Analyzing Growing Plants from 4D Point Cloud Data

Yangyan Li[1] Xiaochen Fan[1] Niloy J. Mitra[2] Daniel Chamovitz[3] Daniel Cohen-Or[3] Baoquan Chen[1,4,*]
[1]Shenzhen,VisuCA Key Lab/SIAT [2]University College London [3]Tel Aviv University [4]Shandong University

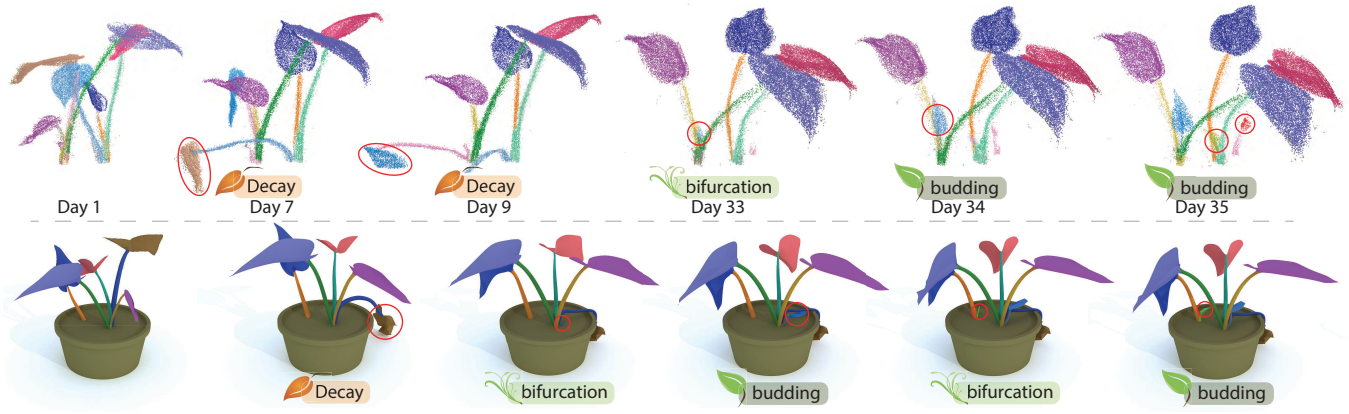


Figure 1: *(Top) Dishlia growth time lapse point cloud over 5 weeks, with classified organs and detected budding, bifurcation and decay events. (Bottom) The extracted events are then used to bring a static plant model to life with both motion and growth.*

Abstract

Studying growth and development of plants is of central importance in botany. Current quantitative are either limited to tedious and sparse manual measurements, or coarse image-based 2D measurements. Availability of cheap and portable 3D acquisition devices has the potential to automate this process and easily provide scientists with volumes of accurate data, at a scale much beyond the realms of existing methods. However, during their development, plants grow new parts (e.g., vegetative buds) and bifurcate to different components — violating the central incompressibility assumption made by existing acquisition algorithms, which makes these algorithms unsuited for analyzing growth. We introduce a framework to study plant growth, particularly focusing on accurate localization and tracking topological events like budding and bifurcation. This is achieved by a novel forward-backward analysis, wherein we track robustly detected plant components back in time to ensure correct spatio-temporal event detection using a locally adapting threshold. We evaluate our approach on several groups of time lapse scans, often ranging from days to weeks, on a diverse set of plant species and use the results to animate static virtual plants or directly attach them to physical simulators.

Keywords: growth analysis, 4D point cloud, event detection

Links: DL PDF WEB VIDEO DATA

ACM Reference Format
Li, Y., Fan, X., Mitra, N., Chamovitz, D., Cohen-Or, D., Chen, B. 2013. Analyzing Growing Plants from 4D Point Cloud Data. ACM Trans. Graph. 32, 6, Article 157 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508368 http://doi.acm.org/10.1145/2508363.2508368.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric Algorithms; I.4.6 [Image Proc. and Computer Vision]: Segmentation—Point Classification;

1 Introduction

Studying growth processes in organic life forms has a long history in science. Traditionally, such studies rely on manual recordings of growth stages, or image-based measurements taken at sparse intervals. Such workflows are tedious, prone to measurement bias, and difficult to scale to large-scale observations, both in space and time. Advances in affordable 3D acquisition devices provide new opportunities.

Plant growth is fundamentally different from animal growth. Most animals are born with all their body organs, which grow and mature with age. In contrast, plants grow and develop throughout their life cycle, constantly producing new tissues and structures [Chen and Laux 2012]. Studying such developments involves detecting specific growth events (e.g., budding of a leaf, see Figure 2), quantifying these events, and tracking their subsequent evolution over time. Beyond difficulties arising from motion, the key challenge is to track the continuous shape changes in geometry and topology, albeit at a very slow rate, due to growth, and possibly due to decay. This is different from typical motion capture setups studying human movements.

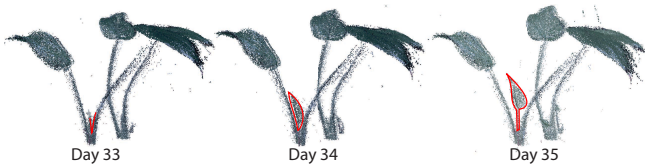


Figure 2: *Challenges in detecting bifurcation/budding events from point cloud. We observe that the bifurcation on day 33 and the budding on day 34 are subtle and hard to identify; the organs mature and become easier to detect later on day 35.*

Reconstructing Detailed Dynamic Face Geometry from Monocular Video

Pablo Garrido[1] * Levi Valgaerts[1] † Chenglei Wu[1,2] ‡ Christian Theobalt[1] §
[1]Max Planck Institute for Informatics [2]Intel Visual Computing Institute

Figure 1: *Two results obtained with our method.* Left: *The input video.* Middle: *The tracked mesh shown as an overlay.* Right: *Applying texture to the mesh and overlaying it with the input video using the estimated lighting to give the impression of virtual face make-up.*

Abstract

Detailed facial performance geometry can be reconstructed using dense camera and light setups in controlled studios. However, a wide range of important applications cannot employ these approaches, including all movie productions shot from a single principal camera. For post-production, these require dynamic monocular face capture for appearance modification. We present a new method for capturing face geometry from monocular video. Our approach captures detailed, dynamic, spatio-temporally coherent 3D face geometry without the need for markers. It works under uncontrolled lighting, and it successfully reconstructs expressive motion including high-frequency face detail such as folds and laugh lines. After simple manual initialization, the capturing process is fully automatic, which makes it versatile, lightweight and easy-to-deploy. Our approach tracks accurate sparse 2D features between automatically selected key frames to animate a parametric blend shape model, which is further refined in pose, expression and shape by temporally coherent optical flow and photometric stereo. We demonstrate performance capture results for long and complex face sequences captured indoors and outdoors, and we exemplify the relevance of our approach as an enabling technology for model-based face editing in movies and video, such as adding new facial textures, as well as a step towards enabling everyone to do facial performance capture with a single affordable camera.

CR Categories: I.3.7 [COMPUTER GRAPHICS]: Three-Dimensional Graphics and Realism; I.4.1 [IMAGE PROCESSING]: Digitization and Image Capture—Scanning; I.4.8 [IMAGE PROCESSING]: Scene Analysis;

Keywords: Facial Performance Capture, Monocular Tracking, Temporally Coherent Optical Flow, Shading-based Refinement

*e-mail: pgarrido@mpi-inf.mpg.de
†e-mail: valgaerts@mpi-inf.mpg.de
‡e-mail: chenglei@mpi-inf.mpg.de
§e-mail: theobalt@mpi-inf.mpg.de

Links: DL PDF WEB VIDEO

1 Introduction

Optical performance capture methods can reconstruct faces of virtual actors in videos to deliver detailed dynamic face geometry. However, existing approaches are expensive and cumbersome as they can require dense multi-view camera systems, controlled light setups, active markers in the scene, and recording in a controlled studio (Sec. 2.2). At the other end of the spectrum are computer vision methods that capture face models from monocular video (Sec. 2.1). These captured models are extremely coarse, and usually only contain sparse collections of 2D or 3D facial landmarks rather than a detailed 3D shape. Recently, Valgaerts et al. [2012] presented an approach for detailed performance capture from binocular stereo. However, 3D face models of a quality level needed for movies and games cannot yet be captured from monocular video.

In this work, we aim to push the boundary and application range further and move towards monocular video. We propose a new method to automatically capture *detailed* dynamic face geometry from *monocular* video filmed under general lighting. It fills an important algorithmic gap in the spectrum of facial performance capture techniques between expensive controlled setups and low-quality monocular approaches. It opens up new application possibilities for professional movie and game productions by enabling facial performance capture on set, directly from the primary camera. Finally, it is a step towards democratizing face capture technology for everyday users with a single inexpensive video camera.

A 3D face model for a monocular video is also a precondition for many relevant video editing tasks (Sec. 2.3). Examples in-

ACM Reference Format
Garrido, P., Valgaerts, L., Wu, C., Theobalt, C. 2013. Reconstructing Detailed Dynamic Face Geometry from Monocular Video. ACM Trans. Graph. 32, 6, Article 158 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508380 http://doi.acm.org/10.1145/2508363.2508380.

DOI: http://doi.acm.org/10.1145/2508363.2508380

Inverse Dynamic Hair Modeling with Frictional Contact

Alexandre Derouet-Jourdan Florence Bertails-Descoubes Gilles Daviet Joëlle Thollot

INRIA and Laboratoire Jean Kuntzmann (Grenoble University, CNRS), France*

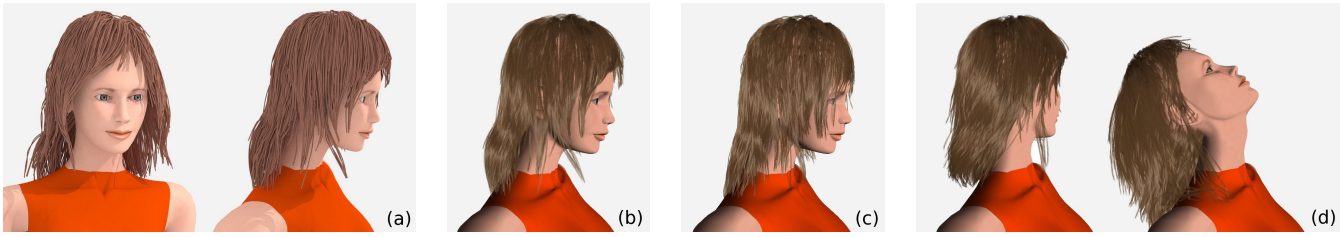


Figure 1: *Our inversion method makes the hair synthesis pipeline consistent: (a) Raw hair geometry (a set of polylines) resulting from the manual design or the automatic capture of a static hairstyle (here, a capture from [Herrera et al. 2012]); (b) Input geometry is automatically converted into a dynamic hair model (a set of super-helices) at equilibrium under gravity and frictional hair-body and hair-hair contact forces; Unlike classical hair simulators (c) which ignore surrounding forces when initializing the hairstyle and are thus prone to undesired sagging, our simulator (b) exactly matches the original hair geometry at initial state and (d) yields a realistic, character-specific hair animation.*

Abstract

In the latest years, considerable progress has been achieved for accurately acquiring the geometry of human hair, thus largely improving the realism of virtual characters. In parallel, rich physics-based simulators have been successfully designed to capture the intricate dynamics of hair due to contact and friction. However, at the moment there exists no consistent pipeline for converting a given hair geometry into a realistic physics-based hair model. Current approaches simply initialize the hair simulator with the input geometry in the absence of external forces. This results in an undesired sagging effect when the dynamic simulation is started, which basically ruins all the efforts put into the accurate design and/or capture of the input hairstyle. In this paper we propose the first method which consistently and robustly accounts for surrounding forces — gravity and frictional contacts, including hair self-contacts — when converting a geometric hairstyle into a physics-based hair model. Taking an arbitrary hair geometry as input together with a corresponding body mesh, we interpret the hair shape as a static equilibrium configuration of a hair simulator, in the presence of gravity as well as hair-body and hair-hair frictional contacts. Assuming that hair parameters are homogeneous and lie in a plausible range of physical values, we show that this large underdetermined inverse problem can be formulated as a well-posed constrained optimization problem, which can be solved robustly and efficiently by leveraging the frictional contact solver of the direct hair simulator. Our method was successfully applied to the animation of various hair geometries, ranging from synthetic hairstyles manually designed by an artist to the most recent human hair data automatically reconstructed from capture.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: Hair modeling, inverse statics, frictional contact

*contact:{alexandre.derouet-jourdan,florence.descoubes}@inria.fr

Links: DL PDF WEB VIDEO CODE

1 Introduction

Realistically synthesizing the shape and motion of hair is crucial for representing convincing virtual humans. In the two related fields of Computer Graphics, namely *hairstyling* and *hair animation*, considerable progress has been achieved these latest years. Yet, these two areas of research have remained fairly disconnected from each other. For the sake of flexibility and control, hairstyling is generally performed using a purely geometric process, either through the interactive editing of geometric primitives, or, more recently, with the help of automatic image-based capture methods. Resulting hairstyles are then represented as raw geometric data, with no connection to physics. In contrast, hair animation is often considered as a passive and complex phenomenon that can be captured realistically using physics-based simulation. Before simulation, the hairstyle is generally initialized with a rest pose in the absence of external forces and is typically prone to *sagging* when the simula-

ACM Reference Format
Derouet-Jourdan, A., Bertails-Descoubes, F., Daviet, G., Thollot, J. 2013. Inverse Dynamic Hair Modeling with Frictional Contact. ACM Trans. Graph. 32, 6, Article 159 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508398 http://doi.acm.org/10.1145/2508363.2508398.

DOI: http://doi.acm.org/10.1145/2508363.2508398

On-set Performance Capture of Multiple Actors With A Stereo Camera

Chenglei Wu[1,2] * Carsten Stoll[1] † Levi Valgaerts[1] ‡ Christian Theobalt[1] §
[1]Max Planck Institute for Informatics [2]Intel Visual Computing Institute

Figure 1: *One of our performance capture results. Left to right: input video, reconstructed geometry, edited video with virtual logo added.*

Abstract

State-of-the-art marker-less performance capture algorithms reconstruct detailed human skeletal motion and space-time coherent surface geometry. Despite being a big improvement over marker-based motion capture methods, they are still rarely applied in practical VFX productions as they require ten or more cameras and a studio with controlled lighting or a green screen background. If one was able to capture performances directly on a general set using only the primary stereo camera used for principal photography, many possibilities would open up in virtual production and previsualization, the creation of virtual actors, and video editing during post-production. We describe a new algorithm which works towards this goal. It is able to track skeletal motion and detailed surface geometry of one or more actors from footage recorded with a stereo rig that is allowed to move. It succeeds in general sets with uncontrolled background and uncontrolled illumination, and scenes in which actors strike non-frontal poses. It is one of the first performance capture methods to exploit detailed BRDF information and scene illumination for accurate pose tracking and surface refinement in general scenes. It also relies on a new foreground segmentation approach that combines appearance, stereo, and pose tracking results to segment out actors from the background. Appearance, segmentation, and motion cues are combined in a new pose optimization framework that is robust under uncontrolled lighting, uncontrolled background and very sparse camera views.

CR Categories: I.3.7 [COMPUTER GRAPHICS]: Three-Dimensional Graphics and Realism; I.4.1 [IMAGE PROCESSING]: Digitization and Image Capture—Scanning; I.4.8 [IMAGE PROCESSING]: Scene Analysis;

Keywords: Performance Capture, Skeletal Motion Estimation, Shape Refinement, Bidirectional Reflectance Distribution Function

*e-mail: chenglei@mpi-inf.mpg.de
†e-mail: stoll@mpi-inf.mpg.de
‡e-mail: valgaerts@mpi-inf.mpg.de
§e-mail: theobalt@mpi-inf.mpg.de

Links: DL PDF WEB VIDEO

1 Introduction

Marker-less performance capture methods enable the reconstruction of detailed motion, dynamic geometry, and the appearance of real world scenes from multiple video recordings, for instance, reconstructing the full body or face of an actor, [Bradley et al. 2010; de Aguiar et al. 2008; Vlasic et al. 2008; Gall et al. 2009]. Despite the ability to capture richer and more expressive models than marker-based capture methods, marker-less methods are yet to be found in many practical feature film productions. One of the reasons for this is that most existing marker-less methods require studios with controlled lighting, controlled background, and a multitude of cameras. The benefit of being able to capture detailed models of actors in natural motion and natural apparel without markers is constrained in application by the remaining requirement to capture the actors in a separate green-screen controlled stage and not on set or on location. The ability to capture detailed moving 3D models of actors on the actual production set rather than a separate stage would broadly benefit movie and VFX production.

Currently, performances of real actors in a scene are frequently composited with virtual renditions of actors during post-processing. One example is the movie *Pirates of the Caribbean*, where real actors in a scene wear marker suits. The Imocap system is used to track their skeletal motion from the primary camera and a few satellite cameras and, in post-production, the actors in marker suits are replaced with virtual renditions. This common example shows the importance and tremendous difficulty of the task, since even the skeletal tracking alone required substantial manual marker labeling by an operator. On a real production set, it is difficult to effectively place additional satellite cameras for tracking as the environment

ACM Reference Format
Wu, C., Stoll, C., Valgaerts, L., Theobalt, C. 2013. On-set Performance Capture of Multiple Actors With A Stereo Camera. ACM Trans. Graph. 32, 6, Article 161 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508418 http://doi.acm.org/10.1145/2508363.2508418.

Inverse Volume Rendering with Material Dictionaries

Ioannis Gkioulekas
Harvard University

Shuang Zhao
Cornell University

Kavita Bala
Cornell University

Todd Zickler
Harvard University

Anat Levin
Weizmann Institute

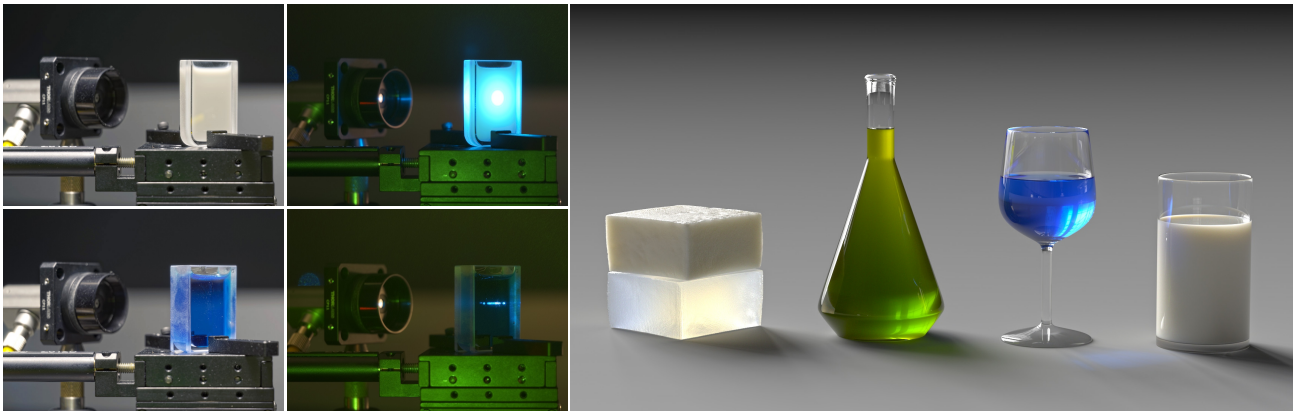


Figure 1: *Acquiring scattering parameters. Left: Samples of two materials (milk, blue curacao) in glass cells used for acquisition. Middle: Samples illuminated by a trichromatic laser beam. The observed scattering pattern is used as input for our optimization. Right: Rendering of materials in natural illumination using our acquired material parameter values.*

Abstract

Translucent materials are ubiquitous, and simulating their appearance requires accurate physical parameters. However, physically-accurate parameters for scattering materials are difficult to acquire. We introduce an optimization framework for measuring bulk scattering properties of homogeneous materials (phase function, scattering coefficient, and absorption coefficient) that is more accurate, and more applicable to a broad range of materials. The optimization combines stochastic gradient descent with Monte Carlo rendering and a material dictionary to invert the radiative transfer equation. It offers several advantages: (1) it does not require isolating single-scattering events; (2) it allows measuring solids and liquids that are hard to dilute; (3) it returns parameters in physically-meaningful units; and (4) it does not restrict the shape of the phase function using Henyey-Greenstein or any other low-parameter model. We evaluate our approach by creating an acquisition setup that collects images of a material slab under narrow-beam RGB illumination. We validate results by measuring prescribed nano-dispersions and showing that recovered parameters match those predicted by Lorenz-Mie theory. We also provide a table of RGB scattering parameters for some common liquids and solids, which are validated by simulating color images in novel geometric configurations that match the corresponding photographs with less than 5% error.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Raytracing;

Keywords: scattering, inverse rendering, material dictionaries

Links: DL PDF WEB

ACM Reference Format
Gkioulekas, I., Zhao, S., Bala, K., Zickler, T., Levin, A. 2013. Inverse Volume Rendering with Material Dictionaries. ACM Trans. Graph. 32, 6, Article 162 (November 2013), 13 pages. DOI = 10.1145/2508363.2508377 http://doi.acm.org/10.1145/2508363.2508377.

DOI: http://doi.acm.org/10.1145/2508363.2508377

1 Introduction

Scattering plays a critical role in the appearance of most materials. Much effort has been devoted to modeling and simulating its visual effects, giving us precise and efficient scattering simulation algorithms. However, these algorithms produce images that are only as accurate as the material parameters given as input. This creates a need for acquisition systems that can faithfully measure the scattering parameters of real-world materials.

Collecting accurate and repeatable measurements of scattering is a significant challenge. For homogeneous materials—which is the primary topic of this paper—scattering at any particular wavelength is described by two scalar values and one angular function. The *scattering coefficient* σ_s and *absorption coefficient* σ_a represent the fractions of light that are scattered and absorbed, and the *phase function* $p(\theta)$ describes the angular distribution of scattering. Measurement is difficult because a sensor almost always observes the combined effects of many scattering and absorption events, and these three factors cannot be easily separated. Indeed, for deeply-scattering geometries, similarity theory [Wyman et al. 1989] proves that one can analytically derive distinct parameter-sets that nonetheless produce indistinguishable images.

Most existing acquisition systems address the measurement challenge using a combination of two strategies (e.g., [Hawkins et al. 2005; Narasimhan et al. 2006; Mukaigawa et al. 2010]). First, they manipulate lighting and/or materials to isolate single-scattering effects; and second, they "regularize" the recovered scattering parameters by relying on a low-parameter phase function model, such as the Henyey-Greenstein (HG) model. These approaches can provide accurate results, but both of the employed strategies have severe limitations. The single-parameter HG model limits applicability to materials that it represents well; and this excludes some common natural materials [Gkioulekas et al. 2013]. Meanwhile, isolating single scattering relies on either: (a) diluting the sample [Hawkins et al. 2005; Narasimhan et al. 2006], which cannot be easily applied to solids or to liquids that have unknown dispersing media; or, (b) using structured lighting patterns [Mukaigawa et al. 2010], which provide only approximate isolation [Holroyd and Lawrence 2011; Gupta et al. 2011] and therefore induce errors in measured scattering parameters that are difficult to characterize.

Inverse Bi-scale Material Design

Hongzhi Wu Julie Dorsey Holly Rushmeier
Computer Graphics Group, Yale University

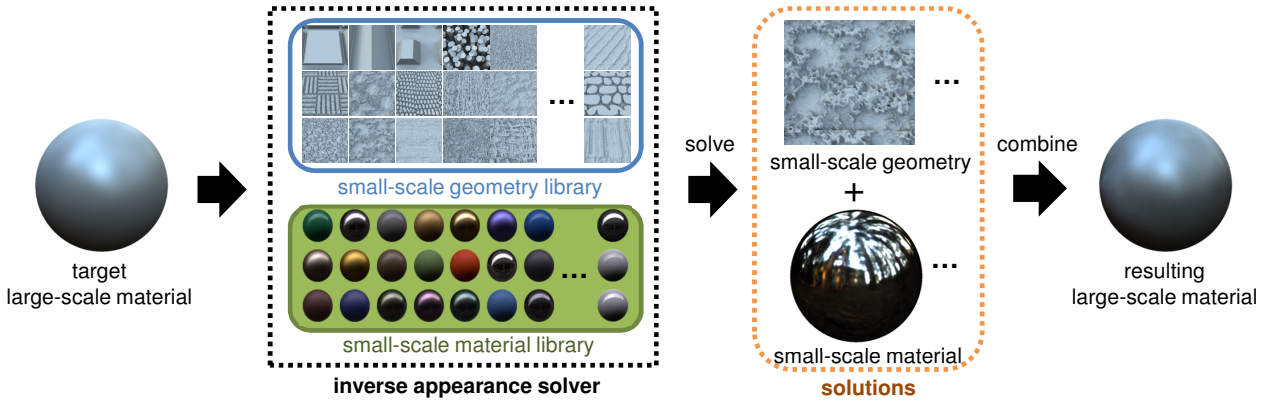


Figure 1: *Given a target large-scale material, our inverse appearance solver efficiently searches through precomputed small-scale material and geometry libraries. Then it returns a few small-scale materials and geometries as results, which approximate the target large-scale appearance when combined.*

Abstract

One major shortcoming of existing bi-scale material design systems is the lack of support for inverse design: there is no way to directly edit the large-scale appearance and then rapidly solve for the small-scale details that approximate that look. Prior work is either too slow to provide quick feedback, or limited in the types of small-scale details that can be handled. We present a novel computational framework for inverse bi-scale material design. The key idea is to convert the challenging inverse appearance computation into efficient search in two precomputed large libraries: one including a wide range of measured and analytical materials, and the other procedurally generated and height-map-based geometries. We demonstrate a variety of editing operations, including finding visually equivalent details that produce similar large-scale appearance, which can be useful in applications such as physical fabrication of materials.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: inverse appearance computation, bi-scale, microfacet

Links: DL PDF WEB VIDEO

ACM Reference Format
Wu, H., Dorsey, J., Rushmeier, H. 2013. Inverse Bi-scale Material Design. ACM Trans. Graph. 32, 6, Article 163 (November 2013), 10 pages. DOI = 10.1145/2508363.2508394 http://doi.acm.org/10.1145/2508363.2508394.

DOI: http://doi.acm.org/10.1145/2508363.2508394

1 Introduction

Deriving large-scale materials from small-scale details is a powerful appearance modeling technique. From a physical point of view, the large-scale look is determined by averaging the appearance of small-scale details [Bruneton and Neyret 2012]. Previous work has succeeded in modeling a wide range of real-world materials, from velvet [Westin et al. 1992], brushed metal [Ashikmin et al. 2000], to woven fabrics [Zhao et al. 2011].

Recently, researchers further pushed the idea by introducing interactive bi-scale material design [Wu et al. 2011; Iwasaki et al. 2012]: the user can manipulate both the small-scale material and geometry; the corresponding change on the large-scale appearance is then rapidly computed. However, one major limitation in existing work is that, the editing operations can only be applied to the small-scale details. There is no way to directly design the large-scale appearance, and then automatically find small-scale details to approximate it. Tedious trial-and-error on the small-scale is needed to achieve even simple edits on the large-scale appearance.

To address the above problem, we would like to build an inverse bi-scale material design system, which turns out to be highly challenging: first, there is a large number of degrees of freedom in the small-scale materials and geometries, making it difficult to efficiently solve for the optimal result in this huge space; second, due to the visibility factor, the effect of small-scale geometry over the large-scale appearance is both non-linear and non-local; moreover, inverse bi-scale material computation is an ill-posed problem, as there might be multiple small-scale materials and geometries, which correspond to similar appearance on the large scale [Han et al. 2007]. Previous papers in this field (e.g., [Weyrich et al. 2009]) solve sophisticated non-linear optimization problems, a process that is far from interactive. In addition, existing approaches (e.g., [Ershov et al. 2004]) handle very limited types of materials and geometry. It is non-trivial to extend these frameworks to more general cases.

This paper presents a novel appearance computational framework for inverse bi-scale material design. Inspired by previous work

Joint Importance Sampling of Low-Order Volumetric Scattering

Iliyan Georgiev[1,2] Jaroslav Křivánek[3] Toshiya Hachisuka[4] Derek Nowrouzezahrai[1,5] Wojciech Jarosz[1]

[1]Disney Research Zürich [2]Saarland University [3]Charles University, Prague [4]Aarhus University [5]Université de Montréal

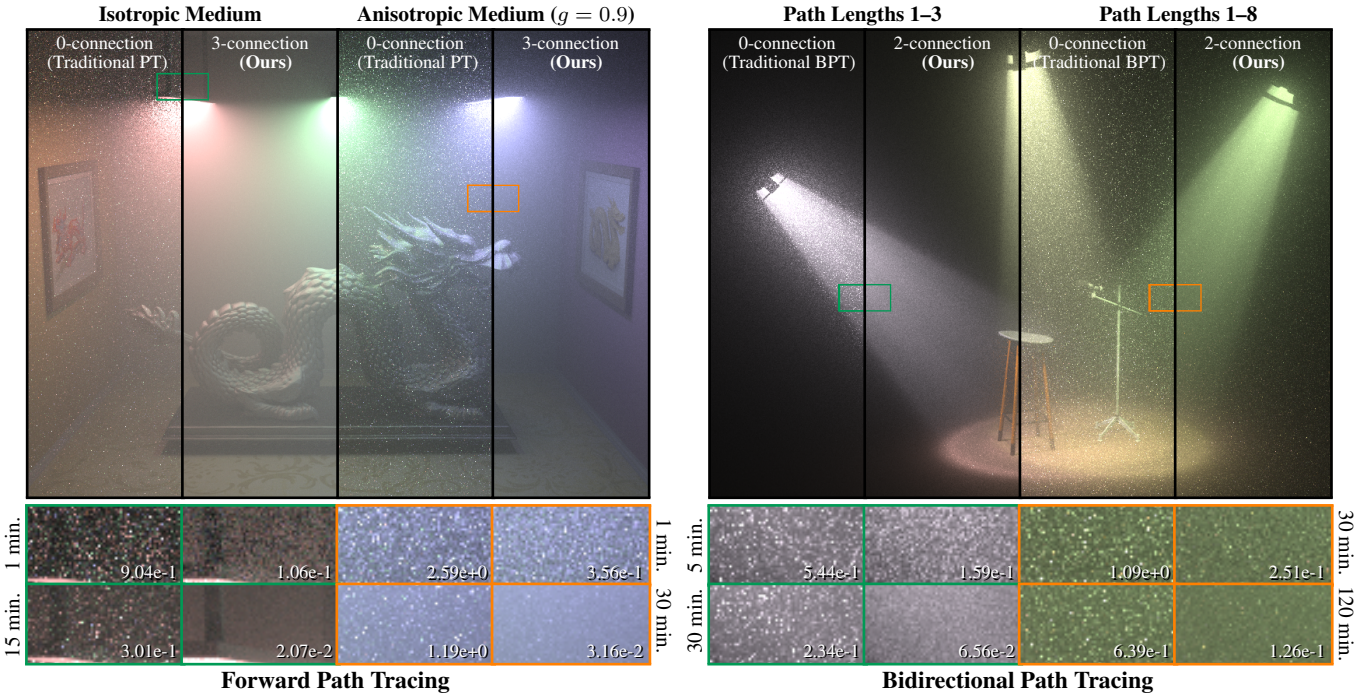


Figure 1: *We present novel importance sampling techniques for constructing paths in participating media and apply them to unidirectional and bidirectional path tracing. By generalizing traditional shadow connections (0-connection) to longer, importance sampled connection subpaths (2- and 3-connection) we obtain* $5\times$ *to* $37\times$ *reduction in RMS error (see zoom-ins), corresponding to* $25\times$ *to* $1444\times$ *reduction in render time.*

Abstract

Central to all Monte Carlo-based rendering algorithms is the construction of light transport paths from the light sources to the eye. Existing rendering approaches sample path vertices *incrementally* when constructing these light transport paths. The resulting probability density is thus a product of the *conditional* densities of each local sampling step, constructed without explicit control over the form of the final *joint* distribution of the complete path. We analyze why current incremental construction schemes often lead to high variance in the presence of participating media, and reveal that such approaches are an unnecessary legacy inherited from traditional surface-based rendering algorithms. We devise *joint importance sampling* of path vertices in participating media to construct paths that explicitly account for the product of all scattering and geometry terms along a sequence of vertices instead of just locally at a single vertex. This leads to a number of practical importance sampling routines to explicitly construct single- and double-scattering subpaths in anisotropically-scattering media. We demonstrate the benefit of our new sampling techniques, integrating them into several path-based rendering algorithms such as path tracing, bidirectional path tracing, and many-light methods. We also use our sampling routines to generalize deterministic shadow connections to *connection subpaths* consisting of two or three random decisions, to efficiently simulate higher-order multiple scattering. Our algorithms significantly reduce noise and increase performance in renderings with both isotropic and highly anisotropic, low-order scattering.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Raytracing

Keywords: global illumination, participating media, importance sampling, path tracing, bidirectional path tracing, many-light rendering methods

Links: DL PDF WEB

ACM Reference Format
Georgiev, I., KÅáriv¬¥anek, J., Hachisuka, T., Nowrouzezahrai, D., Jarosz, W. 2013. Joint Importance Sampling of Low-Order Volumetric Scattering. ACM Trans. Graph. 32, 6, Article 164 (November 2013), 14 pages. DOI = 10.1145/2508363.2508411 http://doi.acm.org/10.1145/2508363.2508411.

0730-0301/13/11-ART164 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508411

Wave-Ray Coupling for Interactive Sound Propagation in Large Complex Scenes

Hengchin Yeh* Ravish Mehra† Zhimin Ren* Lakulish Antani* Dinesh Manocha*
Ming Lin*
University of North Carolina at Chapel Hill

Figure 1: *Our hybrid technique is able to model high-fidelity acoustic effects for large, complex indoor or outdoor scenes at interactive rates: (a) building surrounded by walls, (b) underground parking garage, and (c) reservoir scene in Half-Life 2.*

Abstract

We present a novel hybrid approach that couples geometric and numerical acoustic techniques for interactive sound propagation in complex environments. Our formulation is based on a combination of spatial and frequency decomposition of the sound field. We use numerical wave-based techniques to precompute the pressure field in the near-object regions and geometric propagation techniques in the far-field regions to model sound propagation. We present a novel two-way pressure coupling technique at the interface of near-object and far-field regions. At runtime, the impulse response at the listener position is computed at interactive rates based on the stored pressure field and interpolation techniques. Our system is able to simulate high-fidelity acoustic effects such as diffraction, scattering, low-pass filtering behind obstruction, reverberation, and high-order reflections in large, complex indoor and outdoor environments and Half-Life 2 game engine. The pressure computation requires orders of magnitude lower memory than standard wave-based numerical techniques.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling; I.3.5 [Computer Graphics]: Applications—Sound rendering; I.6.8 [Simulation and Modeling]: Types of simulation—Animation;

Keywords: sound propagation, interactive applications, source simulation, scattering

Links: DL PDF WEB

*{hyeh,zren,lakulish,dm,lin}@cs.unc.edu
†ravish.mehra07@gmail.com

ACM Reference Format
Yeh, H., Mehra, R., Ren, Z., Antani, L., Manocha, D., Lin, M. 2013. Wave-Ray Coupling for Interactive Sound Propagation in Large Complex Scenes. ACM Trans. Graph. 32, 6, Article 165 (November 2013), 11 pages. DOI = 10.1145/2508363.2508420 http://doi.acm.org/10.1145/2508363.2508420.

DOI: http://doi.acm.org/10.1145/2508363.2508420

1 Introduction

Sound propagation techniques are used to model how sound waves travel in the space and interact with various objects in the environment. Sound propagation algorithms are used in many interactive applications, such as computer games or virtual environments, and offline applications, such as noise prediction in urban scenes, architectural acoustics, virtual prototyping, etc.. Realistic sound propagation that can model different acoustic effects, including diffraction, interference, scattering, and late reverberation, can considerably improve a user's immersion in an interactive system and provides spatial localization [Blauert 1983].

The acoustic effects can be accurately simulated by numerically solving the acoustic wave equation. Some of the well-known solvers are based on the boundary-element method, the finite-element method, the finite-difference time-domain method, etc. However, the time and space complexity of these solvers increases linearly with the volume of the acoustic space and is a cubic (or higher) function of the source frequency. As a result, these techniques are limited to interactive sound propagation at low frequencies (e.g. 1-2KHz) [Raghuvanshi et al. 2010; Mehra et al. 2013], and may not scale to large environments.

Many interactive applications use geometric sound propagation techniques, which assume that sound waves travels like rays. This is a valid assumption when the sound wave travels in free space or when the size of intersecting objects is much larger than the wavelength. As a result, these geometric techniques are unable to simulate many acoustic effects at low frequencies, including diffraction, interference, and higher-order wave effects. Many hybrid combinations of numeric and geometric techniques have been proposed, but they are limited to small scenes or offline applications.

Main Results: In this paper, we present a novel hybrid approach that couples geometric and numerical acoustic techniques to perform interactive and accurate sound propagation in complex scenes. Our approach uses a combination of spatial decomposition and frequency decomposition, along with a novel two-way wave-ray coupling algorithm. The entire simulation domain is decomposed into different regions, and the sound field is computed separately by geometric and numerical techniques for each region. In the vicinity of objects whose sizes are comparable to the simulated wavelength (near-object regions), we use numerical wave-based methods to simulate all wave effects. In regions away from objects (far-field

PiCam: An Ultra-Thin High Performance Monolithic Camera Array

Kartik Venkataraman*, Dan Lelescu, Jacques Duparré, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis
Pelican Imaging Corporation

Shree Nayar
Columbia University

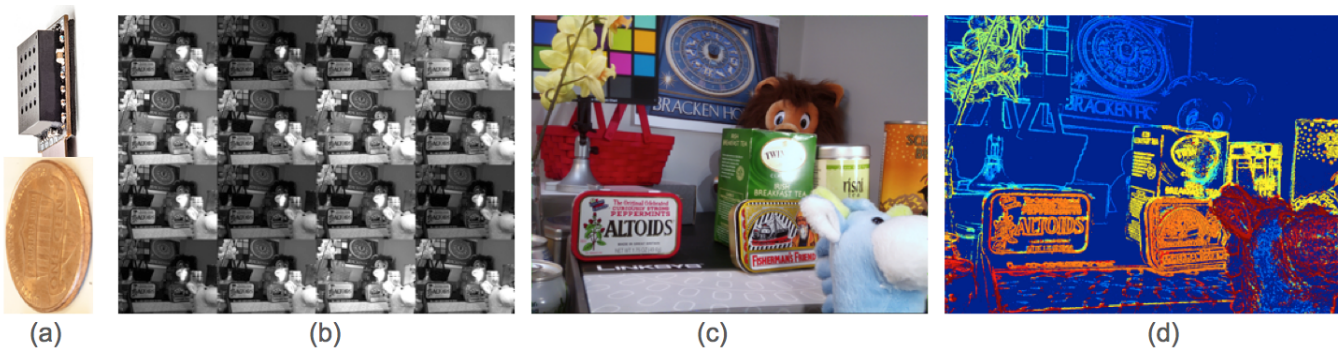


Figure 1: *From left to right - (a) The PiCam Camera Array Module (b) Raw* 4×4 *array images each* 1000×750 *pixels (c) Parallax corrected and superresolved high resolution 8MP Image; (d) A high resolution filtered 8MP depth map*

Abstract

We present *PiCam* (Pelican Imaging Camera-Array), an ultra-thin high performance monolithic camera array, that captures light fields and synthesizes high resolution images along with a range image (scene depth) through integrated parallax detection and superresolution. The camera is passive, supporting both stills and video, low light capable, and small enough to be included in the next generation of mobile devices including smartphones. Prior works [Rander et al. 1997; Yang et al. 2002; Zhang and Chen 2004; Tanida et al. 2001; Tanida et al. 2003; Duparré et al. 2004] in camera arrays have explored multiple facets of light field capture - from viewpoint synthesis, synthetic refocus, computing range images, high speed video, and micro-optical aspects of system miniaturization. However, none of these have addressed the modifications needed to achieve the strict form factor and image quality required to make array cameras practical for mobile devices. In our approach, we customize many aspects of the camera array including lenses, pixels, sensors, and software algorithms to achieve imaging performance and form factor comparable to existing mobile phone cameras.

