Christian Richardt

Spatiotemporal Video Editing and Processing

Welcome back from the break, let's continue our course on user-centric computational videography with a look at research papers on spatiotemporal video editing and processing from the last 10 years.

Interactive multi-perspective imagery

[Lieng et al., Eurographics 2012]

- Compute SfM on input photos and/or video(s)
- User marks 'portal' in an image
- Warp other images to fit portal
- Interpolate smooth warping
- Enables interactive exploration

At the beginning of the course, I briefly mentioned the fairly basic video transitions offered by most video editing software and hinted at **advanced video transitions** in the research literature that are visually more interesting.

Lieng et al. proposed one such system, in which photos and videos can be navigated in a multi-perspective fashion.

After computing structure-from-motion from the input photos and videos, the user marks an area such as a door or a passageway as a "portal".

Images from other viewpoints are then warped to fit the portal, and smoothly interpolated to enable interactive exploration of a scene – and looking around corners.

Reference:

Henrik Lieng, James Tompkin and Jan Kautz
Interactive Multi-perspective Imagery from Photos and Videos
Computer Graphics Forum (Proceedings of Eurographics), **2012**, *31*(2), 285–293

DOI: http://dx.doi.org/10.1111/j.1467-8659.2012.03007.x

URL: http://vecg.cs.ucl.ac.uk/Projects/InteractiveMultiPerspective/

Video source:

Supplemental video "mpi_showreel_final".

Video transitions of places

[Tompkin et al., TAP 2013]

- Perceptual evaluation of 7 video transition types
- User preference depends on view change:
  - Considerable view change: full 3D static transition
  - Slight view change: warp transitions
- But no single transition is universally applicable

Tompkin et al. performed a perceptual evaluation of seven types of video transitions for different places to study which transitions are most preferred, and which visual artefacts are most undesirable.

The most involved transitions use a reconstruction of scene geometry, which in practice often has limited quality and can therefore introduce artefacts.

Nevertheless, they discovered a strong preference for full 3D static transitions when the viewpoint is changing considerably, such as in this video, and a preference for warp transitions in the slight view change case.

However, no video transition was found to be universally applicable.

Reference:

James Tompkin, Min H. Kim, Kwang In Kim, Jan Kautz and Christian Theobalt
Preference and artifact analysis for video transitions of places
ACM Transactions on Applied Perception, **2013***, 10*(3), 13:1–19

DOI: http://dx.doi.org/10.1145/2501601

URL: http://gvv.mpi-inf.mpg.de/projects/VideoTransitionsOfPlaces/

Video source:

Composited from individual videos for “scene 1, considerable view change” from the authors’ project website.

DuctTake: Spatiotemporal video compositing

[Rüegg et al., Eurographics 2013]

- Find transition with coarse-to-fine graph cut through video volume
- Handles spatial cuts, temporal blends, and complex mixtures

An interesting take on video transitions is the DuctTake system by Rüegg et al. that finds transitions between two input videos using a graph cut through the video volume.

But first, the input videos are aligned using an efficient block-based matching scheme with per-frame homographies.

Corresponding frames in the two videos are then also matched in their motion blur as well as colours.

The input videos are finally blended together along the computed cutting boundary, and the resulting video is cropped to remove empty areas.

This system supports a range of different effects, such as spatial cuts between two videos, for example to create impossible reflections …

Reference:

Jan Rüegg, Oliver Wang, Aljoscha Smolic and Markus Gross
DuctTake: Spatiotemporal Video Compositing
Computer Graphics Forum (Proceedings of Eurographics), **2013**, *32*(2), 51–61

DOI: http://dx.doi.org/10.1111/cgf.12025

URL: http://zurich.disneyresearch.com/~owang/pub/ducttake.html

Video source:

Supplemental video “ducttake_define1_small”.

DuctTake: Spatiotemporal video compositing

[Rüegg et al., Eurographics 2013]

- Find transition with coarse-to-fine graph cut through video volume
- Handles spatial cuts, temporal blends, and complex mixtures

… or finding temporal transition that lets the camera appear to move through the closed window.

Reference:

Jan Rüegg, Oliver Wang, Aljoscha Smolic and Markus Gross
DuctTake: Spatiotemporal Video Compositing
Computer Graphics Forum (Proceedings of Eurographics), **2013**, *32*(2), 51–61

DOI: http://dx.doi.org/10.1111/cgf.12025

URL: http://zurich.disneyresearch.com/~owang/pub/ducttake.html

Video source:

Supplemental video “ducttake_define2_small”.

