promote the constraint-guided energy:

$$\beta(u) = \mathrm{e}^{0.01 \cdot u} - 1.0$$

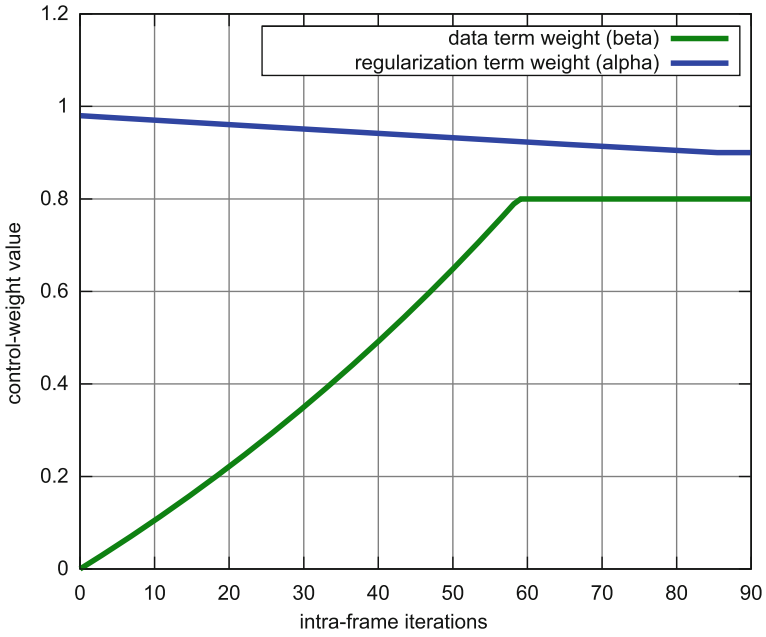

Fig. 4.24 Weight-control update rules. The cage stiffness is relaxed and the deformation prior is enforced across iterations. The evolution of α is depicted in *green color*, and the evolution of β across iterations is depicted in *blue color*

where 0.01 is the growth rate and u is the current iteration number. Concurrently, the weight-control α enforcing the global shape prior is setup to 1.0 at each frame and its importance slightly decreases across intra-frame iteration to gradually relax the deformation stiffness prior, following an exponential decay penalizing the shape resistance behavior:

$$\alpha(u) = 0.99 \cdot \mathrm{e}^{-0.001 \cdot u}$$

where -0.001 is the decay constants, and u is the current iteration number. The value of α is clamped between 0 and 0.8, and the one for β between 1.0 and 0.85 to maintain a coherent balance between energy terms by keeping the shape prior always superior to the data term (see Fig. 4.24). The convergence speed and fitting accuracy are closely dependent of these weight-control update rules.

4.3.6 Experimental Results

Real-World Datasets. We have implemented a stand-alone prototype to demonstrate the robustness and usefulness of our techniques. All of our experiments are executed on a workstation machine with 4 Go of memory and a GeForce 8800

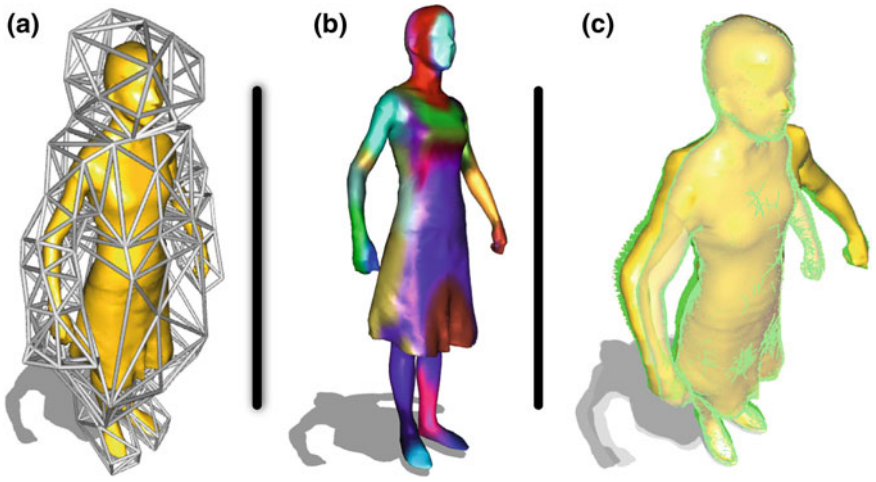

Fig. 4.25 Handle-aware rigidities for detail-preserving registration. An input generic cage is roughly arranged in order to enclose fairly a laser-scanner template having complex skirt topology (**a**), Cage-handles with biharmonic fields produce a sense handle-aware overlapping control over the surface. Each colored submesh corresponds to the proximity-rigidity influence (**b**), The starting pose of the skeleton-free template is superposed with its registered configuration after few movements of samba dance (**c**)

GTX graphic card. We train our algorithm by aligning several real-world datasets of Vlasic et al. [136]. We illustrate the capabilities of the techniques to produce automatic animator-friendly registration (see Fig. 4.25). In all of our experiments, we use the same generic humanoid-type cage of abstraction, represented as a user-specified closed two-manifold.

Computational Timing. This common cage structure composed of approximatively 141 cage-handles is roughly adapted around the input template by a non-skilled artist in a shrink-wrapping fashion. The enclosed static template model is reconstructed using a 3D laser scanner and sampled into 10,002 vertices with 20 K triangular faces. To evaluate the computational process, we measure the timing statistic of each step in the processing cycle. At each iteration, the target point cloud reconstruction takes 20 s, optimizations included. Output reconstructed point clouds have a approximately between 17 K vertices and 30 K vertices. The system spends 126 ms to match the source template with the target point cloud, and 9.2 s to prune and rearrange the correspondences (for a total average of 4, 000 final correspondences at the first iteration, according to our pruning strategy). We fix the number of source vertex candidates to 10 for the nearest neighbor search for each points in the target. The separable solver process takes 9 s to retrieve cage-handles locations and a constant time of 38 ms is required to generate the corresponding enclosed mesh. Our algorithm spends the vast majority of processing efforts in the non-linear optimization process. Even if the results are yet encouraging with few intra-frame iterations

and non-optimal adjustments, much more iterations and finest weight-control are required for accurate results. We would like to point out the importance to have a good weighting system to drive a robust deformation. Finally, quantitative comparison with other similar skeleton-free registration techniques is not directly feasible because implementations are not publicly available and could be quite tricky and time-consuming to reproduce.

4.4 Beyond The Cage: Toward Captured RGB-Images

In this section, we propose a potential extension of our system (not yet incorporated added in our current system) in order to take into account not only the contribution of silhouette cues to form a visual hull geometric prior, but also the captured RGB-Images that could form interesting image-based linear constraints to be injected into the current system. The main idea is to sketch a strategy not just to bet on deforming constraints using geometric information like point cloud, but also to directly rely on image informations.

Good Image Features. Inspired from previous techniques like *3D Scene Flow* where the 3D motion field of points is simply built from the projection of 2D motion field of points in an image plane of the camera, we propose to extract sparse features constraints from texturedness captured images. Such features could be useful in order to obtain more information that is not accessible in the case of an unsuccessful silhouette segmentation. Firstly, the principle of extracting good features to track frame to frame was proposed by Shi et al. [137]. In particular, authors explores how to measure feature dissimilarity based on appearance feature changes. Moreover, *optical flow* is an approximation of the local image motion based upon local derivatives in a given sequence of images. For instance, Lucas et al. [138] present the KLT (Kanade-Lucas-Tomasi) tracker, a feature tracking algorithm that track optimal features in two consecutive images, given the assumption of brightness constancy of small regions, using a coarse to fine optical flow estimation using a pyramidal implementation.

