



SIGGRAPH2015
Xroads of Discovery

The 42nd International Conference and Exhibition
on Computer Graphics and Interactive Techniques



User-Centric Computational Videography

Christian Richardt

Intel VCI, MPI Informatik

James Tompkin

Harvard University

Jiamin Bai

Light, Co.

Christian Theobalt

MPI Informatik

Show this slide before the course.

Welcome to our course on User-Centric Computational Videography, or “what users really want to get out of video”.

My name is Christian Richardt, and I am a postdoctoral researcher at the Intel Visual Computing Institute and Max Planck Institute for Informatics in Saarbrücken, Germany.

My co-presenters today are James Tompkin from Harvard University, Jiamin Bai from Light and Christian Theobalt from the Max Planck Institute for Informatics.

This is the first SIGGRAPH course on video processing in six years, so in the next three hours and a bit, we’ll be covering a fairly broad spectrum of techniques from the last decade to show you the progress made in this field.

The course website is:

http://gvv.mpi-inf.mpg.de/teaching/uccv_course_2015/

Veni, vidi, video, vimeo

(I came, I saw, I captured, I uploaded)



VISUAL
COMPUTING
INSTITUTE

mpi
max planck institut
informatik

- Digital video is largely ubiquitous
 - Virtually every mobile device has at least one video camera
 - Vast online video databases: weeks of videos uploaded per minute
- But video is harder than photos:
 - Video is a dimension trickier
 - People are generally bad at video capture
 - Bad video is worse than bad images
 - More difficult to edit

Consumers need better tools to author/edit/browse videos

2015-08-13

Christian Richardt – User-Centric Computational Videography

2

Let me start by adapting one of Julius Caesar's most famous quotes to the current day: veni, vidi, video, vimeo – I came, I saw, I captured and I uploaded.

What this is hinting at is that digital video is becoming more and more ubiquitous, just like digital photos have become over the last few years.

Most mobile devices now come with at least one high-resolution video camera, and fast mobile internet enables the upload of weeks of video to online video communities such as YouTube, Vine and Instagram, every single minute.

However,

- Video is harder than photos, mostly because of the added time dimension that multiplies the amount of data that needs to be processed.
- Video is harder to capture than photos, as handheld mobile devices often produce shaky and wobbly footage, sometimes with poor lighting and afflicted by Vertical Video Syndrome.
- Video is harder to get right than photos. While a bad photo can be pretty bad, a bad video can be so much more uncomfortable to look at than bad photos.
- And video is harder to edit than photos despite a huge range of consumer and professional video editing tools.

To us, it is clear that consumers clearly need better tools for authoring, editing and browsing videos. In this course, we look at the progress made so far on this topic, and discuss current trends in the software industry as well as in research.

User-Centric Computational Videography

- = all aspects of consumer video capture, editing and viewing
- Users have high-level goals, but generally not much skill or time
 - Techniques need to produce good results as automatically as possible
 - Huge gap of skill and effort between casual and professional users
 - Offer some optional possibilities for user input
- Goal: improve the quality and flexibility of:
 - Capture / recording
 - Editing / authoring
 - Viewing / exploring

2015-08-13

Christian Richardt – User-Centric Computational Videography

3

This is what we understand by the term “User-Centric Computational Videography”.

It covers all aspects of handling consumer videos, from their capture with handheld digital video cameras, over video editing and authoring, all the way to how to make large collections of videos easily explorable.

Users often have some high-level goals, such as creating a high-quality holiday video, but may lack the skill and time needed to achieve these goals manually with existing tools.

Casual users prefer techniques that produce good results as automatically as possible, so that they don't need to provide large amounts of manual input.

Professional users, on the other hand, are more happy to provide input, which also often leads to better results than produced by automatic approaches.

This huge gap of skill and effort between casual and professional users means that techniques for casual users need to compensate for this lack of user input to achieve satisfactory results.

If the automatic approach fails, providing a small amount of user input may be all that is required to achieve the desired goal.

We want to help users accomplish their goals, which requires improvements in the quality and flexibility of video capturing and recording, video editing and authoring, and video viewing and exploration techniques.

User wishlist

- Capture / recording:
 - Stabilising shaky, poorly framed videos
 - Denoising, deflickering, deblurring, HDR, colour grading
- Editing / authoring:
 - Combining multiple video clips into one
 - Editing, adding and removing objects and environments
- Viewing / exploring:
 - Exploring and navigating online community video collections
 - Visualising the spatial or temporal overlaps between videos

Let's look at what users want in terms of each of these activities.