Semi-automated video morphing

[Liao et al., EGSR 2014]

- Goal: Smooth transitions between different videos
- Some manual correspondences
- Temporal synchronisation of input videos to align them
- Optimise morph in spatio-temporal halfway domain

[Liao et al., EGSR 2014]

2015-08-13 Christian Richardt – User-Centric Computational Videography 6

A different sort of video transition is video morphing, such as this semi-automatic work by Liao et al.

The goal here is to produce ultra smooth transitions between different yet sufficiently similar videos, by warping and blending them cleverly.

This approach requires a few manual correspondences, which are shown as yellow circles in the two input videos on the left.

In a first step, the input videos are synchronised temporally to align the motions in the videos.

Then comes the core of the technique: the morphing approach based on a halfway domain between the two input video volumes.

While this might seem pretty involved, it produces very good results, as shown in these examples.

Reference:

Jing Liao, Rodolfo S. Lima, Diego Nehab, Hugues Hoppe and Pedro V. Sander
Semi-Automated Video Morphing
Computer Graphics Forum (Proceedings of Eurographics Symposium on Rendering), **2014**, *33*(4), 51–60

DOI: http://dx.doi.org/10.1111/cgf.12412

URL: http://research.microsoft.com/en-us/um/people/hoppe/proj/videomorph/

Video source:

Supplemental video.

Cartoon-style motion cues in video

[Collomosse et al., Graphical Models 2005]

- Artistic rendering of motion
- Track features in video, and recover depth ordering
- Augmentation cues:
  - streak lines, ghosting, blurring
- Deformation cues:
  - squash-and-stretch, drag effects

[Collomosse et al., Graphical Models 2005]

Another video effect that requires and exploits spatiotemporal information is **motion visualisation**.

Collomosse et al. chose to render motions using cartoon-style artistic rendering.

They first track features in the input video, and then recover a depth ordering of multiple motion layers.

This enables them to insert augmentation cues behind each moving object, for example to show streak lines like in the example video, but they also support other effects such as ghosting and motion blurring.

Reference:

John P. Collomosse, David Rowntree and Peter M. Hall
Rendering cartoon-style motion cues in post-production video
Graphical Models, **2005**, *67*(6), 549–564
DOI: http://dx.doi.org/10.1016/j.gmod.2004.12.002

Video source:

Video clip “metro_streaks” from John Collomosse’s PhD dissertation.

http://personal.ee.surrey.ac.uk/Personal/J.Collomosse/pubs/thesissupplm/

Cartoon-style motion cues in video

[Collomosse et al., Graphical Models 2005]

- Artistic rendering of motion
- Track features in video, and recover depth ordering
- Augmentation cues:
  - streak lines, ghosting, blurring
- Deformation cues:
  - squash-and-stretch, drag effects

[Collomosse et al., Graphical Models 2005]

In addition to augmentation cues, they also show deformation cues, that squash-and-stretch objects, or drag them out to visualise fast motions.

Reference:

John P. Collomosse, David Rowntree and Peter M. Hall
Rendering cartoon-style motion cues in post-production video
Graphical Models, **2005**, *67*(6), 549–564
DOI: http://dx.doi.org/10.1016/j.gmod.2004.12.002

Video source:

Video clip “wand_cartoon” from John Collomosse’s PhD dissertation.

http://personal.ee.surrey.ac.uk/Personal/J.Collomosse/pubs/thesissupplm/

Computational time-lapse video

[Bennett & McMillan, SIGGRAPH 2007]

- Sample (non-)uniformly from input video frames
  - min-error cost: best change-preserving approximation
  - min-change cost: most uniform video frames
  - Optionally with a bias towards uniform sampling

[Bennett & McMillan, SIGGRAPH 2007]

In their work “Computational time-lapse video”, Bennett & McMillan process long videos into time-lapse videos.

However, selecting output video frames uniformly from the input video can easily miss important information, such as the occasional car driving along.

This is addressed by a non-uniform sampling scheme that can optimise different costs.

The “min-error” cost minimises the approximation error between the sampled video frames and the entire input video.

This mostly extracts the moving cars from the long video rather than the slowly moving clouds, which are favoured by the “min-change” cost that considers the cars to be outliers.