An extension of the classical optical flow technique called *Brox Optical Flow* is also proposed by Brox et al. [139]. More recently, other efficient techniques like SIFT [140] or SURF [141] was introduced to provide better quality features as output. Scale-invariant feature transform (or SIFT) is a classical algorithm in *Computer Vision* to automatically detect and extract local features in images, and SURF (or Speeded Up Robust Feature) is a robust local feature detector. The main limitation is that KLT is a sparse optical flow that does not provide directly a dense one. In addition, the collection of images features can be noisy and resulting matching can content a lot of outliers that are difficult to filter and remove without heavy computation or post-processing on the features. Finally, the robust matching of image-based features remains one of the bottlenecks in *Computer Vision*.

Images-based Correspondences. The resulting sparse optical flow (or SIFT) can be seen as set of temporal coherence hard constraints (see Fig. 4.26). Thus, we propose

Fig. 4.26 SIFT temporal matching: Source image at t (*left*), target image at $t+1$ for the same camera (*middle*), and the resulting estimated set of sparse SIFT features between the two consecutive frames (*right*)

a formulation to incorporate projection constraints based on extracted image-based feature correspondences. Even if we consider a set of image pairwise correspondences, let's explain the case of a single one for the sake of clarity. Each image-based matches between two consecutive frames from the same camera can be defined as the couple formed by a source and target image pixel coordinates. In more details, the source image feature belonging to RGB-image at time t is denoted by its pixel 2D coordinates $\mathrm{p}_a : \left(u^t, v^t\right) \in \mathbb{N}^2$ and the target image feature belonging to the RGB-image from the same calibrated camera at time $t+1$ is denoted by its pixel 2D coordinates $\mathrm{p}_b : \left(u^{t+1}, v^{t+1}\right) \in \mathbb{N}^2$. In fact, by relying on view-dependent projection constraints of unknown 3D points, we force the enclosed mesh vertex that reprojects onto the source pixel to move to a specific unknown 3D position that lies along a ray passing through the target pixel position and having the camera center as its origin. Then, the projection constraints is imposed in the image plane by the optical flow (or SIFT) velocities rather than in 3D. An 2D-3D spacetime correspondences is defined by the quadruplet $(\mathbf{e}, l, \mathrm{p}_a, \mathrm{p}_b)$ where l is the index of the lth camera, e is the index eth enclosed mesh vertex to be deformed and that reprojects into the involved source pixel. The vertex e must be the visible one for the lth camera.

Incorporating Projection Constraints. Then, we formulate bilinear equations that could be easily incorporated in the current system. Since the target position in the image plane is p_b and the projection operator of the lth camera mapping a 3D point into a 2D pixel coordinates is denoted with $\Pi_l(\cdot) : \mathbb{R}^3 \mapsto \mathbb{N}^2$, the reprojection equation of 3D point in images would lead to:

$$\Pi_l(\mathbf{e}) - \mathrm{p}_b = 0$$

In other to incorporate this constraint formulation in our current framework, we need to rewrite the above bilinear equations in term of the cage-based parameters to be estimated.
Given the fact that,

$$\Pi_l(\mathbf{e}) = R_l \cdot \mathbf{e} + T_l$$

where R_l is a 3D square matrix representing the rotational part and T_l is a column 3D vector representing the translation part of the projection matrix associated to lth camera, by mathematical development, we obtain:

$$\begin{aligned}\left(R_l^x - \mathrm{p}_b^u \cdot R_l^z\right) \cdot \mathcal{H}(\mathbf{e}) \cdot \mathcal{C} &= T_l^z \cdot \mathrm{p}_b^u - T_i^x \\ \left(R_l^y - \mathrm{p}_b^v \cdot R_l^z\right) \cdot \mathcal{H}(\mathbf{e}) \cdot \mathcal{C} &= T_l^z \cdot \mathrm{p}_b^v - T_i^y\end{aligned}$$

where $\mathcal{C}$ is the vector that stores cage parameters to estimate, and $\mathcal{H}(\mathbf{e})$ is the vector that stakes *Harmonic Coordinates* associated to model vertex $\mathbf{e}$.

4.5 Discussions

In this section, we discuss the properties of our proposed low-dimensional registration technique. In particular the benefits of reusable registered curve-handles are detailed. Finally, current limitations of the techniques are clearly uncovered.

Abstraction and Generality. The shape-aware abstraction coupled with near-conformal rigging has sufficient flexibility to potentially drive the scalable fitting of versatile geometry template from multimodal sensor. In particular, our subspace-based registration is not limited to single-component template thanks to its detached form and allows the registration of versatile geometric input templates such polygon soup, degenerated geometries such as uncleaned non-manifold or disconnect multi-component template. The intermediate markerless visual hull reconstruction is well-adapted for textureless surface. Our technique is also potentially well-adapted to perform geometric fusion of various sensors. Building animation-friendly registration framework on top of the low-dimensional cage-time structure avoids heavy downsampling or processing of dense point clouds.

Low-Dimensional Registration. The cage parametrization is a well-suited low dimensional subspace for capturing the latent space of human activities observed in images, similarly to skeletons. Our model reduction can relatively deal with the registration of dynamic effects. Furthermore, the purely geometric tessellation of cage allows as-conformal-as possible data reduction, and this is at the main importance to skin target point clouds. Rather than focusing exclusively on surface registration, we pay also attention to the registration of template-independent sparse temporal

curve to compress the data in purely geometric manner. The Laplacian relationship of cage-handles mixed with their biharmonic influences convey a meaningful sense of handle-aware biharmonics scalar fields, controlling the local elastic-rigidity properties over the enclosed surface. Then, the registered surface is restricted to evolve along the geodesics imposed by the space of natural ambient warping.

Elasticity versus Plasticity. The kinematic-free cage connectivity offers sufficient degrees of freedom and the associated rigging functions maintain a controllable proximity influence. On top of that, the Laplacian operator on the cage acts as a linear elasticity tensor between cage-handles and consequently portrays the global elasticity of overlapping biharmonic deformation fields. Meanwhile, cage-handles enriched with biharmonic fields impose meaningful overlapping elasto-rigid influence control and enforce the model registration to be locally smooth and softly rigid along the geodesics, without strong prior on piecewise rigidity. We refer to the theory of plasticity to invoke the change of the underlying animatable structure. In few words, the plasticity is the characteristic of surface to be deformed definitively under external flexural stress. The balance between the linear elasticity and the flexural tensor is adjusted by control-weights controlling the modulus of elasticity and rigidity.

Reusable Registered Handle Curves. Our ICP-like subspace optimization offers a controllable, meaningful, and animator-friendly tool in order to acquire frame-to-frame consistent surface with an unsupervised space-time correspondences strategy. The non-rigid nature of the parameterization allows volume changes of the template, necessarily to gradually approximate surface motion and volume variations subtle by loose-fitting apparel. Besides, our iterative updating framework naturally overcomes the problem of large translation and rotation problem, without the need of user-specified feature correspondences. Since cage-based paradigm is concerning ambient space deformation, our registration technique is versatile and can be employed to enclose disconnected polygons soup, particles or mesh-less points as source template- as all cage-based techniques inherently empowered. Therefore, the estimated cage configuration can be reused to perform scalable shape manipulation over the registered shape, shape interpolation or other type of cage-based transfer applications. The main benefit of our framework remains in its iterative form, preserving the life-likeness of input wrinkled template.