In terms of video capture, the most important task for software tools is to get the most out of the recorded footage, to stabilise shaky and wobbly videos, improve the framing of videos if possible, denoise, deflicker and deblur them, and perhaps reconstruct high-dynamic-range video, or fix the captured colours.

In our course, we consider these things as necessary preprocessing tasks. This means we will only cover them briefly to provide some background, and we will mostly focus on the other two areas.

Video editing and authoring covers all aspects of creating a new video, usually by combining multiple video clips, photos and music.

The long-term goal of users is full flexibility in their edits: they want to edit objects directly by cutting and pasting them in a different place and perhaps time, and they want to change the look of videos by modifying the environments and lighting in existing videos.

Increasingly important becomes the task of how videos can be viewed and explored.

Online video communities comprise millions of videos, but the only way to search them is using textual keywords, instead of visualising the spatial or temporal overlaps between videos.

Status quo

- Capture/recording:
 - Video stabilization using computer vision techniques
 - Video denoising + deblurring using computational photography
- Editing/authoring:
 - More robust correspondences within and between videos, e.g. optical flow, SIFT flow, feature matching, PatchMatch
- Viewing/exploring:
 - Improved tools for content exploration using video-based rendering
 - Structuring media collections using semi-automatic clustering

So where are we as a community relative to these goals?

Video capture has become more flexible as more robust video stabilisation techniques have found their way into consumer devices, and computational photography techniques for denoising and deblurring videos will probably soon follow them.

Video editing has benefited from improved and more robust correspondence techniques that are applied within and between videos, which allow the exploitation of content relationships for new empowering video experiences.

And video exploring stands to be revolutionized by novel tools for interactive content exploration that are based on automatically structured media collections.

Goals in a Nutshell

Minimal interaction

High quality

Intuitive

Fast



Video as a
creative medium
for everyone.

In a nutshell, users want video editing and processing tools that require minimal user interaction, produce high-quality results, are intuitive to use and fast.

Techniques that satisfy all of these goals will unlock video as a creative medium **for everyone**.

This is the central message of our course.

Course syllabus

- Introduction and welcome
[Richardt, 0:10]
- Background:
State-of-the-art video tools
[Richardt, 0:20]
- Timeline editing
[Bai, 0:20]
- Editing using lightweight models
[Theobalt, 0:20]
- Model-based video editing
[Theobalt, 0:20]
- Break
[0:15]
- Spatiotemporal video editing
and processing
[Richardt, 0:20]
- Motion editing in videos:
cinemagraphs & cliplets
[Bai, 0:20]
- Exploring videos
[Tompkin, 0:20]
- Exploring videos in contexts
[Tompkin, 0:20]
- Closing and Q & A
[all, 0:10]

2015-08-13

Christian Richardt – User-Centric Computational Videography

7

Over the next 3 hours, we are covering the following topics.

We start with a look at existing state-of-the-art video editing tools, what functionality they provide and what is missing.

Jiamin then dives into timeline editing, which is concerned with the temporal arrangement of videos, their synchronisation, alignment and summarisation.

After that, Christian Theobalt discusses model-based video editing techniques with lightweight and more involved models, that range from video cut-and-paste to inverse rendering for manipulation of 3D objects and environments.

We'll then have a short break of 15 minutes, to get up and stretch a bit, before I continue with spatiotemporal video editing and processing. This part covers advanced video transitions, motion visualisation and magnification, hyperlapses and more.

Next, Jiamin will talk about motion editing in videos in the form of cinemagraphs and cliplets.

After that, James will give an overview of existing video exploration tools that allow for browsing videos or collections of videos by themselves, or within the context of other videos.

There will be time for some questions after each part, and also at the very end of the course.

Course syllabus

- Introduction and welcome
[Richardt, 0:10]
- **Background:**
State-of-the-art video tools
[Richardt, 0:20]
- Timeline editing
[Bai, 0:20]
- Editing using lightweight models
[Theobalt, 0:20]
- Model-based video editing
[Theobalt, 0:20]
- Break
[0:15]
- Spatiotemporal video editing
and processing
[Richardt, 0:20]
- Motion editing in videos:
cinemagraphs & cliplets
[Bai, 0:20]
- Exploring videos
[Tompkin, 0:20]
- Exploring videos in contexts
[Tompkin, 0:20]
- Closing and Q & A
[all, 0:10]

Let's start by looking into existing state-of-the-art video tools.

Christian Richardt

Background: State-of-the-Art Video Tools

2015-08-13

Christian Richardt – User-Centric Computational Videography

9

And in particular, I would like to start by first briefly looking into tools for creating videos, and I will then focus the remainder of my time on common tools and techniques for video editing and processing.