An additional term can also be added to bias the selected video frames towards a more uniform distribution over time.

Reference:

Eric P. Bennett and Leonard McMillan
Computational time-lapse video
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2007**, *26*(3), 102

DOI: http://10.1145/1276377.1276505

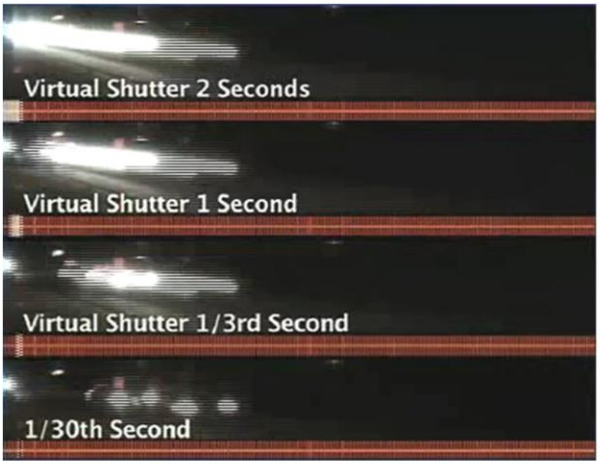
Photo source:

Cropped from Figure 1 in their paper.

Computational time-lapse video

[Bennett & McMillan, SIGGRAPH 2007]

- Combine frames using virtual shutters to create new virtual exposures:
  - Min/max/median
  - Extended exposure
  - Motion tails



In addition to non-uniform sampling, Bennett and McMillan propose virtual shutter effects that for example extend the effective exposure time for each output frame beyond physical limits.

This enables two-second exposures (top) in a 30 fps video, which normally has a maximum exposure time of 33 ms (bottom).

In this example, longer light streaks represent faster cars.

Reference:

Eric P. Bennett and Leonard McMillan
Computational time-lapse video
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2007***, 26*(3), 102

DOI: http://10.1145/1276377.1276505

Video source:

Recording of the authors' SIGGRAPH 2007 presentation, available from the ACM Digital Library.

(A copy of the supplemental video could not be found online. Kindly contact Christian Richardt if you have a copy. Thanks.)

Motion magnification

[Liu et al., SIGGRAPH 2005]

- Want to amplify imperceptible motions
- Compute feature trajectories with sub-pixel precision
- Cluster correlated trajectories into motion layers (+outliers)
- Interpolate dense optical flow
- Warp pixels using scaled flow
- Fill holes using inpainting

Another way to visualise motions is **motion magnification**, which aims to amplify hardly visible motions.

Liu et al. achieve this with a direct approach that essentially estimates and amplifies per-pixel optical flow (this has been dubbed a “Lagrangian” approach by follow-up work).

They start by stabilising the input video to remove camera shake.

For this, they compute sub-pixel-accurate 2D feature point trajectories by detecting Harris corners, matching them using sum-of-squared-differences (SSD), and performing sub-pixel refinement using Lucas-Kanade.

They fit a per-frame affine transform to the best matches to cancel out any camera shake.

The stabilised feature trajectories are then refined, and clustered into motion layers based on their correlation with each other.

This puts trajectories in the same cluster even if their motion is not identical, but for example linked due to physical processes such as vibration of an object. (All outliers end up in the same layer.)

The feature trajectories are then interpolated across each video frame to obtain per-pixel optical flow.

This flow is then scaled and used for warping the input video frame, filling any holes that appear with inpainting.

Reference:

Ce Liu, Antonio Torralba, William T. Freeman, Frédo Durand and Edward H. Adelson
Motion magnification