Vision-Oriented Reliability. The proposed technique brings an inanimate mesh template to life by removing topological problem arising with a finite number of cameras. For instance, the detached registration allows visually pleasant and plausible deformation of fine-detail laser surface driven by video-based targets. Such targets are inconsistent and rough time-varying topology caused by an insufficient resolution of the visual hull construction. In addition, our purely sensor-based geometric fitting is not limited to casual multi-view fusion, and could be largely generalizable to multimodal sensors such as kinect®, time-of-flight or structured light scanner able to produce oriented point clouds as targets for our system.

Applicability, Genericity and Reusing. The input cage connectivity is shared between different multi-view sequences and could be employed for inter-sequence

registration. The compact set of registered animator-friendly parameters allows flexible reuse and space-time editing of the sensor-based animation reconstruction. For instance, the resulting cage-based parameters are a powerful information enabling restitution of multiresolution high-quality deformation at real-time with low-cost memory. These properties could benefit to the film-game convergence, as well as cloud-based game through the easy transmission of temporal cage parameters over the network. In addition, the kinematic-free nature of the parametrization does not limit the registration to piecewise rigid or quasi-articulated deformation but extend its applicability to physically-plausible organic motion and non-articulated surface in motion. Intuitively, our non-rigid parametrization offers several advantages over skeleton-based approach to capture the global pose of life-like surfaces that often stretch locally, and this without secondary deformation.

Markerless Reconstruction and Unsupervised Correspondences. The intermediate markerless visual hull reconstruction is well-adapted for textureless surface. Then, a geometric fusion of sensor-based information is required to generate target information that can be injected into a sparse linear least-squares formulation. Furthermore, the minimization framework is flexible enough to incorporate additional linearizable energy terms. Above all, our technique does not require user-specified constraints, rigid clusterization (landmarks) or smoothed template geometry.

Dimension Reduction for Registration. Our formulation reduces the dimensionality of the registration phase while preserving as-faithfully-as possible geometric details that are baked in the specified static template. Building an animation-friendly registration framework on the top of the low-dimensional cage-time structure avoids heavy downsampling of dense point clouds. The deformation restriction imposed by the compact set of parameters described by the coarse polytope prevents local self-intersection. The biharmonic rigging enforces rigidity and smoothness. Control weights adaptation also contribute to pilot this deformation restriction to be biharmonic.

Current Limitations. The inherent limitation of cage-based registration is once again the well-known burden process to arrange cage-handles in a shape-aware manner. This process is realized at the default pose without prior knowledge of the motion or skilled artists. It is worth to mention also the time-consuming process and memory overhead related to the computation of biharmonic coordinates. Actually, this detached underlying parametrization may fail under drastic topological changes. More severely, our approach is subject to other limitations such as non-scale invariance under extra-large motion.

At the bind pose, the need to setup a suitable normal-offseted cage to enclose the mesh is hard to maintain because common geometry interference during the cage adaptation, and may affect the computation of rigging weights. Our method is subject to error accumulation in the presence of unfiltered outlier correspondences and may provoke failures over long sequence. Even if we combat the inherent problem of ICP techniques to fall into local minima, our cage-based registration can conduct to suboptimal results and over-fitting. During our experiments, we notice the difficulty to

control the surface bending precisely when the embedding is too coarse over highly-deformed zones. This extremely dimension reduction implies by the abstraction of a meaningful cage may lead to a solution that does not necessarily correspond to a fair registration in the presence of locally small deformations. Finally, our template-based approach does not handle changes in topology in that current state, where the underlying cage and input template present the disadvantage of having a fixed topology.

4.6 Conclusions

Cage-based Shape Capture. We introduce a new approach called *Cage-based Shape Capture*, offering a simultaneous registration of surface and cage capture. In this chapter, to the best of our knowledge we propose the first approach trying to turn markerless multi-view video into a high-fidelity cage-based registration. Our reduced deformable model restricts the dynamic geometric processing to the closed domain and allows physically plausible deformation. Then, the registration benefits from an iterative feedback between revised *correspondence estimation* and gradual *cage-based surface fitting*.

Cage-based Registration. Our approach is a first insightful step toward the hard problem of automatic template-based registration for highly non-rigid dynamic shape using low-dimensional cage-based encoding. At the first sight, the low-dimensional nature of the underlying parameterization allows us to simultaneous register and compress unstructured input animated geometry. On second thoughts, the implicit Laplacian-of-Biharmonics shape control prevents triangle fold-over or the candy-wrapper effects during the registration process under unsupervised correspondences. Similarly, the main advantage of our iterative optimization remains in the simultaneous cross-reconstruction of dynamic shape and skin-detached registration of reusable curves, retaining template wrinkles details as much as possible. Thus, our system registers non-rigid shape variations while preserves the life-likeness of the input template. Our technique capture consistent surface parameters using an unsupervised correspondence strategy. Thus, cage-based control lies at the heart of our video-based registration method to pilot dynamic life-like template-based capture. Because our techniques are not tailored to work over long sequence, our technique belongs to *registration* and not the *mesh tracking* category.

Future Work. Registering cage-based surfaces opens new directions in the outstanding area of video-based mesh tracking, by allowing the use of non-rigid underlying structures able to track articulated-free animatable surface. Thus, we will focus to realize fine-tuned non-rigid registration for multi-view shape completion. Additionally, we hope to develop tailored ICP-like for determining registration over longer sequences. Other compelling future work would be to considerate the scenario of incomplete, noisy point cloud or versatile targets inputs. We wish to generalize our point clouds fitting technique to a generic human template that do necessarily not

correspond to the observed actor. Several avenues remain for longer-term future work. More importantly, we hope to extend our current pipeline to a template-less cage-free registration approach at video-rate from freely-moving cameras, filming multi-characters interacting in the scene. Toward this direction, we are interested in the temporally coherent completion of life-like dynamic shapes without orientation field estimation and intermediate geometric reconstruction from casual images.

Subsequently, the proposed scheme invites us to pursue ongoing efforts in designing accurate numerical objective functions that converge with less iterations. Rethinking the registration-by-deformation phase by incorporating photometric cues or different kind of linear constraints will be a must to improve the data-driven fitting. A more careful study of confidence weights and comparison with other cage-based coordinates could also be intriguing to run. Ultimately, a major challenge for years to come in this area is to propose an effective non-iterative procedure for off-hours cage-based recording of topology-evolving surface. Assuming no knowledge of a template offers the possibility to introduce an animator-friendly surface registration that undergo large deformations with drastic topological changes.