Video capture

- Digital video is now ubiquitous
- 1.2 billion smartphones sold in 2014 alone
- HD video camera in every device
- Video capture is easy
- Storage is cheap
- Mobile internet is affordable



2015-08-13

Christian Richardt – User-Centric Computational Videography

10

As I mentioned, digital video is now ubiquitous.

Last year alone, more than 1.2 BILLION smartphones were sold around the world, which is more than the number of traditional cameras sold in the last 15 years.

One of the main reasons why people buy smartphones is that they come with at least one high-resolution, HD video camera, and often even two: one on the back of the device and one facing the user.

The technological advances of recent years have made it possible for everyone to carry a high-powered digital video camera in their pockets.

This makes video capture really easy, as the camera is always at arm's length.

This changes everything, as you can see in this comparison of the papal inaugurations of 2005 and 2013.

The storage of the recorded videos on the device, on people's computers and in the cloud is also pretty cheap, and getting cheaper.

And lastly, the speed and affordability of mobile internet is constantly improving, so that more and more videos will be uploaded and shared on community video websites.

Sources:

Gartner Says Smartphone Sales Surpassed One Billion Units in 2014

<http://www.gartner.com/newsroom/id/2996817>

CIPA Camera Sales Figures 2014

<http://www.cipa.jp/stats/documents/common/cr200.pdf>

Photo collage by NBC News:

<https://instagram.com/p/W2FCksR9-e/>

Creating videos from photos



"Turn your photos and video clips into professional-quality videos"



"Ever combines music with your images to create videos for sharing"



"Make beautiful videos instantly with pictures, texts & music"



"Instantly create a beautiful video online with photos, videos & music"



"Instant movies from your best moments and the music you love"



"Make amazing videos in seconds. Just add photos, videos, music, and text"

2015-08-13

Christian Richardt – User-Centric Computational Videography

11

There are also many different solutions for creating videos from still photos, some music and maybe some video clips using mobile apps or simply online.

Animoto "turn[s] your photos and video clips into professional-quality videos."

"Ever combines music with your images to create videos for sharing."

Picovico "Make[s] beautiful videos instantly with pictures, texts & music."

Proshow Web "Instantly create[s] a beautiful video online with photos, videos & music."

Slidely creates "Instant movies from your best moments and the music you love."

And lastly, Stupeflix "Make[s] amazing videos in seconds. Just add photos, videos, music, and text."

Some of these services are essentially glorified slide-show generators, but others are actually quite good and create exciting little video clips to share with friends and family.

Websites (in alphabetic order):

<https://animoto.com/>

<https://www.evver.com/>

<http://www.picovico.com/>

<http://web.photodex.com/>

<http://slide.ly/>

<https://stupeflix.com/>

Creating videos from photos



VISUAL
COMPUTING
INSTITUTE

mpi
max planck institut
informatik



"Turn your photos and video clips into professional quality videos"



"Ever combines music with images to create videos for business"



"Make beautiful videos instantly with pictures, texts & music"



"Instantly create a beautiful video online with photos, videos & music"



"Photo Story helps you tell engaging stories using your photos."



"Instant movies from your best moments and the music you love"

"Make amazing videos in seconds. Just add photos, videos, music, and text"

2015-08-13

Christian Richardt – User-Centric Computational Videography

12

And just a few hours ago [on Thu 13 Aug 2015], Microsoft announced Photo Story for Windows Phone that "helps you tell engaging stories using your photos."

Website:

<https://www.microsoft.com/store/apps/photo-story/9nblggh1z351>

Finding paths through the world's photos

[Snavely et al., SIGGRAPH 2008]

- Interactive exploration of community photo collections
- Display photos on planar proxy
- Graph of all photos with edges encoding transition costs
- Path planning using Dijkstra
- Smooth the resulting path
- Stabilise appearance



2015-08-13

Christian Richardt – User-Centric Computational Videography

13

Aside from these commercial offerings, there are also interesting research systems that can create videos from a collection of photos.

In this system by Snavely et al., photos of a popular location, such as the Pantheon in Rome, are made interactively explorable.

The system automatically detects camera orbits around points of interest and stabilises relevant photos using planar proxy geometry through the orbit's axis of revolution.

One particularly interesting mode in this system finds a visually interesting path through a user's photographs, augmented by internet photos, as shown in the video.

The system builds a graph of all photos, and encodes transition costs between photos as edge weights.

The transitions trade off image distortion, field of view and resolution with the goal of producing a smooth camera path through the scene.