Our contributions to the post-processing of images from camera arrays include a cost function for parallax detection that integrates across multiple color channels, and a regularized image restoration (superresolution) process that takes into account all the system degradations and adapts to a range of practical imaging conditions. The registration uncertainty from the parallax detection process is integrated into a Maximum-a-Posteriori formulation that synthesizes an estimate of the high resolution image and scene depth. We conclude with some examples of our array capabilities such as post-capture (still) refocus, video refocus, view synthesis to demonstrate motion parallax, 3D range images, and briefly address future work.

CR Categories: I.3.7 [Computer Graphics]: Digitization and Image Capture—Applications I.4.4 [Image Processing and Computer Vision]: Restoration—Inverse filtering I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Range data;

*kartik@pelicanimaging.com

Keywords: plenoptic acquisition, computational camera, light field, array camera, parallax detection, superresolution, depth map

Links: DL PDF WEB VIDEO

1 Introduction

A fundamental problem in designing cameras for mobile devices such as smartphones and tablets is the industrial design constraints that have to be met. For any given XY sensor format there is a corresponding optical format which, through standard optical design constraints, defines the lower limit of the camera module z-height automatically. This is a key industrial design constraint.

We describe an approach to solve the form factor problem by replacing a single aperture in the camera with a number of smaller apertures (lens array) with largely overlapping fields of view. With the smaller optical format (aperture), the focal length and therefore the camera module z-height, are correspondingly reduced. The sensor array together with the lens array are packaged to form an integrated camera array with a thin form factor. We refer to this basic camera array module as *PiCam* (Figure 2). The choice of aperture size and number of apertures in the *PiCam* is guided by the need to regain the resolution that is suppressed in the aperture size reduction through a post-capture superresolution process. The *PiCam* camera architecture specifications are presented in Section 3.

ACM Reference Format
Venkataraman, K., Lelescu, D., Duparr¬¥e, J., McMahon, A., Molina, G., Chatterjee, P., Mullis, R., Nayar, S. 2013. PiCam: An Ultra-Thin High Performance Monolithic Camera Array. ACM Trans. Graph. 32, 6, Article 166 (November 2013), 13 pages. DOI = 10.1145/2508363.2508390 http://doi.acm.org/10.1145/2508363.2508390.

Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles

Achuta Kadambi[1]*, Refael Whyte[2,1], Ayush Bhandari[1], Lee Streeter[2],
Christopher Barsi[1], Adrian Dorrington[2], Ramesh Raskar[1]
[1]Massachusetts Institute of Technology, Boston USA [2]University of Waikato, Waikato NZ

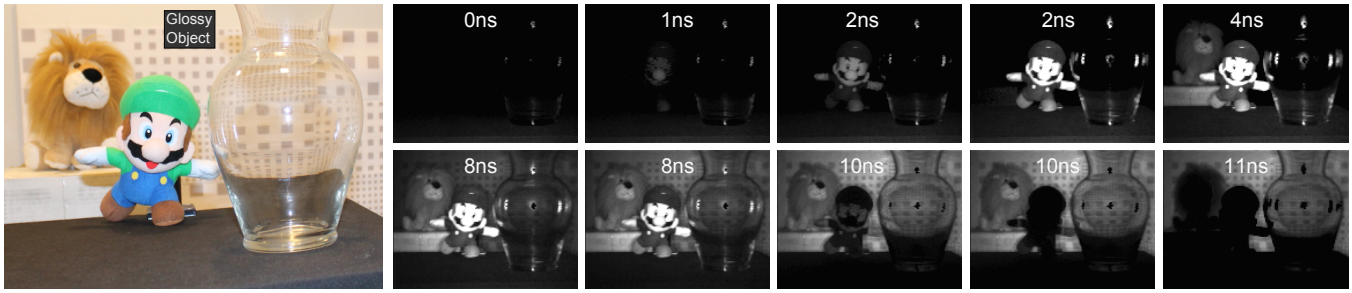


Figure 1: *Using our custom time of flight camera, we are able to visualize light sweeping over the scene. In this scene, multipath effects can be seen in the glass vase. In the early time-slots, bright spots are formed from the specularities on the glass. Light then sweeps over the other objects on the scene and finally hits the back wall, where it can also be seen through the glass vase (8ns). Light leaves, first from the specularities (8-10ns), then from the stuffed animals. The time slots correspond to the true geometry of the scene (light travels 1 foot in a nanosecond, times are for round-trip). Please see http://media.mit.edu/∼achoo/lightsweep for animated light sweep movies.*

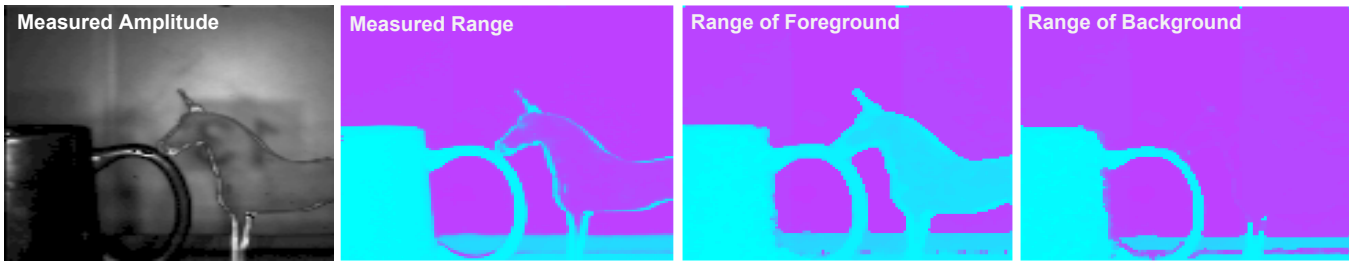


Figure 2: *Recovering depth of transparent objects is a hard problem in general and has yet to be solved for Time of Flight cameras. A glass unicorn is placed in a scene with a wall behind (left). A regular time of flight camera fails to resolve the correct depth of the unicorn (center-left). By using our multipath algorithm, we are able to obtain the depth of foreground (center-right) or of background (right).*

Abstract

Time of flight cameras produce real-time range maps at a relatively low cost using continuous wave amplitude modulation and demodulation. However, they are geared to measure range (or phase) for a single reflected bounce of light and suffer from systematic errors due to multipath interference.

We re-purpose the conventional time of flight device for a new goal: to recover per-pixel sparse time profiles expressed as a sequence of impulses. With this modification, we show that we can not only address multipath interference but also enable new applications such as recovering depth of near-transparent surfaces, looking through diffusers and creating time-profile movies of sweeping light.

Our key idea is to formulate the forward amplitude modulated light propagation as a convolution with custom codes, record samples by introducing a simple sequence of electronic time delays, and perform sparse deconvolution to recover sequences of Diracs that correspond to multipath returns. Applications to computer vision include ranging of near-transparent objects and subsurface imaging through diffusers. Our low cost prototype may lead to new insights regarding forward and inverse problems in light transport.

CR Categories: I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—3D/stereo scene analysis

Keywords: Time of Flight (ToF) cameras, sparse deconvolution, multipath interference, femtophotography, time-coded illumination

Links: DL PDF

1. Introduction

Commercial time of flight (ToF) systems achieve ranging by amplitude modulation of a continuous wave. While ToF cameras provide a single optical path length (range or depth) value per pixel the scene may actually consist of multiple depths, e.g., a transparency in front of a wall. We refer to this as a mixed pixel. Our goal is to re-

*achoo@mit.edu

ACM Reference Format
Kadambi, A., Whyte, R., Bhandari, A., Streeter, L., Barsi, C., Dorrington, A., Raskar, R. 2013. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. ACM Trans. Graph. 32, 6, Article 167 (November 2013), 10 pages. DOI = 10.1145/2508363.2508428 http://doi.acm.org/10.1145/2508363.2508428.

0730-0301/13/11-ART167 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508428

QEx: Robust Quad Mesh Extraction

Hans-Christian Ebke* RWTH Aachen University
David Bommes† INRIA Sophia Antipolis
Marcel Campen* RWTH Aachen University
Leif Kobbelt* RWTH Aachen University

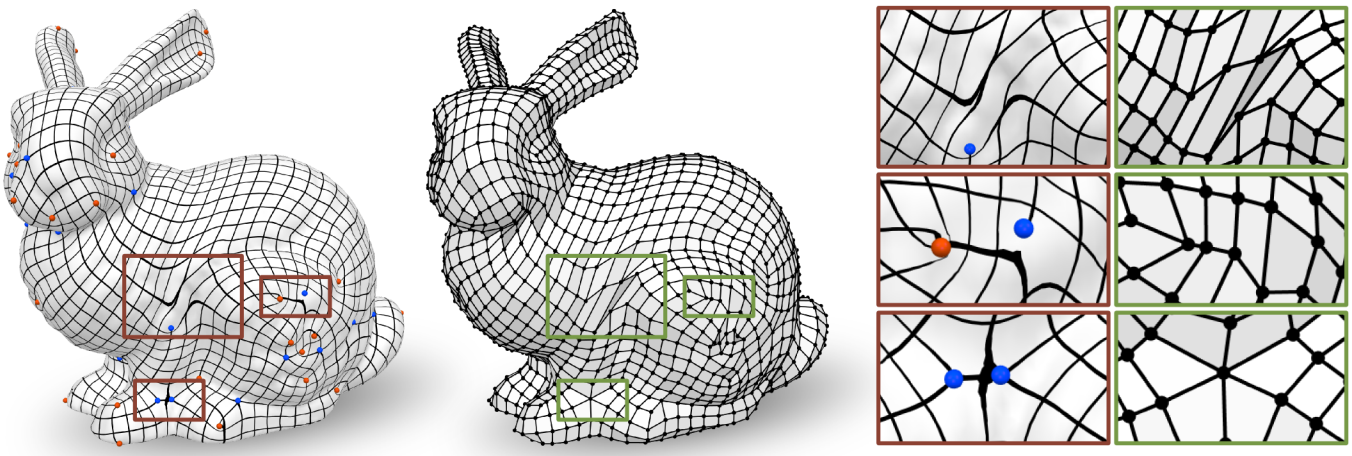


Figure 1: *Parametrization based quad meshing methods cannot guarantee integer-grid maps without fold-overs and degeneracies. The bunny on the left is textured using an integer-grid map generated by the Mixed Integer Quadgrangulation method without stiffening. It contains a number of fold-overs some of which are magnified on the right. The quad mesh generated using QEx, our fold-over tolerant quad extractor is displayed in the center.—Previously, degenerate integer-grid maps such as this one were considered useless and state of the art quad meshing methods spend a considerable amount of run time to fix them.*

Abstract

The most popular and actively researched class of quad remeshing techniques is the family of *parametrization based quad meshing methods*. They all strive to generate an *integer-grid map*, i.e. a parametrization of the input surface into $\mathbb{R}^2$ such that the canonical grid of integer iso-lines forms a quad mesh when mapped back onto the surface in $\mathbb{R}^3$. An essential, albeit broadly neglected aspect of these methods is the *quad extraction* step, i.e. the materialization of an actual quad mesh from the mere "quad texture". Quad (mesh) extraction is often believed to be a trivial matter but quite the opposite is true: numerous special cases, ambiguities induced by numerical inaccuracies and limited solver precision, as well as imperfections in the maps produced by most methods (unless costly countermeasures are taken) pose significant challenges to the quad extractor. We present a method to sanitize a provided parametrization such that it becomes numerically consistent even in a limited precision floating point representation. Based on this we are able to provide a comprehensive and sound description of how to perform quad extraction robustly and without the need for any complex tolerance thresholds or disambiguation rules. On top of that we develop a novel strategy to cope with common local fold-overs in the parametrization. This allows our method, dubbed *QEx*, to generate all-quadrilateral meshes where otherwise holes, non-quad polygons or no output at all would have been produced. We thus enable the practical use of an entire class of maps that was previously considered defective. Since state of the art quad meshing methods spend a significant share of their run time solely to prevent local fold-overs, using our method it is now possible to obtain quad meshes significantly quicker than before. We also provide `libQEx`, an open source C++ reference implementation of our method and thus significantly lower the bar to enter the field of quad meshing.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

Keywords: quad extraction, quad meshing, integer-grid maps

Links: DL PDF WEB CODE

1 Introduction

All algorithms in the very popular and actively researched class of parametrization based quad meshing approaches share one common trait: they strive to generate an *integer-grid map* [Bommes et al. 2013a] which refers to parametrizations which map the input mesh into $\mathbb{R}^2$ in such a way that the canonical grid of integer iso-lines forms a quad mesh when mapped back onto the surface in $\mathbb{R}^3$. While these parametrizations implicitly define a quad mesh, it is necessary to materialize an explicit polygonal quad mesh for virtually all applications. We call this process *quad extraction*.

Judging from the lack of attention the descriptions of all state of the art parametrization based quad meshing methods pay to the quad extraction post-process, one is led to conclude that quad extraction is a trivial matter. However, typical integer-grid parametrizations exhibit a plethora of pitfalls that cause naïve quad extractors to fail. Transition functions pedantically have to be kept track of and nu-

*e-mail:{ebke,campen,kobbelt}@cs.rwth-aachen.de
†e-mail:david.bommes@inria.fr

ACM Reference Format
Ebke, H., Bommes, D., Campen, M., Kobbelt, L. 2013. QEx: Robust Quad Mesh Extraction. ACM Trans. Graph. 32, 6, Article 168 (November 2013), 10 pages. DOI = 10.1145/2508363.2508372 http://doi.acm.org/10.1145/2508363.2508372.

0730-0301/13/11-ART168 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508372

Real-time 3D Reconstruction at Scale using Voxel Hashing

Matthias Nießner[1,3] Michael Zollhöfer[1] Shahram Izadi[2] Marc Stamminger[1]
[1]University of Erlangen-Nuremberg [2]Microsoft Research Cambridge [3]Stanford University

Figure 1: *Example output from our reconstruction system without any geometry post-processing. Scene is about* $20m$ *wide and* $4m$ *high and captured online in less than 5 minutes with live feedback of the reconstruction.*

Abstract

Online 3D reconstruction is gaining newfound interest due to the availability of real-time consumer depth cameras. The basic problem takes live overlapping depth maps as input and incrementally fuses these into a single 3D model. This is challenging particularly when real-time performance is desired without trading quality or scale. We contribute an online system for large and fine scale volumetric reconstruction based on a memory and speed efficient data structure. Our system uses a simple spatial hashing scheme that compresses space, and allows for real-time access and updates of implicit surface data, without the need for a regular or hierarchical grid data structure. Surface data is only stored densely where measurements are observed. Additionally, data can be streamed efficiently in or out of the hash table, allowing for further scalability during sensor motion. We show interactive reconstructions of a variety of scenes, reconstructing both fine-grained details and large scale environments. We illustrate how all parts of our pipeline from depth map pre-processing, camera pose estimation, depth map fusion, and surface rendering are performed at real-time rates on commodity graphics hardware. We conclude with a comparison to current state-of-the-art online systems, illustrating improved performance and reconstruction quality.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and Scanning

Keywords: real-time reconstruction, scalable, data structure, GPU

Links: DL PDF

ACM Reference Format
Nießner, M., Zollhöfer, M., Izadi, S., Stamminger, M. 2013. Real-time 3D Reconstruction at Scale using Voxel Hashing. ACM Trans. Graph. 32, 6, Article 169 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508374 http://doi.acm.org/10.1145/2508363.2508374.

DOI: http://doi.acm.org/10.1145/2508363.2508374

1 Introduction

While 3D reconstruction is an established field in computer vision and graphics, it is now gaining newfound momentum due to the wide availability of depth cameras (such as the Microsoft Kinect and Asus Xtion). Since these devices output live but noisy depth maps, a particular focus of recent work is *online* surface reconstruction using such consumer depth cameras. The ability to obtain reconstructions in *real-time* opens up various interactive applications including: augmented reality (AR) where real-world geometry can be fused with 3D graphics and rendered live to the user; autonomous guidance for robots to reconstruct and respond rapidly to their environment; or even to provide immediate feedback to users during 3D scanning.

Online reconstruction requires incremental *fusion* of many overlapping depth maps into a single 3D representation that is continuously refined. This is challenging particularly when real-time performance is required without trading fine-quality reconstructions and spatial scale. Many state-of-the-art online techniques therefore employ different types of underlying data structures accelerated using graphics hardware. These however have particular trade-offs in terms of reconstruction speed, scale, and quality.

Point-based methods (e.g., [Rusinkiewicz et al. 2002; Weise et al. 2009]) use simple unstructured representations that closely map to range and depth sensor input, but lack the ability to directly reconstruct connected surfaces. High-quality online scanning of small objects has been demonstrated [Weise et al. 2009], but larger-scale reconstructions clearly trade quality and/or speed [Henry et al. 2012; Stückler and Behnke 2012]. *Height-map* based representations [Pollefeys et al. 2008; Gallup et al. 2010] support efficient compression of connected surface data, and can scale efficiently to larger scenes, but fail to reconstruct complex 3D structures.

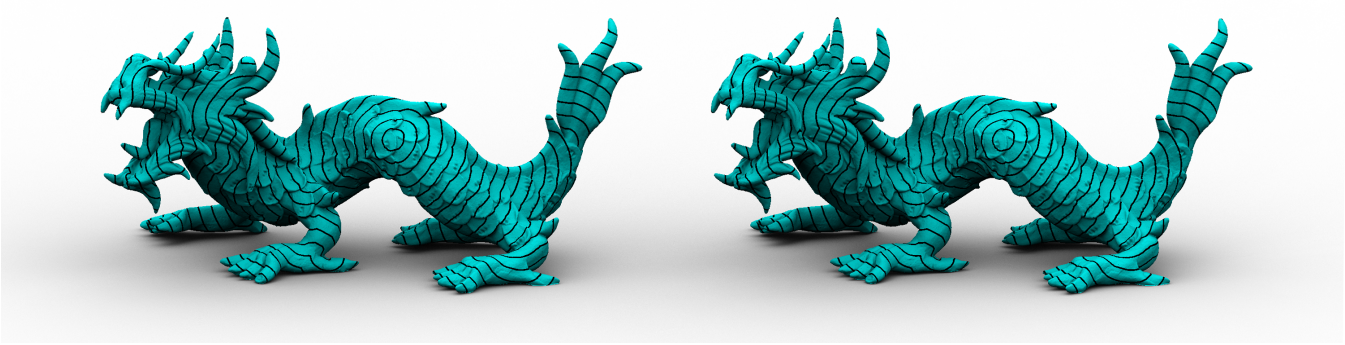
For active sensors, implicit *volumetric* approaches, in particular the method of Curless and Levoy [1996], have demonstrated compelling results [Curless and Levoy 1996; Levoy et al. 2000; Zhou and Koltun 2013], even at real-time rates [Izadi et al. 2011; Newcombe et al. 2011]. However, these rely on memory inefficient regular voxel grids, in turn restricting scale. This has led to either *moving volume* variants [Roth and Vona 2012; Whelan et al. 2012], which stream voxel data out-of-core as the sensor moves, but still constrain the size of the active volume. Or *hierarchical* data structures that subdivide space more effectively, but do not parallelize efficiently given added computational complexity [Zeng et al. 2012; Chen et al. 2013].

Saddle Vertex Graph (SVG): A Novel Solution to the Discrete Geodesic Problem

Xiang Ying
Nanyang Technological University

Xiaoning Wang
Nanyang Technological University

Ying He*
Nanyang Technological University



Exact polyhedral distance: ICH 127s; MMP 144s Our result: 1.96s, max err. 0.11%, rms err. 0.05%, mean err. 0.04%

Figure 1: *It takes* 39.1 *seconds to construct an approximate SVG for the* 1.5*M-face Dragon on an Nvidia Tesla K20 GPU. Then any subsequent computation of the single source geodesic distance takes less than* 2.0*s on a* 2.66*GHz Intel Xeon machine using a single core. Therefore, our method is highly desirable to the applications that require frequent geodesic computations. Moreover, our method guarantees the computed geodesic distance is a metric, which distinguishes itself from all the other approximate geodesic algorithms.*

Abstract

This paper presents the Saddle Vertex Graph (SVG), a novel solution to the discrete geodesic problem. The SVG is a sparse undirected graph that encodes complete geodesic distance information: a geodesic path on the mesh is equivalent to a shortest path on the SVG, which can be solved efficiently using the shortest path algorithm (e.g., Dijkstra algorithm). The SVG method solves the discrete geodesic problem from a local perspective. We have observed that the polyhedral surface has some interesting and unique properties, such as the fact that the discrete geodesic exhibits a strong local structure, which is not available on the smooth surfaces. The richer the details and complicated geometry of the mesh, the stronger such local structure will be. Taking advantage of the local nature, the SVG algorithm breaks down the discrete geodesic problem into significantly smaller sub-problems, and elegantly enables information reuse. It does not require any numerical solver, and is numerically stable and insensitive to the mesh resolution and tessellation. Users can intuitively specify a model-independent parameter K, which effectively balances the SVG complexity and the accuracy of the computed geodesic distance. More importantly, the computed distance is guaranteed to be a metric. The experimental results on real-world models demonstrate significant improvement to the existing approximate geodesic methods in terms of both performance and accuracy.

*Corresponding author. e-mail: yhe@ntu.edu.sg

ACM Reference Format
Ying, X., Wang, X., He, Y. 2013. Saddle Vertex Graph (SVG): A Novel Solution to the Discrete Geodesic Problem. ACM Trans. Graph. 32, 6, Article 170 (November 2013), 12 pages.
DOI = 10.1145/2508363.2507379 http://doi.acm.org/10.1145/2508363.2507379.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

Keywords: Discrete geodesic, saddle vertex graph, shortest path

Links: DL PDF

1 Introduction

The discrete geodesic problem has attracted a great deal of attention since Mitchell, Mount and Papadimitriou published their seminal paper in 1987. The MMP algorithm [Mitchell et al. 1987] computes the single-source geodesic distance in $O(n^2 \log n)$ time, where n is the number of vertices of the input mesh. In 1990, Chen and Han [1990] improved the time complexity to $O(n^2)$, which remains the best-known complexity. However, extensive experiments have shown that the CH algorithm often runs much slower than the MMP algorithm. Xin and Wang realized that such slowness is due to the extremely large number of operations to the useless windows (a data structure to that will be explained later), so they modified the CH algorithm by adopting window filtering and maintaining windows in a priority queue. Although the improved CH (or ICH) algorithm has $O(n^2 \log n)$ time complexity, it outperforms both the MMP and the CH algorithms.

So this raises an interesting question: why can the theoretical time complexity not indicate the real performance in the discrete geodesic problem? The exact geodesic algorithms, such as MMP, CH and ICH, all adopt the same computational framework: each edge is partitioned into a set of intervals, called windows, and each window encodes the geodesic distance from the source to any point in the interval. The windows are maintained in either a sorted data structure (like a priority queue in the MMP and ICH algorithms) or a hierarchical data structure (like a tree in the CH algorithm). In each iteration, the algorithms propagate one window across a triangle by computing the window's children windows. The algorithms terminate when all windows have been processed. Thus, window complexity is a key factor to measure the real performance of the algorithm. Mitchell et al. [1987] and Chen and Han [1990] noted

PolyCut: Monotone Graph-Cuts for PolyCube Base-Complex Construction

Marco Livesu
Universita' di Cagliari

Nicholas Vining
University of British Columbia

Alla Sheffer
University of British Columbia

James Gregson
University of British Columbia

Riccardo Scateni
Universita' di Cagliari

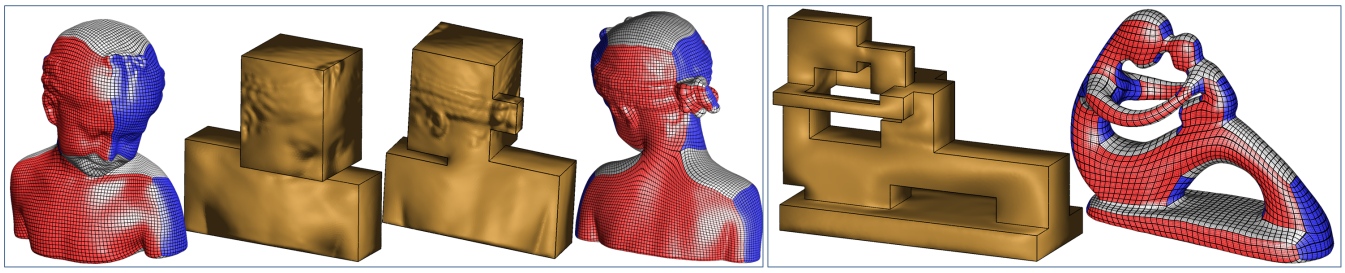


Figure 1: *PolyCut generated PolyCubes combine a small singularity count with low distortion. (Bimba con Nostra mesh provided courtesy of INRIA by the AIM@SHAPE model repository; fertility idol provided courtesy of Frank ter Haar by the AIM@SHAPE model repository.)*

Abstract

PolyCubes, or orthogonal polyhedra, are useful as parameterization base-complexes for various operations in computer graphics. However, computing quality PolyCube base-complexes for general shapes, providing a good trade-off between mapping distortion and singularity counts, remains a challenge. Our work improves on the state-of-the-art in PolyCube computation by adopting a graph-cut inspired approach. We observe that, given an arbitrary input mesh, the computation of a suitable PolyCube base-complex can be formulated as associating, or labeling, each input mesh triangle with one of six signed principal axis directions. Most of the criteria for a desirable PolyCube labeling can be satisfied using a multi-label graph-cut optimization with suitable *local* unary and pairwise terms. However, the highly constrained nature of PolyCubes, imposed by the need to align each chart with one of the principal axes, enforces additional *global* constraints that the labeling must satisfy. To enforce these constraints, we develop a constrained discrete optimization technique, *PolyCut*, which embeds a graph-cut multi-label optimization within a hill-climbing local search framework that looks for solutions that minimize the cut energy while satisfying the global constraints. We further optimize our generated PolyCube base-complexes through a combination of distortion-minimizing deformation, followed by a labeling update and a final PolyCube parameterization step. Our *PolyCut* formulation captures the desired properties of a PolyCube base-complex, balancing parameterization distortion against singularity count, and produces demonstrably better PolyCube base-complexes then previous work.

ACM Reference Format
Livesu, M., Vining, N., Sheffer, A., Gregson, J., Scateni, R. 2013. PolyCut: Monotone Graph-Cuts for PolyCube Base-Complex Construction. ACM Trans. Graph. 32, 6, Article 171 (November 2013), 12 pages. DOI = 10.1145/2508363.2508388 http://doi.acm.org/10.1145/2508363.2508388.

0730-0301/13/11-ART171 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508388

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—;

Keywords: PolyCube, mesh parameterization, mesh segmentation

Links: DL PDF

1 Introduction

PolyCubes, or orthogonal polyhedra, are volumes bounded by axis-aligned planes. Recent research has highlighted the advantages of using PolyCubes as base-complexes for parametrizing closed surfaces and volumes for applications including surface texture mapping [Tarini et al. 2004; Yao and Lee 2008], hexahedral meshing [Gregson et al. 2011; Xia et al. 2010], trivariate spline fitting [Wang et al. 2007], and volumetric texturing [Chang and Lin 2010]. However, construction of quality PolyCubes that provide an optimal trade-off between parametrization distortion and chart or singularity counts remains a challenge. Consequently, most of the works above rely on manually or semi-manually constructed PolyCubes.

We improve on the state-of-the-art in automatic PolyCube base-complex computation by recasting the problem as one of mesh segmentation, or labeling, followed by PolyCube geometry extraction and subsequent parameterization. We observe that, by construction, a PolyCube parameterization re-orients triangle normals on the input mesh, aligning each normal with a signed principal axis direction. These per-triangle alignment choices define a segmentation of the surface mesh into charts, which map to axis-aligned faces of the PolyCube. We therefore formulate PolyCube base-complex extraction as a labeling computation which associates each input mesh triangle with one of six possible orientations minimizing the labeling cost subject to global constraints arising from the requirements to minimize distortion and ensure PolyCube validity. To compute the desired labeling we use a discrete optimization framework, we name *PolyCut*, that embeds multi-label graph-cut optimization within a local search algorithm that resolve these global constraints.

We demonstrate our new method on a large number of inputs and compare our results to the state-of-the art [Lin et al. 2008; He et al. 2009; Gregson et al. 2011], highlighting the improvement in terms of parameterization distortion, compactness, and singularity counts

A Compact Random-Access Representation for Urban Modeling and Rendering

Zhengzheng Kuang Bin Chan Yizhou Yu Wenping Wang
The University of Hong Kong

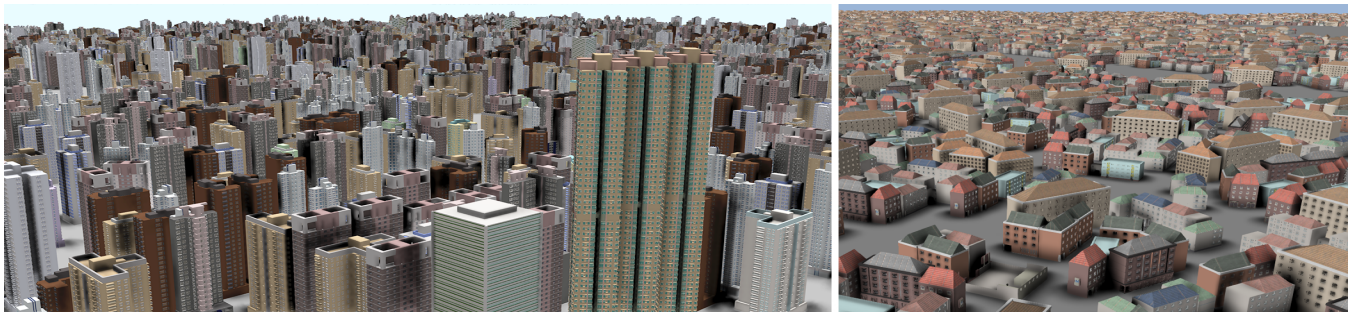


Figure 1: *Left: a city scene with 5,061 distinct buildings represented using our proposed Non-Uniform Textures (NUTs). The size of this digital model is 42 MB, compressible to 10.6MB PNG files. This model is rendered at 49 fps @ 1280×720 on an NVidia GTX580 GPU. An equivalent mesh-based model has 671 million triangles (37GB). Right: a much larger European-style city with 40,400 distinct buildings. This model consists of NUTs (229MB) and a complementary mesh (300MB) for roof structures. The total PNG-compressed NUT file size is 60.5MB. It is rendered at 30 fps. An equivalent mesh model has 1.36 billion triangles, or 61GB.*

Abstract

We propose a highly memory-efficient representation for modeling and rendering urban buildings composed predominantly of rectangular block structures, which can be used to completely or partially represent most modern buildings. With the proposed representation, the data size required for modeling most buildings is more than two orders of magnitude less than using the conventional mesh representation. In addition, it substantially reduces the dependency on conventional texture maps, which are not space-efficient for defining visual details of building facades. The proposed representation can be stored and transmitted as images and can be rendered directly without any mesh reconstruction. A ray-casting based shader has been developed to render buildings thus represented on the GPU with a high frame rate to support interactive fly-by as well as street-level walk-through. Comparisons with standard geometric representations and recent urban modeling techniques indicate the proposed representation performs well when viewed from a short and long distance.

CR Categories: I.3.2 [Computer Graphics]: Graphics Systems—Distributed/network graphics I.3.6 [Computer Graphics]: Methodology and Techniques—Graphics data structures and data types I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Raytracing;

Keywords: urban modeling, texture mapping, ray casting

Links: DL PDF VIDEO

ACM Reference Format
Kuang, Z., Chan, B., Yu, Y., Wang, W. 2013. A Compact Random-Access Representation for Urban Modeling and Rendering. ACM Trans. Graph. 32, 6, Article 172 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508424 http://doi.acm.org/10.1145/2508363.2508424.

DOI: http://doi.acm.org/10.1145/2508363.2508424

1 Introduction

Urban visualization is a challenging task because handling the large number of buildings involved demands high storage capacity, transmission speed and rendering performance. Although systems such as Google Earth enable interactive browsing of urban environments, the visual quality is limited by coarse 3D models and low-resolution textures, making it inadequate for many applications that require highly detailed models and high-resolution rendering.

While detailed building models are desired for an improved viewing experience, the required data size can be prohibitively large. For instance, a detailed model of a multi-story building can easily consume tens of megabytes of memory. Representing a moderate-sized city model with thousands of such buildings would require tens or even hundreds of gigabytes. The conventional mesh representation is commonly used for 3D models because it can approximate arbitrary shapes and be rendered efficiently on the GPU. Nevertheless, it incurs excessive memory consumption.

Urban buildings are often composed of rectangular elements, corners with a right angle, horizontal or vertical edges and faces as well as chamfers and fillets. We wish to explore these characteristics to develop an efficient representation to drastically reduce the amount of data that needs to be stored, processed or transmitted for large-scale urban visualization.

Many existing procedural urban modeling techniques assume repeated instances of the same set of building elements or facade textures across a city. Such an assumption is often invalid in reality. In this paper we adopt a more realistic setting where there exist a large number of distinct buildings in a city, a condition under which many other methods would not perform well. We propose a new building representation, called *Non-Uniform Texture* , or *NUT* for short, that facilitates the individual representation of each building. The NUT completely relies on an image-based format without involving any mesh data structures. A building or a block of aligned buildings can be represented as a NUT in a very compact manner. The NUT can be directly transmitted to a client's GPU to be rendered efficiently without being converted to any other format first. The memory footprint of a NUT is typically more than two orders of magnitude smaller than that of its mesh counterpart. Our exper-

Content-Adaptive Image Downscaling

Johannes Kopf
Microsoft Research

Ariel Shamir
The Interdisciplinary Center

Pieter Peers
College of William & Mary

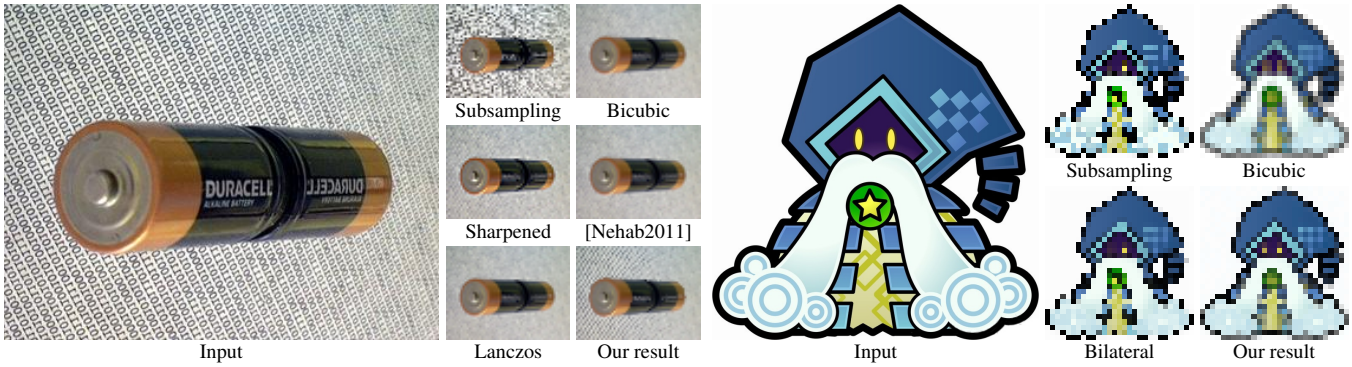


Figure 1: *Previous* content-insensitive *downscaling methods have to compromise between preserving sharpness while introducing aliasing artifacts (e.g., subsampling), or preventing aliasing at the expense of smoothing out fine details and edges (e.g., Bicubic, Lanczos, etc.). Our new* content-adaptive *algorithm provides a more balanced result, that is crisp and contains neither noise nor ringing, and mostly avoids aliasing artifacts. ("Merlon" input image © Nintendo Co., Ltd.)*

Abstract

This paper introduces a novel *content-adaptive* image downscaling method. The key idea is to optimize the shape and locations of the downsampling kernels to better align with local image features. Our content-adaptive kernels are formed as a bilateral combination of two Gaussian kernels defined over space and color, respectively. This yields a continuum ranging from smoothing to edge/detail preserving kernels driven by image content. We optimize these kernels to represent the input image well, by finding an output image from which the input can be well reconstructed. This is technically realized as an iterative maximum-likelihood optimization using a constrained variation of the Expectation-Maximization algorithm. In comparison to previous downscaling algorithms, our results remain crisper without suffering from ringing artifacts. Besides natural images, our algorithm is also effective for creating *pixel art* images from vector graphics inputs, due to its ability to keep linear features sharp *and* connected.

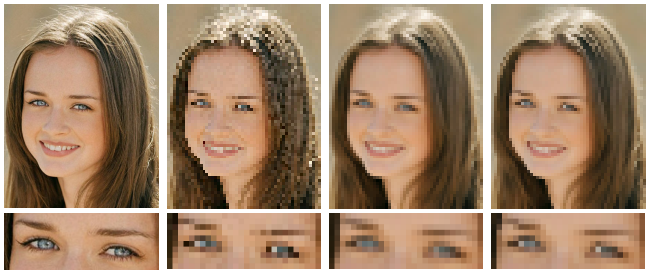
CR Categories: I.4.1 [Computer Graphics]: Image Processing and Computer Vision—Sampling

Keywords: Images, Downscaling

Links: DL PDF WEB CODE

1 Introduction

Downscaling is perhaps *the* most commonly used image operation today. We rarely view a photo we just took at its original resolution anymore. Instead, it is instantly reduced from its original multi-megapixel size to much smaller dimensions to be viewed on a camera viewfinder, on a computer or mobile screen, or on the web.



(a) Input (b) Subsampling (c) Bicubic (d) Our result

Figure 2: *Balancing sharpness and antialiasing: the* subsampled *result is sharp everywhere but suffers severely from noise and aliasing. On the other hand, the* bicubic *result, over-smoothes detail in the face.* Our algorithm *avoids both problems and produces a crisp, noise-free image that exhibits only minimal aliasing.*

The de facto standard for image downscaling are linear filters, originating from the signal processing community [Wolberg 1990]. Here, the image is first convolved with a low-pass kernel to reduce the bandwidth before it is resampled to the final resolution. The filtering pushes the image below the Nyquist frequency and prevents aliasing, but as a side effect the result might suffer from loss of fine details and blurring of sharp edges (Figure 2c). Sharpening these images or using kernels that more closely model a *sinc* filter (e.g., Lanczos) can cause ringing (Figure 3), while simply subsampling the image *without* prefiltering typically leads to strong aliasing artifacts (Figure 2b). Because all these methods are *content-invariant* (i.e., they use invariant kernels), the tradeoff between preserving detail and preventing aliasing is global.