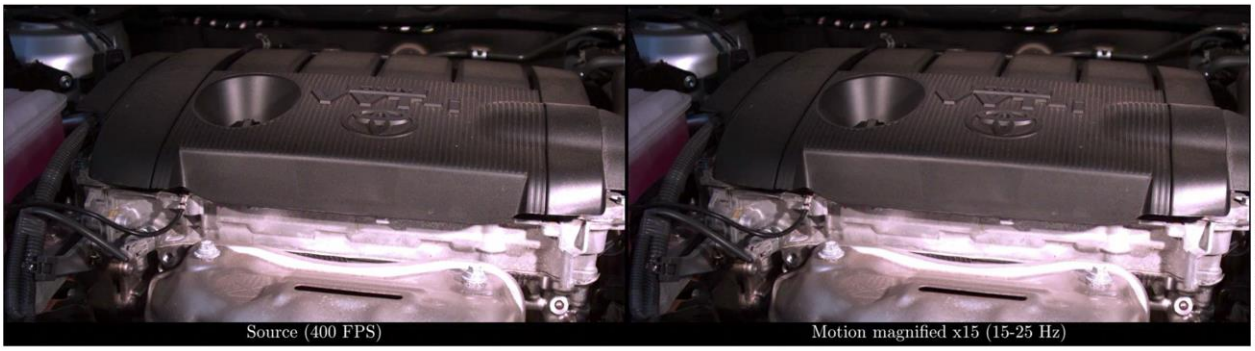
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2005***, 24*(3), 519–526
DOI: http://dx.doi.org/10.1145/1073204.1073223
URL: http://people.csail.mit.edu/celiu/motionmag/motionmag.html

Video source:
Supplemental video.

Phase-based video motion processing

[Wadhwa et al., SIGGRAPH 2013]

- Decompose video using complex steerable pyramids
- Temporally filter phases, then amplify/attenuate them as desired
- Finally reconstruct the video to amplify imperceptible motion



More recently, Wadhwa et al. proposed an approach that does not need to estimate optical flow explicitly, but operates directly on the phase information contained in videos (a so-called “Eulerian” approach).

They decompose input videos using complex steerable pyramids into a sort-of localised Fourier domain, where they can directly filter the phase information over time, and amplify or attenuate it, as desired.

Finally, they reconstruct the video to obtain the desired result.

This approach can magnify motions easily by more than an order of magnitude, and has great noise characteristics, as noise in the input video is not amplified, but simply translated.

Reference:

Neal Wadhwa, Michael Rubinstein, Frédo Durand and William T. Freeman
Phase-based video motion processing
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2013***, 32*(4), 80:1–10

DOI: http://dx.doi.org/10.1145/2461912.2461966

URL: http://people.csail.mit.edu/nwadhwa/phase-video/

Additional previous work:

Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand and William Freeman
Eulerian video magnification for revealing subtle changes in the world
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2012***, 31*(4), 65:1–8

DOI: http://dx.doi.org/10.1145/2185520.2185561
URL: http://people.csail.mit.edu/mrub/vidmag/

Video source:
Example video “car_engine_result” on the paper’s website.

Hyperlapse from Instagram

[Instagram 2014; Karpenko et al., 2011]

- Record video and gyroscope information during capture
- Preview stabilised video in real time at different speeds
- Adaptively zoom into video to stabilise central portion

[Instagram 2014]

[Instagram 2014]

Let's move from amplifying motions to removing motion as much as possible, specifically in the context of **hyperlapses**, which are time-lapse videos with smoothly moving cameras.

This is a fairly new area in video processing, although it is of course closely related to existing video stabilisation techniques.

Instagram's free hyperlapse app on iOS records videos together with information from the phone's gyroscopes so that the output video can be stabilised without having to estimate motion from the input video.

This algorithm is based on work by Alex Karpenko et al. at Stanford.

After video recording is finished, the app offers multiple different speed-up factors, and provides a real-time preview of each result.

The stabilisation result is obtained by cropping the central region from the input video frames, and the amount of cropping depends on the strength of scene motion.

The shakier a video is, the more cropping is required to produce the stabilised output video.

The benefit of using the gyro is that the scene is always stabilised with respect to the global reference frame and not any foreground objects. However, requiring extra information also means that this approach cannot be applied to arbitrary existing videos.

References:

Alex Karpenko

The technology behind hyperlapse from instagram

Instagram Engineering Blog, August **2014**

URL: http://instagram-engineering.tumblr.com/post/95922900787/hyperlapse

Alexandre Karpenko, David Jacobs, Jongmin Baek and Marc Levoy
Digital Video Stabilization and Rolling Shutter Correction using Gyroscopes
Stanford University, **2011** (CTSR 2011-03)