The optimal path through the photos is found using the shortest path through this graph, as computed by Dijkstra's algorithm.

The resulting path is then smoothed and the appearance of photos is stabilised to create a visually smooth experience.

Source:

Noah Snavely, Rahul Garg, Steven M. Seitz and Richard Szeliski

Finding paths through the world's photos

ACM Transactions on Graphics (Proceedings of SIGGRAPH), 2008, 27(3), 15

<http://dx.doi.org/10.1145/1360612.1360614>

<http://phototour.cs.washington.edu/findingpaths/>

Exploring photobios

[Kemelmacher-Shlizerman et al., SIGGRAPH 2011]

- Goal: create smooth face animation from photos, aka 'photobios'
- Automatic 3D face alignment
- Build Face Graph with transition costs
- Get animation from shortest path
- Cross dissolve produces motion effect
- Implemented as "face movies" in Picasa



[Kemelmacher-Shlizerman et al., SIGGRAPH 2011]

In a similar vein, Kemelmacher-Shlizerman et al. explore photobios by creating smooth animations from photographs of a single person, such as George W. Bush.

Their approach first automatically aligns faces in photos by detecting facial landmarks, estimating the 3D pose by fitting a template face model, and then warping each photo to a frontal view.

They then build a graph of face photos, similar to the previous paper, in which edges measure the distance between faces in terms of their 3D pose, appearance and time of capture.

The frames of the animation are again obtained from the shortest path through this graph.

In a final step, these frames are put together with cross-dissolve transitions that produce an appealing effect of motion.

Source:

Ira Kemelmacher-Shlizerman, Eli Shechtman, Rahul Garg and Steven M. Seitz

Exploring photobios

ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2011**, 30(4), 61:1–10

<http://dx.doi.org/10.1145/2010324.1964956>

<http://grail.cs.washington.edu/photobios/>

Time-lapse Mining from Internet Photos

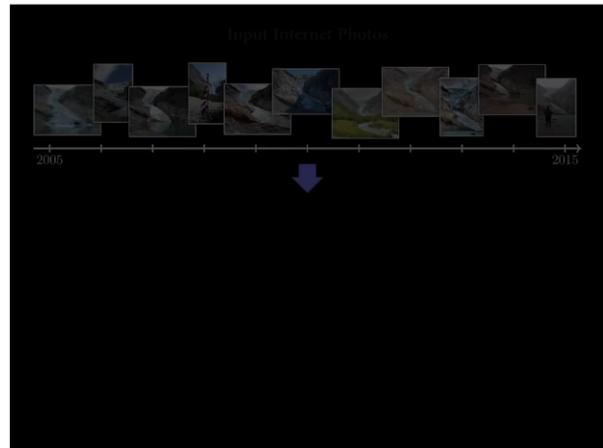
[Martin-Brualla et al., SIGGRAPH 2015]



VISUAL
COMPUTING
INSTITUTE

max planck institut
informatik

- Cluster 86 million internet photos by geolocation
- Reconstruct landmarks (SfM)
- Align photos to reference view
- Stabilise appearance using temporal Huber regularisation
- Discovered 11K+ time lapses



[Martin-Brualla et al., SIGGRAPH 2015]

2015-08-13

Christian Richardt – User-Centric Computational Videography

15

And most recently, on Tuesday, Martin-Brualla et al. proposed a new system for automatically mining time-lapse videos from internet photos.

They first take 86 million geo-tagged internet photos and cluster them into “landmarks” according to their location.

These landmarks are reconstructed using structure-from-motion, and a reference frame is identified, to which all other photos with sufficient overlap are aligned.

This geometric alignment computes a depth map for the reference frame using plane-sweep multi-view stereo (medians of NCC + MRF graph cut).

The depth map is used for warping the other photos to the reference viewpoint, after which any cracks are inpainted.

Finally, the appearance of the aligned photos is stabilised using an optimisation with a Huber loss function.

This system explored 755K 3D reconstructions for 120K landmarks, and finally produced 11K time lapses of 3K landmarks.