In this work we present a new *content-adaptive* downscaling algorithm. As in classic methods, the output pixels are computed as a weighted sum of the input pixels. This can be interpreted as associating an averaging kernel to every output pixel. Our key idea is to adapt the shape of these kernels in order to better align them with local image features (Figure 2d). Following previous techniques such as bilateral filtering [Tomasi and Manduchi 1998] and mean shift [Comaniciu et al. 2002], we use kernels that are a combination of a spatial Gaussian kernel ensuring locality, and a color space Gaussian kernel for alignment with the image content. The simple parametric form of these kernels achieves a good trade-off between

ACM Reference Format
Kopf, J., Shamir, A., Peers, P. 2013. Content-Adaptive Image Downscaling. ACM Trans. Graph. 32, 6, Article 173 (November 2013), 8 pages. DOI = 10.1145/2508363.2508370 http://doi.acm.org/10.1145/2508363.2508370.

0730-0301/13/11-ART173 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508370

“Mind the Gap”: Tele-Registration for Structure-Driven Image Completion

Hui Huang[⋆∗] Kangxue Yin[⋆] Minglun Gong[†] Dani Lischinski[‡] Daniel Cohen-Or[§] Uri Ascher[✳] Baoquan Chen[⋆◇∗]
[⋆]Shenzhen VisuCA Key Lab / SIAT [†]Memorial University of Newfoundland [‡]The Hebrew University of Jerusalem
[§]Tel-Aviv University [✳]The University of British Columbia [◇]Shandong University

Figure 1: *Given several pieces extracted from the original image and casually placed together (left), our method applies* ***tele-registration*** *to align them (middle), and then uses* ***structure-driven*** *image completion to fill the gaps (right).*

Abstract

Concocting a plausible composition from several non-overlapping image pieces, whose relative positions are not fixed in advance and without having the benefit of priors, can be a daunting task. Here we propose such a method, starting with a set of sloppily pasted image pieces with gaps between them. We first extract salient curves that approach the gaps from non-tangential directions, and use likely correspondences between pairs of such curves to guide a novel *tele-registration* method that simultaneously aligns all the pieces together. A *structure-driven* image completion technique is then proposed to fill the gaps, allowing the subsequent employment of standard in-painting tools to finish the job.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—[Curve, surface, solid, and object representations]

Keywords: image completion, image editing, registration, stitching, texture synthesis

Links: DL PDF WEB

1 Introduction

Image completion is a challenging task, as it attempts to conjure visual detail inside a missing portion of an image. Many completion techniques have been proposed in the computer graphics and image processing literature during the past decade, some of which maturing to the point of being featured in commercial image editing products, such as Adobe Photoshop.

In virtually all of the existing techniques, however, the shape and position of the missing regions (holes) in the image is fixed, and the focus is on filling them with visually plausible content, while ensuring visual continuity across the boundaries with the known image regions. In this work we consider a significantly more difficult version of the problem, where the exact relative placement of the non-overlapping input image parts is not provided, adding a whole new dimension to consider. An example demonstrating this problem is shown in Figure 1, where the input is an assembly of sloppily pasted image pieces and the result is a complete, natural looking picture. This challenging scenario arises, for example, when creating a digital photomontage [Agarwala et al. 2004] from input images that might not be registered, attempting to copy-and-paste an object (either between two different images or within the same image) [Pérez et al. 2003], or creating a panorama from input images which might not overlap with each other [Poleg and Peleg 2012], or might even depict completely different places on earth.

Different steps of the proposed approach are illustrated in Figures 2-4 using a synthetic example. Given a set of image pieces that are casually pasted together, with gaps left between the pieces, our goal is to first align the pieces relative to each other despite their lack of overlap, and then fill the remaining gaps. We begin with detecting salient curves inside each image piece (Section 3 and Figure 2), and then attempt to find for each curve a matching curve from an adjacent piece, across the gap (Section 4.1, Figure 3). The matched salient curves are used to construct a vector field surrounding all the pieces (Section 4.2). Next, we use this *ambient vector field* to find a similarity transformation (any combination of translation, rotation, and uniform scaling) for each piece, such that the corresponding pairs of salient curves line up (Section 4.3), and construct smooth *bridging curves* that connect such pairs across gaps. Finally, we fill the gaps using structure-driven synthesis (Section 5, Figure 4), while any remaining inside/outside holes are completed using standard inpainting tools, e.g., [Barnes et al. 2009; Darabi et al. 2012].

In summary, the main contributions of our method are two-fold. The first is our novel *tele-registration* method, which simultaneously optimizes the positions of multiple disjoint non-overlapping image pieces with respect to one another. The tele-registration process is automatic, requiring the user only to provide an approximate initial placement of the pieces, which is why we refer to this pro-

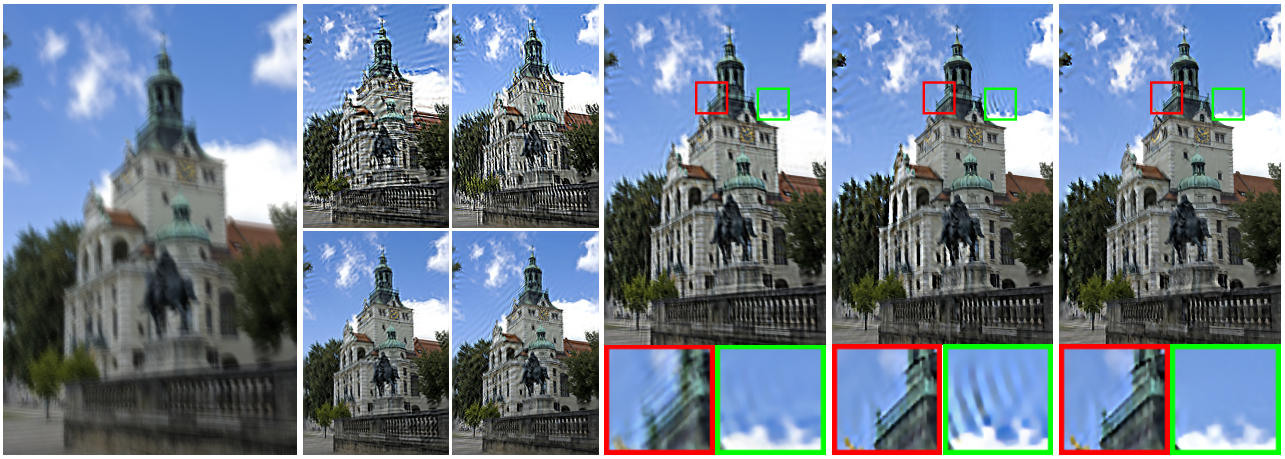
∗Corresponding authors: Hui Huang (hhzhiyan@gmail.com), Baoquan Chen (baoquan.chen@gmail.com)

ACM Reference Format
Huang, H., Yin, K., Gong, M., Lischinski, D., Cohen-Or, D., Ascher, U., Chen, B. 2013. Mind the Gap: Tele-Registration for Structure-Driven Image Completion. ACM Trans. Graph. 32, 6, Article 174 (November 2013), 10 pages. DOI = 10.1145/2508363.2508373 http://doi.acm.org/10.1145/2508363.2508373.

DOI: http://doi.acm.org/10.1145/2508363.2508373

A No-Reference Metric for Evaluating the Quality of Motion Deblurring

Yiming Liu[1]* Jue Wang[2] Sunghyun Cho[2] Adam Finkelstein[1] Szymon Rusinkiewicz[1]
[1]Princeton University [2]Adobe Research



(a) (b) (c) (d) (e)

Figure 1: *We develop a no-reference metric for evaluating the perceptual quality of image motion deblurring results. The metric can be used for fusing multiple deblurring results (b) of the same input image (a) to generate one with the best quality (e). (c) and (d) are the result of simply averaging all deblurring results and the result of a naive fusion method, respectively. See text in Sec. 6.3 for more details. Original image courtesy digital cat@Flickr.*

Abstract

Methods to undo the effects of motion blur are the subject of intense research, but evaluating and tuning these algorithms has traditionally required either user input or the availability of ground-truth images. We instead develop a metric for automatically predicting the perceptual quality of images produced by state-of-the-art deblurring algorithms. The metric is learned based on a massive user study, incorporates features that capture common deblurring artifacts, and does not require access to the original images (i.e., is "no-reference"). We show that it better matches user-supplied rankings than previous approaches to measuring quality, and that in most cases it outperforms conventional full-reference image-similarity measures. We demonstrate applications of this metric to automatic selection of optimal algorithms and parameters, and to generation of fused images that combine multiple deblurring results.

CR Categories: I.3.0 [Computer Graphics]: General—;

Keywords: image quality metric, no-reference, percetually-validated, deblurring

Links: DL PDF WEB

*Part of this work was done when the first author was an intern at Adobe Research.

ACM Reference Format
Liu, Y., Wang, J., Cho, S., Finkelstein, A., Rusinkiewicz, S. 2013. A No-Reference Metric for Evaluating the Quality of Motion Deblurring. ACM Trans. Graph. 32, 6, Article 175 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508391 http://doi.acm.org/10.1145/2508363.2508391.

0730-0301/13/11-ART175 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508391

1 Introduction

The wide availability and ever-increasing sophistication of modern image processing and computational photography algorithms has brought about a need to evaluate their results. For instance, for a task such as image deblurring, a realistic characterization of image quality and the presence or absence of artifacts is necessary to select between different methods, as well as to choose parameters for each algorithm. Lacking an automated method for image quality assessment, many systems resort to asking the user. This, however, becomes increasingly impractical if dozens of algorithms and hundreds of parameter settings must be compared. While a large-scale user study (using, for example, the Amazon Mechanical Turk) might be able to compare many combinations of algorithms and parameters, it would be unrealistic to use this methodology for *every* image that is processed. Finally, the lack of a high-quality "ground truth" image in most applications precludes the use of traditional metrics for image comparison, such as peak signal to noise ratio (PSNR), structural similarity index (SSIM), or visual information fidelity (VIF).

We explore a methodology in which image quality and artifacts are scored according to some function of features that are computed over the image. The function is learned based on thousands of training examples provided in a massive online user study. By picking the right features, and collecting enough user input, we are able to obtain a metric that generalizes over images and over the algorithms used to process them. We call this learned function a *perceptually-validated metric*, as it is not built upon the underlying psycho-physical mechanisms of the human visual system, but rather can score image quality and artifacts consistently with human ratings.

In this paper, we address the problem of motion deblurring or blind deconvolution: undoing the (unknown) motion blur introduced by camera shake. This is a problem that has been studied intensively over the past several years [Fergus et al. 2006; Yuan et al. 2007; Shan et al. 2008; Cho and Lee 2009; Krishnan and Fergus 2009;

Structure-Preserving Image Smoothing via Region Covariances

Levent Karacan* Erkut Erdem† Aykut Erdem‡

Department of Computer Engineering
Hacettepe University

Input image Structure component Texture component

Figure 1: *Our approach makes use of region covariances in decomposing an image into coarse and fine components. The coarse component correspond to prominent structures beneath the image, whereas the fine component only includes texture. Our smoothing method successfully captures the grain of the figures and the rocky texture while preserving the edges in the extracted structure (source image © reibai).*

Abstract

Recent years have witnessed the emergence of new image smoothing techniques which have provided new insights and raised new questions about the nature of this well-studied problem. Specifically, these models separate a given image into its structure and texture layers by utilizing non-gradient based definitions for edges or special measures that distinguish edges from oscillations. In this study, we propose an alternative yet simple image smoothing approach which depends on covariance matrices of simple image features, aka the region covariances. The use of second order statistics as a patch descriptor allows us to implicitly capture local structure and texture information and makes our approach particularly effective for structure extraction from texture. Our experimental results have shown that the proposed approach leads to better image decompositions as compared to the state-of-the-art methods and preserves prominent edges and shading well. Moreover, we also demonstrate the applicability of our approach on some image editing and manipulation tasks such as image abstraction, texture and detail enhancement, image composition, inverse halftoning and seam carving.

CR Categories: I.4.3 [Image Processing and Computer Vision]: Enhancement—Smoothing

Keywords: image smoothing, structure extraction, texture elimination, region covariances

*e-mail: karacan@cs.hacettepe.edu.tr
†e-mail: erkut@cs.hacettepe.edu.tr
‡e-mail: aykut@cs.hacettepe.edu.tr

Links: DL PDF WEB VIDEO CODE

1 Introduction

Natural images provide rich visual information about the world we live in, and typically contain various objects organized in a meaningful configuration. For instance, consider the image given in Figure 1, which shows a historical site consisting of highly textured figures on a rocky surface. While our visual system is very successful at extracting the prominent structures beneath the image without getting distracted by the texture, enabling a machine to perform the same task, i.e. decomposing the image into its structure and texture components, poses great challenges.

From a computational point of view, image decomposition can be expressed as an estimation problem in which a given image is separated into two components that respectively correspond to coarse and fine scale image details. Historically, Gaussian filter is the oldest and the most commonly used smoothing operator [Witkin 1984; Burt and Adelson 1983]. It provides a linear scale-space representation of an image where the input image is smoothed at a con-

ACM Reference Format
Karacan, L., Erdem, E., Erdem, A. 2013. Structure-Preserving Image Smoothing via Region Covariances. ACM Trans. Graph. 32, 6, Article 176 (November 2013), 11 pages. DOI = 10.1145/2508363.2508403 http://doi.acm.org/10.1145/2508363.2508403.

0730-0301/13/11-ART176 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508403

Cost-effective Printing of 3D Objects with Skin-Frame Structures

Weiming Wang[†,‡] Tuanfeng Y. Wang[†] Zhouwang Yang[†*] Ligang Liu[†] Xin Tong[§]
Weihua Tong[†] Jiansong Deng[†] Falai Chen[†] Xiuping Liu[‡]
[†]University of Science and Technology of China [‡]Dalian University of Technology [§]Microsoft Research Asia

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

Figure 1: *Given an input Horse model (a), our method generates a skin-frame structure (b), which is approximate to the model, to minimize the cost of material used in printing it. The frame structure is designed to meet various constraints by an optimization scheme. In (b) we remove the front part of the skin in order to show the internal structure of frame. (c) is the photo of an printed model by removing part of its skin to see the internal struts. (d) is the photo of the printed model generated by our method. A small red drawing pin is put under the object as a size reference in (c) and (d) respectively. The material usage in (d) is only 15.0% of that of a solid object.*

Abstract

3D printers have become popular in recent years and enable fabrication of custom objects for home users. However, the cost of the material used in printing remains high. In this paper, we present an automatic solution to design a skin-frame structure for the purpose of reducing the material cost in printing a given 3D object. The frame structure is designed by an optimization scheme which significantly reduces material volume and is guaranteed to be physically stable, geometrically approximate, and printable. Furthermore, the number of struts is minimized by solving an ℓ_0 sparsity optimization. We formulate it as a multi-objective programming problem and an iterative extension of the preemptive algorithm is developed to find a compromise solution. We demonstrate the applicability and practicability of our solution by printing various objects using both powder-type and extrusion-type 3D printers. Our method is shown to be more cost-effective than previous works.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling; I.3.8 [Computer Graphics]: Applications;

Keywords: 3D printing, fabrication, frame structure, sparsity optimization

Links: DL PDF

*Corresponding author: yangzw@ustc.edu.cn (Zhouwang Yang)

ACM Reference Format
Wang, W., Wang, T., Yang, Z., Liu, L., Tong, X., Tong, W., Deng, J., Chen, F., Liu, X. 2013. Cost-effective Printing of 3D Objects with Skin-Frame Structures. ACM Trans. Graph. 32, 6, Article 177 (November 2013), 10 pages. DOI = 10.1145/2508363.2508382 http://doi.acm.org/10.1145/2508363.2508382.

DOI: http://doi.acm.org/10.1145/2508363.2508382

1 Introduction

Additive manufacturing (3D printing) enables fabrication of physical objects from digital models where the printed objects are created by laying down successive layers of material [3DSystems 2012; Shapeways 2012]. During the last few years, research on computational techniques of 3D printing has received considerable attention for assisting users to generate desired manufacturable objects [Stava et al. 2012; Luo et al. 2012; Chen et al. 2013; Prévost et al. 2013].

However, reducing the material used in printing, which is an important problem due to its high cost, has not been well studied. The straightforward approach used in commercial printer packages [Shapeways 2012] is to uniformly hollow the 3D object by extruding the outer surface and creating a scaled-down version on its inside. The user has to choose a scaling factor (thickness of object) based on experience. A large factor may lead to material waste while a small factor could cause structural stability problem. Thus it is technically nontrivial for the hollowing method to simultaneously match the goals of saving material and maintaining physical stability in 3D printing.

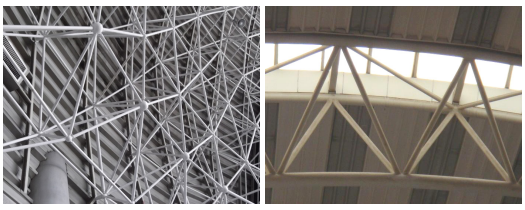


Figure 2: *Common frame structures used in architecture.*

In this paper, we present an automatic method to minimize material cost of the object in 3D fabrication. The key idea is to 'hollow' the object by creating a lightweight frame structure (see Figure 2), made of a mesh of nodes and thin cylindrical struts with large voids among them inside the object (see Figure 1 (b)). Frame structures benefit 3D printing in two aspects. First, the mass of object could

Sphere-Meshes: Shape Approximation using Spherical Quadric Error Metrics

Jean-Marc Thiery Émilie Guy Tamy Boubekeur

Telecom ParisTech – CNRS LTCI – Institut Mines-Telecom

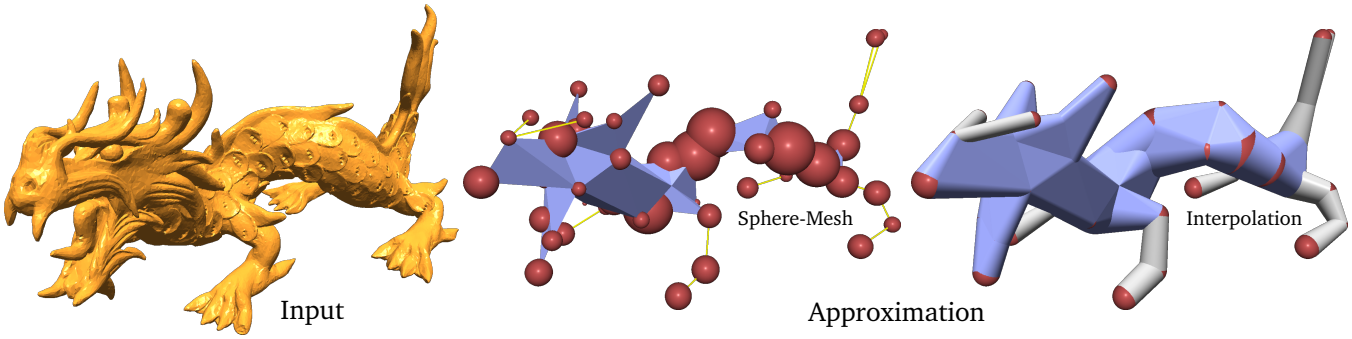


Figure 1: *Starting from a surface mesh (left, 200k triangles), we propose an approximation algorithm which generates a sphere-mesh (middle) defining an extremely simplified model (here with 50 spheres) as a sphere interpolation over a set of edges and polygons (right). The input mesh is shown in orange (left), the sphere-mesh is displayed in the middle (spheres in red, edges in yellow, triangles in blue) and the interpolated sphere-mesh geometry is shown on the right (edge interpolation in grey, triangle interpolation in blue).*

Abstract

Shape approximation algorithms aim at computing simple geometric descriptions of dense surface meshes. Many such algorithms are based on mesh decimation techniques, generating coarse triangulations while optimizing for a particular metric which models the distance to the original shape. This approximation scheme is very efficient when enough polygons are allowed for the simplified model. However, as coarser approximations are reached, the intrinsic piecewise linear point interpolation which defines the decimated geometry fails at capturing even simple structures. We claim that when reaching such extreme simplification levels, highly instrumental in shape analysis, the approximating representation should explicitly and progressively model the volumetric extent of the original shape. In this paper, we propose *Sphere-Meshes*, a new shape representation designed for extreme approximations and substituting a *sphere* interpolation for the classic point interpolation of surface meshes. From a technical point-of-view, we propose a new shape approximation algorithm, generating a sphere-mesh at a prescribed level of detail from a classical polygon mesh. We also introduce a new metric to guide this approximation, the *Spherical Quadric Error Metric* in $\mathbb{R}^4$, whose minimizer finds the sphere that best approximates a set of tangent planes in the input and which is sensitive to surface orientation, thus distinguishing naturally between the *inside* and the *outside* of an object. We evaluate the performance of our algorithm on a collection of models covering a wide range of topological and geometric structures and compare it against alternate methods. Lastly, we propose an application to deformation control where a sphere-mesh hierarchy is used as a convenient rig for altering the input shape interactively.

CR Categories: Computing Methodologies [Computer Graphics]: Shape Modeling—Mesh Models; Computing Methodologies [Computer Graphics]: Shape Modeling—Shape Analysis.

Keywords: shape approximation, simplification

Links: DL PDF

1 Introduction

Approximating 3D shapes using a minimal set of geometric primitives offers a wide range of applications, from shape analysis to interactive modeling. Starting from a dense surface mesh, simplification methods are a popular class of algorithms to generate such approximations. They can essentially be classified in three categories: (i) *clustering* methods [Rossignac and Borrel 1993][Lindstrom 2000][Schaefer and Warren 2003][Cohen-Steiner et al. 2004], which decompose the original surface into a collection of regions and substitute each region with a single representative (e.g., point or face), (ii) *decimation* methods [Hoppe et al. 1993][Garland and Heckbert 1997], which iteratively remove surface samples and relocate their neighbors to optimize for the original shape and (iii) *resampling* methods [Turk 1992; Alliez et al. 2003; Yan et al. 2009] which compute a new, potentially coarser, point distribution on the surface and establish a new connectivity.

Alternatively, one can also consider the volume bounded by the surface and provide a simplification by means of the Medial Axis Transform (MAT) [Blum 1967] for instance, leading to a representation with simplified volumetric structures [Amenta et al. 2001; Dey and Zhao 2004; Chazal and Lieutier 2005; Sud et al. 2007; Miklos et al. 2010].

In this paper, we introduce *sphere-meshes*, an approximation model

ACM Reference Format
Thiery, J., Guy, É., Boubekeur, T. 2013. Sphere-Meshes: Shape Approximation using Spherical Quadric Error Metrics. ACM Trans. Graph. 32, 6, Article 178 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508384 http://doi.acm.org/10.1145/2508363.2508384.

DOI: http://doi.acm.org/10.1145/2508363.2508384

Sparse Localized Deformation Components

Thomas Neumann[1,2], Kiran Varanasi[3,4], Stephan Wenger[2], Markus Wacker[1], Marcus Magnor[2], Christian Theobalt[3]

[1]HTW Dresden, Germany, [2]Computer Graphics Lab, TU Braunschweig, Germany,
[3]Max-Planck-Institut Informatik, Saarbrücken, Germany, [4]Technicolor Research, France

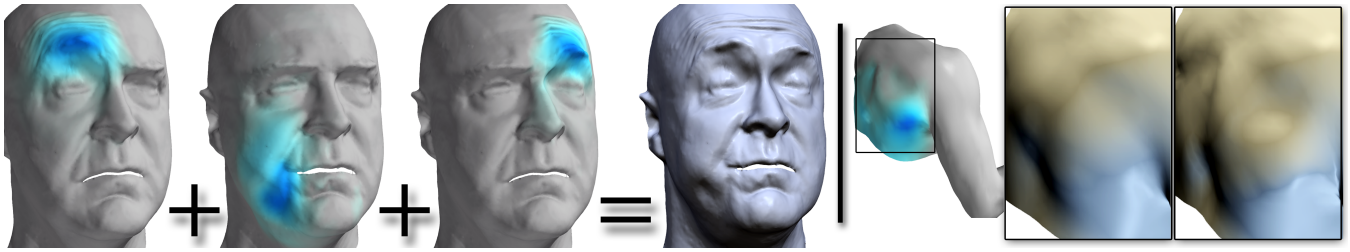


Figure 1: *Our method automatically decomposes any mesh animations like performance captured faces (left) or muscle deformations (right) into sparse and localized deformation modes (shown in blue).* Left: *a new facial expression is generated by summing deformation components. Our method automatically separates spatially confined effects like separate eyebrow motions from the data.* Right: *Our algorithm extracts individual muscle and bone deformations. The deformation components can then be used for convenient editing of the captured animation. Here, the deformation component of the clavicle is over-exaggerated to achieve an artistically desired look.*

Abstract

We propose a method that extracts sparse and spatially localized deformation modes from an animated mesh sequence. To this end, we propose a new way to extend the theory of sparse matrix decompositions to 3D mesh sequence processing, and further contribute with an automatic way to ensure spatial locality of the decomposition in a new optimization framework. The extracted dimensions often have an intuitive and clear interpretable meaning. Our method optionally accepts user-constraints to guide the process of discovering the underlying latent deformation space. The capabilities of our efficient, versatile, and easy-to-implement method are extensively demonstrated on a variety of data sets and application contexts. We demonstrate its power for user friendly intuitive editing of captured mesh animations, such as faces, full body motion, cloth animations, and muscle deformations. We further show its benefit for statistical geometry processing and biomechanically meaningful animation editing. It is further shown qualitatively and quantitatively that our method outperforms other unsupervised decomposition methods and other animation parameterization approaches in the above use cases.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation;

Keywords: mesh deformation, editing captured animations, data-driven animation, dimensionality reduction

Links: DL PDF

ACM Reference Format
Neumann, T., Varanasi, K., Wenger, S., Wacker, M., Magnor, M., Theobalt, C. 2013. Sparse Localized Deformation Components. ACM Trans. Graph. 32, 6, Article 179 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508417 http://doi.acm.org/10.1145/2508363.2508417.

0730-0301/13/11-ART179 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508417

1 Introduction

Nowadays, time-varying dynamic geometry with very fine dynamic shape detail can be generated and rendered at very high visual fidelity. When creating such content, artists usually rely on a low-dimensional control parametrization, for example a kinematic skeleton rig to control surface motion via joint angles, a muscle system simulating face motion and tissue deformation, or a physics-based simulation model generating realistically deforming cloth or other materials. Despite increasing expressive power of such parametrizations and simulations, producing such realistic animations from scratch is a labor-intensive process, in particular since it is highly non-trivial to design or customize a specific parameterization to a new object to be animated.

Performance capture techniques were thus developed that measure detailed time-varying surface models of the real world in motion from sensor data [Theobalt et al. 2010]. These methods capture highly detailed animation models, mostly as space-time surface mesh sequences. However, their applicability in animation production so far has been strongly limited because a low-dimensional control parametrization for the captured data is missing. Conveniently re-using, editing or otherwise analysing this input data is therefore not possible.

To overcome this problem, some recent work suggested to fit specific parametric models to such input data: simple linear regression models, skeletons or bone transformations, or even pre-modeled facial blendshapes, see Sect. 2. Such methods require additional prior knowledge about the underlying physical process of the animation (e.g. a template shape or a parametric physics model), or additional data needs to be recorded alongside the capture. They are thus object-type specific and not very easy to use on general data. Dimensionality reduction techniques for animation parameterization rely less on specific prior information, and are thus more generally applicable to a broader range of animations. Many of them are, in essence, matrix decomposition techniques that explain observed deformations as a linear combination of factors, or deformation components, that can be controlled by an animator. Many previously used dimensionality reduction techniques, like Principal Component Analysis (PCA), reproduce input data faithfully and maintain certain compression guarantees, but the individual com-

A general and efficient method for finding cycles in 3D curve networks

Yixin Zhuang[1,2] Ming Zou[2] Nathan Carr[3] Tao Ju[2]

[1]National University of Defense Technology, China
[2]Washington University in St. Louis, USA
[3]Adobe, USA

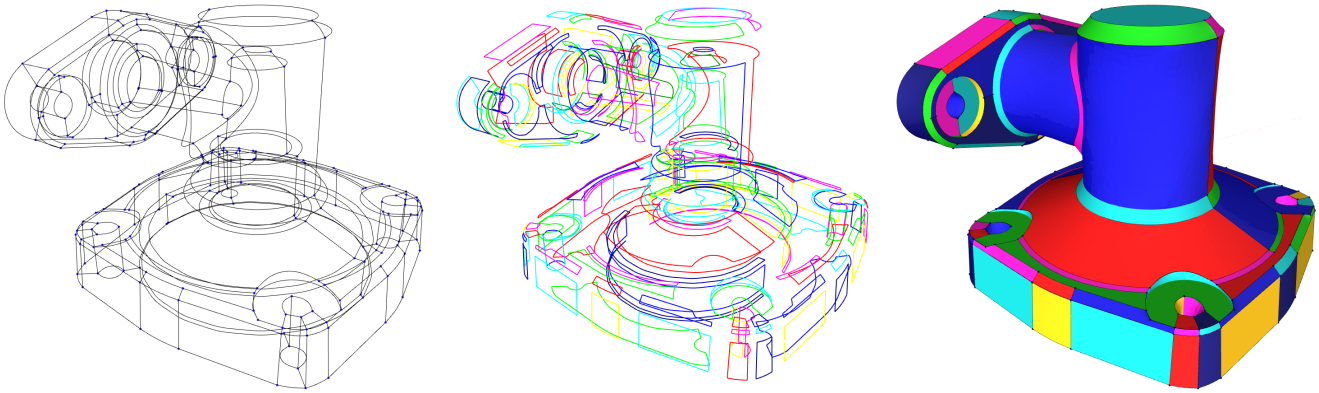


Figure 1: *A curve network (*misc2*) representing a genus-7 mechanical part (left), cycles found by our algorithm (middle), and surface patches generated from the cycles (right). The curve network contains 410 curves. Our algorithm completed in half a second.*

Abstract

Generating surfaces from 3D curve networks has been a longstanding problem in computer graphics. Recent attention to this area has resurfaced as a result of new sketch based modeling systems. In this work we present a new algorithm for finding cycles that bound surface patches. Unlike prior art in this area, the output of our technique is unrestricted, generating both manifold and non-manifold geometry with arbitrary genus. The novel insight behind our method is to formulate our problem as finding local mappings at the vertices and curves of our network, where each mapping describes how incident curves are grouped into cycles. This approach lends us the efficiency necessary to present our system in an interactive design modeler, whereby the user can adjust patch constraints and change the manifold properties of curves while the system automatically re-optimizes the solution.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms

Keywords: curve networks, wireframe, sketch-based modeling, lofting, dynamic programming

Links: DL PDF

ACM Reference Format
Zhuang, Y., Zou, M., Carr, N., Ju, T. 2013. A general and efficient method for finding cycles in 3D curve networks. ACM Trans. Graph. 32, 6, Article 180 (November 2013), 10 pages. DOI = 10.1145/2508363.2508423 http://doi.acm.org/10.1145/2508363.2508423.

DOI: http://doi.acm.org/10.1145/2508363.2508423

1 Introduction

Creating surfaces from a network of 3D curves, often known as *lofting* or *skinning*, is a classical and fundamental problem in computer-aided design (CAD). In a traditional CAD pipeline, engineers start by defining a wireframe of the desired model, which then needs to be turned into a surface representation for analysis, simulation, and manufacturing. More recently, sketching tools [Wesche and Seidel 2001; Bae et al. 2008; Schmidt et al. 2009; Grimm and Joshi 2012] offer intuitive means for drawing 3D curves using 2D input devices, allowing artists to quickly come up with concept designs. Even though these sketches may not be used for defining the final product, surface visualization could help users better appreciate their design [Abbasinejad et al. 2011; Bessmeltsev et al. 2012].

An important first step in lofting is identifying a cycle of curves in the network that bounds an individual surface patch. Although humans can quickly "see" such cycles in a typical curve network created by a designer, the actual process of selecting curves and grouping them into cycles can be rather tedious. The process may be significantly shortened if the computer can suggest most, if not all, of the cycles, requiring only minimal input from the user in the form of quick inspection and small adjustments.

Finding cycles automatically is a challenging task. First of all, there is usually a huge number of cycles to choose from in a curve network [Biondi et al. 1970]. Second, while the complete perceptual model of how humans identify the "good" cycles is not yet well-understood, it is evident that we evaluate the goodness of one cycle not in isolation, but in the context of other cycles. Consider the example in Figure 1 (left): among the many circular rings in the network, only a few of them describe faces of the model while others are cross-sections or ends of cylindrical shafts. Last but not least, the specific application context often imposes additional constraints on the cycles. For example, wireframe designs of mechanical models usually define closed manifolds (often times with high genus, like the one in Figure 1), meaning that each input curve needs to be

Urban Pattern: Layout Design by Hierarchical Domain Splitting

Yong-Liang Yang
KAUST

Jun Wang
KAUST

Etienne Vouga
Columbia University

Peter Wonka
KAUST

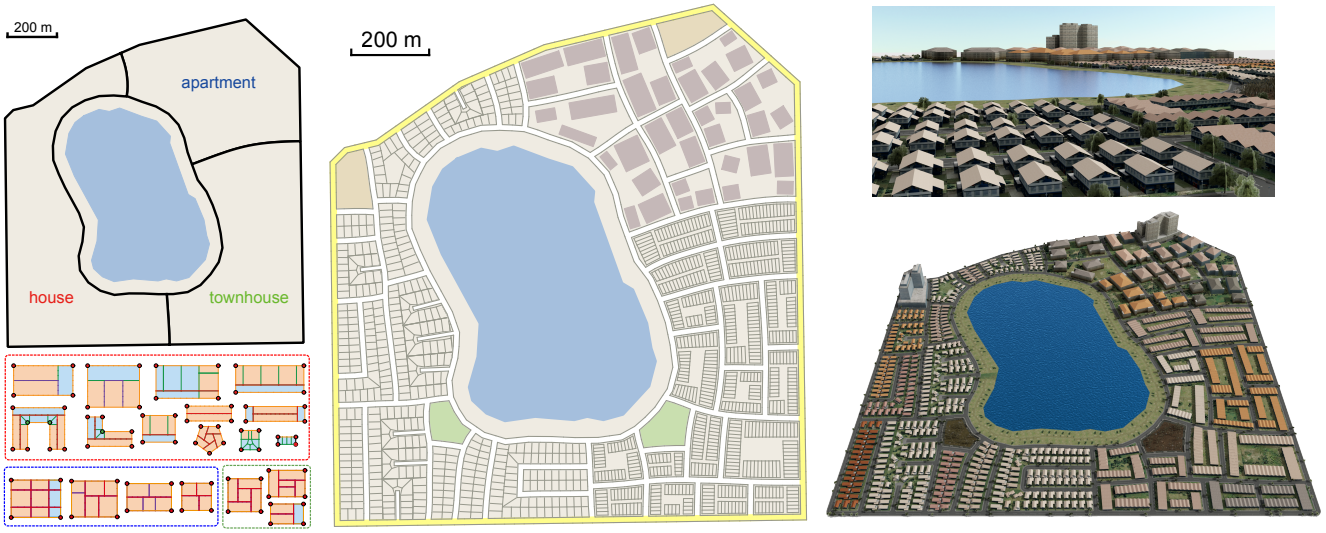


Figure 1: *Starting from a polygonal region and user-prescribed design elements (top left), we hierarchically subdivide the input region using streamline-based and template-based splitting operations. Selected templates are shown on the bottom left, the final layout in the middle, and a simulated 3D construction on the right.*

Abstract

We present a framework for generating street networks and parcel layouts. Our goal is the generation of high-quality layouts that can be used for urban planning and virtual environments. We propose a solution based on hierarchical domain splitting using two splitting types: streamline-based splitting, which splits a region along one or multiple streamlines of a cross field, and template-based splitting, which warps pre-designed templates to a region and uses the interior geometry of the template as the splitting lines. We combine these two splitting approaches into a hierarchical framework, providing automatic and interactive tools to explore the design space.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—;

Keywords: street layouts, parcel generation, computational design, mesh optimization, region splitting

Links: DL PDF VIDEO

ACM Reference Format
Yang, Y., Wang, J., Vouga, E., Wonka, P. 2013. Urban Pattern: Layout Design by Hierarchical Domain Splitting. ACM Trans. Graph. 32, 6, Article 181 (November 2013), 12 pages. DOI = 10.1145/2508363.2508405 http://doi.acm.org/10.1145/2508363.2508405.

DOI: http://doi.acm.org/10.1145/2508363.2508405

1 Introduction

We present a framework for generating street and parcel layouts (see Fig. 1). Besides synthesizing virtual urban environments, our work is useful for planning future developments. In contrast to existing work in urban modeling, e.g., [Chen et al. 2008a; Weber et al. 2009], we aim to generate solutions with high geometric quality that respect some specified polygonal boundary conditions. Because of the slow and chaotic historical process of urban development, many of our oldest cities have inefficient layouts, including labyrinthine streets that meet at odd angles, and land parcels that are oddly shaped and proportioned. Existing work in urban modeling, focused primarily on synthesizing virtual urban environments, generates solutions that mimic many of these problems. For urban planning, these artifacts are not acceptable: we do not want to recreate the problems of the past when planning high-quality layouts for the future. We therefore propose a new framework that generates high-quality, user-controllable street and parcel layouts.

There are two reasons why we believe that this is an interesting problem. First, existing subdivisions are currently generated without computational design tools, which often leads to inefficient solutions in the form of undesirable parcel shapes, parcel sizes, or too much land devoted to roads. Given the limited availability and the high cost of land in most urban areas, the financial impact of better planning is significant. Second, the problem of urban layout generation is an interesting mesh generation problem in geometry that has not been solved with high geometric quality.