URL: http://graphics.stanford.edu/papers/stabilization/

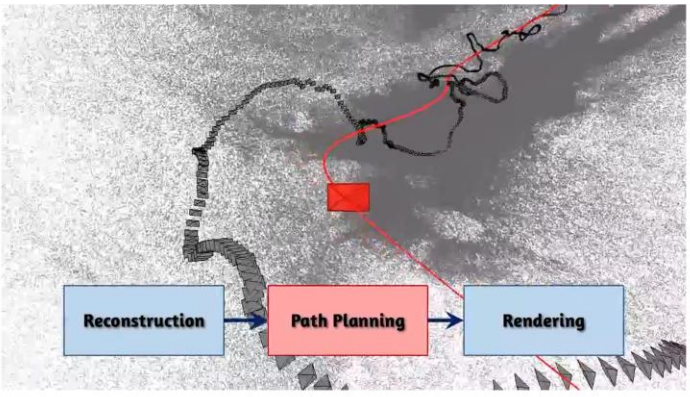
Video sources:

Videos included in “The technology behind hyperlapse from instagram” (see above).

First-person hyper-lapse videos

[Kopf et al., SIGGRAPH 2014]

- Reconstruct scene using chunked SfM (1.4K frames at a time)
- Plan smooth camera path that minimises rendering errors
- Render output frame by stitching 3–5 video frames with spatiotemporal MRF
- Generally good results, even for difficult videos
- Computationally expensive



In the same year, Kopf et al. proposed an approach that doesn't require extra sensor information.

First, the scene is reconstructed using structure-from-motion by cutting the video into several chunks, and merging the reconstructions of all chunks.

Second, a smooth camera path is fitted through all camera positions, and the orientations of the virtual camera are optimised to maximise the rendering quality of the final step.

This last step selects 3 to 5 video frames with high quality scores that cover the output view, and fuses them into the output frame.

This uses a spatiotemporal Markov Random Field formulation with spatiotemporal Poisson blending to account for changes in exposure and white balance.

Although the results look fairly good, this approach is computationally very expensive and can easily take a full day for processing a single video.

Reference:

Johannes Kopf, Michael F. Cohen and Richard Szeliski
First-person Hyper-lapse Videos
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2014**, *33*(4), 78:1–10

DOI: http://dx.doi.org/10.1145/2601097.2601195

URL: http://research.microsoft.com/en-us/um/redmond/projects/hyperlapse/

Video source:

Supplemental video ("technical" video).

Real-time hyperlapse creation

[Joshi et al., SIGGRAPH 2015]

- Select more similar video frames at roughly the right intervals
- Trade off frame matching, velocity & acceleration costs
- Optimise frame selection using banded dynamic programming
- Stabilise selected video frames
- Free apps that run in real time on mobile (Android, WP8) and Windows PCs

[Joshi et al., SIGGRAPH 2015]

So on Tuesday [11 August 2015], Joshi et al. presented a new, more efficient approach that is several orders of magnitude faster and runs in real time.

Instead of performing complex structure-from-motion estimation and spatiotemporal image-based rendering, they select more similar video frames at roughly the right intervals, so that they are easier to stabilise afterwards.

They express this as a banded dynamic programming problem that trades off the costs of matching frames visually, obtaining some desirable velocity as well as minimising acceleration.

This is followed by bundled video stabilisation, which contributes to the great efficiency of the overall approach.

Microsoft currently provides free apps for Android and Windows Phone, as well as Windows PCs.

Reference:

Neel Joshi, Wolf Kienzle, Mike Toelle, Matt Uyttendaele and Michael F. Cohen
Real-Time Hyperlapse Creation via Optimal Frame Selection
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2015***, 34*(4), 63:1–9

DOI: http://dx.doi.org/10.1145/2766954

URL: http://research.microsoft.com/en-us/um/redmond/projects/hyperlapserealtime/

Additional Reference:

Yair Poleg, Tavi Halperin, Chetan Arora and Shmuel Peleg

EgoSampling: Fast-Forward and Stereo for Egocentric Videos

Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), **2015**

URL: http://www.vision.huji.ac.il/egosampling/

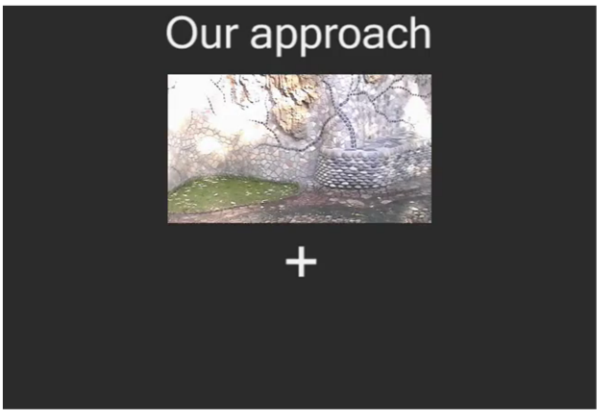
Video source:

Supplemental video.