Source:

Ricardo Martin-Brualla, David Gallup and Steven M. Seitz

Time-lapse Mining from Internet Photos

ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2015**, 34, 62:1–8

<http://dx.doi.org/10.1145/2766903>

<http://grail.cs.washington.edu/projects/timelapse/>

Existing video editing tools

Consumer tools

- Adobe Premiere Elements
- Apple iMovie
- Pinnacle Studio
- Sony Vegas Movie Studio
- Windows Movie Maker
- YouTube Video Editor

Professional tools

- Adobe After Effects
- Adobe Premiere Pro
- Avid Media Composer
- Final Cut Pro
- Nuke
- Sony Vegas Pro

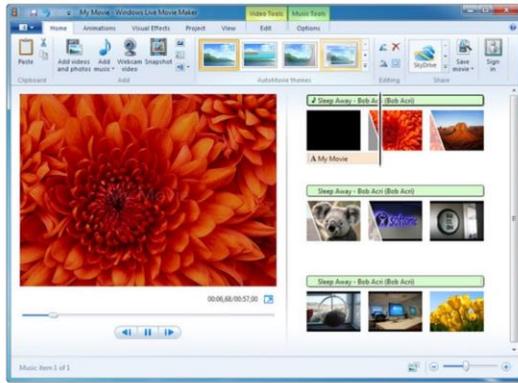
There are quite a few software tools for video editing available, which can be broadly categorised into consumer and professional (or prosumer) tools.

I am sure you will have heard of many of these tools, and most likely also used one or the other of them for your personal videos or at work.

And while these tools cover a huge spectrum of functionality, and also price, they do have some common basic functionality.

Existing video editing tools

Consumer tools



[<http://screenshots.en.sftcdn.net/en/scrn/74000/74149/windows-live-movie-maker-22.jpg>]

Windows Movie Maker

Professional tools



[http://en.wikipedia.org/wiki/File:Final_Cut_Pro_X.jpg]

Final Cut Pro

2015-08-13

Christian Richardt – User-Centric Computational Videography

17

At their core, both consumer and professional tools focus on timeline-based non-linear video editing. This means that multiple videos, photos and music are combined with titles and transitions into an edited video sequence.

Although the interfaces greatly vary in their complexity, they both allow lossless video editing such as trimming, stretching and rearranging of video and audio clips.

Sources:

<http://screenshots.en.sftcdn.net/en/scrn/74000/74149/windows-live-movie-maker-22.jpg>

http://en.wikipedia.org/wiki/File:Final_Cut_Pro_X.jpg

Basic video editing functionality

- Non-linear video editing:
 - One or more video + audio tracks that are composited together
 - Combine videos, music, photos with titles and mattes
 - Losslessly trim, stretch and rearrange audio/video clips
- Audio effects: denoise, high/low pass, reverb, equaliser, ...
- Video transitions: cross dissolve, wipe, iris, zoom, 3D motions, ...
- Video effects: colour correction, blur, grain, keying, warping, ...

Beyond the basic non-linear video editing functionality, these tools also provide a wide range of audio and video effects as well as transitions.

These can be used to alter the look and sound of a video clip in line with an artistic vision, and also to glue together smaller clips in a visually smooth and interesting fashion.

Let's look at some of this functionality in a little more detail.



[Christian Richardt]

Every video editing software will offer some support for video transitions, such as these transitions in Adobe Premiere CS6.

Although there are quite a few transitions to choose from, most of them are fairly basic and gimmicky.

After the break in our course, I will talk a bit about some more interesting video transitions that were proposed in the research literature.

Source:

Premiere CS6 Transitions Demo

created by author [Christian Richardt]

Video stabilisation

Shaky input video



Stabilised video



2015-08-13

Christian Richardt – User-Centric Computational Videography

20

One very common and very useful video processing technique is video stabilisation.

As you know, videos captured with handheld cameras are often shaky, particularly when one is walking while recording a video, such as in this example on the left.

The goal of video stabilisation is to remove the camera shake and create a video with a smoothly moving camera that looks like it was captured with a professional steadycam.

Here are some recent techniques that are not discussed in more detail on the following slides:

Feng Liu, Michael Gleicher, Hailin Jin and Aseem Agarwala

Content-Preserving Warps for 3D Video Stabilization

ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2009**, 28(3), 44:1–9

<http://dx.doi.org/10.1145/1531326.1531350>

<http://pages.cs.wisc.edu/~fliu/project/3dstab.htm>

Amit Goldstein and Raanan Fattal

Video stabilization using epipolar geometry

ACM Transactions on Graphics, **2012**, 31(5), 126:1–10

<http://dx.doi.org/10.1145/2231816.2231824>

<http://www.cs.huji.ac.il/~raananf/projects/stab/>

Shuaicheng Liu, Lu Yuan, Ping Tan and Jian Sun

Bundled camera paths for video stabilization

ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2013**, 32(4), 78:1–10

<http://dx.doi.org/10.1145/2461912.2461995>

<http://www.liushuaicheng.org/SIGGRAPH2013/index.htm>

Yu-Shuen Wang, Feng Liu, Pu-Sheng Hsu and Tong-Yee Lee

Spatially and Temporally Optimized Video Stabilization

IEEE Transactions on Visualization and Computer Graphics, **2013**, 19(8), 1354–1361