In Fig. 11, left (CE1), we visualize a street and parcel layout that was generated with a current state-of-the-art urban modeling tool (CityEngine). We observe two major problems that we wish to overcome. First, its street layout algorithm is based on growing street segments and this strategy fails when more complex bound-

Versatile Surface Tension and Adhesion for SPH Fluids

Nadir Akinci
University of Freiburg

Gizem Akinci
University of Freiburg

Matthias Teschner
University of Freiburg

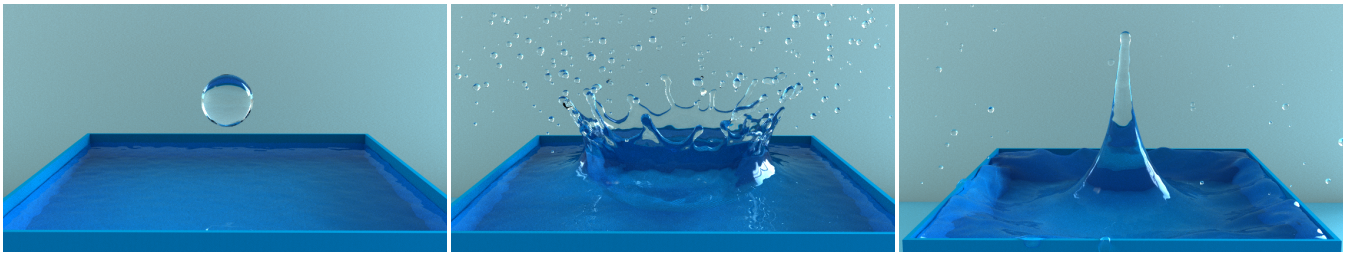


Figure 1: *A water crown emerges as a result of the impact of a water droplet into a filled container. Our surface tension force allows realistic simulation of such natural phenomena.*

Abstract

Realistic handling of fluid-air and fluid-solid interfaces in SPH is a challenging problem. The main reason is that some important physical phenomena such as surface tension and adhesion emerge as a result of inter-molecular forces in a microscopic scale. This is different from scalar fields such as fluid pressure, which can be plausibly evaluated on a macroscopic scale using particles. Although there exist techniques to address this problem for some specific simulation scenarios, there does not yet exist a general approach to reproduce the variety of effects that emerge in reality from fluid-air and fluid-solid interactions. In order to address this problem, we present a new surface tension force and a new adhesion force. Different from the existing work, our surface tension force can handle large surface tensions in a realistic way. This property lets our approach handle challenging real scenarios, such as water crown formation, various types of fluid-solid interactions, and even droplet simulations. Furthermore, it prevents particle clustering at the free surface where inter-particle pressure forces are incorrect. Our adhesion force allows plausible two-way attraction of fluids and solids and can be used to model different wetting conditions. By using our forces, modeling surface tension and adhesion effects do not require involved techniques such as generating a ghost air phase or surface tracking. The forces are applied to the neighboring fluid-fluid and fluid-boundary particle pairs in a symmetric way, which satisfies momentum conservation. We demonstrate that combining both forces allows simulating a variety of interesting effects in a plausible way.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation;

Keywords: physically-based animation, fluid simulation, Smoothed Particle Hydrodynamics, surface tension, adhesion

Links: DL PDF VIDEO

ACM Reference Format
Akinci, N., Akinci, G., Teschner, M. 2013. Versatile Surface Tension and Adhesion for SPH Fluids. ACM Trans. Graph. 32, 6, Article 182 (November 2013), 8 pages. DOI = 10.1145/2508363.2508395 http://doi.acm.org/10.1145/2508363.2508395.

0730-0301/13/11-ART182 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508395

1 Introduction

Surface tension is a ubiquitous effect in daily life. For instance, when pouring water into a glass, the force that keeps liquid molecules together is the surface tension force. It is caused by cohesive forces among neighboring fluid molecules. Inside the fluid, each molecule is pulled equally by its neighbors, resulting in a zero net force. However, as the free surface does not have neighbors on all sides, the molecules in such regions are pulled inwards. Furthermore, surface tension minimizes surface area according to Laplace's law, which causes droplets of water to form a sphere when external forces are excluded. Another effect that is again caused by molecular interaction is adhesion. Adhesion allows fluids to get attracted by other materials. For instance, the unique appearance of dew on plants and the ability of water striders to stay atop water are caused by the interplay of surface tension and adhesion forces. In this paper, we focus on simulating those two molecular interaction related phenomena in the context of computer animation, more specifically for SPH (Smoothed Particle Hydrodynamics) fluids.

One important issue that arises at fluid-air and fluid-solid interfaces in SPH is density underestimation, where densities of the particles are erroneously computed as less than the rest density when the density summation approach is used. Those wrong density values result in negative pressures and cause the particles to cluster, which is known as tensile instability in SPH. This phenomenon can be alleviated by using artificial pressure forces [Monaghan 2000; Macklin and Mueller 2013], which, however, result in spurious surface tensions. For this reason, either a density correction technique [Shepard 1968], or simply not allowing negative pressures are other common practices for avoiding tensile instability. However, this still does not solve the problem of sticking particles at the fluid interface, since the pressure field is still not reconstructed in a physically sensible way. [Akinci et al. 2012b] addresses this issue for fluid-solid interface by pre-computing a single layer of boundary particles for the solid boundaries, which also extends to two-way fluid-solid coupling. In [Schechter and Bridson 2012], ghost SPH particles are dynamically generated at both fluid-solid and fluid-air interfaces.

For modeling surface tension in SPH, additional techniques are generally preferred. These can be listed as: Curvature based external forces on particles (e.g. [Müller et al. 2003]), pairwise forces based on cohesion (e.g. [Becker and Teschner 2007]), a modified SPH formulation [Clavet et al. 2005], and more recently forces based on surface mesh curvature [Yu et al. 2012]. However, there is no sin-

Spatio-temporal Extrapolation for Fluid Animation

Yubo Zhang*
UC Davis

Kwan-Liu Ma†
UC Davis

(a) Coarse Grid Simulation $160 \times 128 \times 64$ (b) Fine Grid Simulation $320 \times 256 \times 128$ (c) Spatio-temporal Extrapolation $320 \times 256 \times 128$

Figure 1: *Our spatio-temporal extrapolation method can enhance the fluid simulation result by applying a black-box grid-based solver on two grid scales and combining the solutions together at each time step. The extrapolation technique produces better motion quality (c) than those from directly applying the solver on the two scales separately (a) and (b).*

Abstract

We introduce a novel spatio-temporal extrapolation technique for fluid simulation designed to improve the results without using higher resolution simulation grids. In general, there are rigid demands associated with pushing fluid animations to higher resolutions given limited computational capabilities. This results in tradeoffs between implementing high-order numerical methods and increasing the resolution of the simulation in space and time. For 3D problems, such challenges rapidly become cost-ineffective. The extrapolation method we present improves the flow features without using higher resolution simulation grids. In this paper, we show that simulation results from our extrapolation are comparable to those from higher resolution simulations. In addition, our method differs from high-order numerical methods because it does not depend on the equation or specific solver. We demonstrate that it is easy to implement and can significantly improve the fluid animation results.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation;

Keywords: extrapolation, fluid animation

*e-mail:ybzhang@ucdavis.edu
†e-mail:ma@cs.ucdavis.edu

Links: DL PDF

1 Introduction

In computer animation, fluid simulation is a fundamental research topic for modeling realistic physically-based flow phenomena. This involves formulating model equations for various phenomena and developing efficient numerical methods to solve them. Presently, it is possible to model the majority of natural flow phenomena. However, the results generated by practical simulation methods remain limited to moderate resolutions and motion quality. Although this can be remedied by velocity up-sampling and procedural synthesis, the added details are limited to local flow features and the macroscopic motion is not significantly improved.

Another approach for tackling this challenge is using high-order numerical methods aimed at achieving improved advection with less numerical dissipation. For example, after Stam [Stam 1999] popularized the first order semi-Lagrangian advection scheme, several improvements [Dupont and Liu 2003; Selle et al. 2008] has been made to achieve second-order advection which introduces less numerical dissipation. There are even higher order schemes which can be applied to advection, such as Weighted Essentially Non-

ACM Reference Format
Zhang, Y., Ma, K. 2013. Spatio-temporal Extrapolation for Fluid Animation. ACM Trans. Graph. 32, 6, Article 183 (November 2013), 8 pages. DOI = 10.1145/2508363.2508401 http://doi.acm.org/10.1145/2508363.2508401.

DOI: http://doi.acm.org/10.1145/2508363.2508401

Interactive Localized Liquid Motion Editing

Zherong Pan[1] Jin Huang[1*] Yiying Tong[2] Changxi Zheng[3] Hujun Bao[1*]
[1]State Key Lab of CAD&CG, Zhejiang University [2]Michigan State University [3]Columbia University

Abstract

Animation techniques for controlling liquid simulation are challenging: they commonly require carefully setting initial and boundary conditions or performing a costly numerical optimization scheme against user-provided keyframes or animation sequences. Either way, the whole process is laborious and computationally expensive.

We introduce a novel method to provide intuitive and interactive control of liquid simulation. Our method enables a user to locally edit selected keyframes and automatically propagates the editing in a nearby temporal region using geometric deformation. We formulate our local editing techniques as a small-scale nonlinear optimization problem which can be solved interactively. With this uniformed formulation, we propose three editing metaphors, including (i) sketching local fluid features using a few user strokes, (ii) dragging a local fluid region, and (iii) controlling a local shape with a small mesh patch. Finally, we use the edited liquid animation to guide an offline high-resolution simulation to recover more surface details. We demonstrate the intuitiveness and efficacy of our method in various practical scenarios.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation;

Keywords: fluid simulation, sketch, deformation

Links: DL PDF

1 Introduction

While fluid simulation has long been used to create realistic fluid animations, generating fluid animations with desired effects remains difficult and time-consuming. Liquid motion simulated with set initial and boundary conditions are intrinsically chaotic, making it hard to produce desired later frames by tuning parameters [Foster and Metaxas 1997a]. In computer graphics applications such as film production, controlling a liquid animation often requires multiple attempts even for experienced animators to obtain the desired results.

As a viable alternative to explicitly tuning parameters and boundary conditions, some previous methods [Treuille et al. 2003; McNamara et al. 2004; Wojtan et al. 2006] proposed to specify *a priori* a set of global fluid shape keyframes and rely on an offline numerical optimization scheme to find the desired fluid animation. However, unlike solid objects, authoring even a single fluid keyframe is laborious, and the manually designed frames tend to be non-volume-preserving and overly smooth, suppressing rich visual details in natural fluid motion. Moreover, the nonlinear optimization process over the dynamics with provided keyframes is computationally expensive. It is very hard, if not impossible, to provide a user fast feedback of resulting effects of the supplied keyframes.

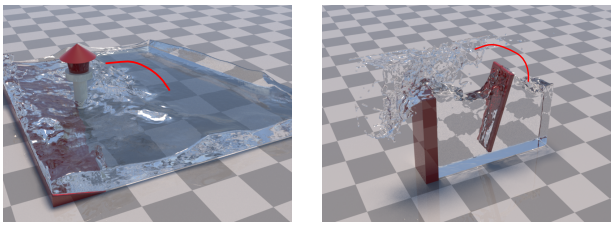


Figure 1: *The user can guide the fluid simulation by sketching (red curves) for the desired motion with fast feedback, crucial to animation prototyping. From the left to the right: a lighthouse at shore in tidal waves; a wave raised to engulf the lighthouse; a washing machine in action; water made to splash into anther chamber at a chosen instant.*

Instead of globally keyframing the entire fluid shape, we propose a method to edit fluid animation in a *local* spatial and temporal region. Limiting user editing in a local region enables an intuitive control process. Indeed, the general philosophy of local control has proven useful in many areas of computer graphics, such as spline curves [Barsky and Beatty 1983], shape editing [Botsch and Kobbelt 2005], and animation design [Cohen 1992]. For fluid control, this means that the user can progressively edit fluid animation toward the desired effects as the fluid simulation advances. Meanwhile, we can limit the size of the formulated optimal control problem, making an interactive feedback possible and hence accelerating the animation design cycles.

In our pipeline, the user starts from either a low-resolution or a downsampled high-resolution fluid simulation. While the simulation advances, one can select any resulting frame for editing. We provide three local editing metaphors: (i) the user can sketch a few strokes to guide the fluid local shapes or silhouettes projected on a view-dependent plane; (ii) for more detailed control, one can select a small fluid region and drag it locally around; and (iii) one can provide a small mesh patch to control local fluid features. All three metaphors boil down to a uniform optimization problem of *geometric* deformation, which deforms the fluid shape at the selected frame in an interactive and volume-preserving fashion. Once the selected frame is edited, its changes are then propagated over its previous frames to ensure temporally coherent results. Through guided simulation over a short local sequence, the newly edited frame also serves as updated initial conditions at the original resolution to affect subsequent fluid simulation. As the simulation progresses, the user incrementally edits the whole animation sequence. Alternatively, the entire sequence can be edited with low-resolution simulation, followed by an offline guided high-resolution simulation (e.g., [Nielsen and Bridson 2011; Yuan et al. 2011]) recovering fine details while still preserving desired motion effects.

Contributions Our approach features the following novel components:

*Corresponding authors: {hj,bao}@cad.zju.edu.cn

ACM Reference Format
Pan, Z., Huang, J., Tong, Y., Zheng, C., Bao, H. 2013. Interactive Localized Liquid Motion Editing. ACM Trans. Graph. 32, 6, Article 184 (November 2013), 10 pages. DOI = 10.1145/2508363.2508429 http://doi.acm.org/10.1145/2508363.2508429.

DOI: http://doi.acm.org/10.1145/2508363.2508429

Physics-based Animation of Large-scale Splashing Liquids

Dan Gerszewski
University of Utah

Adam W. Bargteil
University of Utah

Abstract

Fluid simulation has been one of the greatest successes of physics-based animation, generating hundreds of research papers and a great many special effects over the last fifteen years. However, the animation of large-scale, splashing liquids remains challenging. In this paper, we show that a novel combination of unilateral incompressibility, mass-full FLIP, and blurred boundaries is extremely well-suited to the animation of large-scale, violent, splashing liquids.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation; I.6.8 [Simulation and Modeling]: Types of Simulation—Animation.

Keywords: Unilateral incompressibility, liquid simulation, physics-based animation, fluid-implicit-particle method

Links: DL PDF

1 Introduction

Over the last fifteen years, fluid simulation has emerged as one of the most effective applications of physics-based approaches to animation. The use of fluid simulation for special effects is now commonplace and the topic has received copious attention in the graphics community. Despite tremendous progress, challenges remain. In this paper, we address the problem of large-scale, violent, splashing liquids. We demonstrate that a novel combination of unilateral incompressibility, mass-full FLIP, and blurred boundaries is extremely well-suited to the animation of such liquids and avoids common artifacts such as artificial surface tension, volume loss/gain, and fluid sticking to obstacles.

Because at the spatial and temporal scales we seek to animate liquids compression is negligible, computer graphics researchers have largely focused on simulating incompressible fluids. Even approaches, such as smoothed particle hydrodynamics (SPH), that are naturally suited to simulate compressible flow are often modified for incompressible flow [Solenthaler and Pajarola 2009]. While computationally efficient, incompressibility induces an artificial surface tension that prevents liquid near the surface from mixing with the surrounding air. This mixing is important at large scales, especially during violent splashes, such as after underwater explosions. In this paper, we show that for single-phase fluid simulation, such mixing is effectively modeled with unilateral incompressibility [Narain et al. 2009; Narain et al. 2010], which allows positive divergence while prohibiting negative pressures, thus avoiding the pressure oscillations found in compressible simulation, while removing the artificial surface tension caused by bilateral incompressibility.

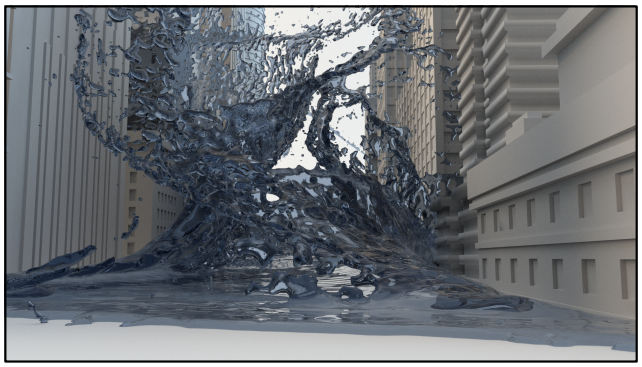


Figure 1: *A city block is flooded.*

We use a variant of the fluid-implicit-particle (FLIP) method as our underlying simulation method. However, our approach, which we call mass-full FLIP, attaches mass to the particles and more closely resembles compressible FLIP [Brackbill and Ruppel 1986] than the incompressible variety [Zhu and Bridson 2005]. Mass-full FLIP is extremely well-suited to the unilateral incompressibility (UIC) solve. In the context of UIC, ensuring conservation of mass becomes difficult—allowing positive divergence can result in significant volume gain. Like SPH methods, mass-full FLIP conserves mass by conserving particles. Additionally, the UIC is most appropriate in highly turbulent simulations where the numerical viscosity associated with semi-Lagrangian and related schemes would be especially inappropriate.

Finally, we treat obstacles and fluid in a unified manner—we discretize obstacles using particles and rasterize their mass onto the background grid using the same trilinear kernel. We additionally employ the variational approach to obstacles endorsed by Batty and colleagues [2007]. Combined with unilateral incompressibility, our treatment of boundaries easily allows liquids to separate from obstacles, avoiding the common visual artifact of liquid gliding along the ceiling.

We demonstrate our approach on several examples, such as the flooding of a city depicted in Figure 1. Side-by-side comparisons with incompressible simulations clearly demonstrate the different behavior afforded by our approach. In general, the more tumultuous the motion, the more different the results. While we do not expect our approach to replace bilateral incompressibility, we believe the rich behavior afforded by it will prove an important tool for animating large-scale splashing liquids.

2 Related Work

The most closely related work to ours from a technical standpoint is the work of Narain and colleagues, who introduced unilateral incompressibility and applied it to two-dimensional crowd simu-

Contact email: {danger,adamb}@cs.utah.edu

ACM Reference Format
Gerszewski, D., Bargteil, A. 2013. Physics-based Animation of Large-scale Splashing Liquids.
ACM Trans. Graph. 32, 6, Article 185 (November 2013), 6 pages. DOI = 10.1145/2508363.2508430
http://doi.acm.org/10.1145/2508363.2508430.

0730-0301/13/11-ART185 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508430

Designing and Fabricating Mechanical Automata from Mocap Sequences

Duygu Ceylan[1] Wilmot Li[2] Niloy J. Mitra[3] Maneesh Agrawala[4] Mark Pauly[1]
[1]EPFL [2]Adobe Research [3]University College London [4]University of California Berkeley

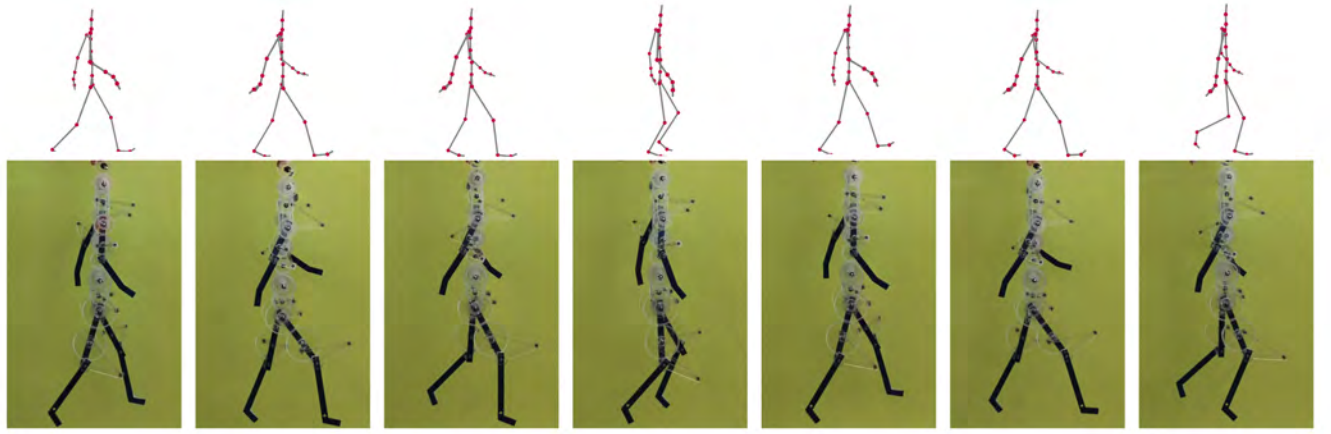


Figure 1: *We present a method for generating the design of a mechanical automaton (bottom row) that approximates an input motion sequence (top row). Our algorithm automatically determines the configuration, dimensions, and layout of all mechanical components.*

Abstract

Mechanical figures that mimic human motions continue to entertain us and capture our imagination. Creating such automata requires expertise in motion planning, knowledge of mechanism design, and familiarity with fabrication constraints. Thus, automaton design remains restricted to only a handful of experts. We propose an automatic algorithm that takes a motion sequence of a humanoid character and generates the design for a mechanical figure that approximates the input motion when driven with a single input crank. Our approach has two stages. The *motion approximation* stage computes a motion that approximates the input sequence as closely as possible while remaining compatible with the geometric and motion constraints of the mechanical parts in our design. Then, in the *layout* stage, we solve for the sizing parameters and spatial layout of all the elements, while respecting all fabrication and assembly constraints. We apply our algorithm on a range of input motions taken from motion capture databases. We also fabricate two of our designs to demonstrate the viability of our approach.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: mechanical assembly, humanoid motion, computational design, fabrication constraints, automaton

Links: DL PDF WEB VIDEO DATA CODE

ACM Reference Format
Ceylan, D., Li, W., Mitra, N., Agrawala, M., Pauly, M. 2013. Designing and Fabricating Mechanical Automata from Mocap Sequences. ACM Trans. Graph. 32, 6, Article 186 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508400 http://doi.acm.org/10.1145/2508363.2508400.

DOI: http://doi.acm.org/10.1145/2508363.2508400

1 Introduction

Mechanical automata are machines that use a combination of interconnected mechanical parts such as cranks, gears, and pulleys to convert a driving force into a specific target motion. Since antiquity, mechanists have created a wide variety of such machines for many different purposes (e.g., clocks, music boxes, fountains). Automata designed to look like human figures performing every-day actions like walking, waving, etc. are especially popular. Famous historical examples include Leonardo Da Vinci's life-sized armor-clad "robot" from 1495 that could sit, stand and move its arms, as well as the"Draughstman-Writer" of Henri Maillardet from the late 18th century, which was the primary inspiration for the automaton in Brian Selznick's book, *The Invention of Hugo Cabret.* Today, wind-up toys in the form of characters are the most common types of mechanical figures. Our longstanding fascination with humanoid automata likely stems from their ability to produce surprisingly complex, lifelike motions from simple input forces through purely mechanical means (see Figure 2).

Despite their widespread appeal, the design of automata is currently restricted to a very small group of experts. Consider the challenges involved in creating a mechanical figure that performs a specific target motion. First, a designer must choose a set of mechanical parts and decide how to connect the parts together to approximate the

Figure 2: *Examples of mechanical automata: Da Vinci's mechanical knight; Maillardet's writer; an aviation chronograph.*

3D Self-Portraits

Hao Li[1] Etienne Vouga[2] Anton Gudym[5] Linjie Luo[3] Jonathan T. Barron[4] Gleb Gusev[5]
[1]University of Southern California [2] Columbia University [3]Adobe Research [4]UC Berkeley [5]Artec Group

Figure 1: *With our system, users can scan themselves with a single 3D sensor by rotating the same pose for a few different views (typically eight, ∼45 degrees apart) to cover the full body. Our method robustly registers and merges different scans into a watertight surface with consistent texture in spite of shape changes during repositioning, and lighting differences between the scans. These surfaces are suitable for applications such as online avatars or 3D printing (the miniature shown here was printed using a ZPrinter 650.)*

Abstract

We develop an automatic pipeline that allows ordinary users to capture complete and fully textured 3D models of themselves in minutes, using only a single Kinect sensor, in the uncontrolled lighting environment of their own home. Our method requires neither a turntable nor a second operator, and is robust to the small deformations and changes of pose that inevitably arise during scanning. After the users rotate themselves with the same pose for a few scans from different views, our system stitches together the captured scans using multi-view non-rigid registration, and produces watertight final models. To ensure consistent texturing, we recover the underlying albedo from each scanned texture and generate seamless global textures using Poisson blending. Despite the minimal requirements we place on the hardware and users, our method is suitable for full body capture of challenging scenes that cannot be handled well using previous methods, such as those involving loose clothing, complex poses, and props.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Digitizing and scanning

Keywords: 3D scanning, non-rigid registration, depth-sensor, human body, texture reconstruction

Links: DL PDF WEB VIDEO

1 Introduction

For many years, acquiring 3D models of real-world objects was a complex task relegated to experts using sophisticated equipment such as laser scanners, carefully calibrated stereo setups, or large arrays of lights and cameras. The recent rise of cheap, consumer-level 3D sensors, such as Microsoft's Kinect, is rapidly *democratizing* the process of 3D scanning: as these sensors become smaller, cheaper, more accurate and robust, they will continue to permeate the consumer market. Within a decade, 3D capability will likely become as standard built-in feature on laptops and home computers as ordinary video cameras are today.

Recent work on software systems for geometry processing have leaped forward to adapt to the revolution in 3D acquisition hardware. Using methods like Kinect Fusion [Newcombe et al. 2011], ordinary users with no domain knowledge can now generate scans of everyday objects with stunning detail and accuracy. However, with the users behind the 3D sensor, it is difficult to use these methods to capture the *3D self-portraits* of the users *on their own* analogous to photographic self-portraits. In this paper, we concern ourselves with the development of a flexible, robust and accurate capture system for 3D self-portraits using a single 3D sensor.

Figure 2: *3D printed miniatures of captured surfaces.*

There are many potential applications for such 3D self-portraits: combined with some algorithms for automatic skinning, these portraits could be used as personalized, *virtual avatars* in video games or video conferencing applications. Users could quickly scan and upload complete 3D portraits of themselves showing off new outfits and styles to social media sites, or create physical action figures of themselves by having the models 3D printed. Since a 3D portrait fully captures a user's measurements, it could be used to accurately preview the fit and drape of clothing ("virtual try-on") when shopping online. By scanning themselves regularly over a period of time, users could both visually and quantitatively track improvements in their health and fitness.

With a single 3D sensor and no other operators helping to move the sensor, users have to rotate themselves to scan all parts of their bodies. This naturally raises two problems. First, incidental changes of

ACM Reference Format
Li, H., Vouga, E., Gudym, A., Luo, L., Barron, J., Gusev, G. 2013. 3D Self-Portraits.
ACM Trans. Graph. 32, 6, Article 187 (November 2013), 9 pages. DOI = 10.1145/2508363.2508407
http://doi.acm.org/10.1145/2508363.2508407.

DOI: http://doi.acm.org/10.1145/2508363.2508407

Anatomy Transfer

Dicko Ali-Hamadi[1,2,3] Tiantian Liu[4] Benjamin Gilles[5,1] Ladislav Kavan[4] François Faure[1, 2,3]
Olivier Palombi[1,3] Marie-Paule Cani [1, 2,3]

[1]INRIA [2]LJK-CNRS [3]Université de Grenoble [4]University of Pennsylvania [5]LIRMM-CNRS

Figure 1: *A reference anatomy (left) is automatically transferred to arbitrary humanoid characters. This is achieved by combining interpolated skin correspondences with anatomical rules.*

Abstract

Characters with precise internal anatomy are important in film and visual effects, as well as in medical applications. We propose the first semi-automatic method for creating anatomical structures, such as bones, muscles, viscera and fat tissues. This is done by transferring a reference anatomical model from an input template to an arbitrary target character, only defined by its boundary representation (skin). The fat distribution of the target character needs to be specified. We can either infer this information from MRI data, or allow the users to express their creative intent through a new editing tool. The rest of our method runs automatically: it first transfers the bones to the target character, while maintaining their structure as much as possible. The bone layer, along with the target skin eroded using the fat thickness information, are then used to define a volume where we map the internal anatomy of the source model using harmonic (Laplacian) deformation. This way, we are able to quickly generate anatomical models for a large range of target characters, while maintaining anatomical constraints.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: Character modeling

Links: DL PDF

ACM Reference Format
Ali-Hamadi, D., Liu, T., Gilles, B., Kavan, L., Faure, F., Palombi, O., Cani, M. 2013. Anatomy Transfer. ACM Trans. Graph. 32, 6, Article 188 (November 2013), 8 pages. DOI = 10.1145/2508363.2508415 http://doi.acm.org/10.1145/2508363.2508415.

DOI: http://doi.acm.org/10.1145/2508363.2508415

1 Introduction

A high level of anatomical precision is necessary in many Computer Graphics applications, from visualizing the internal anatomy for education purposes, to anatomical simulation for feature films, ergonomics, medical, or biomechanical applications (e.g. optimizing muscle energy). Highly realistic animations showing muscles or tendons deforming the skin typically require precise anatomical models. Moreover, the control of the fat distribution is important for achieving the associated secondary dynamics effects. While a lot of research addresses the challenge of fast and accurate simulation, we focus on the upstream part of the pipeline, modeling anatomy.

The current tools available for artists to model anatomical deformations [Maya-Muscle 2013] as well as early academic work [Wilhelms and Van Gelder 1997; Scheepers et al. 1997] extensively rely on user input, essentially amounting to setting up the musculature from scratch. Recent years witnessed huge improvements in anatomically-based simulation, especially in terms of computational efficiency [Patterson et al. 2012]. However, the cost of setting up a 3D anatomical model for a given character remains. This task is very time consuming and tedious, as it requires modeling of the bones, organs, muscles, and connective and fat tissues. With real humans, it is possible to take advantage of 3D imaging, such as MRI [Blemker et al. 2007]. However, this route is difficult or even impossible for fictional characters, ranging from Popeye to Avatar's Na'vi.

A naive idea to solve the problem would be to transfer the anatomy from a reference character to the target in a purely geometric way. It is obvious this route has a number of shortcomings: humanoids are made of bones, viscera, muscles, and fat tissues. Specific anatomical rules need to be preserved in order to generate a plausible anatomical structure: bones should remain straight and symmetric, and the distribution of fat, which may vary from one individual to another, should be taken into account while transferring muscles and viscera. CG characters can also contain non-anatomical or styl-

Augmenting Physical Avatars using Projector-Based Illumination

Amit Bermano[1,2], Philipp Brüschweiler[1,2], Anselm Grundhöfer[1], Daisuke Iwai[3], Bernd Bickel[1], Markus Gross[1,2]

[1]Disney Research Zurich [2]ETH Zurich [3]Osaka University

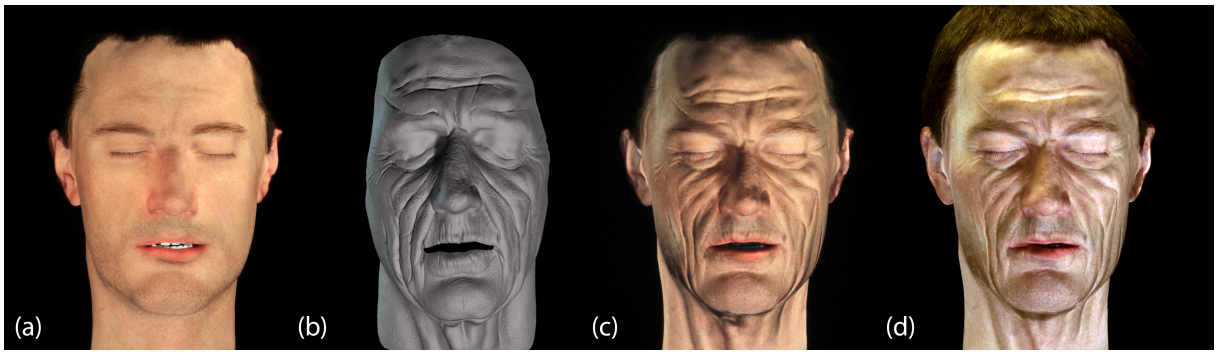


Figure 1: *Our system allows augmentation of a physical avatar (a) with projector-based illumination, significantly increasing its expressiveness. In (b) the target performance is shown. The appearance under controlled and ambient illumination is shown in (c) and (d).*

Abstract

Animated animatronic figures are a unique way to give physical presence to a character. However, their movement and expressions are often limited due to mechanical constraints. In this paper, we propose a complete process for augmenting physical avatars using projector-based illumination, significantly increasing their expressiveness. Given an input animation, the system decomposes the motion into low-frequency motion that can be physically reproduced by the animatronic head and high-frequency details that are added using projected shading. At the core is a spatio-temporal optimization process that compresses the motion in gradient space, ensuring faithful motion replay while respecting the physical limitations of the system. We also propose a complete multi-camera and projection system, including a novel defocused projection and subsurface scattering compensation scheme. The result of our system is a highly expressive physical avatar that features facial details and motion otherwise unattainable due to physical constraints.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling I.4.3 [Image Processing and Computer Vision]: Enhancement;

Keywords: projector-camera systems, animatronics

Links: DL PDF

ACM Reference Format
Bermano, A., Brüschweiler, P., Grundhöfer, A., Iwai, D., Bickel, B., Gross, M. 2013. Augmenting Physical Avatars using Projector-Based Illumination. ACM Trans. Graph. 32, 6, Article 189 (November 2013), 10 pages. DOI = 10.1145/2508363.2508416 http://doi.acm.org/10.1145/2508363.2508416.

0730-0301/13/11-ART189 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508416

1 Introduction

Bringing virtual characters to life is one of the great challenges in computer graphics. While there were tremendous advancements in capturing, animating, and rendering realistic human faces in the past decade, displaying them on traditional screens conveys only a limited sense of physical presence. Animatronic figures or robotic avatars can bridge this gap. However, in contrast to virtual face models, reproducing detailed facial motions on an animatronic head is highly challenging due to physical constraints. Although steady progress in creating highly sophisticated robotic heads that strive to recreate convincing facial motions can be observed, for example those in Disney World's Hall of Presidents or "Geminoids" [Nishio et al. 2007], these achieve only limited expressiveness when compared to a real human being.

Our goal is to significantly increase the expressiveness of such figures, and to allow to animate them and controlling their motion and appearance easily, by adding additional degrees of freedom with projected shading. An animatronic head consists of a deformable skin attached to an underlying rigid articulated structure. The appearance is determined by the material of the skin and its static texture. The articulated structure is driven by a set of motors, and their motion range determines the expressiveness of the figure. While adding additional mechanical components to extend the degrees of freedom would be an obvious choice, in practice this is often prohibitive due to the lack of space inside the head and the extensive cost. Instead, we suggest projected shading to obtain dynamic control of the appearance, and emulate expressive motion and appearance using a combination of low-frequency motion of the animatronic head and high-frequency shading.

In this paper, we present a two-scale model for representing facial motion tailored to animatronic heads, embedded in a multi-projection system. Low-frequency motions that can be reproduced by the physical head are represented as control parameters of actuators. High-frequency details and subtle motions that cannot be reproduced are emulated in texture space. In practice, we face the challenge that the mechanical motion range of the robotic head is significantly smaller than that of a human. However, the formation of facial details is strongly correlated to the underlying low-frequency motion. Given an arbitrary performance capture se-

Fine-Grained Semi-Supervised Labeling of Large Shape Collections

Qi-Xing Huang Hao Su Leonidas Guibas

Stanford University

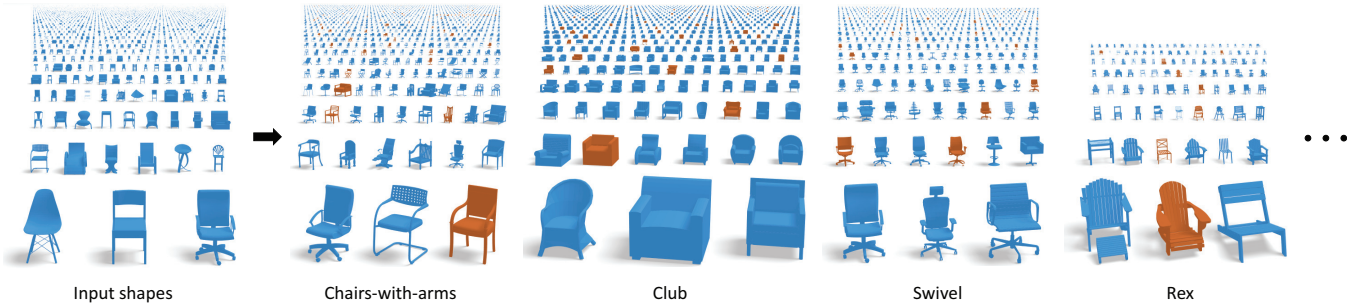


Figure 1: *The proposed approach takes a large set of shapes with sparse and noisy labels as input; it outputs cleaned and complete labels for each shape, facilitating organization and search of the shape collection. Labeled chair sets are shown, with training shapes in orange.*

Abstract

In this paper we consider the problem of classifying shapes within a given category (e.g., chairs) into finer-grained classes (e.g., chairs with arms, rocking chairs, swivel chairs). We introduce a multi-label (i.e., shapes can belong to multiple classes) semi-supervised approach that takes as input a large shape collection of a given category with associated sparse and noisy labels, and outputs cleaned and complete labels for each shape. The key idea of the proposed approach is to jointly learn a distance metric for each class which captures the underlying geometric similarity within that class, e.g., the distance metric for swivel chairs evaluates the global geometric resemblance of chair bases. We show how to achieve this objective by first geometrically aligning the input shapes, and then learning the class-specific distance metrics by exploiting the feature consistency provided by this alignment. The learning objectives consider both labeled data and the mutual relations between the distance metrics. Given the learned metrics, we apply a graph-based semi-supervised classification technique to generate the final classification results.

In order to evaluate the performance of our approach, we have created a benchmark data set where each shape is provided with a set of ground truth labels generated by Amazon's Mechanical Turk users. The benchmark contains a rich variety of shapes in a number of categories. Experimental results show that despite this variety, given very sparse and noisy initial labels, the new method yields results that are superior to state-of-the-art semi-supervised learning techniques.

CR Categories: I.3.5 [Computing Methodologies]: Computer Graphics—Computational Geometry and Object Modeling;

Keywords: shape matching, semi-supervised learning, distance learning, multi-label classification, noisy labels, fine-grained

Links: DL PDF WEB

1 Introduction

Shape classification is a fundamental problem in shape analysis. So far most existing works have focused on classifying shapes into different high-level categories, e.g., cars, chairs, desks, etc. With the emergence of large shape collections, however, even the shapes within each category still exhibit significant variation. For example, chair models from the Trimble 3D Warehouse contain dozens of sub-classes, including chairs-with-arms, swivel chairs, rocking chairs, etc. (See Figure 1). Classifying shapes into these fine-grained classes can benefit a variety of applications such as product search, browsing and exploration of shape variability, and interactive shape modeling.

In this paper, we consider a semi-supervised problem setting, where the given input is a set of man-made shapes together with associated sparse and noisy labels (e.g., models from Trimble 3D Warehouse and their associated text), and the output consists of cleaned and complete labels for each input shape. This problem is particularly challenging due to (1) relatively subtle geometric differences between different classes, (2) the availability of only very sparse and often quite noisy labels, (3) the fact that each shape can be associated with multiple labels, and finally (4) the size of the problem, as a shape collection will typically contain thousands of models.

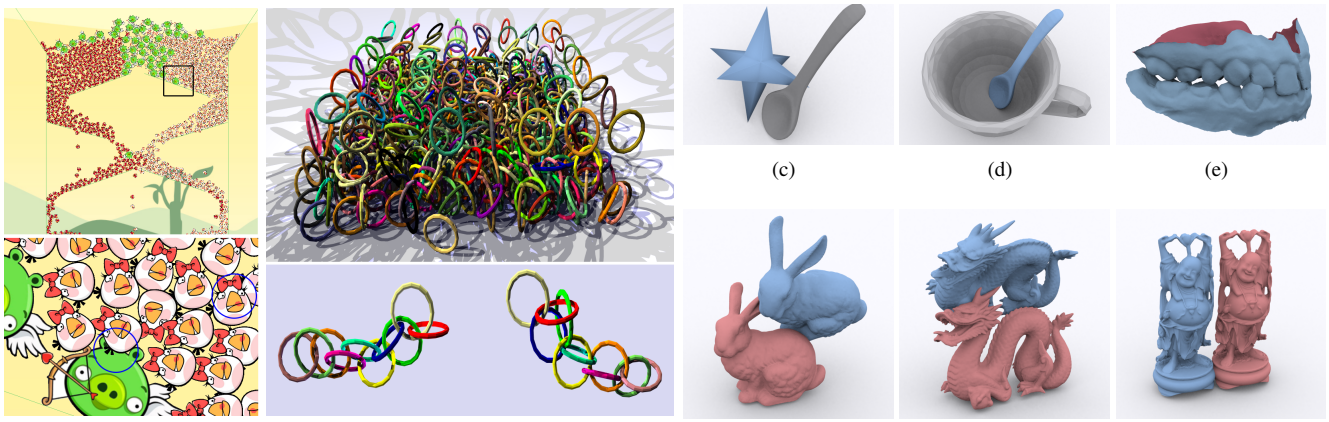
The proposed approach addresses these challenges by combining two simple ideas motivated from recent advances in geometry processing and machine learning. First, inspired by current interest in data-driven shape matching, [Kim et al. 2012; Huang et al. 2012; Kim et al. 2013], we propose to align the input shapes of a given category into a common space, thus implicitly generating a set of correspondences between the shapes. This common space provides us with a convenient framework in which to compare shapes, making features across different shapes more consistent and comparable. For example, the common space allows us to focus at particular neighborhoods of that space and examine local shape variations in those neighborhoods under appropriate similarity metrics. To handle large datasets with high shape variability, we introduce a scalable shape matching approach that is able to simultaneously align many thousands of diverse shapes.