References

1. J. Gallego, J. Salvador, J.R. Casas, M. Pardàs, Joint multi-view foreground segmentation and 3D reconstruction with tolerance loop, in *18th IEEE International Conference on Image Processing* (2011), pp. 997–1000
2. B. Goldlucke, M.A. Magnor, Joint 3D-reconstruction and background separation in multiple views using graph cuts, in *CVPR* (2003), pp. 683–694
3. M. Gong, L. Cheng, Real-time foreground segmentation on gpus using local online learning and global graph cut optimization, in *CPR* (2008), pp. 1–4
4. A. Laurentini, The visual hull concept for silhouette-based image understanding. IEEE Trans. Pattern Anal. Mach. Intell. **16**(2), 150–162 (1994)
5. E. Aganj, J-P. Pons, F. Ségonne, R. Keriven. Spatio-temporal shape from silhouette using four-dimensional delaunay meshing, in *ICCV* (2007), pp. 1–8
6. W. Matusik, C. Buehler, L. McMillan, Polyhedral visual hulls for real-time rendering, in *Proceedings of the 12th Eurographics Workshop on Rendering, Techniques* (2001), pp. 115–126
7. S. Yous, H. Laga, M. Kidode, K. Chihara. Gpu-based shape from silhouettes, in *Graphite '07* (2007), pp. 71–77
8. R. Szeliski. Rapid octree construction from image sequences. CVGIP: Image Underst. **58**(1), 23–32 (1993)
9. K. Müller, A. Smolic, B. Kaspar, T. Rein, P. Eisert, Octree voxel modeling with multi-view texturing in cultural heritage scenarios, (2004)
10. A. Erol, G. Bebis, R.D. Boyle, M. Nicolescu, Visual hull construction using adaptive sampling, in *WACV-MOTION '05* (2005), pp. 234–241
11. D. Knoblauch, F. Kuester, Focused volumetric visual hull with color extraction, in *Proceedings of the 5th International Symposium on Advances in Visual Computing, ISVC* (2009), pp. 208–217
12. D. Knoblauch, F. Kuester, Region-of-interest volumetric visual hull refinement, in *VRST '10* (2010), pp. 143–150
13. W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, L. McMillan, Image-based visual hulls, in *SIGGRAPH '00* (2000), pp. 369–374

14. W. Matusik, C. Buehler, L. McMillan, S.J. Gortler, An efficient visual hull computation algorithm. Technical report (2002), pp. 1–5
15. J.R. Isidoro, Stochastic mesh-based multiview reconstruction. Ph.D. thesis (2004)
16. K. Grauman, G. Shakhnarovich, T. Darrell, A bayesian approach to image-based visual hull reconstruction, in *Proceedings of the, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 03)* (2003)
17. D. Snow, P. Viola, R. Zabih, Exact voxel occupancy with graph cuts, in *CVPR* (2000), pp. 345–352
18. A. Ladikos, S. Benhimane, N. Navab, Efficient visual hull computation for real-time 3d reconstruction using cuda, in *Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition Workshops* (2008)
19. H. Kim, M. Sarim, T. Takai, J.-Y. Guillemaut, A. Hilton, Dynamic 3d scene reconstruction in outdoor environments, in *3DPVT* (2010)
20. B. Mercier, D. Meneveaux, Shape from silhouette: Image pixels for marching cubes. J. WSCG'2005 **13**, 112–118 (2005)
21. P. Milne, F. Nicolls, G. Jager, Visual hull surface estimation, in *PRASA* (2004)
22. C. Liang, K.-Y. Kenneth Wong, Exact visual hull from marching cubes, in *VISAPP, Institute for Systems and Technologies of Information, Control and, Communication* (INSTICC) (2008), pp. 597–604
23. J. Salvador, X. Suau, J.R. Casas, From silhouettes to 3d points to mesh: towards free viewpoint video, in *3DVP* (2010), pp. 19–24
24. A. Hornung, L. Kobbelt, Robust reconstruction of watertight 3d models from non-uniformly sampled point clouds without normal information, in *SGP '06* (2006), pp. 41–50
25. M. Kazhdan, A. Klein, K. Dalal, H. Hoppe, Unconstrained isosurface extraction on arbitrary octrees, in *SGP* (2007), pp. 125–133
26. M. Sormann, C. Zach, J. Bauer, K. Karner, H. Bishof, Watertight multi-view reconstruction based on volumetric graph-cuts, in *SCIA* (2007), pp. 393–402
27. Z. J. Wood, P. Schröder, D. Breen, M. Desbrun, Semi-regular mesh extraction from volumes, in *Proceedings of the Conference on Visualization* (2000), pp. 275–282
28. B. Goldlucke, M.A. Magnor, Space-time isosurface evolution for temporally coherent 3d reconstruction, in *CVPR* (2004), pp. 350–355
29. J. Sreevalsan-Nair, L. Linsen, B. Hamann, Topologically accurate dual isosurfacing using ray intersection. J. Virtual Reality Broadcast. **4**(4), 12 (2007)
30. A.A. Montenegro, L. Velho, P.C.P. Carvalho, J. Jr. Sossai, Polygonization of volumetric reconstructions from silhouettes, in *SIBGRAPI* (2006)
31. H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, W. Stuetzle, Piecewise smooth surface reconstruction, in *SIGGRAPH '94* (1994)
32. M. Goesele, B. Curless, S.M. Seitz, Multi-view stereo revisited, in *CVPR '06* (2006), pp. 2402–2409
33. Y. Liu, Q. Dai, W. Xu, A point-cloud-based multiview stereo algorithm for free-viewpoint video. IEEE Trans. Visual Comput. Graph. **16**(3), 407–18 (2010)
34. T. Tung, S. Nobuhara, T. Matsuyama, Simultaneous super-resolution and 3d video using graph-cuts, in *CVPR* (2008)
35. H. Schirmacher, M. Li, M.A. Magnor, H.-P. Seidel, Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes, in *IEEE Workshop on Multimedia, Signal Processing* (2002), pp. 9–12
36. C.H. Esteban, F. Schmitt, Silhouette and stereo fusion for 3d object modeling, in *Fourth International Conference on 3D Digital Imaging and Modeling* (2004), pp. 46–54
37. Y. Liu, G. Chen, N. Max, C. Hofsetz, P. McGuinness. Visual hull rendering with multi-view stereo refinement, in *WSCG'04* (2004), pp. 261–268
38. P. Song, X. Wu, M.Y. Wang, Volumetric stereo and silhouette fusion for image-based modeling. Vis. Comput. **26**, 1435–1450 (2010)
39. C. Wu, Y. Liu, Q. Dai, B. Wilburn, Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. IEEE Trans. Vis. Comput. Graph. **17**, 1082–1095 (2011)