<http://dx.doi.org/10.1109/TVCG.2013.11>

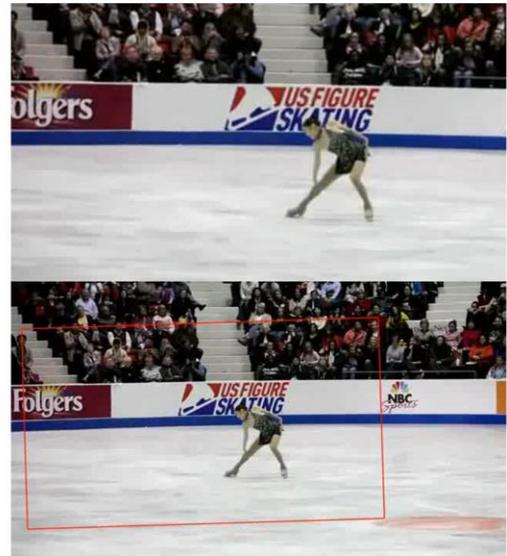
YouTube video stabilisation

[Grundmann et al., CVPR 2011]

- Compute 2D point trajectories
- Fit linear motion model
- Optimise crop for simple paths
- Residual motion suppression against rolling shutter



[Grundmann et al., CVPR 2011]



[Grundmann et al., CVPR 2011]

2015-08-13

Christian Richardt – User-Centric Computational Videography

21

When you upload a shaky video to YouTube, chances are it will suggest to stabilise it for you.

The underlying algorithm is based on a technique by Grundmann et al. at CVPR 2011.

This is a 2D video stabilisation technique that first detects feature points in each video frame, and then matches them to build 2D point trajectories.

Based on these, they fit a linear motion model between consecutive video frames, and optimise for a cropping region that results in simple output point trajectories.

You can see this cropping in effect in the example video on the bottom right.

The cropped video at the top is the stabilised video result.

Source:

Matthias Grundmann, Vivek Kwatra and Irfan Essa

Auto-directed video stabilization with robust L1 optimal camera paths

Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), 2011, 225–232

<http://dx.doi.org/10.1109/CVPR.2011.5995525>

<http://www.cc.gatech.edu/cpl/projects/videstabilization/>

Adobe warp stabiliser

[Liu et al., ToG 2011]

- Factorise 2D trajectory matrix into low-dimensional subspace
- Smooth eigen-trajectories + apply content-preserving warp



[Liu et al., SIGGRAPH 2009]

2015-08-13

Christian Richardt – User-Centric Computational Videography

22

Another well-known stabilisation technique is Adobe's warp stabiliser found in After Effects and Premiere Pro, which is based on work by Liu et al.

This approach also starts with 2D point trajectories, which are put into a matrix that is then factorised into a low-dimensional subspace.

The resulting so-called eigen-trajectories are then smoothed, and the trajectory matrix is reconstituted, and used for content-preserving warping of the input video frames.

This technique works best for mostly static scenes, as in this example, because dynamic objects violate the low-rank assumption used in the matrix factorisation.

Source:

Feng Liu, Michael Gleicher, Jue Wang, Hailin Jin and Aseem Agarwala

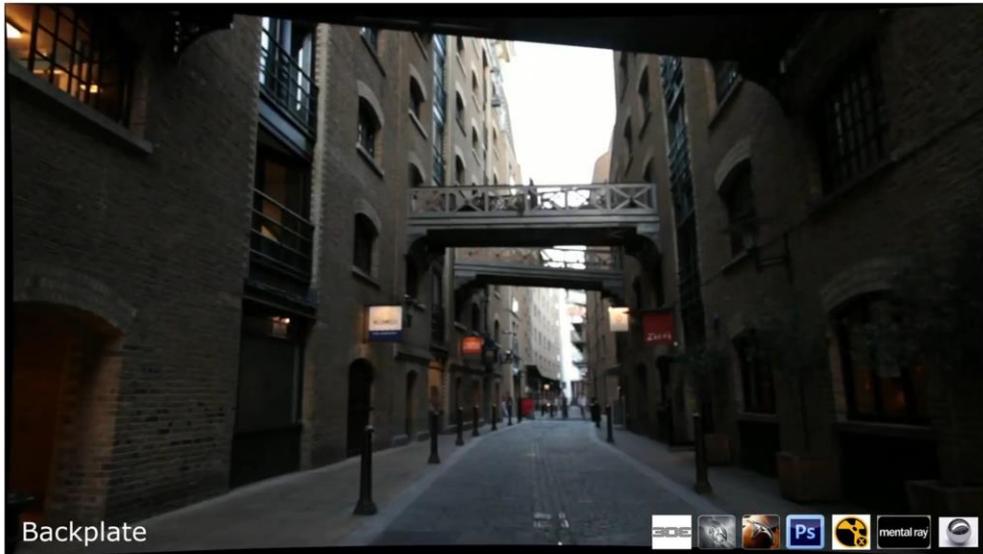
Subspace video stabilization

ACM Transactions on Graphics, 2011, 30(1), 4:1–10

<http://dx.doi.org/10.1145/1899404.1899408>

http://web.cecs.pdx.edu/~fliu/project/subspace_stabilization/

Camera tracking (aka match moving)



2015-08-13

Christian Richardt – User-Centric Computational Videography

23

The 2D point trajectories estimated during camera stabilisation can also be used to reapply camera shake when compositing synthetic objects (CGI) into live-action videos.

This is known as “camera tracking” or “match moving” in the film industry, and is heavily used in movie postproduction.

The key computer-vision technique that underlies all of this is structure-from-motion.

It reconstructs the camera motion and scene structure from sparse point trajectories over time.

And since automatic approaches can fail, point trajectories can also be specified manually.

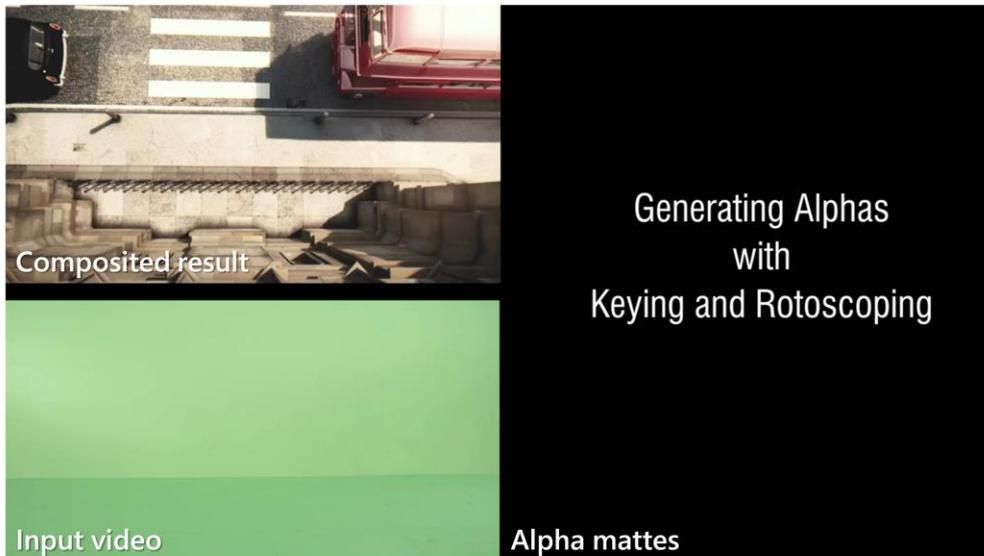
Camera tracking can mostly be found in research systems and professional tools or plugins, such as Voodoo, VideoTrace, VisualSfM, Blender, Maya, Nuke and others (PFTrack, Boujou, Mocha Pro, ...).

Source:

CGI VFX Matchmove Showreel HD: “Jiuk Han Choi – 2014”

<https://youtu.be/bK6vCPcFkfk?t=58s>

Video matting (aka keying) / compositing



2015-08-13

Christian Richardt – User-Centric Computational Videography

24

Another staple of professional video editors is video matting or keying, which is the process of extracting some desirable foreground from the background.

The simplest form assumes a single-coloured background, such as the greenscreen in the bottom left, and then uses a pixel's colour to determine if it is foreground or background or a mixture of the two.

The basic principle is known as “chroma keying”, and the result are the greyscale alpha mattes on the right, where white areas are 100% foreground, black areas are 100% background, and intermediate values appear on the boundary between the two.

Using the mattes, the foreground object can be easily extracted from the input video, and composited into a different video, as shown on the top left.

As you can see, the final video also uses camera tracking to seamlessly position the extracted athletes in the virtual environment.