ACM Reference Format
Huang, Q., Su, H., Guibas, L. 2013. Fine-Grained Semi-Supervised Labeling of Large Shape Collections. ACM Trans. Graph. 32, 6, Article 190 (November 2013), 10 pages. DOI = 10.1145/2508363.2508364 http://doi.acm.org/10.1145/2508363.2508364.

DOI: http://doi.acm.org/10.1145/2508363.2508364

Efficient Penetration Depth Approximation using Active Learning

Jia Pan Xinyu Zhang Dinesh Manocha *
University of North Carolina at Chapel Hill



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1: *Our algorithm computes a global penetration depth between overlapping non-convex and non-manifold objects. (a) Dynamic simulation of angry bird characters falling into a complex chute in Box2D physics engine; (b) rainfall of* $1,000$ *rings in Bullet physics engine; (c) a star and a spoon; (d) a spoon and a cup; (e) multiple contacts between upper and lower teeth (each has more than* $40,000$ *triangles). (f-h) benchmarks consisting of complex models (bunny, dragon and Buddha models have* $70K$*,* $230K$ *and* $1M$ *triangles, respectively). Our PD algorithm takes less than* $0.1 \sim 2$ *milliseconds, with less than 2-3% relative error, for each pair of overlapping objects.*

Abstract

We present a new method for efficiently approximating the *global penetration depth* between two rigid objects using machine learning techniques. Our approach consists of two phases: offline learning and performing run-time queries. In the learning phase, we precompute an approximation of the *contact space* of a pair of intersecting objects from a set of samples in the configuration space. We use active and incremental learning algorithms to accelerate the precomputation and improve the accuracy. During the run-time phase, our algorithm performs a nearest-neighbor query based on translational or rotational distance metrics. The run-time query has a small overhead and computes an approximation to global penetration depth in a few milliseconds. We use our algorithm for collision response computations in Box2D or Bullet game physics engines and complex 3D models and observe more than an order of magnitude improvement over prior PD computation techniques.

CR Categories: I.3.3 [Computer Graphics]: Computational Geometry and Object Modeling—Geomtric algorithms, languages and systems

Keywords: Contact Space, Penetration Depth, Support Vector Machine, Active Learning, Dynamic Simulation

Links: DL PDF WEB

*e-mail: {panj,zhangxy,dm}@cs.unc.edu

ACM Reference Format
Pan, J., Zhang, X., Manocha, D. 2013. Efficient Penetration Depth Approximation using Active Learning. ACM Trans. Graph. 32, 6, Article 191 (November 2013), 12 pages. DOI = 10.1145/2508363.2508305 http://doi.acm.org/10.1145/2508363.2508305.

DOI: http://doi.acm.org/10.1145/2508363.2508305

1 Introduction

Measuring the extent of inter-penetration between two intersecting objects is an important problem in physically-based simulation and geometric computing. A widely used metric to quantify inter-penetration is the *penetration depth* (PD), which is defined as the minimum amount of motion transformation required to separate two intersecting objects. This transformation may correspond to only translation (translational PD) or both translation and rotation (generalized PD). PD computation is widely used for contact resolution in dynamic simulation [Baraff and Witkin 2001], tolerance verification for virtual prototyping [Kim et al. 2002a], force computation in haptic rendering [Wang et al. 2012b], motion planning in robotics [Zhang et al. 2007b], etc. Many well known game physics engines such as Box2D [Catto 2010] and Bullet [Coumans 2010] also perform PD computation for collision response.

The time complexity of exact PD computation in 3D can be as high as $\mathcal{O}(m^3n^3)$ for translational PD and $\mathcal{O}(m^6n^6)$ for generalized PD, where m and n are the number of triangles in two non-convex input models. Given the high combinatorial complexity of exact PD computation, many approximate algorithms have been proposed. At a broad level, prior methods can be classified into techniques based on local feature analysis, constrained optimization, distance fields, convex-decomposition, and point-based approximations. Many of these algorithms compute local PD based on overlapping features, and may not provide a reliable solution for general non-convex models (see Figure 2). The accuracy of local PD algorithms varies based on the relative configuration of the two objects and on the heuristics used to estimate the extent and direction of penetration [Heidelberger et al. 2004; Tang et al. 2012]. Methods based on convex decomposition or point-based approximations can compute a more reliable measure of global penetration, but they are too slow for interactive applications.

Main Results: We present a novel algorithm to approximate global PDs between rigid objects for interactive applications using machine learning classification techniques. Our approach is applicable to all rigid models and can compute translational and gener-

Projective Analysis for 3D Shape Segmentation

Yunhai Wang[⋆∗] Minglun Gong[†⋆] Tianhua Wang[‡⋆] Daniel Cohen-Or[⋇] Hao Zhang[§] Baoquan Chen[⋆♯∗]
[⋆]Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University
[⋇]Tel-Aviv University [§]Simon Fraser University [♯]Shangdong University

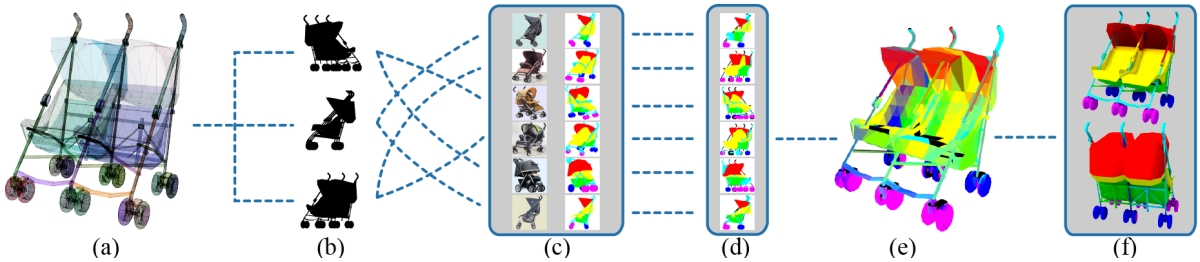


Figure 1: *Our projective analysis treats an input 3D model (a) as a collection of projections (b), which are labeled (d) based on selected images (c) from a pre-labeled image database. Back-projecting 2D labels onto the 3D model forms a probability map (e), which allows us to infer the final shape segmentation and labeling (f). Note how the labeling of the twin stroller is inferred from the images of single strollers.*

Abstract

We introduce *projective analysis* for semantic segmentation and labeling of 3D shapes. The analysis treats an input 3D shape as a collection of 2D projections, labels each projection by transferring knowledge from existing labeled images, and back-projects and fuses the labelings on the 3D shape. The image-space analysis involves matching projected binary images of 3D objects based on a novel *bi-class Hausdorff distance*. The distance is topology-aware by accounting for internal holes in the 2D figures and it is applied to *piecewise-linearly warped* object projections to compensate for part scaling and view discrepancies. Projective analysis simplifies the processing task by working in a lower-dimensional space, circumvents the requirement of having complete and well-modeled 3D shapes, and addresses the data challenge for 3D shape analysis by leveraging the massive available image data. A large and dense labeled set ensures that the labeling of a given projected image can be inferred from closely matched labeled images. We demonstrate semantic labeling of imperfect (e.g., incomplete or self-intersecting) 3D models which would be otherwise difficult to analyze without taking the projective analysis approach.

CR Categories: I.3.7 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems;

Keywords: Projective shape analysis, semantic segmentation and labeling, bilateral symmetric Hausdorff distance, shape matching

Links: DL PDF WEB DATA CODE

∗Corresponding authors: Yunhai Wang (cloudseawang@gmail.com), Baoquan Chen (baoquan.chen@gmail.com)

ACM Reference Format
Wang, Y., Gong, M., Wang, T., Cohen-Or, D., Zhang, H., Chen, B. 2013. Projective Analysis for 3D Shape Segmentation. ACM Trans. Graph. 32, 6, Article 192 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508393 http://doi.acm.org/10.1145/2508363.2508393.

DOI: http://doi.acm.org/10.1145/2508363.2508393

1 Introduction

Human visual perception of 3D shapes is based on 2D observations [Fleming and Singh 2009], which are typically obtained by projecting a 3D object into multiple views. A collection of the multi-view projective images together forms an understanding of the 3D object. This is the basic premise of multi-view 3D shape reconstruction and recognition [Ferrari et al. 2004], where 2D data as sparse as object silhouettes can be highly effective [Laurentini 1994]. Silhouettes turn out to be one of the most important visual cues in object recognition [Koenderink 1984], while binary images provide enriched shape characterizations. One of the most successful global shape descriptors for 3D retrieval is based on the multi-view light field descriptor (LFD) [Chen et al. 2003], which is computed from projected contour and image data.

In this paper, we propose projective analysis of 3D shapes beyond multi-view object reconstruction or recognition. We focus on the higher-level and more delicate task of semantic segmentation and labeling of 3D shapes. The core idea is to transfer labels from available 2D data by selecting and back-projecting the inferred labels onto a 3D shape. Rather than merely transforming a global shape analysis problem from 3D to 2D [Chen et al. 2003], we perform fine-grained shape matching in the projective space and establish connections between 2D and 3D parts to allow label transfer.

Potential gains offered by projective analysis are three-fold. First, analyzing projected images rather than 3D geometry can circumvent the requirement of having complete and well-modeled 3D shapes with quality surface tessellations, without losing the ability to discriminate or characterize the projected shapes. Second, working in a low-dimensional space, from 3D to 2D, simplifies the 3D shape segmentation task. Last but not least, the approach makes it possible to tap into and leverage the massive availability of image data, e.g., those from online photographs.

In recent years, there has been growing interest in data-driven analysis [Kalogerakis et al. 2010; Sidi et al. 2011] of 3D shapes to address the challenge of learning shape semantics. The success of data-driven analysis rests directly on the quality of the utilized data. Quality datasets as a whole should be sufficiently large in number, as well as dense and rich in variability, so as to cover the variability in the input. At an individual level, each shape should possess an adequate representation quality (e.g., complete or watertight) to

3D Wikipedia: Using online text to automatically label and navigate reconstructed geometry

Bryan C. Russell[1] Ricardo Martin-Brualla[2] Daniel J. Butler[2] Steven M. Seitz[2] Luke Zettlemoyer[2]
[1]Intel Labs [2]University of Washington

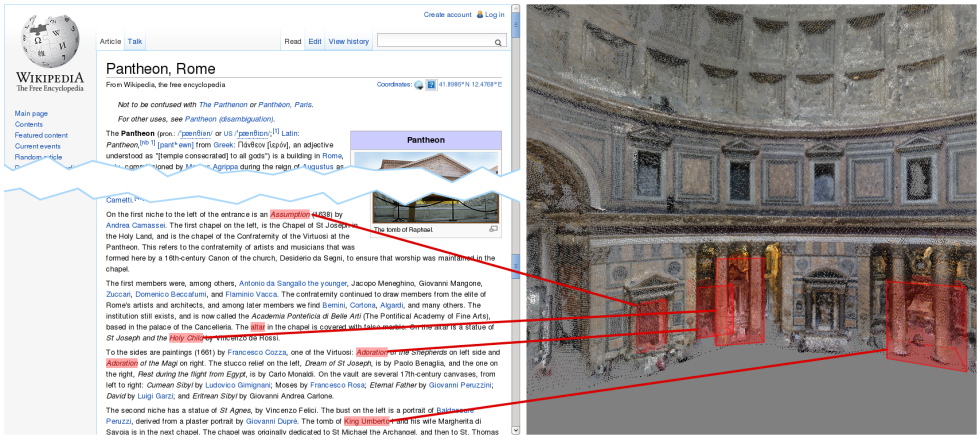


Figure 1: *Given a reference text describing a specific site, for example the Wikipedia article above for the Pantheon, we* ***automatically*** *create a labeled 3D reconstruction, with objects in the model linked to where they are mentioned in the text. The user interface enables coordinated browsing of the text with the visualization (see video).*

Abstract

We introduce an approach for analyzing Wikipedia and other text, together with online photos, to produce *annotated* 3D models of famous tourist sites. The approach is completely automated, and leverages online text and photo co-occurrences via Google Image Search. It enables a number of new interactions, which we demonstrate in a new 3D visualization tool. Text can be selected to move the camera to the corresponding objects, 3D bounding boxes provide anchors back to the text describing them, and the overall narrative of the text provides a temporal guide for automatically flying through the scene to visualize the world as you read about it. We show compelling results on several major tourist sites.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities I.2.7 [Artificial Intelligence]: Natural Language Processing—Text analysis I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Modeling and recovery of physical attributes

Keywords: image-based modeling and rendering, Wikipedia, natural language processing, 3D visualization

Links: DL PDF

ACM Reference Format
Russell, B., Martin-Brualla, R., Butler, D., Seitz, S., Zettlemoyer, L. 2013. 3D Wikipedia: Using online text to automatically label and navigate reconstructed geometry. ACM Trans. Graph. 32, 6, Article 193 (November 2013), 10 pages. DOI = 10.1145/2508363.2508425 http://doi.acm.org/10.1145/2508363.2508425.

0730-0301/13/11-ART193 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508425

1 Introduction

Tourists have long relied on guidebooks and other reference texts to learn about and navigate sites of interest. While guidebooks are packed with interesting historical facts and descriptions of site-specific objects and spaces, it can be difficult to fully *visualize* the scenes they present. The primary cues come from images provided with the text, but coverage is sparse and it can be difficult to understand the spatial relationships between each image viewpoint. For example, the Berlitz and Lonely Planet guides [Berlitz International 2003; Garwood and Hole 2012] for Rome each contain just a single photo of the Pantheon, and have a similar lack of photographic coverage of other sites. Even online sites such as Wikipedia, which do not have space restrictions, have similarly sparse and disconnected visual coverage.

Instead of relying exclusively on static images embedded in text, suppose you could create an interactive, photorealistic visualization, where, for example, a Wikipedia page is shown next to a detailed 3D model of the described site. When you select an object (e.g., “Raphaels tomb”) in the text, it flies you to the corresponding location in the scene via a smooth, photorealistic transition. Similarly, when you click on an object in the visualization, it highlights the corresponding descriptive text on the Wikipedia page. Our goal is to create such a visualization **completely automatically** by analyzing the Wikipedia page itself, together with many photos of the site available online (Figure 1).

Automatically creating such a visualization presents a formidable challenge. The text and photos, in isolation, provide only very indirect cues about the structure of the scene. Although we can easily gather text describing the world, automatically extracting the names of objects (e.g.,“Raphaels tomb” or “Coronation of the Virgin”) is not trivial. For example, we know a noun phrase often describes an entity, which could be an object in the scene. However, it could also name the artist that created the object, or some other unrelated concept. Given the correct names, even more challenging is deter-

Inverse Image Editing: Recovering a Semantic Editing History from a Before-and-After Image Pair

Shi-Min Hu[1] Kun Xu[1] Li-Qian Ma[1] Bin Liu[1] Bi-Ye Jiang[1] Jue Wang[2]
[1]TNList, Tsinghua University, Beijing [2]Adobe Research

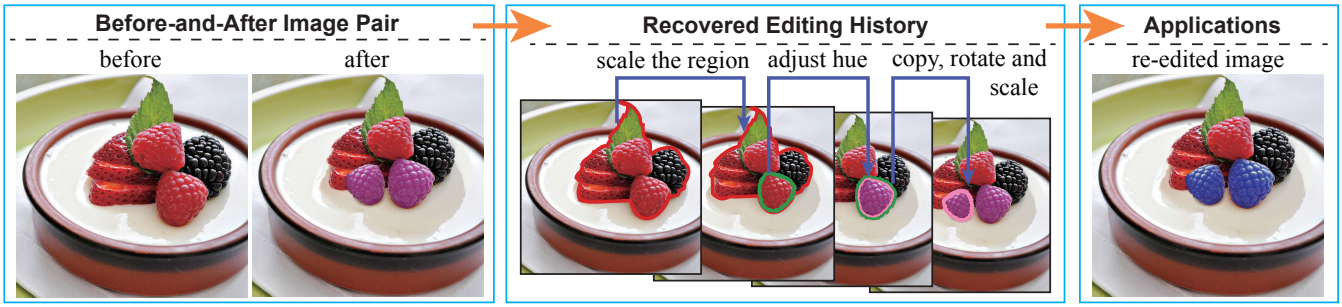


Figure 1: *Given a source image and an edited copy (left), our system automatically recovers a semantic editing history (middle), which can be used for various applications, such as re-editing (right). In this case, the second editing step of the recovered history, involving hue modification, is altered to change the berries to a difference color. Image courtesy of Andrea Lein.*

Abstract

We study the problem of *inverse image editing*, which recovers a semantically-meaningful editing history from a source image and an edited copy. Our approach supports a wide range of commonly-used editing operations such as cropping, object insertion and removal, linear and non-linear color transformations, and spatially-varying adjustment brushes. Given an input image pair, we first apply a dense correspondence method between them to match edited image regions with their sources. For each edited region, we determine geometric and semantic appearance operations that have been applied. Finally, we compute an optimal editing path from the region-level editing operations, based on predefined semantic constraints. The recovered history can be used in various applications such as image re-editing, edit transfer, and image revision control. A user study suggests that the editing histories generated from our system are semantically comparable to the ones generated by artists.

CR Categories: I.3.4 [Computer Graphics]: Graphics Utilities—Graphics editors; I.4.9 [Image Processing and Computer Vision]: Applications

Keywords: image editing history, inverse image editing, history recovery, region matching

Links: DL PDF WEB

ACM Reference Format
Hu, S., Xu, K., Ma, L., Liu, B., Jiang, B., Wang, J. 2013. Inverse Image Editing: Recovering a Semantic Editing History from a Before-and-After Image Pair. ACM Trans. Graph. 32, 6, Article 194 (November 2013), 11 pages. DOI = 10.1145/2508363.2508371 http://doi.acm.org/10.1145/2508363.2508371.

0730-0301/13/11-ART194 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508371

1 Introduction

In image editing, a series of operations are often performed to accomplish an editing task. For instance, the user may first select an object in the image, apply geometric transforms to adjust its shape and position, and then apply various color adjustments to enhance its appearance. This process is repeated if multiple image regions need to be touched up to achieve an editing goal, resulting in a long and sometimes complicated editing history.

Having a clean and complete editing history available is required in many graphics and file management applications, such as image editing revision control [Chen et al. 2011], automatic tutorial generation [Grabler et al. 2009] and editing visualization [Heer et al. 2008]. Retention of an editing history also leads to new possibilities in image editing, such as: adjusting the history to produce results with different variations from the source image; transferring the history to another image, etc. These editing goals are much harder to achieve without having the editing history.

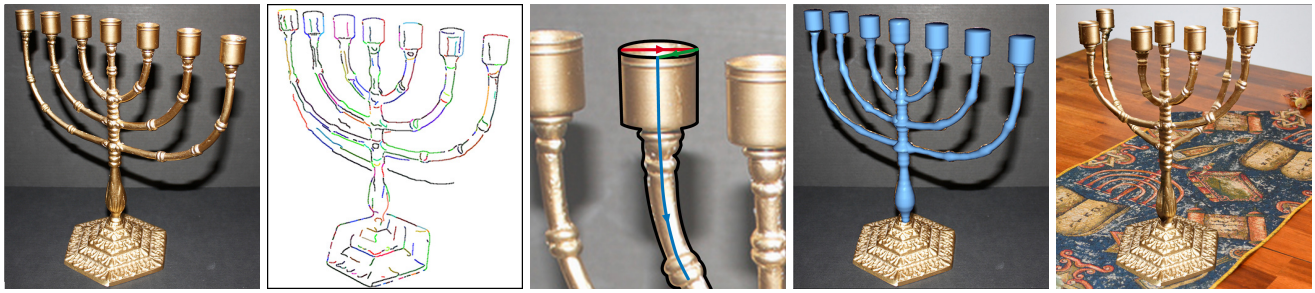
Unfortunately, although existing software provides powerful editing tools, there is no universal, efficient solution to encode, store, transmit and re-use an editing history. For instance, Adobe Photoshop [1] only allows a *partial* history to be saved in either the image file or the command log file. It is often the case that a *complete* history is not available after the editing task is accomplished. Furthermore, image editing is a trial and error process especially for amateurs. It can easily take dozens of operations to explore different ideas, fix errors and fine tweak parameters before producing a desired output. As a result, the raw editing history produced by a novice user is often long, redundant, less understandable, and may not be directly applicable in some applications that requires a *clean* history.

In this work, we study the problem of recovering a *clean* and *semantically meaningful* editing history given a source image and an edited version, which we call *inverse image editing*. This task involves several technical challenges. Firstly, we need to discover which objects or regions have been edited. Secondly, we need to determine how each object or region has been edited. Finally, a semantically meaningful editing path needs to be generated from

[1]http://www.photoshop.com/

3-Sweep: Extracting Editable Objects from a Single Photo

Tao Chen[1,2,3*] Zhe Zhu[1] Ariel Shamir[3] Shi-Min Hu[1] Daniel Cohen-Or[2]
[1]TNList, Department of Computer Science and Technology, Tsinghua University
[2]Tel Aviv University [3]The Interdisciplinary Center



(a) (b) (c) (d) (e)

Figure 1: *3-Sweep Object Extraction. (a) Input image. (b) Extracted edges. (c) 3-Sweep modeling of one component of the object. (d) The full extracted 3D model. (e) Editing the model by rotating each arm in a different direction, and pasting onto a new background. The base of the object is transferred by alpha matting and compositing.*

Abstract

We introduce an interactive technique for manipulating simple 3D shapes based on extracting them from a single photograph. Such extraction requires understanding of the components of the shape, their projections, and relations. These simple cognitive tasks for humans are particularly difficult for automatic algorithms. Thus, our approach combines the cognitive abilities of humans with the computational accuracy of the machine to solve this problem. Our technique provides the user the means to quickly create editable 3D parts— human assistance implicitly segments a complex object into its components, and positions them in space. In our interface, three strokes are used to generate a 3D component that snaps to the shape's outline in the photograph, where each stroke defines one dimension of the component. The computer reshapes the component to fit the image of the object in the photograph as well as to satisfy various inferred geometric constraints imposed by its global 3D structure. We show that with this intelligent interactive modeling tool, the daunting task of object extraction is made simple. Once the 3D object has been extracted, it can be quickly edited and placed back into photos or 3D scenes, permitting object-driven photo editing tasks which are impossible to perform in image-space. We show several examples and present a user study illustrating the usefulness of our technique.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems;

Keywords: Interactive modeling, Photo manipulation

Links: DL PDF

*This work was performed while Tao Chen was a postdoctoral researcher at TAU and IDC, Israel.

ACM Reference Format
Chen, T., Zhu, Z., Shamir, A., Hu, S., Cohen-Or, D. 2013. 3-Sweep: Extracting Editable Objects from a Single Photo. ACM Trans. Graph. 32, 6, Article 195 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508378 http://doi.acm.org/10.1145/2508363.2508378.

DOI: http://doi.acm.org/10.1145/2508363.2508378

1 Introduction

Extracting three dimensional objects from a single photo is still a long way from reality given the current state of technology, as it involves numerous complex tasks: the target object must be separated from its background, and its 3D pose, shape and structure should be recognized from its projection. These tasks are difficult, and even ill-posed, since they require some degree of semantic understanding of the object. To alleviate this problem, complex 3D models can be partitioned into simpler parts, but identifying object parts also requires further semantic understanding and is difficult to perform automatically. Moreover, having decomposed a 3D shape into parts, the relations between these parts should also be understood and maintained in the final composition.

In this paper we present an interactive technique to extract 3D man-made objects from a single photograph, leveraging the strengths of both humans and computers (see Figure 1). Human perceptual abilities are used to partition, recognize and position shape parts, using a very simple interface based on triplets of strokes, while the computer performs tasks which are computationally intensive or require accuracy. The final object model produced by our method includes its geometry and structure, as well as some of its semantics. This allows the extracted model to be readily available for intelligent editing, which maintains the shape's semantics.

Our approach is based on the observation that many man-made objects can be decomposed into simpler parts that can be represented by a generalized cylinder, cuboid or similar primitives. The key idea of our method is to provide the user with an interactive tool to guide the creation of 3D editable primitives. The tool is based on a rather simple modeling gesture we call *3-sweep*. This gesture allows the user to explicitly define the three dimensions of the primitive using three sweeps. The first two sweeps define the first and second dimension of a 2D profile and the third, longer, sweep is used to define the main curved axis of the primitive.

As the user sweeps the primitive, the program dynamically adjusts the progressive profile by sensing the pictorial context in the photograph and automatically snapping to it. Using such 3-sweep operations, the user can model 3D parts consistent with the object in the photograph, while the computer automatically maintains global

PatchNet:
A Patch-based Image Representation for Interactive Library-driven Image Editing

Shi-Min Hu[1] * Fang-Lue Zhang[1] Miao Wang[1] Ralph R. Martin[2] Jue Wang[3]
[1] Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing
[2] Cardiff University [3] Adobe Research

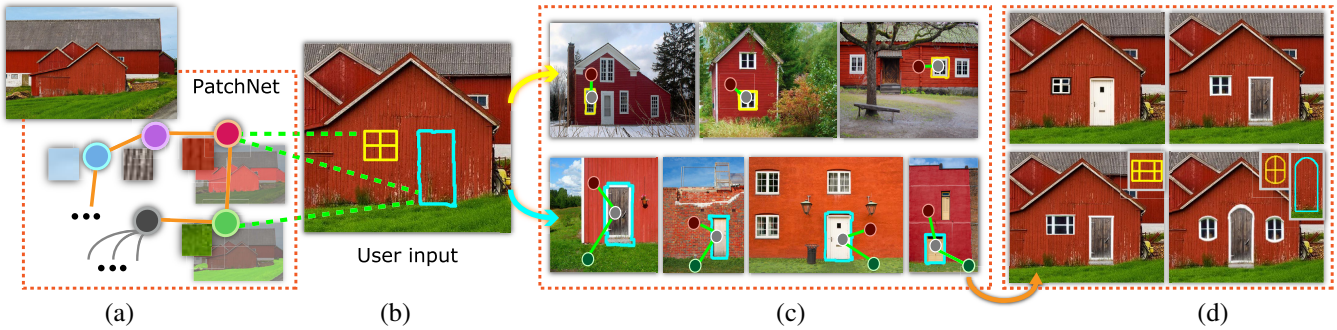


Figure 1: *PatchNet supports interactive library-based image editing. (a) Input image and its PatchNet representation. (b) The user draws a rough sketch to specify an object synthesis task. (c) Using PatchNet, the system searches a large image library in a few seconds to find the best candidate regions meeting editing constraints. (d) The user selects candidate regions to synthesize output as desired, or modifies the sketch to synthesize different object structures (lower-right). Image courtesy of Flickr user LarryBiker.*

Abstract

We introduce *PatchNets*, a compact, hierarchical representation describing structural and appearance characteristics of image regions, for use in image editing. In a PatchNet, an image region with coherent appearance is summarized by a graph node, associated with a single representative patch, while geometric relationships between different regions are encoded by labelled graph edges giving contextual information. The hierarchical structure of a PatchNet allows a coarse-to-fine description of the image. We show how this PatchNet representation can be used as a basis for interactive, library-driven, image editing. The user draws rough sketches to quickly specify editing constraints for the target image. The system then automatically queries an image library to find semantically-compatible candidate regions to meet the editing goal. Contextual image matching is performed using the PatchNet representation, allowing suitable regions to be found and applied in a few seconds, even from a library containing thousands of images.

CR Categories: I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques; K.7.m [Computing Methodologies]: Image Processing and Computer Vision—Applications

Keywords: PatchNet, image representation, patch synthesis, interactive image editing, contextual features

Links: DL PDF WEB

*Corresponding author. E-mail:shimin@tsinghua.edu.cn.

ACM Reference Format
Hu, S., Zhang, F., Wang, M., Martin, R., Wang, J. 2013. PatchNet: A Patch-based Image Representation for Interactive Library-driven Image Editing. ACM Trans. Graph. 32, 6, Article 196 (November 2013), 12 pages. DOI = 10.1145/2508363.2508381 http://doi.acm.org/10.1145/2508363.2508381.

0730-0301/13/11-ART196 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508381

1 Introduction

Patch-based synthesis methods have recently emerged as a powerful tool for various image and video editing tasks [Kwatra et al. 2003; Barnes et al. 2009; Darabi et al. 2012; Xiao et al. 2011]. Existing patch-based interactive editing systems usually require the user to provide semantic guidance or constraints to the low-level patch synthesis algorithms to achieve semantically meaningful results [Hu et al. 2013]. In particular, for multiple source synthesis, when new objects are placed in a target image by cloning source regions from other library images, the user must manually specify the source region to copy. This quickly becomes tedious if the number of library images is high, as the user must manually search through the library to find useful source image regions. This could potentially be automated by using dense correspondence algorithms such as NRDC [HaCohen et al. 2011], but applying such algorithms to the entire library would be extremely slow, and would need redoing for each new input image. An alternative approach would be to use image search methods to find the most similar images in the library to the input image, giving a smaller reference data set for a particular input image. However, image search based on global features may not yield optimal results when local portions of images are to be used.

The fundamental difficulty in extending patch-based synthesis methods to a large library lies in the fact that patches only describe local image appearance, while a high-level representation of image structure that would allow efficient, semantic search of the library is missing. Although compact image representations have been extensively used in computer vision for applications such as image classification and object recognition, such representations are typically built upon highly abstracted features (e.g. a bag of visual words), and so cannot be directly applied to *local* image editing tasks in which the object scale is relatively small.

In this paper we address the problem of how to efficiently leverage a large image library for interactive image editing. We present PatchNets, a compact, patch-based image representation that captures characteristics of both the *global structure* and the *local appearance* of an image. As shown in Fig. 2, a PatchNet is a graph model in which each node represents a contiguous, homogeneous

A Sparse Control Model for Image and Video Editing

Li Xu Qiong Yan Jiaya Jia*

Department of Computer Science and Engineering
The Chinese University of Hong Kong

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

Figure 1: *Very sparse control samples for image and video editing. Applications that can be benefitted include, but are not limited to, (a) tone adjustment, (b) colorization, (c) non-photorealistic rendering, and (d) re-colorization. In the two rows, we show input images with editing samples and our corresponding results respectively. Input image (c) courtesy of Patrick Smith.*

Abstract

It is common that users draw strokes, as control samples, to modify color, structure, or tone of a picture. We discover inherent limitation of existing methods for their implicit requirement on where and how the strokes are drawn, and present a new system that is principled on minimizing the amount of work put in user interaction. Our method automatically determines the influence of edit samples across the whole image jointly considering spatial distance, sample location, and appearance. It greatly reduces the number of samples that are needed, while allowing for a decent level of global and local manipulation of resulting effects and reducing propagation ambiguity. Our method is broadly beneficial to applications adjusting visual content.

Keywords: image editing, video editing, image processing, matting, smoothing, cutout, colorization, segmentation

Links: DL PDF WEB CODE

*e-mail: {xuli, qyan, leojia}@cse.cuhk.edu.hk

ACM Reference Format
Xu, L., Yan, Q., Jia, J. 2013. A Sparse Control Model for Image and Video Editing. ACM Trans. Graph. 32, 6, Article 197 (November 2013), 10 pages. DOI = 10.1145/2508363.2508404 http://doi.acm.org/10.1145/2508363.2508404.

DOI: http://doi.acm.org/10.1145/2508363.2508404

1 Introduction

Many recent image and video processing methods are performed with the input of user set control samples for spatially-variant editing. For example, colorization [Levin et al. 2004] reconstructs chrominance channels based on a few color strokes. Interactive tone adjustment is achieved in [Lischinski et al. 2006] in a similar manner. Other representative tools include natural image matting [Wang and Cohen 2005; Levin et al. 2008a; Levin et al. 2008b], material editing [Pellacini and Lawrence 2007; An and Pellacini 2008], and white balance correction [Boyadzhiev et al. 2012]. All these methods share the characteristics of making results comply with image structures, in order to preserve edges in regional adjustment. The control sample strategy eschews blind parameter tuning employed in early image processing.

Albeit important and general, optimal adaptive editing from very sparse control points remains a major challenge due to the ambiguity for pixels *distant from* or *in between* the samples. The simple example in Fig. 1(d) fully exhibits the difficulty, where the user-drawn sparse strokes are expected to turn the left-bottom yellow tulip to red and all other yellow ones to orange.

Intriguingly, previous global and local methods have their respective limitations to produce correct results for this simple example. All-pixel-pair (a.k.a. global) methods [An and Pellacini 2008; Xu et al. 2009a; Xu et al. 2009b] that relate each pixel to all samples can quickly propagate edit across the image. But they suffer from mixing the two types of edits (white and gray) on many pixels. Inevitably, unpleasant color blending is produced, as will be detailed in Section 3. Local-pixel-pair (a.k.a. local) approaches [Levin et al. 2004; Lischinski et al. 2006], contrarily, need to edit tulips densely, which involve much more user interaction.

Apparently, this problem is about density and distribution of user

WYSIWYG Computational Photography via Viewfinder Editing

Jongmin Baek[1], Dawid Pająk[2], Kihwan Kim[2], Kari Pulli[2], and Marc Levoy[1]

[1]Stanford University [2]NVIDIA Research

Figure 1: *Viewfinder editing and appearance-based metering.* **(a)** *In the proposed system, the user interacts with the camera viewfinder as if it were a live canvas, by making sparse strokes to select regions and to specify edits. The edits propagate spatiotemporally as the camera and the scene objects move, and the edits are applied in the subsequent frames of the viewfinder, which are tonemapped HDR images, to provide WYSIWYG experience.* **(b)** *After making a local tonal edit, the user presses the shutter to trigger a high-resolution capture. Here we show the resulting tonemapped HDR composite without any edits applied, for reference. The edit mask computed from the user strokes is shown in the inset.* **(c)** *The result with edits applied is shown. This approximately corresponds to what the user sees on the screen just as he presses the shutter. Our appearance-based metering acquired an exposure stack at (0.645 ms, 5.555 ms, 11.101 ms) by optimizing the quality of the final result based on the user's global and local edits.* **(d)** *The regions indicated in (c), magnified.* **(e)** *We captured another stack with a histogram-based HDR metering at (0.579 ms, 9.958 ms, 23.879 ms) and applied the same post-processing pipeline. As the standard HDR metering considers equally all the pixels in the scene, it allocates too much effort to capture the dark areas that were not as important to the user, leading to a longer capture times that invite ghosting (top) and higher noise in mid-tones (bottom).*

Abstract

Digital cameras with electronic viewfinders provide a relatively faithful depiction of the final image, providing a WYSIWYG experience. If, however, the image is created from a burst of differently captured images, or non-linear interactive edits significantly alter the final outcome, then the photographer cannot directly see the results, but instead must imagine the post-processing effects. This paper explores the notion of *viewfinder editing*, which makes the viewfinder more accurately reflect the final image the user intends to create. We allow the user to alter the local or global appearance (tone, color, saturation, or focus) via stroke-based input, and propagate the edits spatiotemporally. The system then delivers a real-time visualization of these modifications to the user, and drives the camera control routines to select better capture parameters.

Keywords: image editing, burst photography, HDR metering

Links: DL PDF

1 Introduction

Early photographers could not directly see their photographs as they were taking them, but had to imagine the results as a function of various imaging parameters such as exposure, focus, even choice of film and paper. Digital cameras with real-time digital displays, which show a good preview image, have made photography much easier in this respect, providing a WYSIWYG (what you see is what you get) interface for the camera. In particular, framing the image and choosing the timing of capture is made easier and more fun as the user sees right away what the outcome will be. However, when applying many computational photography techniques, the user still cannot see an approximation of the result, but needs to imagine the end result. One example is combining an exposure burst into an HDR image and tonemapping it back to LDR for display, In addition, many photographers edit their photographs after capture, using tools such as Photoshop or Lightroom. However, users must compose the shot in the field *before* knowing the effect such edits might have on the result. The capture process thus remains separated from the image editing process, potentially leading to inadequate data acquisition (wrong composition or insufficient SNR) or excessive data acquisition (longer capture time, exacerbated handshake or motion blur, and increased storage and transmission cost).

At the same time, mobile devices with a digital camera, display, adequate computational power, and touch interface are becoming increasingly commonplace. More photographs are captured by mobile devices now than ever, and many of them are edited directly on device and shared from that device, without ever being uploaded to PCs. This phenomenon is reflected in the recent focus on camera control and image processing on mobile platforms [Adams et al. 2010b; Ragan-Kelley et al. 2012] and in the popularity of photography apps on smartphones. The trend so far has been to harness this increased processing power to enable desktop-like workflow on mobile devices. We instead use these processing capabilities to introduce the notion of *viewfinder editing*, which allows the users to perform edits live on the viewfinder prior to capture. We bring the WYSIWYG interface to computational photography applications, allowing the users to directly see the effects of interactive edits or stack photography on the viewfinder. Using this interface we also

ACM Reference Format
Baek, J., Pająk, D., Kim, K., Pulli, K., Levoy, M. 2013. WYSIWYG Computational Photography via Viewfinder Editing. ACM Trans. Graph. 32, 6, Article 198 (November 2013), 10 pages. DOI = 10.1145/2508363.2508421 http://doi.acm.org/10.1145/2508363.2508421.

0730-0301/13/11-ART198 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508421

Image-Based Rendering in the Gradient Domain

Johannes Kopf
Microsoft Research

Fabian Langguth
TU Darmstadt

Daniel Scharstein
Middlebury College

Richard Szeliski
Microsoft Research

Michael Goesele
TU Darmstadt

(a) Input (b) Depth map (c) Standard rendering (d) Gradient domain (e) Result image

Figure 1: *Standard image-based rendering synthesizes novel views of a scene by reprojecting the input image (a) using a coarse estimated depth map (b). The technique cannot handle scenes with reflections and has problems in untextured regions, e.g., holes show up in the reprojection (c). We perform image-based rendering in the gradient domain. Our technique synthesizes horizontal and vertical gradient fields F_x and F_y, as well as an approximate solution S (d). The final image is obtained through Poisson integration (e). Our technique naturally handles reflections and untextured regions.*

Abstract

We propose a novel image-based rendering algorithm for handling complex scenes that may include reflective surfaces. Our key contribution lies in treating the problem in the gradient domain. We use a standard technique to estimate scene depth, but assign depths to image gradients rather than pixels. A novel view is obtained by rendering the horizontal and vertical gradients, from which the final result is reconstructed through Poisson integration using an approximate solution as a data term. Our algorithm is able to handle general scenes including reflections and similar effects *without* explicitly separating the scene into reflective and transmissive parts, as required by previous work. Our prototype renderer is fully implemented on the GPU and runs in real time on commodity hardware.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Viewing algorithms

Keywords: Image-based rendering, gradient domain processing

Links: DL PDF WEB VIDEO CODE

1 Introduction

Traditional computer graphics can produce photo realistic renderings of scenes, but it requires meticulous modeling and fine tuning of every scene detail including geometry, texture, and lighting, as well as sophisticated and often slow global light transport simulations. In contrast, image-based rendering takes a lighter approach by starting from a set of captured photos of a real scene, recovering the camera parameters and coarse 3D proxies, and then reprojecting the input photos using the proxies as seen from a novel viewpoint. This technique can maintain the photographic quality of the input images without requiring explicit modeling of fine scene details. Image-based rendering methods are widely in use today, particularly for navigating urban and street level scenes.