40. Y. Liu, Q. Dai, W. Xu, Graph-cuts fusion of distance fidelity maps for volumetric multi-view stereo. J. Electron. (China) **18**(3), 449–454 (2009)
41. S. Tran, L.S. Davis, 3d surface reconstruction using graph cuts with surface constraints, in *ECCV* (2006), pp. 219–231
42. C.H. Esteban, F. Schmitt, A snake approach for high quality image-based 3d object modeling, in *Variational, Geometric and Level Set Methods in Computer Vision* (2003), pp. 241–248
43. C.H. Esteban, F. Schmitt, Silhouette and stereo fusion for 3d object modeling. Comput. Vis. Image Underst. **96**(3), 367–392 (2004)
44. S. Paris, F.X. Sillion, L. Quan, A surface reconstruction method using global graph cut optimization. Int. J. Comput. Vision **66**(21), 41–161 (2004)
45. G. Vogiatzis, C. Hernandez, R. Cipolla, Reconstruction in the round using photometric normals and silhouettes, in *CVPR* (2006), pp. 1847–1854
46. G. Vogiatzis, P.H.S. Torr, R. Cipolla, Multi-view stereo via volumetric graph-cuts (2005)
47. A. Ladikos, S. Benhimane, N. Navab, Multi-view reconstruction using narrow-band graph-cuts and surface normal optimization, in *BMVC* (2008)
48. P. Paalanen, J.-K. Kamarainen, Narrow baseline GLSL multiview stereo, in *3DPVT* (2010)
49. S.N. Sinha, P. Mordohai, M. Pollefeys, Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh, in *ICCV 2007* (2007)
50. A. Hornung, L. Kobbelt, Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding, in *CVPR '06* (2006), pp. 503–510
51. F. Cuzzolin, A. Sarti, S. Tubaro, Invariant action classification with volumetric data (2005)
52. S. Y. Cheng, M.M. Trivedi, Hand pose estimation using expectation-constrained-maximization from voxel data, in *CVRR*, (Technical report) (2005)
53. E. de Aguiar, C. Theobalt, M. Magnor, H. Theisel, H.-P. Seidel, M3 : Marker-free model reconstruction and motion tracking from 3d voxel data (2004)
54. E. de Aguiar, C. Theobalt, M. Magnor, H.P. Seidel, Reconstructing human shape and motion from multi-view video, in *CVMP* (2005), pp. 42–49
55. K. Li, Q. Dai, W. Xu, Markerless shape and motion capture from multi-view video sequences. IEEE Trans. Circuits Syst. Video Technol. **21**(3), 320–334 (2011)
56. B. Berendsen, X. Luo, W. Hürst, R.C. Veltkamp, Volumetric modeling of 3d human pose from multiple video (2007)
57. S. Ando, W. Xiaojun, A. Suzuki, K. Wakabayashi, H. Koike, Human pose estimation for image monitoring (Image Processing Technologies for Image Monitoring Services), in *Special Feature* (2007)
58. J. Starck, A. Hilton, Model-based multiple view reconstruction of people, in *ICCV '03* (2003)
59. C. Curio, M.A. Giese, Combining view-based and model-based tracking of articulated human movements, in *WACV-MOTION* (2005)
60. F. Caillette, A. Galata, T. Howard, Real-time 3-d human body tracking using learnt models of behaviour. Comput. Vis. Image Underst. **109**, 112–125 (2008)
61. B. Michoud, E. Guillou, H.B. Pulido, S. Bouakaz, Real-time marker-free motion capture from multiple cameras, in *ICCV* (2007)
62. T. Yang, Y. Zhang, M. Li, D. Shao, X. Zhang, A multi-camera network system for markerless 3d human body voxel reconstruction, in *ICIG '09* (2009), pp. 706–711
63. P.J. Besl, N.D. McKay, A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
64. S. Rusinkiewicz, Marc Levoy, Efficient variants of the ICP algorithm, in *Third International Conference on 3D Digital Imaging and Modeling (3DIM)* (2001)
65. A. Segal, D. Haehnel, S. Thrun, Generalized-icp science and systems, in *Proceedings of Robotics* (2009)
66. T. Jost, H. Hügli, Fast icp algorithms for shape registration, in *Proceedings of the 24th DAGM Symposium on, Pattern Recognition* (2002), pp. 91–99
67. S. Corazza, L. Mündermann, E. Gambaretto, G. Ferrigno, T.P. Andriacchi, Markerless motion capture through visual hull, articulated icp and subject specific model generation. Int. J. Comput. Vision **1–2**, 156–169 (2010)

68. W. Luo, T. Yamasaki, K. Aizawa, Articulated human motion capture from segmented visual hulls and surface reconstruction, in *APSIPA* (Annual summit and conference) (2010)
69. S. Pellegrini, K. Schindler, D. Nardi, A generalisation of the icp algorithm for articulated bodies, in *BMVC* (2008)
70. D. Anguelov, D. Koller, P. Srinivasan, S. Thrun, H.-C. Pang, and J. Davis, The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces, in *Advances in Neural Information Processing Systems* (NIPS 2004) (2005)
71. D. Anguelov, D. Koller, H.-C. Pang, P. Srinivasan, S. Thrun, Recovering articulated object models from 3d range data, in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI '04* (2004)
72. W. Chang, M. Zwicker, Automatic registration for articulated shapes, in *Computer Graphics Forum (Proceedings of SGP 2008)* (2008)
73. Y. Pekelny, C. Gotsman, Articulated object reconstruction and markerless motion capture from depth video. Comput. Graph. Forum **27**(2), 399–408 (2008)
74. Q. Zheng, A. Sharf, A. Tagliasacchi, B. Chen, H. Zhang, A. Sheffer, D. Cohen-Or, Consensus skeleton for non-rigid space-time registration. Comput. Graph.Forum (Special Issue of Eurographics) **29**(2), 635–644 (2010)
75. W. Chang, M. Zwicker, Global registration of dynamic range scans for articulated model reconstruction. ACM Trans. Graph. **30**, 26:1–26:15 (2011)
76. L. Sigal, A.O. Balan, M.J. Black, Combined discriminative and generative articulated pose and non-rigid shape estimation, In *NIPS'07* (2007)
77. L.-J. Chu, C.-P. Chen, C.-S. Chen, Y.-P. Hung, 3d human motion tracking with soft-joint constrained icp, in *IPPR Conference on Computer Vision Graphics and Image Processing (CVGIP)* (2008)
78. S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, A.Y. Wu, An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM **45**(6), 89—923 Nov (1998)
79. M. Greenspan, G. Godin, A nearest neighbor method for efficient ICP, in *Proceedings of Third International Conference on 3D Digital Imaging and Modeling* (2001), pp. 161–168
80. M. Connor, P. Kumar, Fast construction of k-nearest neighbor graphs for point clouds. IEEE Trans. Visual Comput. Graphics **16**(4), 599–608 (2010)
81. K.-L. Low, Linear least-squares optimization for point-to-plane icp surface registration (2004)
82. J. Starck, A. Hilton, Spherical matching for temporal correspondence of non-rigid surfaces, in *ICCV* (2005), pp. 1387–1394
83. T. Tung, T. Matsuyama, Dynamic surface matching by geodesic mapping for 3d animation transfer, in *CVPR* (2010)
84. N. Ahmed, C. Theobalt, C. Rössl, S. Thrun, H.-P. Seidel, Dense correspondence finding for parametrization-free animation reconstruction from video, in *CVPR*(2008)
85. A. Doshi, J. Starck, A. Hilton, An empirical study of non-rigid surface feature matching of human from 3d video. J. Virtual Reality Broadcast. **7**(3), 1–11 (2010)
86. A. Doshi, A. Hilton, J. Starck, An empirical study of non-rigid surface feature matching, in *CVMP* (2008)
87. O. Enqvist, K. Josephson, F. Kahl, Optimal correspondences from pairwise constraints, in *IEEE 12th International Conference on Computer Vision* (2009), pp. 1295–1302
88. L. Zhang, S.-I. Choi, S.-Y. Park, Robust icp registration using biunique correspondence, in *3DIMPVT* (2011), pp. 80–85
89. C. Stoll, Z. Karni, C. Rössl, H. Yamauchi, H.-P. Seidel, Template deformation for point cloud fitting, in *SPBG* (2006), pp. 27–35
90. I.-C. Yeh, C.-H. Lin, O. Sorkine, T.-Y. Lee, Template-based 3d model fitting using dual-domain relaxation. IEEE Trans. Visual Comput. Graphics **99**, 1178–1190 (2010)
91. N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, H.-P. Seidel, Technical section: Estimating body shape of dressed humans. Comput. Graph. **33**(3), 211–216 (2009)
92. D.C. Schneider, P. Eisert, Fast nonrigid mesh registration with a data-driven deformation prior, in *NORDIA09* (2009)