Source:

Olympia London Opener "Keying and Rotoscoping" by EFEX GmbH

<https://vimeo.com/46360293>

Video segmentation (aka rotoscoping)



VISUAL
COMPUTING
INSTITUTE



[Elyarch, <https://vimeo.com/132583941>]

2015-08-13

Christian Richardt – User-Centric Computational Videography

25

However, automatic video matting techniques rarely work perfectly, so significant manual effort is invested in “rotoscoping”, or what researchers call “video segmentation”.

The goal is again to separate a video into foreground and background layers, but the approach is different: each individual layer is typically traced by hand.

To save time, this is usually only done in a few keyframes, and the masks are then interpolated between the keyframes.

If the result is not satisfactory, additional keyframes can be traced, and the interpolation refined.

The example shown here is even more challenging, as the camera performs a focus pull that affects different scene layers differently.

This means that the man and the woman are never both in focus at the same time, which also needs to be reflected in the alpha mattes.

Source:

“Rotoscoping breakdown” by Elyarch

<https://vimeo.com/132583941>

Colour grading



[Chris Hall, <https://vimeo.com/88300860>]

2015-08-13

Christian Richardt – User-Centric Computational Videography

26

The last family of techniques that I would like to mention is colour grading, which is essentially artistic colour correction.

As you can see in these examples, a skilled colourist can significantly change the mood of a video, and even the time of day, as shown in this example (and also in the 2015 movie “Mad Max: Fury Road”).

These last few techniques can produce amazing results when used by professionals, but they often require significant manual input from a skilled artist.

This means that they are not currently usable by consumers, who would no doubt love to benefit from them.

Source:

Chris Hall – Chris Hall Color Demo Reel 2014

<https://vimeo.com/88300860>

Limitations of existing video editing tools



VISUAL
COMPUTING
INSTITUTE



[Krähenbühl et al., SIGGRAPH Asia 2009]

2015-08-13

Christian Richardt – User-Centric Computational Videography

27

One example of a technique that currently cannot be found in existing video editing tools is video retargeting.

It is concerned with the adjustment of video content to different display sizes without losing important visual content.

Showing wide-screen videos on devices with a different aspect ratio either leads to letterboxing, so wasting screen real estate, or cropping the video, which may crop important parts of the video.

Techniques such as this work by Krähenbühl et al., optimise an image-based warp to preserve salient objects as much as possible while squashing visually unimportant background regions.

Source:

Philipp Krähenbühl, Manuel Lang, Alexander Hornung and Markus Gross

A System for Retargeting of Streaming Video

ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia), **2009**, 28(5), 126:1–10

<http://dx.doi.org/10.1145/1618452.1618472>

<http://www.disneyresearch.com/publication/video-retargeting/>

Additional References:

Michael Rubinstein, Ariel Shamir and Shai Avidan

Improved seam carving for video retargeting

ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2008**, 27(3), 16

<http://dx.doi.org/10.1145/1360612.1360615>

Shih-Syun Lin, Chao-Hung Lin, I-Cheng Yeh, Shu-Huai Chang, Chih-Kuo Yeh and Tong-Yee Lee

Content-Aware Video Retargeting Using Object-Preserving Warping

IEEE Transactions on Visualization and Computer Graphics, **2013**, 19(10), 1677–1686

<http://dx.doi.org/10.1109/TVCG.2013.75>

Limitations of existing video editing tools



2015-08-13

Christian Richardt – User-Centric Computational Videography

28

There are some tasks that none of the existing tools can perform out of the box, such as this vision of content-aware video editing shown at Adobe's MAX conference last year (2014).

Here, a user rearranges a video by simply dragging the horses around.

The building blocks of this vision are video segmentation and video inpainting, both of which are available in existing video tools, but we are still far from the robust and fast implementations required to pull off this sort of video editing automatically.

This vision is a great example of the kind of editing tasks that User-Centric Computational Videography will enable in the future.

Source:

<http://max.adobe.com/sessions/max-online/>

Session "The Future", 00:16:08–00:16:15

Course syllabus

- Introduction and welcome
[Richardt, 0:10]
- Background:
State-of-the-art video tools
[Richardt, 0:20]
- **Timeline editing**
[Bai, 0:20]
- Editing using lightweight models
[Theobalt, 0:20]
- Model-based video editing
[Theobalt, 0:20]
- Break
[0:15]
- Spatiotemporal video editing
and processing
[Richardt, 0:20]
- Motion editing in videos:
cinemagraphs & cliplets
[Bai, 0:20]
- Exploring videos
[Tompkin, 0:20]
- Exploring videos in contexts
[Tompkin, 0:20]
- Closing and Q & A
[all, 0:10]

And with this, I would like to take any questions you might have, while Jiamin gets ready to talk about timeline editing.