A great challenge for all image-based rendering methods is to reliably recover scene depth in poorly textured areas. The resulting uncertainty typically requires strong regularization and can lead to rendering artifacts. Another problem is that traditional image-based rendering methods recover only a single depth value per input pixel. Thus, reflective or glossy scenes, which require modeling multiple separate layers at different depths, cannot be handled, and lead to ghosting. Sinha et al. [2012] addressed this issue for scenes with reflections by decomposing each input photo into up to two additive layers, each with their own estimated depth. The technique, however, has difficulties recovering accurate depths and discerning reflective and non-reflective scene parts, which leads to various artifacts.

In this paper, we introduce a new approach to image-based rendering that operates in the gradient domain. We use a standard regularized multi-view stereo algorithm to recover the scene depth. However, we interpret the result as the depth of the image *gradients* rather than *pixels*. We then project the gradients to their new locations as seen from the novel viewpoint and reconstruct the image through Poisson integration. To provide a data term for the Poisson solver, we directly render the effect of a gradient moving within the image. The details of our method are described in Sections 3–6.

This approach has many advantages over traditional image-based rendering. Most importantly, any depth estimation technique works best in the vicinity of strong gradients that can be reliably found in neighboring images. In less textured regions, where depth estimation is hardest, the gradients are nearly zero and thus do not contribute much to our results. In contrast, previous methods tend to produce artifacts in such regions. In addition, image-based rendering in the gradient domain is particularly well suited for handling reflections and other non-Lambertian effects as well as reasoning about occlusions. This is due to the fact that gradients in natural images are sparse, and, hence, the gradients of two mixed layers can be easily separated. (In fact, Levin et al. [2004] have used this observation to unmix two reflective layers in a single image.)

ACM Reference Format
Kopf, J., Langguth, F., Scharstein, D., Szeliski, R., Goesele, M. 2013. Image-Based Rendering in the Gradient Domain. ACM Trans. Graph. 32, 6, Article 199 (November 2013), 9 pages.
DOI = 10.1145/2508363.2508369 http://doi.acm.org/10.1145/2508363.2508369.

0730-0301/13/11-ART199 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508369

Data-driven Hallucination of Different Times of Day from a Single Outdoor Photo

Yichang Shih
MIT CSAIL

Sylvain Paris
Adobe

Frédo Durand
MIT CSAIL

William T. Freeman
MIT CSAIL

Input image at "blue hour" (just after sunset) | A database of time-lapse videos | Hallucinate at night

Figure 1: *Given a single input image (courtesy of Ken Cheng), our approach hallucinates the same scene at a different time of day, e.g., from blue hour (just after sunset) to night in the above example. Our approach uses a database of time-lapse videos to infer the transformation for hallucinating a new time of day. First, we find a time-lapse video with a scene that resembles the input. Then, we locate a frame at the same time of day as the input and another frame at the desired output time. Finally, we introduce a novel example-based color transfer technique based on local affine transforms. We demonstrate that our method produces a plausible image at a different time of day.*

Abstract

We introduce "time hallucination": synthesizing a plausible image at a different time of day from an input image. This challenging task often requires dramatically altering the color appearance of the picture. In this paper, we introduce the first data-driven approach to automatically creating a plausible-looking photo that appears as though it were taken at a different time of day. The time of day is specified by a semantic time label, such as "night".

Our approach relies on a database of time-lapse videos of various scenes. These videos provide rich information about the variations in color appearance of a scene throughout the day. Our method transfers the color appearance from videos with a similar scene as the input photo. We propose a *locally affine model* learned from the video for the transfer, allowing our model to synthesize new color data while retaining image details. We show that this model can hallucinate a wide range of different times of day. The model generates a large sparse linear system, which can be solved by off-the-shelf solvers. We validate our methods by synthesizing transforming photos of various outdoor scenes to four times of interest: daytime, the golden hour, the blue hour, and nighttime.

CR Categories: I.4.3 [Computing Methodologies]: Image Processing and Computer Vision—Enhancement

Keywords: Time hallucination, time-lapse videos

ACM Reference Format
Shih, Y., Paris, S., Durand, F., Freeman, W. 2013. Data-driven Hallucination of Different Times of Day from a Single Outdoor Photo. ACM Trans. Graph. 32, 6, Article 200 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508419 http://doi.acm.org/10.1145/2508363.2508419.

DOI: http://doi.acm.org/10.1145/2508363.2508419

Links: DL PDF

1 Introduction

Time of day and lighting conditions are critical for outdoor photography (e.g. [Caputo 2005] chapter "Time of Day"). Photographers spend much effort getting to the right place at the perfect time of day, going as far as dangerously hiking in the dark because they want to reach a summit for sunrise or because they can come back only after sunset. In addition to the famous golden or magical hour corresponding to sunset or sunrise ([Rowell 2012] chapter "The Magical Hour"), the less-known "blue hour" can be even more challenging because it takes place after the sun has set or before it rises ([Rowell 2012] chapter "Between Sunset and Sunrise") and actually only lasts a fraction of an hour when the remaining light scattered by the atmosphere takes a deep blue color and its intensity matches that of artificial lights. Most photographers cannot be at the right place at the perfect time and end up taking photos in the middle of the day when lighting is harsh. A number of heuristics can be used to retouch a photo with photo editing software and make it look like a given time of day, but they can be tedious and usually require manual local touch-up. In this paper, we introduce an automatic technique that takes a single outdoor photo as input and seeks to hallucinate an image of the same scene taken at a different time of day.

The modification of a photo to suggest the lighting of a different time of day is challenging because of the large variety of appearance changes in outdoor scenes. Different materials and different parts of a scene undergo different color changes as a function of reflectance, nearby geometry, shadows, etc. Previous approaches have leveraged additional physical information such as an external 3D model [Kopf et al. 2008] or reflectance and illumination inferred from a collection of photos of the same scene [Laffont et al. 2012; Lalonde et al. 2009].

In contrast, we want to work from a single input photograph and allow the user to request a different time of day. In order to deal with the large variability of appearance changes, we use two main strategies: we densely match our input image with frames from a

Automatic Noise Modeling for Ghost-free HDR Reconstruction

Miguel Granados Kwang In Kim James Tompkin Christian Theobalt*
MPI für Informatik, Saarbrücken, Germany

Figure 1: *Our method receives a set of images taken with different exposure times (smaller images) and reconstructs a ghost-free high dynamic range image (larger images; tone mapped). The* acrobat *sequence on the left was captured hand-held with in-camera exposure bracketing. To our knowledge, our method is the first in the literature to reconstruct plausible HDR images of both highly dynamic scenes (*left*) and highly cluttered scenes (*right*) with both small and large displacements with little or no manual intervention.*

Abstract

High dynamic range reconstruction of dynamic scenes requires careful handling of dynamic objects to prevent ghosting. However, in a recent review, Srikantha et al. [2012] conclude that "there is no single best method and the selection of an approach depends on the user's goal". We attempt to solve this problem with a novel approach that models the noise distribution of color values. We estimate the likelihood that a pair of colors in different images are observations of the same irradiance, and we use a Markov random field prior to reconstruct irradiance from pixels that are likely to correspond to the same static scene object. Dynamic content is handled by selecting a single low dynamic range source image and hand-held capture is supported through homography-based image alignment. Our noise-based reconstruction method achieves better ghost detection and removal than state-of-the-art methods for cluttered scenes with large object displacements. As such, our method is broadly applicable and helps move the field towards a single method for dynamic scene HDR reconstruction.

CR Categories: I.4.8 [Image Processing And Computer Vision]: Scene Analysis—Photometry, Time-varying imagery;

Keywords: HDR deghosting, camera noise, motion detection

Links: DL PDF

*e-mail:{granados,kkim,jtompkin,theobalt}@mpii.de

ACM Reference Format
Granados, M., Kim, K., Tompkin, J., Theobalt, C. 2013. Automatic Noise Modeling for Ghost-free HDR Reconstruction. ACM Trans. Graph. 32, 6, Article 201 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508410 http://doi.acm.org/10.1145/2508363.2508410.

DOI: http://doi.acm.org/10.1145/2508363.2508410

1 Introduction

It is difficult to acquire high dynamic range images (HDR) of dynamic scenes without introducing *ghosting*. Even when using modern cameras with automatic exposure bracketing, the *inter-frame* capture time between input images can be long enough to cause significant object displacement between images of dynamic scenes (Fig. 1). Early HDR research implicitly assumed that both the camera pose and the scene remained static during the acquisition of a set of low dynamic range (LDR) images [Burt and Kolczynski 1993; Mann and Picard 1995]. When these techniques average images of dynamic scenes, they introduce ghosting artifacts (Fig. 8, right). Specialized HDR cameras have also been built, but these are expensive and are not widely available [Tocci et al. 2011].

Deghosting has been addressed in the literature through three different strategies: 1) aligning the scene before color averaging, 2) performing joint alignment and reconstruction using one reference image from the LDR set, and 3) detecting regions with moving objects and excluding their images from the average. All of these strategies fail under challenging real-life conditions. After performing an experimental validation of state-of-the-art deghosting methods, Srikantha et al. [2012] conclude that "there is no single best method and the selection of an approach depends on the user's goal".

1) Scene alignment Bogoni [2000], Kang et al. [2003], and Zimmer et al. [2011] perform a dense alignment of the images using optical flow prior to color averaging. Although optical flow methods can correct short displacements caused by camera shake and moving objects, they typically fail to estimate large displacements, and have difficulties with disocclusions occurring in highly cluttered and highly dynamic scenes. Flow estimation is an active area of research and has many limitations, and the success of these deghosting methods depends on the availability of accurate flow fields.

2) Joint alignment and reconstruction Sen et al. [2012] perform simultaneous alignment and HDR reconstruction. Their method defines a reference image to which all other images are patch-wise aligned. Ill-exposed regions in the reference are filled using an adaptation of the bi-directional similarity func-

Patch-Based High Dynamic Range Video

Nima Khademi Kalantari[1] Eli Shechtman[2] Connelly Barnes[2,3] Soheil Darabi[2] Dan B Goldman[2] Pradeep Sen[1]

[1]University of California, Santa Barbara [2]Adobe [3]University of Virginia

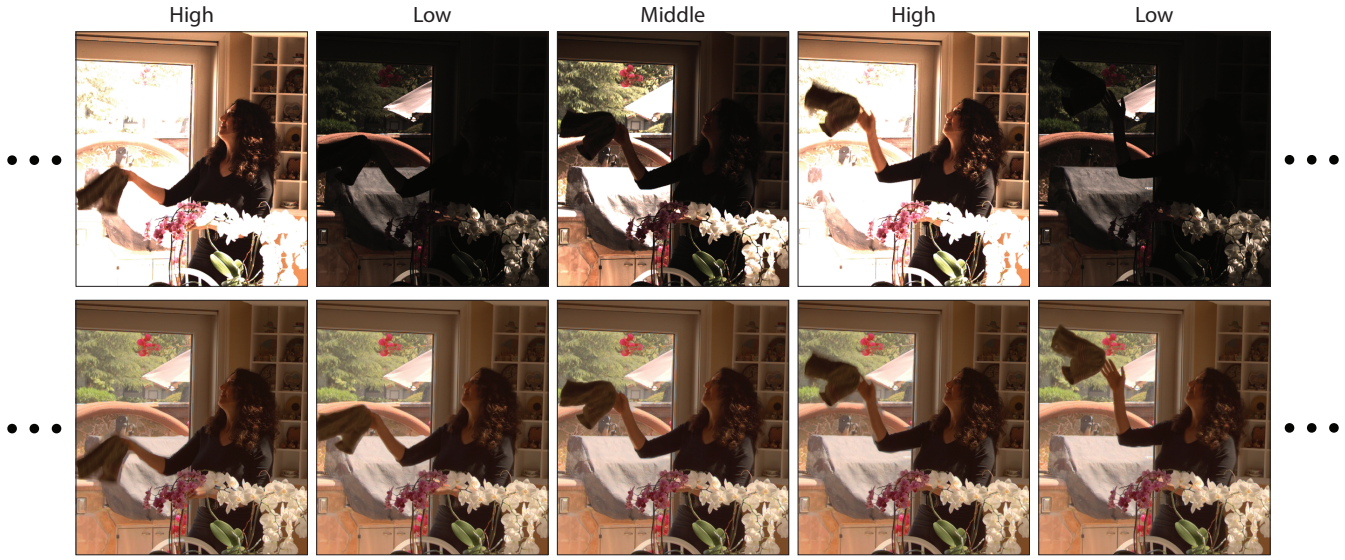


Figure 1: *(top row) Input video acquired using an off-the-shelf camera, which alternates between three exposures separated by two stops. (bottom row) Our algorithm reconstructs the missing LDR images and generates an HDR image at each frame. The HDR video result for this* `ThrowingTowel3Exp` *scene can be found in the supplementary materials. This layout is adapted from Kang et al. [2003].*

Abstract

Despite significant progress in high dynamic range (HDR) imaging over the years, it is still difficult to capture high-quality HDR video with a conventional, off-the-shelf camera. The most practical way to do this is to capture alternating exposures for every LDR frame and then use an alignment method based on optical flow to register the exposures together. However, this results in objectionable artifacts whenever there is complex motion and optical flow fails. To address this problem, we propose a new approach for HDR reconstruction from alternating exposure video sequences that combines the advantages of optical flow and recently introduced patch-based synthesis for HDR images. We use patch-based synthesis to enforce similarity between adjacent frames, increasing temporal continuity. To synthesize visually plausible solutions, we enforce constraints from motion estimation coupled with a search window map that guides the patch-based synthesis. This results in a novel reconstruction algorithm that can produce high-quality HDR videos with a standard camera. Furthermore, our method is able to synthesize plausible texture and motion in fast-moving regions, where either patch-based synthesis or optical flow alone would exhibit artifacts. We present results of our reconstructed HDR video sequences that are superior to those produced by current approaches.

CR Categories: I.4.1 [Computing Methodologies]: Image Processing and Computer Vision—Digitization and Image Capture

Keywords: High dynamic range video, patch-based synthesis

Links: DL PDF WEB

1 Introduction

High dynamic range (HDR) imaging is now popular and becoming more widespread. Most of the research to date, however, has focused on improving the capture of still HDR images, while HDR video capture has received considerably less attention. This is a serious deficit, since high-quality HDR video would significantly improve our ability to capture dynamic environments as our eyes perceive them. The reason for this lack of progress is that the bulk of HDR video research has focused on specialized HDR camera systems (e.g., [Nayar and Mitsunaga 2000; Unger and Gustavson 2007; Tocci et al. 2011; SpheronVR 2013; Kronander et al. 2013]). Unfortunately, the high cost and general unavailability of these cameras make them impractical for the average consumer.

On the other hand, still HDR photography has leveraged the fact that a typical consumer camera can acquire a set of low dynamic range (LDR) images at different exposures, which can then be merged into a single HDR image [Mann and Picard 1995; Debevec and Malik 1997]. However, most of the methods that address artifacts in dynamic scenes (e.g., [Zimmer et al. 2011; Sen et al. 2012]) only produce still images and cannot be used for HDR video.

ACM Reference Format
Kalantari, N., Shechtman, E., Barnes, C., Darabi, S., Goldman, D., Sen, P. 2013. Patch-Based High Dynamic Range Video. ACM Trans. Graph. 32, 6, Article 202 (November 2013), 8 pages.
DOI = 10.1145/2508363.2508402 http://doi.acm.org/10.1145/2508363.2508402.

DOI: http://doi.acm.org/10.1145/2508363.2508402

Animating Human Lower Limbs Using Contact-Invariant Optimization

Igor Mordatch[1] Jack M. Wang[2,3] Emanuel Todorov[1] Vladlen Koltun[3,4]

[1]University of Washington [2]University of Hong Kong [3]Stanford University [4]Adobe Research

Abstract

We present a trajectory optimization approach to animating human activities that are driven by the lower body. Our approach is based on contact-invariant optimization. We develop a simplified and generalized formulation of contact-invariant optimization that enables continuous optimization over contact timings. This formulation is applied to a fully physical humanoid model whose lower limbs are actuated by musculotendon units. Our approach does not rely on prior motion data or on task-specific controllers. Motion is synthesized from first principles, given only a detailed physical model of the body and spacetime constraints. We demonstrate the approach on a variety of activities, such as walking, running, jumping, and kicking. Our approach produces walking motions that quantitatively match ground-truth data, and predicts aspects of human gait initiation, incline walking, and locomotion in reduced gravity.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation;

Keywords: human motion, human simulation, optimization

Links: DL PDF

1 Introduction

Creating realistic human motion from compact high-level objectives is a fundamental problem in character animation. For example, we would like to generate realistic walking for a given human character by simply specifying the desired velocity, or realistic jumping by specifying the desired height. While animations can always be created by a skilled artist given sufficient time and effort, a technique that can generate custom realistic motions can accelerate the process by providing a high-quality animation that can be refined by the artist. Such a technique can also be used for assisting novice animators and for pre-visualization in production settings.

In this paper, we present a method for *de novo* synthesis of human motion. Our method produces animations for given sparse objectives: walking emerges when the objective specifies a velocity for the torso, running emerges when the specified velocity is higher, jumping emerges when a target height is given, and kicking is synthesized when the character's foot must have certain velocity at a given point in space. The primary goal of our work is to enable automatic synthesis of realistic lower body motion for human activities that are driven by the lower body. Such activities include walking, running, jumping, and kicking.

ACM Reference Format
Mordatch, I., Wang, J., Todorov, E., Koltun, V. 2013. Animating Human Lower Limbs Using Contact-Invariant Optimization. ACM Trans. Graph. 32, 6, Article 203 (November 2013), 8 pages.
DOI = 10.1145/2508363.2508365 http://doi.acm.org/10.1145/2508363.2508365.

DOI: http://doi.acm.org/10.1145/2508363.2508365

To synthesize high-fidelity lower limb motion we use a trajectory optimization approach that is based on spacetime constraints. Trajectory optimization techniques have been applied to human characters, but have either required objectives that depend on task-specific motion capture data [Liu et al. 2005] or failed to produce realistic results [Erez and Todorov 2012]. The problem is challenging because spacetime constraints on three-dimensional humanoid models lead to high-dimensional search spaces and complex nonlinear constraints that are difficult to satisfy. Furthermore, ground contact forces that change abruptly with body kinematics lead to discontinuous objective functions prone to poor local minima.

Nevertheless, we show that high-fidelity lower limb motions for a range of activities can be generated by a single trajectory optimization formulation that does not rely on prior task-specific motion data. We achieve this using a reformulation of contact-invariant optimization (CIO) [Mordatch et al. 2012b]. CIO is an optimization framework that smoothes out discontinuities in the objective and allows a single continuous optimization to search over the space of possible motion trajectories and contact patterns. While CIO has been previously used on reduced character models, we apply the framework to human characters with equations of motion that capture the detailed dynamics of the body. In order to enable fine-grained optimization over contact timings, we generalize and simplify CIO by eliminating the quantization of contact patterns that previous formulations relied on.

Since the actuation mechanism of the humanoid model has significant impact on the realism of synthesized motion [Liu et al. 2005; Wang et al. 2012], we actuate the lower body joints with Hill-type musculotendon units (MTUs). The MTUs regularize the synthesized torque patterns and enable the use of a biologically-plausible effort model. Unlike prior work that utilized MTUs and biologically-inspired objectives to generate human locomotion [Wang et al. 2012], we do not rely on a task-specific control structure. As a result, our approach is not limited to steady-state walking and running.

We restrict our attention to the fidelity of lower limb motion since a simplified model of the musculoskeletal system for the lower body is readily available [Wang et al. 2012]. The ideas we present may also enable de novo synthesis of human motion in which the entire body moves with high fidelity to real-world human movement, but this likely requires developing computationally efficient musculoskeletal models for the full body, perhaps building on the work of Lee et al. [2009]. We leave such development to future work.

Our work demonstrates that a single trajectory optimization formulation can produce high-fidelity lower body motion for a range of human activities, without the need for prior motion data or task-specific control structures. We evaluate the presented approach by synthesizing high-fidelity locomotion in a variety of conditions, as well as jumping and kicking motions.

2 Related Work

One natural approach to human motion synthesis is to design controllers for humanoid models in forward dynamical simulations. Early efforts focused on hand-designing controllers for specific activities, resulting in characters with impressive movement reper-

Evaluating the Distinctiveness and Attractiveness of Human Motions on Realistic Virtual Bodies

Ludovic Hoyet* Kenneth Ryall*† Katja Zibrek* Hwangpil Park‡ Jehee Lee‡ Jessica Hodgins†§
Carol O'Sullivan*†‡

*Trinity College Dublin †Disney Research §Carnegie Mellon University ‡Seoul National University

Figure 1: *Example frames from distinctive (leftmost pairs) and non-distinctive walking, jogging and dancing motion clips.*

Abstract

Recent advances in rendering and data-driven animation have enabled the creation of compelling characters with impressive levels of realism. While data-driven techniques can produce animations that are extremely faithful to the original motion, many challenging problems remain because of the high complexity of human motion. A better understanding of the factors that make human motion recognizable and appealing would be of great value in industries where creating a variety of appealing virtual characters with realistic motion is required. To investigate these issues, we captured thirty actors walking, jogging and dancing, and applied their motions to the same virtual character (one each for the males and females). We then conducted a series of perceptual experiments to explore the distinctiveness and attractiveness of these human motions, and whether characteristic motion features transfer across an individual's different gaits. Average faces are perceived to be less distinctive but more attractive, so we explored whether this was also true for body motion. We found that dancing motions were most easily recognized and that distinctiveness in one gait does not predict how recognizable the same actor is when performing a different motion. As hypothesized, average motions were always amongst the least distinctive and most attractive. Furthermore, as 50% of participants in the experiment were Caucasian European and 50% were Asian Korean, we found that the latter were as good as or better at recognizing the motions of the Caucasian actors than their European counterparts, in particular for dancing males, whom they also rated more highly for attractiveness.

CR Categories: I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Animation;

Keywords: human animation, perception, motion capture

*{hoyetl, ryallk, zibrekk, Carol.OSullivan}@scss.tcd.ie
†jkh@disneyresearch.com
‡{hwangpilpark, jehee}@mrl.snu.ac.kr

Links: DL PDF

1 Introduction

Animating realistic human motion is a challenging problem. The complex biomechanical and physiological processes that drive motion are very difficult to understand and replicate, so for many applications real human motion is captured and retargeted to a virtual human model. However, while such data-driven animation can produce extremely realistic animation, it also has several drawbacks in practice. One such disadvantage could be that the style of the captured person's motion could be quite distinctive, and therefore easily recognized when applied to one or more characters (e.g., in a group or crowd). It would also be undesirable to use motion that might be unappealing or unattractive to some or all of the target audience.

In order to create sufficient variety of motion in an environment given a limited repertoire of human motions, insights into the perception of distinctiveness of such movements would be very valuable. It has previously been found that humans find it difficult to distinguish between the motions of multiple walking people [McDonnell et al. 2008; Pražák and O'Sullivan 2011], but it is not clear if this is true for other gaits and actions apart from walking. Some questions that remain unanswered are: How many motions is it necessary to capture from one actor, or how many actors are needed to ensure that enough variety is present? Do distinctive features transfer across different actions, i.e., can you recognize a person from his/her walk, run, or more stylistic motions such as dancing. When synthesizing new motions, either by editing the original motion to satisfy certain constraints, or by procedurally modifying the motion

ACM Reference Format
Hoyet, L., Ryall, K., Zibrek, K., Park, H., Lee, J., Hodgins, J., O'Sullivan, C. 2013. Evaluating the Distinctiveness and Attractiveness of Human Motions on Realistic Virtual Bodies. ACM Trans. Graph. 32, 6, Article 204 (November 2013), 11 pages. DOI = 10.1145/2508363.2508367 http://doi.acm.org/10.1145/2508363.2508367.

0730-0301/13/11-ART204 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508367

The Line of Action: an Intuitive Interface for Expressive Character Posing

Martin Guay* Marie-Paule Cani Rémi Ronfard
Laboratoire Jean Kuntzmann - Université de Grenoble - Inria

Figure 1: *Expressive character poses created in a few seconds each, by sketching intuitive lines of action.*

Abstract

The line of action is a conceptual tool often used by cartoonists and illustrators to help make their figures more consistent and more dramatic. We often see the *expression* of characters—may it be the dynamism of a super hero, or the elegance of a fashion model—well captured and amplified by a single *aesthetic* line. Usually this line is laid down in early stages of the drawing and used to describe the body's *principal* shape. By focusing on this simple abstraction, the person drawing can quickly adjust and refine the overall pose of his or her character from a given viewpoint. In this paper, we propose a mathematical definition of the line of action (LOA), which allows us to automatically align a 3D virtual character to a user-specified LOA by solving an optimization problem. We generalize this framework to other types of lines found in the drawing literature, such as secondary lines used to place arms. Finally, we show a wide range of poses and animations that were rapidly created using our system.

CR Categories: I.3.6 [Methodology and Techniques]: Interaction techniques— [I.3.7]: Computer Graphics—Animation

Keywords: Line of action, sketch-based animation, character animation.

Links: DL PDF

1 Introduction

Because humans have been drawing and sketching for centuries, many researchers—from within and from outside the computer graphics community—have proposed and argued for sketching as an intuitive and natural interface for both modeling and editing 3D virtual objects. By designing these interfaces closer to the creative process involved when sketching, the cognitive workload, *i.e.* the mental steps required to achieve a specific task, can be efficiently reduced. Looking at these stages—*i.e.* the ones involved in the creative sketching process [Goldschmidt 1991; Cherlin et al. 2005]—we usually find laid down in early stages, *principal descriptive primitives*—circles, lines, ovals, etc.—that give the overall perspective, shape and mass of objects and characters. For modeling 3D objects, primitive shapes such as cylinders [Gingold et al. 2009] and abstract forms such as contours [Igarashi et al. 1999; Karpenko and Hughes 2006] have been successfully used to ease the sketching process by making coarse shape design easier.

When it comes to character posing however—a task commonly done by manipulating an articulated skeleton used to parametrize the character's geometry—sketch-based modeling research has mostly relied on stick figures as *the* 2D representation of the character's skeletal parametrization [Davis et al. 2003]. Although drawing 2D stick figures accelerates posing, it requires sketching multiple, intersecting strokes for specifying the individual limbs, and getting expressive poses requires skills.

Inspired by the practice of cartoonists [Lee and Buscema 1978; Blair 1994; Hart 1997; Brooks and Pilcher 2001], we explore the alternative of a single, smooth stroke—the line of action (LOA)—as an abstraction of the character's body. We claim that the LOA is an *intuitive* interface that helps creating more *expressive* poses. Controlling the whole body with a single stroke makes early design easier and less time consuming. And the resulting poses are more expressive as they often exhibit an aesthetic curved shape that conveys the full body expression more clearly in a given *viewpoint*. Examples of expressive poses rapidly created with lines of action are shown in Fig.1. They mimic the poses of a fashion model, of two dancers, a full body swing and a super hero on the run.

Since the line of action is an abstraction of the body and the correspondence between both is not explicitly given, posing a 3D virtual character from an arbitrary LOA is a challenge. In this paper, we give a formal definition of the LOA, enabling us to solve character posing by solving an optimization problem. Our method includes an automatic way of determining the *correspondence* between the line of action and a subset of the character's bones. We then propose extensions to the LOA concept for drawing secondary lines and we address the well known depth ambiguities in a new, but simple way;

*martin.guay@inria.fr

ACM Reference Format
Guay, M., Cani, M., Ronfard, R. 2013. The Line of Action: an Intuitive Interface for Expressive Character Posing. ACM Trans. Graph. 32, 6, Article 205 (November 2013), 8 pages. DOI = 10.1145/2508363.2508397 http://doi.acm.org/10.1145/2508363.2508397.

0730-0301/13/11-ART205 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508397

Flexible Muscle-Based Locomotion for Bipedal Creatures

Thomas Geijtenbeek*
Utrecht University

Michiel van de Panne
University of British Columbia

A. Frank van der Stappen
Utrecht University

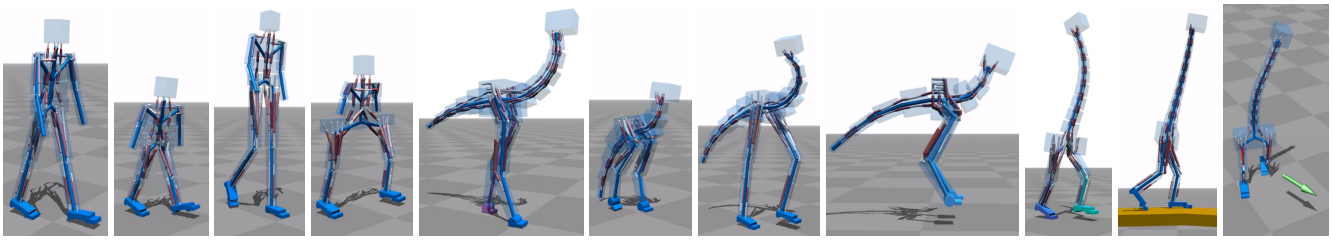


Figure 1: *Physics-based simulation of locomotion for a variety of creatures driven by 3D muscle-based control. The synthesized controllers can locomote in real time at a range of speeds, be steered to a target heading, and can traverse variable terrain.*

Abstract

We present a muscle-based control method for simulated bipeds in which both the muscle routing and control parameters are optimized. This yields a generic locomotion control method that supports a variety of bipedal creatures. All actuation forces are the result of 3D simulated muscles, and a model of neural delay is included for all feedback paths. As a result, our controllers generate torque patterns that incorporate biomechanical constraints. The synthesized controllers find different gaits based on target speed, can cope with uneven terrain and external perturbations, and can steer to target directions.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: physics-based animation, musculoskeletal simulation

Links: DL PDF

1 Introduction

Physics-based simulation is an established technique for the automatic generation of interactive natural-looking motion. To extend this approach to actively controlled virtual characters has been a longstanding research goal, in which tremendous progress has been made in recent years. Locomotion controllers have been developed that robustly deal with changes in character morphology, external perturbations and uneven terrain.

Unfortunately, in many cases the resulting motions are still not as natural as we would like. One common approach that can help improve the quality of the simulated motions is to use motion capture data as part of the control strategy. However, such methods are limited to characters and motions for which data is available. Furthermore, the biomechanical constraints that are implicit in captured motions are not preserved during the motion editing or motion retargeting that is often required to leverage limited motion data. Another approach for improving the motion quality has been to use optimization to help shape the motion, such as optimizing for minimal energy as well as task objectives. However, in the absence of biomechanical constraints, optimization objectives may lead to unnatural torque patterns or require cumbersome manual tuning. Commonly implemented joint limits and torque limits remain a crude approximation of the motion constraints that are implicit in articulated figures driven by musculotendon units.

More recently, emerging from biomechanics research, researchers have begun to develop methods that include biomechanical constraints into the simulation. Using such an approach, the natural gaits of various animals can be simulated without the need for motion data. However, the principal focus to date has been on modeling human motion, and the solutions remain limited in their locomotion abilities and robustness.

In this paper, we make the following contributions:

- We develop a control method and optimization strategy for the simulated locomotion of fully 3D bipedal characters, including imaginary creatures, that are driven entirely by simulated muscle-based actuation. The method produces robust locomotion at given speeds to target directions and does not require pre-existing motion data. The characters can further cope with modest variations in terrain.
- We introduce muscle routing optimization as an important feature that enables and simplifies the design of muscle-based control strategies for a variety of character morphologies. Instead of needing an exact musculoskeletal model, our method requires only an approximate template of where muscles are attached and routed. The specific geometry is then optimized within the specified ranges allowed by the template, along with the parameters related to the muscle control. This approach enables the discovery of efficient muscle routings for creature models for which there exist no real-world data to draw from.
- We make use of a muscle-based approximation to Jacobian transpose control as a core component of our framework. This enables a more creature-generic and motion-generic control architecture and is applied to the majority of joints in our creature models.

*e-mail:t.geijtenbeek@uu.nl

ACM Reference Format
Geijtenbeek, T., van de Panne, M., van der Stappen, A. 2013. Flexible Muscle-Based Locomotion for Bipedal Creatures. ACM Trans. Graph. 32, 6, Article 206 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508399 http://doi.acm.org/10.1145/2508363.2508399.

0730-0301/13/11-ART206 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508399

Robust Realtime Physics-based Motion Control for Human Grasping

Wenping Zhao*† Jianjie Zhang* Jianyuan Min* Jinxiang Chai*
*Texas A&M University †University of Science and Technology of China

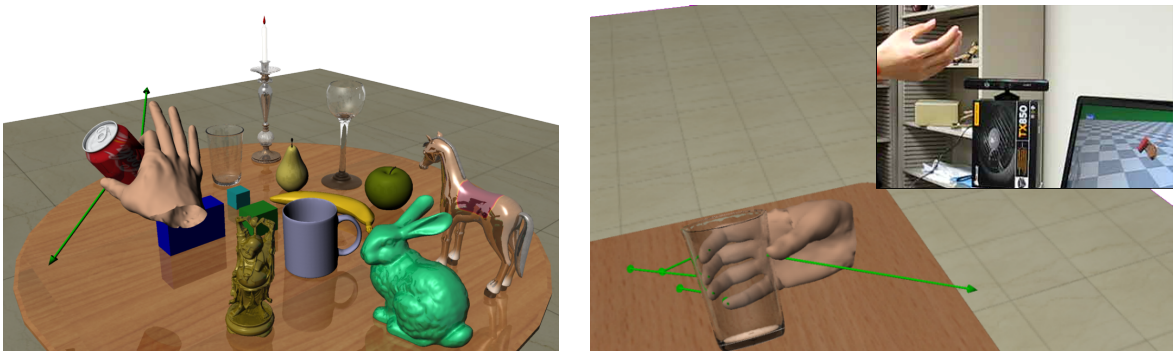


Figure 1: *Realtime generation of physics-based motion control for human grasping: (left) automatic grasping of objects with different shapes, weights, frictions, and spatial orientations; (right) performance interfaces: acting out the desired grasping motion in front of a single Kinect.*

Abstract

This paper presents a robust physics-based motion control system for realtime synthesis of human grasping. Given an object to be grasped, our system automatically computes physics-based motion control that advances the simulation to achieve realistic manipulation with the object. Our solution leverages prerecorded motion data and physics-based simulation for human grasping. We first introduce a data-driven synthesis algorithm that utilizes large sets of prerecorded motion data to generate realistic motions for human grasping. Next, we present an online physics-based motion control algorithm to transform the synthesized kinematic motion into a physically realistic one. In addition, we develop a performance interface for human grasping that allows the user to act out the desired grasping motion in front of a single *Kinect* camera. We demonstrate the power of our approach by generating physics-based motion control for grasping objects with different properties such as shapes, weights, spatial orientations, and frictions. We show our physics-based motion control for human grasping is robust to external perturbations and changes in physical quantities.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—interaction techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—animation

Keywords: Hand grasping and manipulation, data-driven animation, physics-based simulation, performance interfaces

Links: DL PDF WEB VIDEO

ACM Reference Format
Zhao, W., Zhang, J., Min, J., Chai, J. 2013. Robust Realtime Physics-based Motion Control for Human Grasping. ACM Trans. Graph. 32, 6, Article 207 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508412 http://doi.acm.org/10.1145/2508363.2508412.

DOI: http://doi.acm.org/10.1145/2508363.2508412

1 Introduction

Human hands are capable of grasping an astounding variety of objects of different shapes, weights, frictions, and spatial orientations with little effort. However, creating physically realistic animation for human grasping is a nontrivial task. An ideal grasping action must take into account the geometry and dynamic characteristics of the object to be grasped and the selection of contact between the object and the fingers, thumb, and palm of the hand. Recent years have seen some significant advances in this area [Pollard and Zordan 2005; Kry and Pai 2006; Li et al. 2007; Liu 2009; Kyota and Saito 2012]. Still, the ultimate goal of building an automated realtime system that is capable of grasping a wide variety of objects with different geometry and physical quantities remains unsolved. With this goal in mind, we propose an algorithm motivated by the following principles:

Physical realism. Natural appearance is a must because people are extremely adept at judging whether an animated character appears realistic or not. A synthesized grasping motion that accomplishes an intended task might be judged as unacceptable if it appears jerky, moves like a robotic hand, or contains any unpleasant visual artifacts such as hand-object penetration. In addition, we require output animation to be physically plausible, which ensures human grasping takes into account dynamic aspects of objects crucial to human-object interaction.

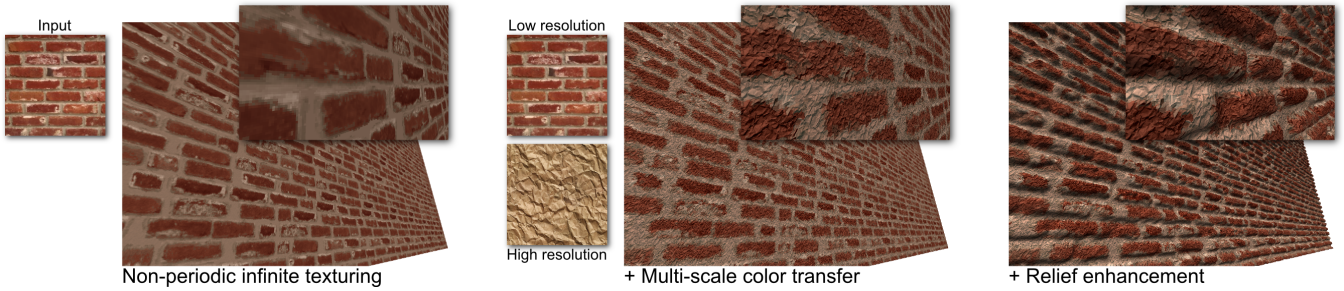
Scalability. A lifelike human character must possess a rich repertoire of grasping actions and display a wide range of variation within the same action. This inevitably requires grasping objects of different shapes, weights, frictions, and spatial orientations using a rich set of grip modes (*e.g.*, the power and pinch grip).

Control. An ideal animation system should empower the user with an easy-to-use interface. This is challenging since we aim to develop a system that allows a random user to generate a desired grasping action quickly and easily–with virtually no learning curve.