93. R.R. Paulsen, R. Larsen, Anatomically plausible surface alignment and reconstruction, in *Proceedings of Theory and Practice of, Computer Graphics* (2010)
94. M. Wand, P. Jenke, Q.-X. Huang, M. Bokeloh, L.J. Guibas, A. Schilling, Reconstruction of deforming geometry from time-varying point clouds, in *SGP* (2007), pp. 49–58
95. A. Tevs, M. Bokeloh, M. Wand, A. Schilling, H.-P. Seidel, Isometric registration of ambiguous and partial data, in *CVPR* (2009)
96. I. Eckstein, J.-P. Pons, Y. Tong, C.-C. J. Kuo, M. Desbrun, Generalized surface flows for mesh processing, in *Proceedings of the fifth Eurographics symposium on Geometry processing* (2007), pp. 183–192
97. Q.-X. Huang, B. Adams, M. Wicke, L.J. Guibas, Non-rigid registration under isometric deformations. Comput. Graph. Forum **27**(5), 1449–1457 (2008)
98. B. Amberg, S. Romdhani, T. Vetter, Optimal step nonrigid icp algorithms for surface registration, in *CVPR* (2007)
99. R. Sagawa, K. Akasaka, Y. Yagi, H. Hamer, L. van Gool, Elastic convolved icp for the registration of deformable objects, in *Proceedings of IEEE 12th International Conference on Computer Vision Workshops (3DIM2009)* (2009), pp. 1558–1565
100. H. Abdelmunim, A.A. Farag, Elastic shape registration using an incremental free form deformation approach with the icp algorithm, in *Canadian Conference on Computer and Robot Vision (CRV)* (2011), pp. 212–218
101. T. Windheuser, U. Schlickewei, Frank R. Schmidt, D.Cremers, Large-scale integer linear programming for orientation-preserving 3d shape matching, in *Computer Graphics Forum (Proceedings Symposium Geometry Processing)* (2011)
102. D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: shape completion and animation of people. ACM Trans. Graph. **24**(3), 408–416 (2005)
103. B. Allen, B. Curless, Z. Popović, The space of human body shapes: reconstruction and parameterization from range scans, in *ACM SIGGRAPH* (2003), pp. 587–594
104. B. Allen, B. Curless, Z. Popović, A. Hertzmann, Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis, in *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer, animation* (2006), pp. 147–156
105. H. Chen, B. Bhanu, Global-to-local non-rigid shape registration, in *ICPR '06* (2006), pp. 57–60
106. D. Hähnel, S. Thrun, W. Burgard,. An extension of the icp algorithm for modeling nonrigid objects with mobile robots, in *Proceedings of 18th International Joint Conference on, Artificial Intelligence (IJCAI-03)* (2003), pp. 915–920
107. N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, H.-P. Seidel. A statistical model of human pose and body shape. in *Proceedings of Eurographics 2008 Computer Graphics Forum* (2009)
108. B. Brown, S. Rusinkiewicz, Global non-rigid alignment of 3d scans. ACM Trans. Graph. (Proc. SIGGRAPH) **26**(3), 1–8 (2007)
109. N. Thorstensen, R. Keriven, Non-rigid shape matching using geometry and photometry. in *ACCV* (2009), pp. 644–654
110. D. Münch, B. ît Combès, S. Prima, A modified icp algorithm for normal-guided surface registration. Proc. SPIE **7623** (2010)
111. L.-P. Morency, T. Darrell, Stereo tracking using icp and normal flow constraint, in *Proceedings of 16th International Conference on Pattern Recognition*, vol. 4 (2002)
112. Z.-Q Cheng, W. Jiang, G. Dang, R.R. Martin, J. Li, H. Li, Y. Chen, Y. Wang, B. Li, K. Xu, S. Jin, Non-rigid registration in 3d implicit vector space, in *SMI '10* (2010). pp. 37–46
113. C. Papazov, D. Burschka, Deformable 3d shape registration based on local similarity transforms. Comput. Graph. Forum **30**(5), 1493–1502 (2011)
114. T. Popa, I. South-Dickinson, D. Bradley, A. Sheffer, W. Heidrich. Globally consistent space-time reconstruction. Comput. Graph. Forum (Proc. SGP) **7729**, 133–147 (2010)
115. H. Li, M. Pauly, First steps toward the automatic registration of deformable scans. (Technical report, ETH Zurich, 2007)
116. A. Sharf, D.A. Alcantara, T. Lewiner, C. Greif, A. Sheffer, N. Amenta, D. Cohen-Or, Space-time surface reconstruction using incompressible flow. ACM Trans. Graph. (2008)

117. N.J. Mitra, S. Flory, M. Ovsjanikov, N. Gelfand, L. Guibas, H. Pottmann, Dynamic geometry registration, in *Symposium on Geometry Processing* (2007), pp. 173–182
118. M. Liao, Q. Zhang, H. Wang, R. Yang, M. Gong, Modeling deformable objects from a single depth camera, in *ICCV* (2009), pp. 167–174
119. Z.-Q. Cheng, H. Li, J. Li, Y. Chen, Y. Wang, B. Li, X. Kai, G. Dang, S. Jin, Robust non-rigid registration of large-difference deformed modelss, in *CASA* (2010)
120. H. Li, R.W. Sumner, M. Pauly, Global correspondence optimization for non-rigid registration of depth scans. Comput. Graph. Forum (Proc. SGP'08) **25**(5), 1459–1468 (2008)
121. H. Li, B. Adams, L.J. Guibas, M. Pauly, Robust single-view geometry and motion reconstruction. ACM Trans. Graph. (Proceedings SIGGRAPH, Asia 2009) **28**(5), 175–185 (2009)
122. A. Tevs, A. Berner, M. Wand, I. Ihrke, M. Bokeloh, J. Kerber, H.-P. Seidel, Animation cartography—intrinsic reconstruction of shape and motion. ACM Trans. Graph. **31**, 12–27 (2012)
123. C. Budd, A, Hilton, Temporal alignment of 3d video sequences using shape and appearance, in *CVMP* (2010), pp. 114–122
124. M. Klaudiny, A. Hilton, Cooperative patch-based 3d surface tracking, in *CVMP* (2011), pp. 67–76
125. P. Huang, C. Budd, A. Hilton, Global temporal registration of multiple non-rigid surface sequences, in *CVPR* (2011), pp. 3473–3480
126. C. Budd, P. Huang, A. Hilton, Hierarchical shape matching for temporally consistent 3d video, in |textit3DIMPVT (2011), pp. 172–179
127. H. Li, L. Luo, D. Vlasic, P. Peers, J. Popović, M. Pauly, S. Rusinkiewicz, Temporally coherent completion of dynamic shapes. ACM Trans. Graph. **31**(1), 11 (2012)
128. J. Süssmuth, M. Zollhöfer, G. Greiner, Animation transplantation. J. Vis. Comput. Animation, **21**, 173–182 (2010)
129. J. Tong, J. Zhou, L. Liu, Z. Pan, H. Yan, Scanning 3d full human bodies using kinects. IEEE Trans. Vis. Comput. Graph. (Proceedings of IEEE Virtual Reality) **18**(4), 643–650 (2012)
130. M. Desbrun, M. Meyer, P. Schröder, A.H. Barr, Implicit fairing of irregular meshes using diffusion and curvature flow, in *SIGGRAPH* (1999), pp. 317–324
131. M. Botsch, O. Sorkine, On linear variational surface deformation methods. IEEE Trans. Vis. Comput. Graph. **14**(1), 213–230 Jan (2008)
132. I. Wald, S. Boulos, P. Shirley, Ray tracing deformable scenes using dynamic bounding volume hierarchies. ACM Trans. Graph. 26(1), 18 (2007)
133. I. Wald, On fast construction of sah-based bounding volume hierarchies, in *Proceedings of the 2007 IEEE Symposium on Interactive Ray Tracing, RT '07* (2007), pp. 33–40
134. A. Jacobson, I. Baran, J. Popović, O. Sorkine, Bounded biharmonic weights for real-time deformation. ACM Trans. Graph. **30**, 165:1–165:8 (2011)
135. O.K.-C. Au, H. Fu, C.-L. Tai, D. Cohen-Or, Handle-aware isolines for scalable shape editing. ACM Trans. Graph. **26**, 83 (2007)
136. D. Vlasic, I. Baran, W. Matusik, J. Popović, Articulated mesh animation from multi-view silhouettes. ACM Trans. Graph. **27**(3), 1–97 Aug (2008)
137. J. Shi, C. Tomasi. Good features to track, in *CVPR'94* (1994), pp. 593–600
138. B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in *IJCAI'81* (1981), pp. 674–679
139. T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in *ECCV (4)'04* (2004), pp. 25–36
140. D.G. Lowe, Object recognition from local scale-invariant features, in *ICCV '99* (1999)
141. H. Bay, A. Ess, T. Tuytelaars, L. van Gool, Speeded-up robust features (surf). Comput. Vis. Image Underst. **110**, 346–359 (2008)

Chapter 5
Conclusions

To put it into a nutshell, this last chapter reviews in short the key ideas of each piece of work presented in this book. Thereby, limitations of our work are fairly detailed. Nevertheless, the highlight of our novel contributions are also emphasized. Finally, the most salient conclusions are summarized, and possible directions for future work are outlined.

5.1 Summary of Contributions

In this section, we offer a comprehensive summary of the contributions detailed in all previous chapters. In this book, we opt for a *Graphics-to-Vision* strategy. Then, this dissertation naturally offers the following contributions in *interaction*, *graphics*, *vision*, *animation* and *rendering*.

Inverse-based Interactive Scalable Modeling. In this book, we propose to inject the *inverse kinematics* paradigm at the heart of a user-driven modeling system. We propose the first interactive *inverse kinematic cage-based modeling* technique. The key idea is to convey a desired scalable deformation in term of resolution-free and reusable compact series of low-dimensional purely-geometric parameters. Thus, we propose a novel hybrid reduced model. This model is made of skin-detached parameters to achieve deformation-via-inversion with smoothness properties. We show that this versatile variational concept can be applied on different types of underlying deformable subspaces, not limited to flexible cages as demonstrated on an articulated skeleton.

Firstly, we show a new cage inversion technique ables to estimate the corresponding cage pose with respect to user-specified deforming constraints on the enclosed mesh. We provide a linear estimation framework for cage-vertex coordinates. The regularization of the deformation is expressed on the cage vertices, yielding a computationally efficient solution which fully benefits from cage-based parameterization. A mix of generalized barycentric and Laplacian coordinates are used inside a

Y. Savoye, *Cage-based Performance Capture*,
Studies in Computational Intelligence 509, DOI: 10.1007/978-3-319-01538-5_5,
© Springer International Publishing Switzerland 2014

linear minimization framework to reconstruct an enclosed mesh. This indirect dual-Laplacian caged-based mesh editing system allows users to create visually pleasing deformations. The linearity of the underlying objective functional makes the processing highly efficient and improves the effectiveness of deformable surfaces computation. Our method offers the opportunity to encode global deformation poses of the shape with respect of local influence and allows animators to re-use the estimation parametrization. Naturally, our framework is not mathematically restricted to a specific coordinate system as-far-as this system is linear and local preserving. The main benefit of our method lies in the minimization framework, fully independent of the model resolution. In essence, we reveal that connectivity information of the default cage already encodes soft kinematic constraints. Further, we would like to incorporate additional constraints for ensuring advanced surface deformations properties (see Fig. 5.1).

Secondly, we propose a generalization of the demonstrated Laplacian regularization for animatable structures. We describe a new and straightforward approach to re-use skeleton of animation with joint-based Laplacian-type regularization and sparse deformation features, in the context of exaggerated skeleton-based character animation. Despite decades of research, interactive character animations still have a lack of flexibility and editability in order to re-use real vertebral motion interactively. In further details, generation of expressive cartoon animation from real data, is a challenging key task for non-photorealistic animation. Hence, a major problem for artists in production is to enhance the expressiveness of classical motion clips by direct manipulation of the underlying skeletal subspace. Breaking physics of rigid motion allows us to obtain cartoon-like effects, independently of the skin layer. Spatial coherence of joints is preserved by differential-aware scaling represented by a quadratic energy function leading to a new *Skeletal Poisson Solver*. The usefulness and flexibility of a Laplacian-type approach is fully demonstrated to emphasize effective high-quality skeletal-based animation pose. In the context of filmmaking, our pipeline reduces the amount of time needed to animate well-articulated expressive character. More importantly, both contributions make complex interactions involving animatable subspaces more accessible to non-professionals and offer efficient processing for video games. Finally, to the best of our knowledge we propose one of the the first

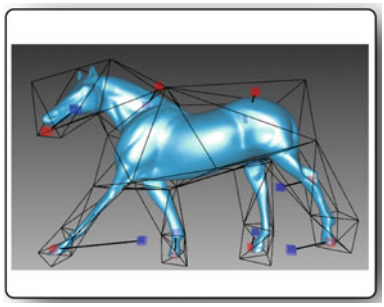

Fig. 5.1 Cage-based subspace modeling

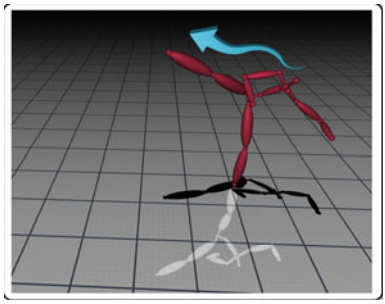

Fig. 5.2 Skeletal-based subspace modeling

scalable modeling via cage inversion process allowing intuitive user interaction, and the first discretization of the elastic energy over a skeletal structure (see Fig. 5.2).

Cage-based Performance Capture and Multi-view Data Reuse. We propose the first technique for cage-based performance capture that allows to represent highly non-rigid dynamic captured shape in a compact and boneless representation. This reparameterization allows to reuse performance capture data with suitable and reusable reduced parameters. Dynamic mesh compression for massive captured data and feature-length animation with free-viewpoint control is also a crucial issue with the emergence of 3DTV. Our technique avoids heavy computation process as required by traditional compression of 3D dynamic mesh sequences, or simplification of animated meshes. Moreover, our extracted parameters could be useful for efficient network transmission or telepresence applications (see Fig. 5.3).

Achieving inverse animation by approximating dynamic meshes using rigid skinning has inspired researchers to convert already aligned video-based mesh sequences into rigid kinematic parameters. In contrast with traditional techniques using the skeleton-based animation paradigm, we describe the very first and seminal framework to convert non-rigid performance capture animation into cage-based animation. In this book, we also propose one of the first techniques for performance capture stylization in term of surface motion exaggeration and video-based texture abstraction, relying on cage-based deformation and cartoon style filtering. There is a key issue for future work toward perceptually-guided animated mesh processing for space-time

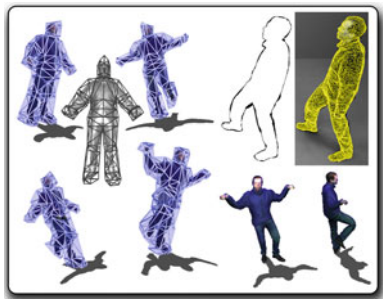

Fig. 5.3 Cage-based perfcap

editing of performance capture data. Markerless highly-detailed performance capture is an impacting technology for vision-based graphics and 3D video.

Firstly, our approach retrieves animation parameters through a single-pass minimization process, without the need of an underlying rigid kinematic structure. Our harmonic inverse cage-based estimation allows rim vertices to preserve silhouette consistency better than inverse skeleton-based estimation. Boneless cage-based approach is more convenient than kinematic hierarchy-based approach to guarantee the deformation fidelity of non-rigid surface to multi-video stream.