Realtime. Realtime animation is essential to many applications such as virtual reality, video games, and interactive animation design. The algorithm must be fast enough so that the interface appears responsive and the user remains engaged in the animation

On-the-Fly Multi-Scale Infinite Texturing from Example

Kenneth Vanhoey, Basile Sauvage, Frédéric Larue, Jean-Michel Dischler*
ICube, Université de Strasbourg, CNRS, France



Abstract

In computer graphics, rendering visually detailed scenes is often achieved through texturing. We propose a method for on-the-fly non-periodic infinite texturing of surfaces based on a single image. Pattern repetition is avoided by defining patches within each texture whose content can be changed at runtime. In addition, we consistently manage multi-scale using one input image per represented scale. Undersampling artifacts are avoided by accounting for fine-scale features while colors are transferred between scales. Eventually, we allow for relief-enhanced rendering and provide a tool for intuitive creation of height maps. This is done using an ad-hoc local descriptor that measures feature self-similarity in order to propagate height values provided by the user for a few selected texels only.
Thanks to the patch-based system, manipulated data are compact and our texturing approach is easy to implement on GPU. The multi-scale extension is capable of rendering finely detailed textures in real-time.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Viewing Algorithms I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture I.4.6 [Image processing and computer vision]: Segmentation—Region growing, partitioning

Keywords: Real-time Rendering, Infinite Texturing, Texture Tiling, Multi-scale Textures, Relief Textures, Color Space

Links: DL PDF WEB VIDEO

*e-mails: {Kenneth.Vanhoey,Sauvage,FLarue,Dischler}@unistra.fr

ACM Reference Format
Vanhoey, K., Sauvage, B., Larue, F., Dischler, J. 2013. On-the-Fly Multi-Scale Infinite Texturing from Example. ACM Trans. Graph. 32, 6, Article 208 (November 2013), 10 pages. DOI = 10.1145/2508363.2508383 http://doi.acm.org/10.1145/2508363.2508383.

DOI: http://doi.acm.org/10.1145/2508363.2508383

1 Introduction

Providing efficient solutions for rendering detailed realistic environments in real-time applications, like games or flight/driving simulators, has always been a major focus in computer graphics. Details can be efficiently rendered using textures. But despite improvements of graphics hardware, memory capacity and data streaming techniques, which allowed over the recent years for increased scene complexity, texturing techniques must still fulfill constraints which are difficult to unify in a single approach. Ideally, they should 1) be as fast as possible to avoid penalizing frame rates, 2) use compact texture maps to limit streaming and data transfers that also penalize frame rates, 3) be non-periodic to avoid visual repetition artifacts, 4) produce fine details to avoid undersampling artifacts, 5) be enhanced with relief when details represent geometry like cracks or bumps to improve the rendering quality by accounting for parallax effects.

Meeting the five aforementioned conditions simultaneously is an open problem. Non-periodic real-time rendering of large surfaces, as well as compactness (properties 1, 2 and 3) can partially be solved by Wang tiling [Cohen et al. 2003] or corner tiling [Lagae and Dutré 2006]. However, to break the repetition effect one should provide tiles with multiple different borders or corners, and the number of tiles grows exponentially with that respect. Both compact and finely detailed textures (properties 2 and 4) can be defined thanks to multi-scale textures. However, stochastic tiling is not easy to extend to multi-scale texturing because border conditions must be made consistent throughout scales [Kopf et al. 2006]. Building the relief (property 5) is tedious without *a priori* knowledge if only a single image is available.

In this paper, we describe a new system to model textures that match all previous properties. It improves on stochastic texture tiling. Our system coherently includes the following contributions:

- Real-time non-periodic infinite surface texturing is achieved by modifying texture tiles at runtime, yet neither introducing visual artifacts nor requiring heavy computations. The memory consumption is also reduced compared to state-of-the-art tiling.
- Multi-scale texturing consistently blends colors between multiple layers of texture: our color transfer mechanism avoids popping and ghosting artifacts by respecting the features of each scale.
- Relief enhancement is integrated as a height map texture, which is kept consistent with the color texture.

Anisotropic Spherical Gaussians

Kun Xu[1] Wei-Lun Sun[1] Zhao Dong[2] Dan-Yong Zhao[1] Run-Dong Wu[1] Shi-Min Hu[1]
[1]TNList, Tsinghua University, Beijing [2] Program of Computer Graphics, Cornell University

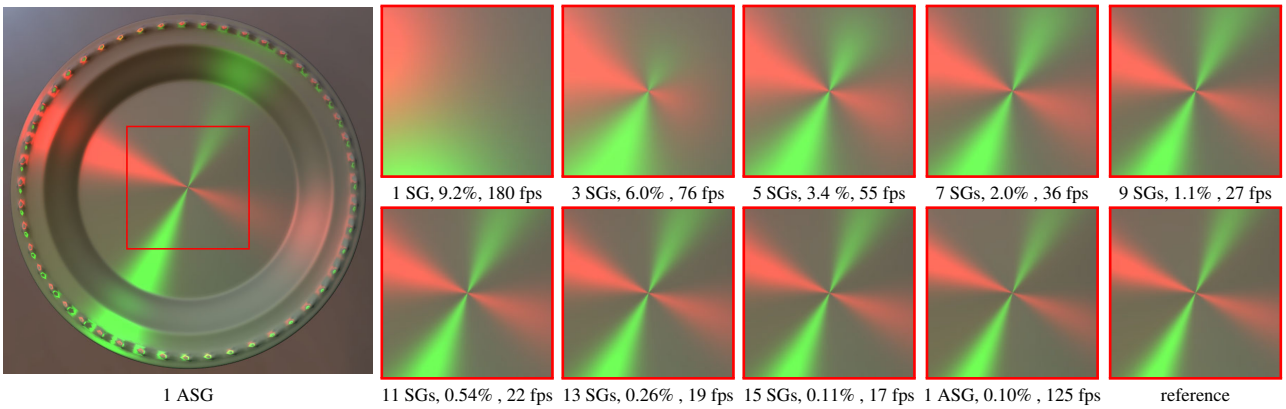


Figure 1: *Comparison of the SG (Spherical Gaussian) based approximation with the ASG (Anisotropic Spherical Gaussian) based approximation in rendering a highly anisotropic metal dish, under an environment light and two local lights. The BRDF of the metal dish is approximated by different number of ASGs or SGs in different images. Notice the superior property of ASGs over SGs. The result generated by 1 ASG already matches the path-traced reference well (with a L^2 error of 0.10%), and achieves a high framerate of 125 fps, while, to achieve a similar quality, more than 10 SGs are required, but with much lower framerates (19 fps for 13 SGs or 17 fps for 15 SGs). The L^2 error and the framerates for each configuration are also given in the corresponding subtitle.*

Abstract

We present a novel *anisotropic Spherical Gaussian* (ASG) function, built upon the *Bingham distribution* [Bingham 1974], which is much more effective and efficient in representing anisotropic spherical functions than Spherical Gaussians (SGs). In addition to retaining many desired properties of SGs, ASGs are also rotationally invariant and capable of representing all-frequency signals. To further strengthen the properties of ASGs, we have derived approximate closed-form solutions for their *integral*, *product* and *convolution* operators, whose errors are nearly negligible, as validated by quantitative analysis. Supported by all these operators, ASGs can be adapted in existing SG-based applications to enhance their scalability in handling anisotropic effects. To demonstrate the accuracy and efficiency of ASGs in practice, we have applied ASGs in two important SG-based rendering applications and the experimental results clearly reveal the merits of ASGs.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: Anisotropic Spherical Gaussians, Spherical Gaussians, Anisotropic BRDFs

Links: DL PDF WEB

ACM Reference Format
Xu, K., Sun, W., Dong, Z., Zhao, D., Wu, R., Hu, S. 2013. Anisotropic Spherical Gaussians. ACM Trans. Graph. 32, 6, Article 209 (November 2013), 11 pages. DOI = 10.1145/2508363.2508386 http://doi.acm.org/10.1145/2508363.2508386.

DOI: http://doi.acm.org/10.1145/2508363.2508386

1 Introduction

Effective and compact representation of spherical function is beneficial for computer graphics applications, especially rendering. To achieve real-time rendering of complex reflectances (BRDFs) under real-world environmental lighting, recent approaches [Tsai and Shih 2006; Wang et al. 2009; Iwasaki et al. 2012a] adopt Spherical Gaussians (SGs) to effectively represent lighting, BRDFs or visibility for computing light transport. The reason why SGs have been chosen is due to their several desirable properties: First, SGs have varying support sizes (bandwidths) and are rotationally invariant, and hence it is convenient to use SGs to represent all-frequency signals, such as lighting and BRDF, and to rotate them freely; secondly, SGs have closed-form solutions for *integral*, *product*, and *convolution*, which are fundamental operators for evaluating rendering integrals [Kajiya 1986] and many other applications [Han et al. 2007; de Rousiers et al. 2012].

SGs are *isotropic*, or circularly symmetric around its lobe axis. Hence, to faithfully represent most real-world lightings or BRDFs, which are *anisotropic* to some degree, a mixture model of n scattered SGs, or simply an *SG Mixture*, is usually required and applied. Yet, since the SG mixture basis is not orthogonal, a product of two n-term SG mixtures has complexity $O(n^2)$ [Tsai and Shih 2006]. Therefore, using SGs to represent anisotropic functions always has to compromise between accuracy and performance, which is an intrinsic limitation.

To address this limitation, we present a novel anisotropic Spherical Gaussian (ASG) function based on the Bingham distribution [Bingham 1974], which can represent anisotropic spherical functions defined in arbitrary local frame (Section 3). To represent complex anisotropic functions, similar to SGs, a mixture model of scattered ASGs (*ASG Mixture*) needs to be applied. Due to the anisotropic nature of ASGs, a much smaller number of ASGs are usually enough to faithfully represent anisotropic functions, which leads to improvements in both accuracy and performance. Such an example is given in Fig. 1, where a single ASG is able to accurately render

GPU-based Out-of-Core Many-Lights Rendering

Rui Wang Yuchi Huo Yazhen Yuan Kun Zhou Wei Hua Hujun Bao*

State Key Lab of CAD&CG, Zhejiang University

Figure 1: *Example scenes rendered using our approach on an NVIDIA GTX 680 GPU with 2GB of memory. The left image is a museum scene, which consists of 117.1 million triangles and 32.4 million lights. The total storage sizes of geometry and lights are 14.3 GB and 3.75 GB respectively. The middle image shows an airport scene with two Boeing 777 models that has total 669.3 million (46.3 GB) triangles and 18 million (2.1 GB) lights. The right image is a carnival scene. There are 17.1 million (1.76 GB) triangles and 256 (29.6 GB) million lights. Our method takes 3m55s, 10m15s and 1m22s to shade the museum, the airport and the carnival scenes respectively, and requires an additional 1m20s, 7m25s and 1m14s to build acceleration structures on these lights and geometry.*

Abstract

In this paper, we present a GPU-based out-of-core rendering approach under the many-lights rendering framework. Many-lights rendering is an efficient and scalable rendering framework for a large number of lights. But when the data sizes of lights and geometry are both beyond the in-core memory storage size, the data management of these two out-of-core data becomes critical and challenging. In our approach, we formulate such a data management as a graph traversal optimization problem that first builds out-of-core lights and geometry data into a graph, and then guides shading computations by finding a shortest path to visit all vertices in the graph. Based on the proposed data management, we develop a GPU-based out-of-GPU-core rendering algorithm that manages data between the CPU host memory and the GPU device memory. Two main steps are taken in the algorithm: the out-of-core data preparation to pack data into optimal data layouts for the many-lights rendering, and the out-of-core shading using graph-based data management. We demonstrate our algorithm on scenes with out-of-core detailed geometry and out-of-core lights. Results show that our approach generates complex global illumination effects with increased data access coherence and has one order of magnitude performance gain over the CPU-based approach.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Rendering;

Keywords: out-of-core, global illumination, many-lights, GPU

Links: DL PDF WEB

*e-mail:bao@cad.zju.edu.cn

ACM Reference Format
Wang, R., Huo, Y., Yuan, Y., Zhou, K., Hua, W., Bao, H. 2013. GPU-based Out-of-Core Many-Lights Rendering. ACM Trans. Graph. 32, 6, Article 210 (November 2013), 10 pages. DOI = 10.1145/2508363.2508413 http://doi.acm.org/10.1145/2508363.2508413.

0730-0301/13/11-ART210 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508413

1 Introduction

Global illumination effects, such as soft shadows and interreflections, greatly affect the quality of computer synthesized images. These effects provide visual cues that enhance realism. Over the past three decades, many methods have been developed to achieve realistic, high-fidelity renderings: radiosity, ray-tracing, and many-lights approaches are a few. When applying these methods to extremely large scenes that do not fit in memory, the data management strategy arises as a critical problem. When these rendering methods run on modern graphics hardware, the problem is worse due to the limit of onboard GPU memory, and the data transfer overhead from the CPU to the GPU.

In this paper, we focus on extending the many-lights rendering framework to handle massive scenes with out-of-core geometry and complex lighting. In this framework, scenes with diffuse and low-gloss materials can be rendered by approximating direct and indirect illumination from a large number of virtual point lights (VPLs) [Keller 1997]. By viewing the integration of these lights as surface sample-light interactions of the light transport matrix, the rendering is formulated into a matrix sampling problem [Walter et al. 2005; Walter et al. 2006; Hašan et al. 2007], where elements of the matrix are first sampled to obtain representative lights and then these representative lights are integrated to samples (pixels). These two steps require different kinds of data: lights and geometry. At first, samples and lights are utilized to obtain representative lights and then geometry data are required in evaluating visibilities. Such a data access pattern is different from those in previous out-of-core rendering approaches, e.g., visualization or ray-tracing, which usually only need to manage out-of-core geometry data. When both of lights and geometry become too large to be stored in the memory, the data management of these two kinds of data in many-lights rendering may bring conflicts. Our results show that a simple or naive data management in out-of-core many-lights rendering will dominate the rendering time and result in one magnitude slower performance. Thus, a new and dedicated data management is necessitated for out-of-core many-lights rendering to not only handle the geometry but also lights .

To reduce the overall I/O overhead, we formulate the data man-

Linear Efficient Antialiased Displacement and Reflectance Mapping

Jonathan Dupuy[1,3 †] Eric Heitz[2,1 †] Jean-Claude Iehl[3] Pierre Poulin[1] Fabrice Neyret[2] Victor Ostromoukhov[3 *]
[1]LIGUM, Dept. I.R.O., Université de Montréal [2]INRIA/LJK [3]LIRIS, Université Lyon 1
†Jonathan Dupuy and Eric Heitz are joint first authors

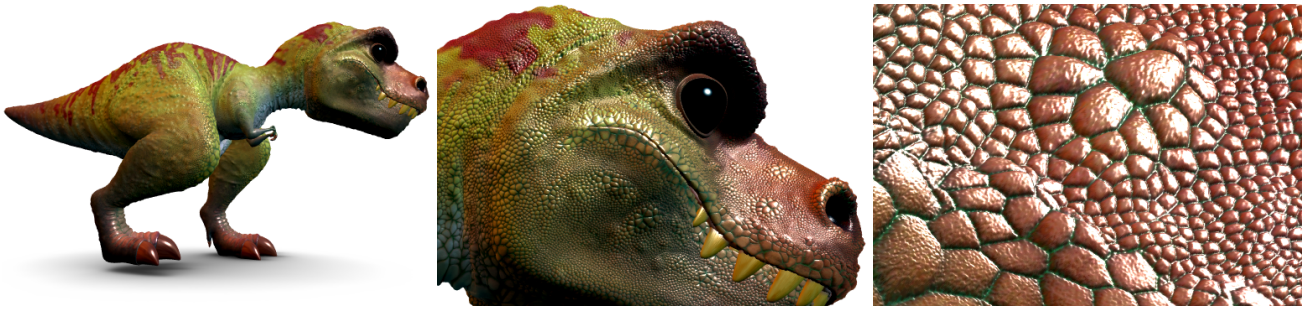


Figure 1: *A high-quality animated production model (Ptex T-rex model © Walt Disney Animation Studios.) rendered in real time under directional and environment lighting using LEADR mapping on an NVidia GTX 480 GPU. The surface appearance is preserved at all scales, using a single shading sample per pixel. Combined with adaptive GPU tessellation, our method provides the fastest, seamless, and antialiased progressive representation for displaced surfaces.*

Abstract

We present Linear Efficient Antialiased Displacement and Reflectance (LEADR) mapping, a reflectance filtering technique for displacement mapped surfaces. Similarly to LEAN mapping, it employs two mipmapped texture maps, which store the first two moments of the displacement gradients. During rendering, the projection of this data over a pixel is used to compute a noncentered anisotropic Beckmann distribution using only simple, linear filtering operations. The distribution is then injected in a new, physically based, rough surface microfacet BRDF model, that includes masking and shadowing effects for both diffuse and specular reflection under directional, point, and environment lighting. Furthermore, our method is compatible with animation and deformation, making it extremely general and flexible. Combined with an adaptive meshing scheme, LEADR mapping provides the very first seamless and hardware-accelerated multi-resolution representation for surfaces. In order to demonstrate its effectiveness, we render highly detailed production models in real time on a commodity GPU, with quality matching supersampled ground-truth images.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: BRDF, microfacet, filtering, LEAN mapping, GPU

Links: DL PDF

* {jdupuy | jciehl | victor.ostromoukhov}@liris.cnrs.fr
{eric.heitz | fabrice.neyret}@inria.fr
poulin@iro.umontreal.ca

ACM Reference Format
Dupuy, J., Heitz, E., Iehl, J., Poulin, P., Neyret, F., Ostromoukhov, V. 2013. Linear Efficient Antialiased Displacement and Reflectance Mapping. ACM Trans. Graph. 32, 6, Article 211 (November 2013), 11 pages. DOI = 10.1145/2508363.2508422 http://doi.acm.org/10.1145/2508363.2508422.

DOI: http://doi.acm.org/10.1145/2508363.2508422

1 Introduction

Rendering applications such as video games commonly employ bump or normal textures (henceforth interchangeably referred to as normal textures) to enhance surface appearance (e.g., [Kilgard 2000]). These textures perturb or modify the normal of a simple underlying surface to emulate geometric variations through shading perturbations. Similarly to albedo textures, normal textures must be filtered for antialiasing purposes. But since shading with linearly filtered normals does not result in proper reflectance filtering, these textures cannot exclusively rely on simple methods such as mipmapping. Recently introduced by Olano and Baker [2010], LEAN mapping is an elegant solution to this problem, that has found widespread adoption because of its effectiveness, efficiency, and simplicity [Baker 2011].

Normal mapping is an inherited paradigm from the 1980's. At that time, geometric models were coarse because of computing capabilities and memory constraints. Textures cheaply enhanced visual details without increasing geometric complexity. The discrepancy between the resolutions of geometry and texture was the key assumption for texture filtering: within the same large and flat triangle, visibility, curvature, and orientation were considered constant. Consequently, filtering in texture space or in geometry space could reasonably be considered equivalent. This assumption does not hold anymore because mesh and texture resolutions can be matched with negligible overhead on modern GPUs.

Filtering appearance of small-scale geometry is thus a critical emerging problem. Indeed, small-scale geometry produces view- and light-dependent effects that include *masking*, *shadowing*, and *projection weighting* (i.e., the cosine term). Filtering this small-scale geometry while neglecting these visual effects violates energy conservation and can result in objectionable aliasing, popping artifacts, and inconsistent appearances throughout scales. Methods for filtering normal maps do not account for these effects, which is why filtering reflectance from normal mapping is not the same problem as filtering reflectance of small-scale geometry [Han et al. 2007; Bruneton and Neyret 2012] (see Figure 2).

We propose a solution to this problem in the important case where the small-scale geometry is generated by *displacement mapping*.

Modeling and Estimation of Internal Friction in Cloth

Eder Miguel, URJC Madrid
Rasmus Tamstorf, Disney Animation Studios
Derek Bradley, Disney Research Zurich
Sara C. Schvartzman, URJC Madrid
Bernhard Thomaszewski, Disney Research Zurich
Bernd Bickel, Disney Research Zurich
Wojciech Matusik, MIT CSAIL
Steve Marschner, Cornell University
Miguel A. Otaduy, URJC Madrid

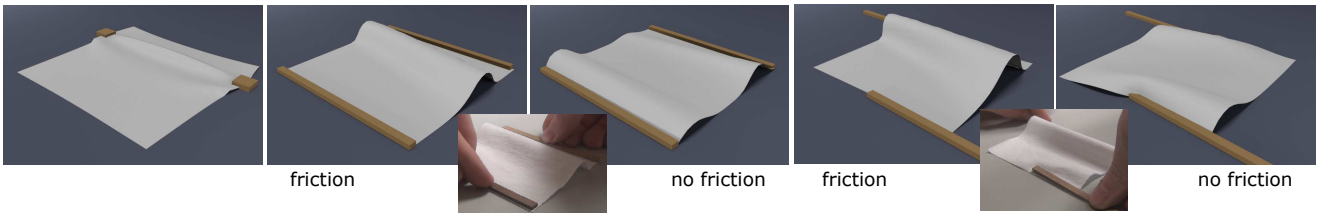


Figure 1: *We produce a wrinkle on a piece of cotton (first from left). With internal friction (second and fourth), it becomes a 'preferred wrinkle' and arises repeatedly under diverse deformations. Without internal friction (third and fifth), folds and wrinkles show no clear similarity. The insets show real-world deformations for the same experiment.*

Abstract

Force-deformation measurements of cloth exhibit significant hysteresis, and many researchers have identified internal friction as the source of this effect. However, it has not been incorporated into computer animation models of cloth. In this paper, we propose a model of internal friction based on an augmented reparameterization of Dahl's model, and we show that this model provides a good match to several important features of cloth hysteresis even with a minimal set of parameters. We also propose novel parameter estimation procedures that are based on simple and inexpensive setups and need only sparse data, as opposed to the complex hardware and dense data acquisition of previous methods. Finally, we provide an algorithm for the efficient simulation of internal friction, and we demonstrate it on simulation examples that show disparate behavior with and without internal friction.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: cloth simulation, friction, hysteresis

Links: DL PDF

1 Introduction

Clothing is a fundamental aspect of our expressiveness, hence computer animation research has put a lot of effort toward realism and efficiency in cloth simulation [Terzopoulos et al. 1987; Volino et al. 1995; Baraff and Witkin 1998; Bridson et al. 2002]. Computer graphics approaches to cloth simulation build on various types of elasticity models [Choi and Ko 2002; Etzmuß et al. 2003; Grinspun et al. 2003; Volino et al. 2009; Thomaszewski et al. 2009], sometimes with the addition of plasticity under large deformations [Bergou et al. 2007]. However, several works in mechanical engineering and computer graphics point out the existence of significant hysteresis between load and unload paths in force-deformation measurements with real cloth, and refer to internal friction, currently missing in computer graphics models, as the phenomenon inducing such hysteresis [Lahey 2002; Ngo-Ngoc and Boivin 2004; Williams 2010].

Ignoring internal friction when estimating elasticity parameters under existing procedures may lead to bias [Volino et al. 2009; Wang et al. 2011; Miguel et al. 2012]. Even if the elastic parameters are chosen to fit the average of loading and unloading behaviors, given observed hysteresis as high as 50% of the average force, ignoring internal friction may induce deformation errors of up to $\pm 25\%$ for a given load. Internal friction also plays a central role in the formation and dynamics of cloth wrinkles. We have observed that internal friction may induce the formation of 'preferred' wrinkles and folds as shown in Fig. 1, persistent deformations as shown in Fig. 14, and history-dependent wrinkles as shown in Fig. 12, and may also make folds and wrinkles settle faster as shown in Fig. 13.

Plasticity models appear to be an alternative way to capture persistent deformations due to large applied loads. However, in their popular form they exhibit no hysteresis under moderate deformations, and they require large reverse loads to eliminate persistent deformations. By contrast, internal friction produces local hysteresis even for small offset deformations, and it requires minor load oscillations to eliminate persistent deformations. These differences make friction more suitable as a source of 'locally persistent wrinkles'. Plasticity and internal friction are two different phenomena, both of which can occur, and we are studying friction because it is a good explanation of the hysteresis commonly seen in cloth measurements made at moderate strains.

We propose to model internal friction in cloth using Dahl's model [Dahl 1968]. Others have already identified Dahl's model as a good match to hysteresis behavior in cloth, including the lack of significant static friction [Lahey 2002; Ngo-Ngoc and Boivin 2004]. However, as we will discuss in Section 3, we also identify features that are not well captured by a standard Dahl model, and propose an augmented reparameterization of the model that includes a strain-dependent definition of hysteresis.

ACM Reference Format
Miguel, E., Tamstorf, R., Bradley, D., Schvartzman, S., Thomaszewski, B., Bickel, B., Matusik, W., Marschner, S., Otaduy, M. 2013. Modeling and Estimation of Internal Friction in Cloth. ACM Trans. Graph. 32, 6, Article 212 (November 2013), 10 pages. DOI = 10.1145/2508363.2508389 http://doi.acm.org/10.1145/2508363.2508389.

0730-0301/13/11-ART212 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508389

An Efficient Construction of Reduced Deformable Objects

Christoph von Tycowicz[1] Christian Schulz[2] Hans-Peter Seidel[2] Klaus Hildebrandt[2]
[1]Freie Universität Berlin [2]Max-Planck-Institut für Informatik

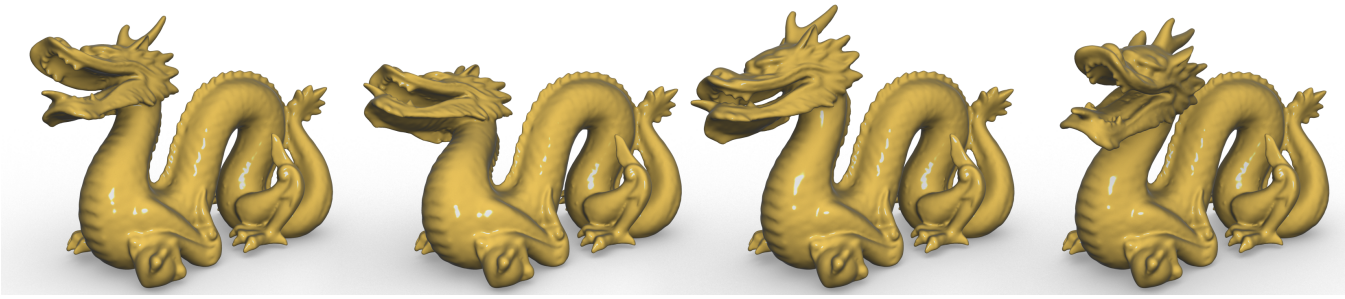


Figure 1: *Nonlinear simulation of a deformable object with 92 k tets computed at over 120 Hz after about 4 mins of preprocessing.*

Abstract

Many efficient computational methods for physical simulation are based on model reduction. We propose new model reduction techniques for the *approximation of reduced forces* and for the *construction of reduced shape spaces of deformable objects* that accelerate the construction of a reduced dynamical system, increase the accuracy of the approximation, and simplify the implementation of model reduction. Based on the techniques, we introduce schemes for real-time simulation of deformable objects and interactive deformation-based editing of triangle or tet meshes. We demonstrate the effectiveness of the new techniques in different experiments with elastic solids and shells and compare them to alternative approaches.

CR Categories: I.6.8 [Simulation and Modeling]: Types of Simulation—Animation I.3.5 [Computer Graphics]: Computation Geometry and Object Modeling—Physically-based modeling;

Keywords: real-time simulation; geometric modeling; model reduction; subset selection; modal analysis; modal derivatives

Links: DL PDF

1 Introduction

Methods for real-time simulation of deformable objects based on model reduction have received much attention in recent years. These schemes construct a low-dimensional approximation of the dynamical system underlying a simulation and thereby achieve a runtime that depends only on the complexity of the low-dimensional system. We focus on two problems of reduced nonlinear simulation: the *subspace construction* and the *efficient approximation of the reduced forces*. We propose new techniques for both problems aiming at accelerating the construction of the approximate dynamical system, increasing the accuracy of the approximation, and simplifying their implementation. Based on this, we implement schemes for real-time simulation of deformable objects and for deformation-based editing of triangular or tetrahedral meshes. Beyond these two applications, the developed techniques are potentially useful for other applications including the acceleration of large simulations and the reduction of constrained spacetime optimization problems, *e.g.* for motion design.

Approximation of reduced forces In addition to dimension reduction, the real-time simulations of deformable objects require a scheme for efficiently approximating the nonlinear reduced forces. The force approximation we consider here, follows the *optimized cubature* introduced by An et al. [2008]. Typically interior forces of a discrete deformable object can be written as a sum whose summands depend only on the deformation of a local neighborhood of the object, *e.g.* a triangle or a tet. The idea is to exploit the correlations between these summands. The dimension reduction restricts the system to a small number of degrees of freedom, which in turn strengthens the correlations. The strategy is to select a small number of summands and to approximate the reduced forces by a linear combination of these summands. The subset and weights are determined through an optimization procedure in which the approximation error on an automatically generated set of training poses is minimized. This is a constrained best subset selection problem. In this paper, we devise a new scheme for efficiently solving this problem, which is based on recent advances in the field of sparse approximation. Our strategy for solving the subset selection problem is substantially different from that used in [An et al. 2008]. They use a greedy strategy that iteratively constructs the selection set by successively adding one entity per iteration. In contrast, our scheme constructs a complete selection set in the first iteration and the whole selection set can be changed in subsequent iterations. We demonstrate in a number of examples that our scheme can produce a significantly smaller approximation error at lower computational costs and is able to achieve a given training error with a smaller selection set.

Subspace construction Subspace construction based on linear modal analysis has become standard practice for the dimension reduction of linear second-order dynamical systems. However, for

ACM Reference Format
von Tycowicz, C., Schulz, C., Seidel, H., Hildebrandt, K. 2013. An Efficient Construction of Reduced Deformable Objects. ACM Trans. Graph. 32, 6, Article 213 (November 2013), 10 pages.
DOI = 10.1145/2508363.2508392 http://doi.acm.org/10.1145/2508363.2508392.

0730-0301/13/11-ART213 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508392

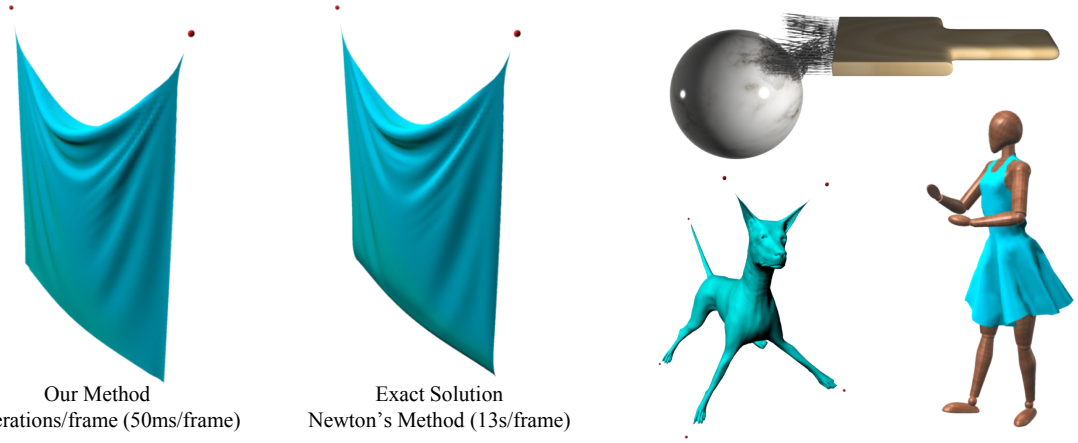
Fast Simulation of Mass-Spring Systems

Tiantian Liu
University of Pennsylvania

Adam W. Bargteil
University of Utah

James F. O'Brien
University of California, Berkeley

Ladislav Kavan*
University of Pennsylvania



Our Method
10 iterations/frame (50ms/frame)

Exact Solution
Newton's Method (13s/frame)

Figure 1: *When used to simulate the motion of a cloth sheet with 6561 vertices our method (left) produces real-time results on a single CPU comparable to those obtained with a much slower off-line method (middle). The method also performs well for one dimensional strands, volumetric objects, and character clothing (right).*

Abstract

We describe a scheme for time integration of mass-spring systems that makes use of a solver based on block coordinate descent. This scheme provides a fast solution for classical linear (Hookean) springs. We express the widely used implicit Euler method as an energy minimization problem and introduce spring directions as auxiliary unknown variables. The system is globally linear in the node positions, and the non-linear terms involving the directions are strictly local. Because the global linear system does not depend on run-time state, the matrix can be pre-factored, allowing for very fast iterations. Our method converges to the same final result as would be obtained by solving the standard form of implicit Euler using Newton's method. Although the asymptotic convergence of Newton's method is faster than ours, the initial ratio of work to error reduction with our method is much faster than Newton's. For real-time visual applications, where speed and stability are more important than precision, we obtain visually acceptable results at a total cost per timestep that is only a fraction of that required for a single Newton iteration. When higher accuracy is required, our algorithm can be used to compute a good starting point for subsequent Newton's iteration.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics—Animation; I.6.8 [Simulation and Modeling]: Types of Simulation—Animation.

*ladislav.kavan@gmail.com

ACM Reference Format
Liu, T., Bargteil, A., O'Brien, J., Kavan, L. 2013. Fast Simulation of Mass-Spring Systems. ACM Trans. Graph. 32, 6, Article 214 (November 2013), 7 pages. DOI = 10.1145/2508363.2508406 http://doi.acm.org/10.1145/2508363.2508406.

0730-0301/13/11-ART214 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508406

Keywords: Time integration, implicit Euler method, mass-spring systems.

Links: DL PDF VIDEO WEB

1 Introduction

Mass-spring systems provide a simple yet practical method for modeling a wide variety of objects, including cloth, hair, and deformable solids. However, as with other methods for modeling elasticity, obtaining realistic material behaviors typically requires constitutive parameters that result in numerically stiff systems. Explicit time integration methods are fast but when applied to these stiff systems they have stability problems and are prone to failure. Traditional methods for implicit integration remain stable but require solving large systems of equations [Baraff and Witkin 1998; Press et al. 2007]. The high cost of solving these systems of equations limits their utility for real-time applications (e.g., games) and slows production work flows in off-line settings (e.g., film and visual effects).

In this paper, we propose a fast implicit solver for standard mass-spring systems with spring forces governed by Hooke's law. We consider the optimization formulation of implicit Euler integration [Martin et al. 2011], where time-stepping is cast as a minimization problem. Our method works well with large timesteps—most of our examples assume a fixed timestep corresponding to the framerate, i.e., $h = 1/30s$. In contrast to the traditional approach of employing Newton's method, we reformulate this minimization problem by introducing auxiliary variables (spring directions). This allows us to apply a block coordinate descent method which alternates between finding optimal spring directions (local step) and finding node positions (global step). In the global step, we solve a linear system. The matrix of our linear system is independent of the current state, which allows us to benefit from a pre-computed sparse Cholesky factorization.

Newton's method is known for its excellent convergence properties. When the iterates are sufficiently close to the optimum, Newton's method exhibits quadratic convergence which out performs block

Simulation and Control of Skeleton-driven Soft Body Characters

Libin Liu* KangKang Yin† Bin Wang† Baining Guo‡

*Tsinghua University †National University of Singapore ‡Microsoft Research Asia

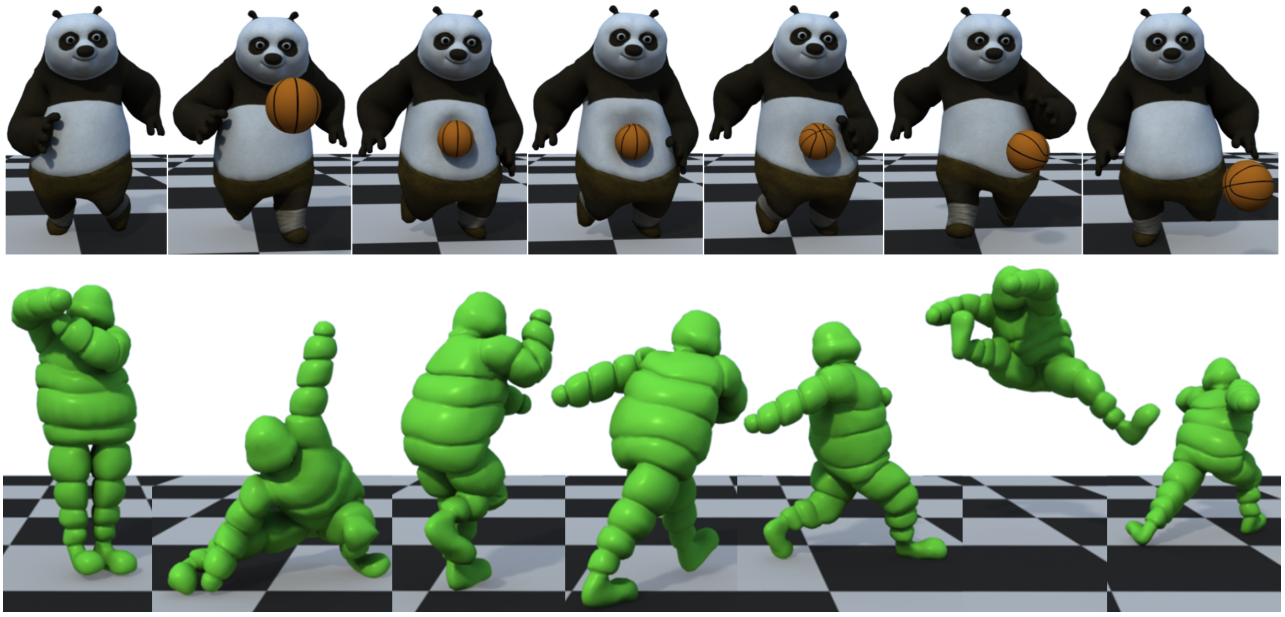


Figure 1: *Our Panda model runs and responds to external perturbations at interactive rates. Our Michelin model does Kung Fu moves.*

Abstract

In this paper we present a physics-based framework for simulation and control of human-like skeleton-driven soft body characters. We couple the skeleton dynamics and the soft body dynamics to enable two-way interactions between the skeleton, the skin geometry, and the environment. We propose a novel pose-based plasticity model that extends the corotated linear elasticity model to achieve large skin deformation around joints. We further reconstruct controls from reference trajectories captured from human subjects by augmenting a sampling-based algorithm. We demonstrate the effectiveness of our framework by results not attainable with a simple combination of previous methods.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

Keywords: physics-based animation, motion control, soft body simulation, finite element method, plasticity model

Links: DL PDF

*e-mail:llb05@mails.tsinghua.edu.cn

†e-mail:{kkyin,wangb}@comp.nus.edu.sg

‡e-mail:bainguo@microsoft.com

ACM Reference Format
Liu, L., Yin, K., Wang, B., Guo, B. 2013. Simulation and Control of Skeleton-driven Soft Body Characters. ACM Trans. Graph. 32, 6, Article 215 (November 2013), 8 pages. DOI = 10.1145/2508363.2508427 http://doi.acm.org/10.1145/2508363.2508427.