Animation Cartoonization. Secondly, recent advances in dynamic surface capture have made the creation of a realistic animation a promising task for modern visual media productions 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 with more exaggerated motion while preserving extreme captured cloth wrinkles is difficult to achieve.

Consequently, another key contribution 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 straightforward and effective algorithm converts realistic spatiotemporal captured surfaces and multi-view data into an exaggerated life-like squash-and-stretch shape evolution coupled with content-aware cartoon-style expressive rendering. Our approach brings real-life animation to a controllable cartoon adaptation in the spirit of the captured visual information. To reach best perceptual effect, we decouple the performance capture stylization on non-rigid cage-based surface motion parameters as-well-as its reconstructed appearance. Finally, injecting cartoon exaggeration for non-rigid surfaces brings a heightened believability to animated toon characters. In the future, we plan to explore more opportunities in *Animation Cartoonization* for dynamic surface motion while preserving the life-like visual appealing. In addition, the problem of transplanting details of captured toward a given inanimate model with different topology is still an open problem. Unlike prior efforts, our work is the first to investigate skeleton-less cartoon filtering for dynamic captured surfaces relying on cage-based shape approximations (see Fig. 5.4).

Cage-based Cross-Reconstruction. In this book, we also introduce the first unsupervised technique to *register non-rigid dynamic surface from multi-view silhouettes* sequence using the cage-based paradigm. Our technique offers two prominent benefits: simultaneous dense surface registration and low-dimensional temporal curve registration, and enforcement of flexible local rigidities with biharmonic fields. Consequently, our template-based technique extracts consistent structural information from a series of unstructured point cloud, or directly from the image level using silhouettes data with an intermediate watertight geometric reconstruction.

Fig. 5.4 Cage-based stylization

Nevertheless, we see an avenue for future work to improve the temporal coherence of our reusable dynamic shape registration over longer sequences.

We propose a novel detail-preserving registration approach with resolution-independent control. Furthermore, our skin-detached surface registration bypasses segmentation or affine fitting to maintain the local plasticity. In particular, we leverage the problem of highly non-rigid spacetime registration by employing an elasto-plastic coarse cage. Thus, we perform scalable handle-aware biharmonic shape registration, relying on the high-level of shape abstraction offered by the space-based paradigm. To the best of our knowledge, our technique is the first to investigate handle-aware elastic overlapping-rigidities for registering life-like dynamic shapes in full-body clothing. The main benefit of our framework remains in the iterative space-based deformation, and preventing triangle fold-over during the registration process (see Fig. 5.5).

Our technique offers controllable, meaningful, and animator-friendly subspaces in order to acquire dynamic surfaces with unsupervised space-time correspondence strategy. Registering cage-based surface opens new directions in the outstanding area of video-based mesh tracking, by allowing the use of non-rigid underlying structure. Our low-dimensional parameterization allows us to simultaneous register and compress unstructured animated geometry. Our new approach is a first step toward the unsupervised template-based registration of highly non-rigid dynamic shapes using low-dimensional space-based encoding. The main advantage of our iterative optimization remains in the simultaneous cross-reconstruction of dynamic

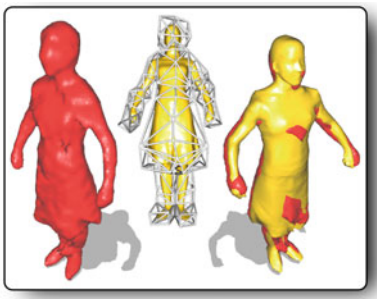

Fig. 5.5 Cage-based registration

shapes, and skin-detached registration of reusable cages. To the best of our knowledge, we propose the first standalone technique to transfer deformations observed in multi-view silhouettes into a life-like and cage-based encoding.

5.2 Overall Discussion

After recalling the original contributions, in this section we will remind the reader the motivation of our work, problems and benefit of proposed solutions. We will continue our discussion by summarizing benefits of our strategy to respond to the tackled subproblems and by detailing the advantage of our core techniques.

Principal Motivations. First of all, we remarked the lack of scalable, global, flexible and reusable parametrization for dynamic surfaces from capture to reuse. Major challenge remains in the surface characteristic with the peculiarity to be entropically non-rigid. Meanwhile, researching in this area is particularly promising because of the high demand in industry for re-using and storing large amount of heterogeneous data, without loosing their natural looking aspect.

Tackled Subproblems. The subproblem of video-based animation, performance capture reuse, and non-rigid surface cross-parametrization by the mean of lower-dimensional space are really hard and challenging tasks. At the heart of corresponding solutions, optimization strategies are crucial and not a trivial aspect. Below, we propose a list of subproblems with demonstrated solutions:

1. Sparse modeling using animatable subspaces: cageIK and Laplacian skeletal editing.
2. Conversion of highly non-rigid surfaces into cage-based dynamic shapes.
3. Content-aware stylization of performance capture data: cage-based shape filtering and appearance depiction.
4. Injective cage-based cross-parametrization from multi-view silhouettes.

Demonstrated Benefits. Even if underlying structures are widespread in performance capture and animation, cage-based parametrizations offer a new transversal tool to obtain non-rigid dimension reduction and cross-parametrizations from multi-view silhouettes. Besides, we observe that cage-based deformation allows to encapsulate life-like surface features in a plausible manner as well as offers global frame-to-frame control of the deformation. Furthermore, our conceptually straightforward constraint-based optimization framework can be easily extended to incorporate other types of linearizable constraints. We introduce cage-based deformation techniques as a useful ingredient to potentially improve the incremental reconstruction across time in the context of video-based animation. In this book, we demonstrate the strength of harmonic-type coordinates with spatial and temporal coherence to retrieve cage parameters along frames. This estimation enhances the flexibility of reusing such reconstructed data. To put in a nutshell, our presented approaches offer promising solution for authoring the reuse of captured surface animation.

5.3 Perspectives and Future Directions

Cage-based Performance Capture. To sum up, we essentially introduce the core formulation of *cage-based conversion* and *cage-based performance capture* with a compelling collection of applications, but not limited to. Limitations will be consciously addressed for future work. According to our referred work, we believe all algorithms are straightforward enough to be widely implemented and tested with previous frameworks. The promising results offered by this work allow new conceivable solutions with sensor-based animation. My on-going research and development will continue to forge a strong unification of *Computer Graphics* and *Computer Vision* techniques for efficient video-based interaction, reconstruction, animation in the context of film-game convergence. Moreover, we expect to combine *Computer Vision* and *Computer Animation* tools more actively to ensure accuracy, efficiency and reusability for video-based reconstruction and associated rendering in the near future. On the top of everything, we have demonstrated that cage-based paradigm can accurately reskin complex captured surface motions without secondary deformation. Strong interest for deformable cages and surface capture make us think that proposed algorithms could be decently improved in the future.

Out of the Cage. According to the numerous need for methods based on reduced subspace control and active geometric structures, I strongly believe that the collection of proposed approaches will provide stimulating and promising venues for follow-up directions. Improvements could be addressed in future works toward *Animation Cartoonization* of captured animation and *Cage-based Registration* for heterogeneous performance capture. Much remains to be done for obtaining an ideal *Cage-based Performance Capture* in three dimensional space, directly from sensors information. I also expect to pursue my ongoing efforts to perform better fine-tuned non-rigid alignment for large organic motion. I also want to achieve more theoretical research work in dynamic geometry processing and skeletal shape deformation, but not limited too.