DOI: http://doi.acm.org/10.1145/2508363.2508427

1 Introduction

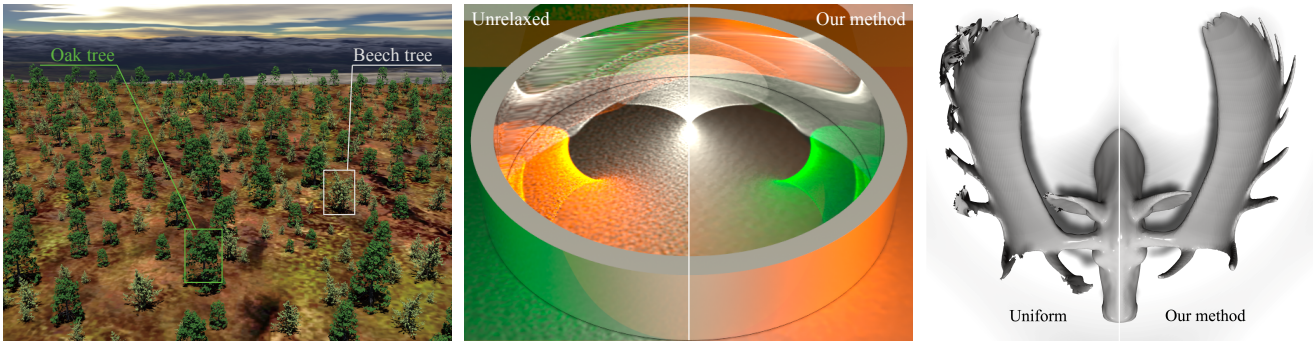
Recent advances in human-like articulated rigid body character control have demonstrated robust results for basic locomotion [Yin et al. 2007; Coros et al. 2010] and highly dynamic motions [Liu et al. 2012; Brown et al. 2013]. The rigidity assumption, however, is not applicable to fat characters, such as a panda with a big belly, whose skin and flesh deformations inevitably affect the underlying skeleton dynamics, especially in highly dynamic tasks. Simulation and control of human-like soft characters is extremely challenging due to coupling between the skeleton and the soft body dynamics, complexity of human skills, large numbers of Degrees of Freedom (DoFs), and large ranges of motion. In the long run, biomechanical approaches that truly model human anatomy are probably needed to completely solve the problem [Lee et al. 2009], but their high modeling and computational costs are prohibitive for graphics applications such as games in the foreseeable future.

We propose an affordable simulation and control framework for human-like soft body characters. Our interactive simulation framework is unique in its ability to conserve momentum, and its simplicity for reimplementation. We use the rigid skeletons for motion control and the surface geometries for deformation, and couple them properly to support two-way interactions. The rich literature on both topics provides basic building blocks for our system. However, a simple integration of prior arts does not work directly. First, joints of human-like characters have large ranges of motion and can rotate more than 90 degrees for example. The flesh around these joints thus experiences large deformations and exerts excessive elastic forces onto the bones when joints bend severely. Since the flesh and the bones are tow-way coupled, these forces will prevent the bones from rotating to their target positions. This is not an issue for simple models such as fish demonstrated in previous work. But for human-like characters we need to properly address

Bilateral Blue Noise Sampling

Jiating Chen[‡*] Xiaoyin Ge[§*] Li-Yi Wei[¶‖] Bin Wang[‡†] Yusu Wang[§] Huamin Wang[§] Yun Fei[‡]
Kang-Lai Qian[‡] Jun-Hai Yong[‡] Wenping Wang[¶]

[‡] Tsinghua University [§] The Ohio State University [¶] The University of Hong Kong [‖] Microsoft Research



(a) *biological distribution* **(b)** *photon density estimation* **(c)** *point cloud sampling + reconstruction*

Figure 1: *Analogous to bilateral filtering [Tomasi and Manduchi 1998], our bilateral sampling method considers both spatial-domain and non-spatial-domain properties. Our method can generate distributions with sample attributes that are not direct functions of the underlying domains, such as more natural biological distribution with domain-independent tree type and size (a), less noisy photon density estimation with arbitrary flux and incoming direction (b), and more accurate (hidden) surface reconstruction from point cloud sampling (c).*

Abstract

Blue noise sampling is an important component in many graphics applications, but existing techniques consider mainly the spatial positions of samples, making them less effective when handling problems with non-spatial features. Examples include biological distribution in which plant spacing is influenced by non-positional factors such as tree type and size, photon mapping in which photon flux and direction are not a direct function of the attached surface, and point cloud sampling in which the underlying surface is unknown a priori. These scenarios can benefit from blue noise sample distributions, but cannot be adequately handled by prior art.

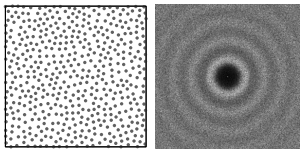
Inspired by bilateral filtering, we propose a bilateral blue noise sampling strategy. Our key idea is a general formulation to modulate the traditional sample distance measures, which are determined by sample position in spatial domain, with a similarity measure that considers arbitrary per sample attributes. This modulation leads to the notion of *bilateral* blue noise whose properties are influenced by not only the uniformity of the sample positions but also the similarity of the sample attributes. We describe how to incorporate our modulation into various sample analysis and synthesis methods, and demonstrate applications in object distribution, photon density estimation, and point cloud sub-sampling.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Antialiasing; I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture;

Keywords: bilateral measure, blue noise, stochastic sampling, point cloud, object distribution, photon mapping

Links: DL PDF

1 Introduction



(a) *samples* **(b)** *spectrum*

Blue noise refers to sample sets that have random and yet uniform distributions. Due to the spatial uniformity and absence of spectral bias, blue noise has found use in a variety of graphics applications such as rendering [Cook 1986; Sun et al. 2013], halftoning [Ulichney 1987; Chang et al. 2009], stippling [Kopf et al. 2006; Balzer et al. 2009; Fattal 2011], animation [Schechter and Bridson 2012], modeling [Lipman et al. 2007; Bowers et al. 2010], and visualization [Li et al. 2010].

The majority of these methods produce samples whose properties are entirely dependent upon the underlying spatial domain. In particular, the position of each sample determines its spacing with other samples, such as stippling an image. However, many common scenarios do not obey this presumption, such as:

Biological distribution [Pommerening 2002; Illian et al. 2008; Pirk et al. 2012], in which plants often exhibit noise-like arrangements and the pair-wise spacing depends on not only plants' positions but also their own attributes that are not directly dependent on positions, such as heights, trunk diameters, and branch spreads.

Photon mapping [Jensen 1996], in which rendering quality depends on the uniformity of photons with similar attributes (such as flux and incoming direction) that are function of the entire scene instead of just photon spatial locations over the attached surface.

Point cloud sampling [Öztireli et al. 2010], in which the goal is

*Jiating Chen and Xiaoyin Ge are joint first authors.
†Bin Wang is the corresponding author.

ACM Reference Format
Chen, J., Ge, X., Wei, L., Wang, B., Wang, Y., Wang, H., Fei, Y., Qian, K., Yong, J., Wang, W. 2013. Bilateral Blue Noise Sampling. ACM Trans. Graph. 32, 6, Article 216 (November 2013), 11 pages.
DOI = 10.1145/2508363.2508375 http://doi.acm.org/10.1145/2508363.2508375.

0730-0301/13/11-ART216 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508375

Halftone QR Codes

Hung-Kuo Chu[1] Chia-Sheng Chang[1] Ruen-Rone Lee[1] Niloy J. Mitra[2]

[1]National Tsing Hua University [2]University College London

Figure 1: *Three halftone QR codes generated by our method. By using a new representation model that minimally binds to the appearance of QR code, our approach is able to combine halftone images with ordinary QR codes without compromising its readability.*

Abstract

QR code is a popular form of barcode pattern that is ubiquitously used to tag information to products or for linking advertisements. While, on one hand, it is essential to keep the patterns machine-readable; on the other hand, even small changes to the patterns can easily render them unreadable. Hence, in absence of any computational support, such QR codes appear as random collections of black/white modules, and are often visually unpleasant. We propose an approach to produce high quality visual QR codes, which we call *halftone QR codes*, that are still machine-readable. First, we build a pattern readability function wherein we learn a probability distribution of what modules can be replaced by which other modules. Then, given a text tag, we express the input image in terms of the learned dictionary to encode the source text. We demonstrate that our approach produces high quality results on a range of inputs and under different distortion effects.

CR Categories: I.4.9 [Image Processing and Computer Vision]: Applications;

Keywords: Non-Photorealistic Rendering, Halftone, QR code

Links: DL PDF WEB VIDEO CODE

ACM Reference Format
Chu, H., Chang, C., Lee, R., Mitra, N. 2013. Halftone QR Codes. ACM Trans. Graph. 32, 6, Article 217 (November 2013), 8 pages. DOI = 10.1145/2508363.2508408 http://doi.acm.org/10.1145/2508363.2508408.

DOI: http://doi.acm.org/10.1145/2508363.2508408

1 Introduction

Quick Response Code, abbreviated as QR code®, is a two-dimensional matrix encoding consisting of black and white squares, called *modules*, forming a machine-readable barcode to tag information onto products. Originally designed by Denso Wave for the automotive industry, QR code has quickly been adapted as a fast and effective way to embed digital content and is extensively used in diverse fields including manufacturing, marketing, etc. While being an excellent machine readable format, visually QR code remains a clutter of black and white squares that can easily disrupt the aesthetic appeal of its parent product.

Since QR codes often take up a non-negligible display area, there is a growing demand for producing visually appealing QR codes. Such codes that incorporate high-level visual features such as colors, letters, illustrations, or logos are referred to as *visual QR codes*. However, creating a visually interesting QR code without compromising its readability is non-trivial. The key challenge arises due to the lack of proper understanding or analytical formulations capturing the stability (i.e., validity) of QR codes under variations in lighting, camera specifications, and even perturbations to the QR codes [DENSO WAVE 2003; Winter 2011]. Patented and ill-documented algorithms employed for reading QR codes cause further difficulties. Consequently, existing approaches are mostly ad hoc and often end up favoring readability at the cost of sacrificing visual quality.

A common strategy to generate visual QR codes relies on inbuilt error correcting capabilities of QR codes to restore from missing or corrupted modules (see Figure 2(a)). In absence of suitable analytical or computational support, such approaches involve tedious trial-and-error runs to produce visual QR codes with little or no control over the final quality. As a result, the resolution and quality of the results are strongly dependent on and restricted by the settings used to generate the QR codes. Another heuristic is to modify a module's appearance while keeping its concentric region untouched, and uniformly blending the neighboring regions with the code modules (see Figure 2(c)). However, due to the tight binding to the appearance of QR code, such blending-based approaches

Interactive By-example Design of Artistic Packing Layouts

Bernhard Reinert[1] Tobias Ritschel[1,2] Hans-Peter Seidel[1]
MPI Informatik[1] MMCI / Saarland University[2]

Figure 1: *Starting from a common layout* (Left)*, the user's objective is inferred from placement of three primitives (push pins), leading to a layout organized vertically by size* (Middle) *and after a different placement additionally by brightness horizontally* (Right).

Abstract

We propose an approach to "pack" a set of two-dimensional graphical primitives into a spatial layout that follows artistic goals. We formalize this process as projecting from a high-dimensional feature space into a 2D layout. Our system does not expose the control of this projection to the user in form of sliders or similar interfaces. Instead, we infer the desired layout of all primitives from interactive placement of a small subset of example primitives. To produce a pleasant distribution of primitives with spatial extend, we propose a novel generalization of Centroidal Voronoi Tesselation which equalizes the distances between boundaries of nearby primitives. Compared to previous primitive distribution approaches our GPU implementation achieves both better fidelity and asymptotically higher speed. A user study evaluates the system's usability.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques;

Keywords: layout inference, packing, distribution, user interface

Links: DL PDF WEB VIDEO DATA

1 Introduction

Arranging sets of primitives into a pleasing spatial packing layout that tightly fills in 2D is tedious and requires expert skills (Fig. 2). While arrangements can serve for recreation and aesthetic purposes, they often seek to convey an underlying message concerning the relation between primitives and serve a didactical purpose. In this work, we propose a system to automate artistic layouts, by inferring the user's high-level intentions from the interaction performed. The interactive exploration of different artistic layouts and primitive relations enabled by our system goes beyond static print or display layouts and helps to improve general layouts, such required for Mind maps [Buzan 1976], tag clouds [Bateman et al. 2008] or any arrangement of graphical 2D primitives.

Fig. 1 shows three steps of a typical interaction using our system: After loading a set of primitives, our system presents a general-purpose layout (Fig. 1, *left*). To change this layout, a simple solution would be to expose many sliders that control what importance to what weight quality would be given. Such high-dimensional parameter spaces are hard to navigate for colloquial users and hamper creative exploration. Our system takes a different approach: We offer the user to move primitives to new positions (Fig. 1, *middle*) and by that to infer the user's intention, leading to a new layout, in this case, where primitives are organized vertically by size. After a second manipulation (Fig. 1, *right*) the layout is organized by brightness horizontally and by size vertically.

To allow such operations we make the following contributions:

- An interactive inverse layout approach to infer a user's packing layout intention from a small number of examples.
- A layout algorithm to evenly distribute primitives with spatial extend in real-time using a GPU.
- A study of packing layout task performance of novice users.

2 Previous Work

Properly distributing primitives in a domain has been a challenge in computer graphics for both, technical and aesthetical reasons. Seeking to place samples that evaluate a function such that aliasing is minimized, Mitchell [1987] argued that samples should have "blue noise" characteristics, that is: the distance to the neighbors should not be smaller than a threshold. Placing primitives for artistic purposes in 2D is widely used for non-photorealistic rendering, e. g., for stippling [Deussen et al. 2000; Hiller et al. 2003], mosaics [Hausner 2001; Kim and Pellacini 2002] or texture synthesis [Lagae and Dutré 2005]. In particular Hiller et al. [2003] who distributes primitives in the plane such that they follow a prescribed density, is a similar case of our system that produces distributions that follow rules inferred from the users' interaction with the distribution itself. Placing primitives in the plane, the usage of Voronoi tessellation is popular to avoid collision [Dalal et al. 2006] and achieve pleasant (temporal) distributions.

ACM Reference Format
Reinert, B., Ritschel, T., Seidel, H. 2013. Interactive By-example Design of Artistic Packing Layouts. ACM Trans. Graph. 32, 6, Article 218 (November 2013), 7 pages. DOI = 10.1145/2508363.2508409 http://doi.acm.org/10.1145/2508363.2508409.

DOI: http://doi.acm.org/10.1145/2508363.2508409

Biharmonic Diffusion Curve Images from Boundary Elements

Peter Ilbery*[1], Luke Kendall[1], Cyril Concolato[2], Michael McCosker[1]
[1]Canon Information Systems Research Australia (CiSRA), [2]Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI

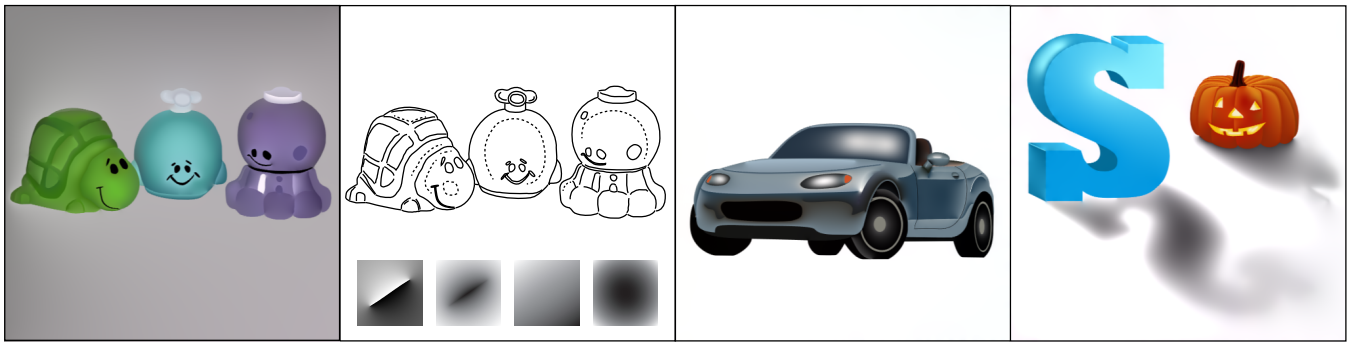


Figure 1: *Boundary element rendering of biharmonic diffusion curve images. From left to right: toys image; sharp-profile (solid) and smooth-profile (dotted) curves of the toys image, with thumbnail images below showing examples of the 4 building block segment fields; car image and pumpkin image. The above images are ©CiSRA; the toys image is a CiSRA re-creation of a photograph taken by C. Concolato.*

Abstract

There is currently significant interest in freeform, curve-based authoring of graphic images. In particular, "diffusion curves" facilitate graphic image creation by allowing an image designer to specify naturalistic images by drawing curves and setting colour values along either side of those curves. Recently, extensions to diffusion curves based on the biharmonic equation have been proposed which provide smooth interpolation through specified colour values and allow image designers to specify colour gradient constraints at curves. We present a Boundary Element Method (BEM) for rendering diffusion curve images with smooth interpolation and gradient constraints, which generates a solved boundary element image representation. The diffusion curve image can be evaluated from the solved representation using a novel and efficient line-by-line approach. We also describe "curve-aware" upsampling, in which a full resolution diffusion curve image can be upsampled from a lower resolution image using formula evaluated corrections near curves. The BEM solved image representation is compact. It therefore offers advantages in scenarios where solved image representations are transmitted to devices for rendering and where PDE solving at the device is undesirable due to time or processing constraints.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms;

Keywords: vector graphics, diffusion curves, boundary elements

Links: DL PDF

*peter.ilbery@cisra.canon.com.au

ACM Reference Format
Ilbery, P., Kendall, L., Concolato, C., McCosker, M. 2013. Biharmonic Diffusion Curve Images from Boundary Elements. ACM Trans. Graph. 32, 6, Article 219 (November 2013), 12 pages.
DOI = 10.1145/2508363.2508426 http://doi.acm.org/10.1145/2508363.2508426.

0730-0301/13/11-ART219 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508426

1 Introduction

Vector graphic images are easily editable and scalable, with compact representations. A key issue for the creation of naturalistic vector graphic images is how an image designer specifies colour variation ("colour gradients") in the image.

Diffusion curves. [Orzan et al. 2008], [Orzan et al. 2013] propose "diffusion curves" for image designers to specify colour variation by drawing curves, setting colour values at sparse points along either side of the curves, and setting "degree of blur" values at sparse points along the curves. A diffusion curves image is generated by an interpolation process followed by a blur process. In the interpolation process, initial image colours are determined at remaining points along curves by 1D interpolation, and then at points away from curves by solving, for each colour component, a partial differential equation (PDE) with boundary constraints provided by the colour values along the curves. In the blur process, degree of blur values are determined throughout the image in a similar fashion to the initial image colours and are then used to apply space-variant blur filtering to transform the initial image colours to final image colours. [Jeschke et al. 2009] describes a fast solver for diffusion curves images; they also describe the interpolation process as solving the 2D Laplace equation with values specified at boundaries, i.e. Dirichlet boundary conditions.

In the absence of blur, diffusion curve images typically have a jump or a gradient discontinuity across the curves; that is, image values have a "sharp-profile" across curves. The diffusion curves blur process may leave the image value profile across a curve unchanged (zero blur) or it may perform some averaging of colours across the curve to form a "smooth-profile". However, the diffusion curves blur process has disadvantages. Typically, at a curve, the blur process does not preserve the average of the left and right colour values an image designer specifies along the curve, including when left and right colours are the same; that is, diffusion curves do not provide smooth interpolation through specified colour values. As well, the blur process has the problem that it can generate colours at a curve having non-zero blur by undesirably drawing colours from both sides, including the far side, of a nearby curve with no blur.

Biharmonic diffusion curves. [Finch et al. 2011] extends the dif-

Near-Eye Light Field Displays

Douglas Lanman David Luebke
NVIDIA Research

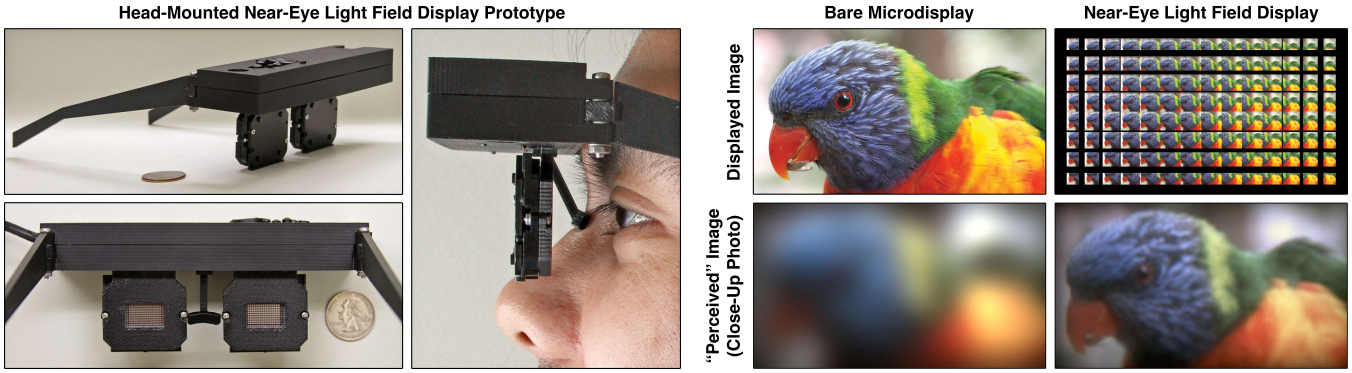


Figure 1: *Enabling thin, lightweight near-eye displays using light field displays. (Left) Our binocular near-eye display prototype is shown, comprising a pair of OLED panels covered with microlens arrays. This design enables a thin head-mounted display, since the black box containing driver electronics could be waist-mounted with longer OLED ribbon cables. (Right) Due to the limited range of human accommodation, a severely defocused image is perceived when a bare microdisplay is held close to the eye. Conventional near-eye displays require bulky magnifying optics to facilitate accommodation. We propose near-eye light field displays as thin, lightweight alternatives, achieving comfortable viewing by synthesizing a light field for a virtual scene located within the accommodation range (here implemented by viewing a microdisplay, depicting interlaced perspectives, through a microlens array). Lorikeet source image courtesy of Robyn Jay.*

Abstract

We propose near-eye light field displays that enable thin, lightweight head-mounted displays (HMDs) capable of presenting nearly correct convergence, accommodation, binocular disparity, and retinal defocus depth cues. Sharp images are depicted by out-of-focus elements by synthesizing light fields corresponding to virtual objects within a viewer's natural accommodation range. We formally assess the capabilities of microlens arrays to achieve practical near-eye light field displays. Building on concepts shared with existing integral imaging displays and light field cameras, we optimize performance in the context of near-eye viewing. We establish fundamental trade-offs between the quantitative parameters of resolution, field of view, and depth of field, as well as the ergonomic parameters of form factor and ranges of allowed eye movement. As with light field cameras, our design supports continuous accommodation of the eye throughout a finite depth of field; as a result, binocular configurations provide a means to address the accommodation-convergence conflict occurring with existing stereoscopic displays. We construct a complete prototype display system, comprising: a custom-fabricated HMD using modified off-the-shelf parts and real-time, GPU-accelerated light field renderers (including a general ray tracing method and a "backward compatible" rasterization method supporting existing stereoscopic content). Through simulations and experiments, we motivate near-eye light field displays as thin, lightweight alternatives to conventional near-eye displays.

Links: DL PDF WEB VIDEO

ACM Reference Format
Lanman, D., Luebke, D. 2013. Near-Eye Light Field Displays. ACM Trans. Graph. 32, 6, Article 220 (November 2013), 10 pages. DOI = 10.1145/2508363.2508366 http://doi.acm.org/10.1145/2508363.2508366.

DOI: http://doi.acm.org/10.1145/2508363.2508366

CR Categories: B.4.2 [Input/Output and Data Communications]: Input/Output Devices—Image Display I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

Keywords: light field displays, head-mounted displays, microlens arrays, accommodation-convergence conflict, virtual reality

1 Introduction

Near-eye displays project images directly into a viewer's eye, encompassing both head-mounted displays (HMDs) and electronic viewfinders. Such displays confront a fundamental problem: the unaided human eye cannot accommodate (focus) on objects placed in close proximity (see Figure 1). As reviewed by Rolland and Hua [2005], a multitude of optical solutions have been proposed since Sutherland [1968] introduced the first graphics-driven HMD. The majority of such designs emulate the behavior of a simple magnifier: synthesizing an enlarged image of a miniaturized display, appearing to be located within the viewer's natural accommodation range. To be of practical utility, a near-eye display should provide high-resolution, wide-field-of-view imagery with compact, comfortable magnifying optics. However, current magnifier designs typically require multiple optical elements to minimize aberrations, leading to bulky eyewear with limited fields of view that have, to date, prohibited widespread consumer adoption.

Conventional displays are intended to emit light isotropically. In contrast, a light field display supports the control of tightly-clustered bundles of light rays, modulating radiance as a function of position and direction across its surface. We consider a simple near-eye architecture: placing a light field display directly in front of a user's eye (or a pair of such displays for binocular viewing). As shown in Figure 1, sharp imagery is depicted by synthesizing a light field for a virtual display (or a general 3D scene) within the viewer's unaided accommodation range. As characterized in this paper, near-eye light field displays provide a means to achieve thin, lightweight HMDs with wide fields of view and to address accommodation-convergence conflict in binocular configurations; however, these benefits come at a cost: spatial resolution is significantly reduced with microlens-based designs, although with com-

Joint View Expansion and Filtering for Automultiscopic 3D Displays

Piotr Didyk Pitchaya Sitthi-Amorn William Freeman Frédo Durand Wojciech Matusik

MIT CSAIL

Figure 1: *The method presented in this paper takes a stream of stereo images as an input and synthesizes additional views required for an automultiscopic display. The output views are also filtered to remove inter-view aliasing. ("Big Buck Bunny" © by Blender Foundation)*

Abstract

Multi-view autostereoscopic displays provide an immersive, glasses-free 3D viewing experience, but they require correctly filtered content from multiple viewpoints. This, however, cannot be easily obtained with current stereoscopic production pipelines. We provide a practical solution that takes a stereoscopic video as an input and converts it to multi-view and filtered video streams that can be used to drive multi-view autostereoscopic displays. The method combines a phase-based video magnification and an interperspective antialiasing into a single filtering process. The whole algorithm is simple and can be efficiently implemented on current GPUs to yield a near real-time performance. Furthermore, the ability to retarget disparity is naturally supported. Our method is robust and works well for challenging video scenes with defocus blur, motion blur, transparent materials, and specularities. We show that our results are superior when compared to the state-of-the-art depth-based rendering methods. Finally, we showcase the method in the context of a real-time 3D videoconferencing system that requires only two cameras.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms,viewing algorithms;

Keywords: automultiscopic displays, view synthesis, interperspective antialiasing

Links: DL PDF WEB VIDEO

ACM Reference Format
Didyk, P., Sitthi-Amorn, P., Freeman, W., Durand, F., Matusik, W. 2013. Joint View Expansion and Filtering for Automultiscopic 3D Displays. ACM Trans. Graph. 32, 6, Article 221 (November 2013), 8 pages.
DOI = 10.1145/2508363.2508376 http://doi.acm.org/10.1145/2508363.2508376.

0730-0301/13/11-ART221 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508376

1 Introduction

Stereoscopic 3D content is becoming more popular as it reaches an increasing number of home users. While most of current TV sets are 3D-enabled, and there is plenty of 3D movies and sports programming available, the adoption of stereoscopic 3D is hampered by the use of 3D glasses required to view the content. Multi-view autostereoscopic (or automultiscopic) displays offer a superior visual experience since they provide both binocular and motion parallax without the use of special glasses. A viewer is not restricted to be in a particular position and many viewers can watch the display at the same time. Furthermore, automultiscopic displays can be manufactured inexpensively, for example, by adding a parallax barrier or a lenticular screen to a standard display.

However, there are three major problems that need to be addressed in order for a multi-view autostereoscopic TV to become a reality. First, current 3D content production pipelines provide only two views, while multi-view stereoscopic displays require images from many viewpoints. Capturing TV-quality scenes with dense camera rigs is impractical because of the size and cost of professional quality cameras. A solution to use view-interpolation to generate these additional views requires an accurate depth and inpainting of missing scene regions. There has been a steady progress in stereo depth reconstruction algorithms, but the quality is not yet good enough for TV broadcast and movies. Handling scenes that include defocus blur, motion blur, transparent materials, and specularities is especially difficult. Second, multi-view autostereoscopic displays require special filtering to remove interperspective aliasing – all image content that is not supported by a given display [Zwicker et al. 2006]. Without performing this step severe ghosting and flickering can be seen. However, in order to properly antialias a multi-view video, a dense light field is necessary. Finally, to assure viewing comfort, image disparities usually have to be modified according to the display type, size, and viewer preference. This disparity retargeting step also requires rerendering the scene with the adjusted disparities.

We propose a method that addresses all these three limitations. Our method takes a stereoscopic stream as an input and produces a correctly filtered multi-view video for a given automultiscopic display as shown in Figure 1. The solution does not require any changes to the current stereoscopic production and content delivery pipelines.

A Metric of Visual Comfort for Stereoscopic Motion

Song-Pei Du[1] Belen Masia[2] Shi-Min Hu[1] Diego Gutierrez[2]
[1] TNList, Tsinghua University [2] Universidad de Zaragoza

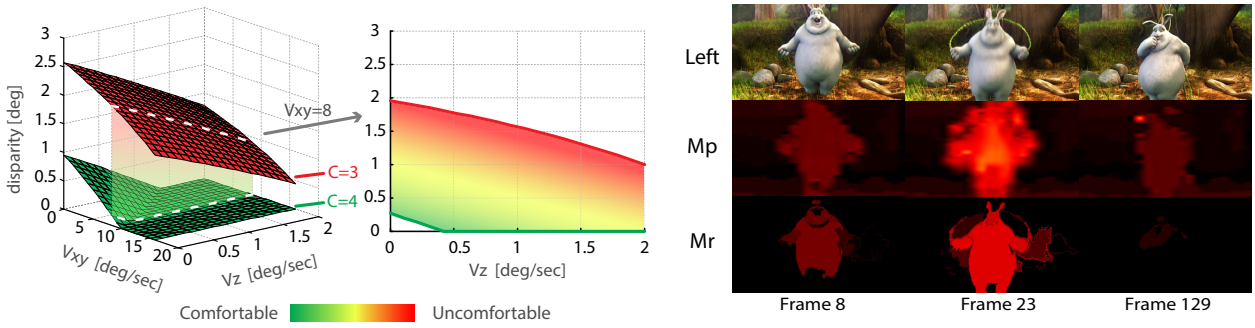


Figure 1: *Our novel metric of visual comfort for stereoscopic content takes into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast. Based on our measured comfort function, we derive a metric to predict the degree of comfort for short stereoscopic videos. Left: Example slice of the comfort zone computed by our comfort function for a spatial frequency of* 1*cpd (bounded by comfort values of 3 and 4). Right: comfort maps computed using our metric on three representative frames of the* bunny *movie (© Blender Foundation). From top to bottom, input frames, per-pixel results, and per-region results (brighter red indicates less comfort). Our metric predicts less comfort with faster movement (frame 23), in agreement with the perceptual experiments.*

Abstract

We propose a novel metric of visual comfort for stereoscopic motion, based on a series of systematic perceptual experiments. We take into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast. We further derive a comfort metric to predict the comfort of short stereoscopic videos. We validate it on both controlled scenes and real videos available on the internet, and show how all the factors we take into account, as well as their interactions, affect viewing comfort. Last, we propose various applications that can benefit from our comfort measurements and metric.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms, viewing algorithms;

Keywords: stereo, motion, visual comfort

Links: DL PDF WEB

1 Introduction

Over the last few years, there has been a renewed interest in stereoscopic displays. Stereoscopic content is generated for movies, games and visualizations for industrial, medical, cultural or educational applications. This has in turn spurred research on aspects of the human visual system that relate to stereo vision [Pollock et al. 2012]. Recent studies analyze the comfort zone for the vergence-accommodation conflict, the influence of luminance on stereo perception, or the depiction of glossy materials, to name just a few [Shibata et al. 2011; Didyk et al. 2012; Templin et al. 2012]. The goal is to understand different aspects of our visual system in order to produce stereo content that guarantees a comfortable viewing experience.

As opposed to natural viewing of the real 3D world, stereoscopic viewing implies conflicting vergence and accommodation cues, which is widely accepted to be a main cause of visual discomfort. However, despite recent advances and the extensive existing literature [Howard and Rogers 2002; Julesz 2006; Didyk et al. 2012; Didyk et al. 2011], some aspects of binocular vision remain largely unexplored. One of the main reasons is the large number of different factors involved, as well as their complex interaction [Cutting and Vishton 1995]. As a consequence, generating stereo content that guarantees a comfortable viewing experience remains a challenging task, often reserved to technicians with a large experience in the field [Lang et al. 2010; Mendiburu 2009].

Thus, one of the goals of stereography is to minimize the discomfort that stereoscopic viewing can cause, and numerous works have been devoted to explaining and characterizing the causes [Kooi and Toet 2004; Lambooij et al. 2009; Shibata et al. 2011]. However, fewer have explored how object *motion* affects this discomfort in stereoscopic viewing. Object motion in stereoscopic movies can in fact be a source of discomfort: Researches and experiments have revealed that visual comfort has a close relationship with some oculomotor functions, including eye movements induced by motion in the scene [Bahill and Stark 1975; Ostberg 1980]. In this work we analyze visual discomfort due to motion in short stereoscopic movies by means of a comprehensive statistical study. Unlike previous work [Yano et al. 2004; Jung et al. 2012], we take into account the interplay of motion velocity both on the screen plane and on the depth axis, as well as *signed* disparity and luminance spatial frequency. Our goal is not only to help understand the phenomena that may lead to visual discomfort; we provide a practical metric to assess existing 3D content as well. This can be used as a guideline for the generation of new stereo content, or to keep navigation pa-

ACM Reference Format
Du, S., Masia, B., Hu, S., Gutierrez, D. 2013. A Metric of Visual Comfort for Stereoscopic Motion. ACM Trans. Graph. 32, 6, Article 222 (November 2013), 9 pages. DOI = 10.1145/2508363.2508387 http://doi.acm.org/10.1145/2508363.2508387.

0730-0301/13/11-ART222 $15.00.
DOI: http://dx.doi.org/10.1145/2508363.2508387

Stereoscopizing Cel Animations

Xueting Liu Xiangyu Mao Xuan Yang Linling Zhang Tien-Tsin Wong

The Chinese University of Hong Kong*

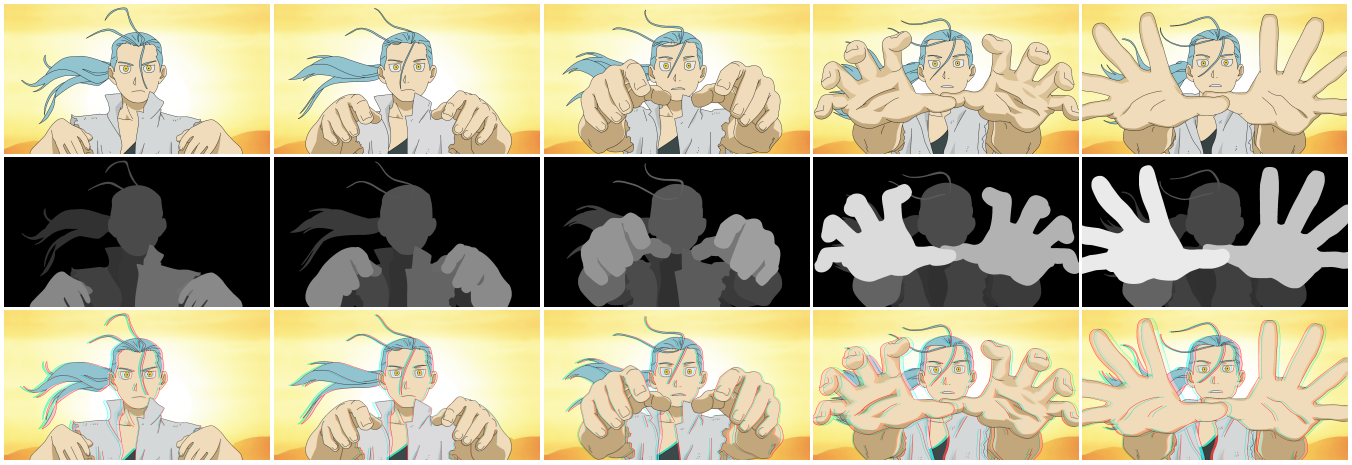


Figure 1: *Stereoscopization of a cel animation. Our method takes an ordinary 2D cel animation (top row) as input, infers the temporal-consistent ordering, and synthesizes the per-frame depth maps (middle row), in order to generate a stereoscopic cel animation (bottom row, presented in the form of anaglyphs). This sequence has 12 frames (*1920×1080*). The frame containing the maximal number of regions has 82 regions. In our experiment, depth ordering takes 12 minutes, and depth synthesis takes 9.6 minutes.*

Abstract

While hand-drawn cel animation is a world-wide popular form of art and entertainment, introducing stereoscopic effect into it remains difficult and costly, due to the lack of physical clues. In this paper, we propose a method to synthesize convincing stereoscopic cel animations from ordinary 2D inputs, without labor-intensive manual depth assignment nor 3D geometry reconstruction. It is mainly automatic due to the need of producing lengthy animation sequences, but with the option of allowing users to adjust or constrain all intermediate results. The system fits nicely into the existing production flow of cel animation. By utilizing the T-junction cue available in cartoons, we first infer the initial, but not reliable, ordering of regions. One of our major contributions is to resolve the temporal inconsistency of ordering by formulating it as a graph-cut problem. However, the resultant ordering remains insufficient for generating convincing stereoscopic effect, as ordering cannot be directly used for depth assignment due to its discontinuous nature. We further propose to synthesize the depth through an optimization process with the ordering formulated as constraints. This is our second major contribution. The optimized result is the spatio-temporally smooth depth for synthesizing stereoscopic effect. Our method has been evaluated on a wide range of cel animations and convincing stereoscopic effect is obtained in all cases.

CR Categories: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Depth cues; J.5 [Computer Application]: Arts and Humanities—Fine arts;

Keywords: Stereopsis, 3D cartoon, and T-junction.

*e-mail:{xtliu,xymao,xyang,llzhang,ttwong}@cse.cuhk.edu.hk

Links: DL PDF

1 Introduction

Traditional cel animation is produced with each frame being drawn manually on celluloids or via computer tablets, and remains a widely used approach (Fig. 2). Unfortunately, it is extremely difficult to introduce stereoscopic effect into cel animations. To our best knowledge, there is only a scarce number of stereoscopic cel animations produced so far. Unlike live-action movies that can be captured with a stereo camera and 3D computer animations that can be computer rendered (e.g. Toy Story and Cyborg 009 [Production I.G. et al. 2012]), hand-drawn cartoons contain no physically correct depth to estimate nor 3D geometrical information to exploit. In fact, frames drawn by cel animators usually contain physically incorrect objects or shapes to maintain aesthetics and style [Rademacher 1999]. Training animators to manually draw stereoscopic pairs of frames is almost infeasible.

As cels may be physically incorrect, 3D geometry reconstruction of the hand-drawn scenes becomes infeasible. Besides, the transition between adjacent frames is typically much larger than that of the live-action videos, thus existing pose estimation and feature track-

ACM Reference Format
Liu, X., Mao, X., Yang, X., Zhang, L., Wong, T. 2013. Stereoscopizing Cel Animations.
ACM Trans. Graph. 32, 6, Article 223 (November 2013), 10 pages. DOI = 10.1145/2508363.2508396
http://doi.acm.org/10.1145/2508363.2508396.

DOI: http://doi.acm.org/10.1145/2508363.2508396