Using photos to enhance videos of a static scene

[Bhat et al., EGSR 2007]

- Transfer photographic qualities and improve video authoring
- Estimate geometry using SfM, dense depth maps using MVS
- Warp photos into viewpoint of video frames using depth
- Stitch with graph cuts and blend with space-time fusion



There are also a few techniques for **editing and improving videos using photos**, such as this work by Bhat et al.

The goal of their work is to transfer desirable photographic qualities from a few photos to improve a video.

For this, they first estimate the scene geometry using structure-from-motion, and compute dense depth maps using a novel multi-view stereo algorithm.

These depth maps are used to warp the photos into the viewpoints of each video frame, where they are stitched and blended with a space-time fusion approach.

Reference:

Pravin Bhat, C. Lawrence Zitnick, Noah Snavely, Aseem Agarwala, Maneesh Agrawala, Brian Curless, Michael Cohen and Sing Bing Kang
Using Photographs to Enhance Videos of a Static Scene
Proceedings of the Eurographics Symposium on Rendering, **2007**, 327–338

DOI: http://dx.doi.org/10.2312/EGWR/EGSR07/327-338

URL: http://grail.cs.washington.edu/projects/videoenhancement/

Related follow-up work:

Ankit Gupta, Pravin Bhat, Mira Dontcheva, Brian Curless, Oliver Deussen and Michael Cohen
Enhancing and Experiencing Spacetime Resolution with Videos and Stills
Proceedings of the International Conference on Computational Photography (ICCP), **2009**

DOI: http://dx.doi.org/10.1109/ICCPHOT.2009.5559006

URL: http://grail.cs.washington.edu/projects/enhancing-spacetime/

Video source:
Supplemental video.

Unwrap mosaics

[Rav-Acha et al., SIGGRAPH 2008]

- Convert video frames into an "unwrap mosaic"
- Edit like a texture map on a deformable 3D surface
- Re-composite into original sequence for final result
- Easy to attach effects layers to deforming objects

[Rav-Acha et al., SIGGRAPH 2008]

The key idea of Unwrap mosaics by Rav-Acha et al. is to convert observations of a deformable surface over many video frames into a single 2D mosaic, such as this one, which can be edited similar to the texture of a deformable 3D surface.

By re-compositing the edited mosaic into the video, the final result is obtained.

This makes it easy to attach effects layers to deforming layers, without reconstructing any explicit 3D geometry.

Reference:

Alex Rav-Acha, Pushmeet Kohli, Carsten Rother and Andrew Fitzgibbon
Unwrap mosaics: a new representation for video editing
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2008***, 27*(3), 17:1–12

DOI: http://dx.doi.org/10.1145/1360612.1360616

URL: http://research.microsoft.com/unwrap/

Video source:

Supplemental video.

Coherent Spatiotemporal RGBZ Videos

[Richardt et al., Eurographics 2012]

- Combine colour and depth videos into coherent RGB+D videos using efficient spatiotemporal upsampling and denoising scheme
- Effects:
  - Video relighting
  - Abstraction and stylisation
  - Background segmentation
  - Stereoscopic 3D

And the last category of spatiotemporal video effects that I would like to discuss are **geometry-based video effects**.

At Eurographics 2012, we proposed a technique for creating high-resolution, temporally coherent RGBZ videos with per-pixel depth by combining videos from a colour camera and a depth sensor such as the Microsoft Kinect.

We achieved this by jointly upsampling and denoising the low-resolution depth information using the high-resolution colour video and spatiotemporal video filtering.

RGBZ videos enable a range of video effects, such as video relighting shown here, but also geometry-based abstraction and stylisation, background segmentation and stereoscopic 3D rendering.

Reference:

Christian Richardt, Carsten Stoll, Neil A. Dodgson, Hans-Peter Seidel and Christian Theobalt
Coherent Spatiotemporal Filtering, Upsampling and Rendering of RGBZ Videos
Computer Graphics Forum (Proceedings of Eurographics), **2012**, *31*(2), 247–256

DOI: http://dx.doi.org/10.1111/j.1467-8659.2012.03003.x

URL: http://www.mpi-inf.mpg.de/resources/rgbz-camera/

Video source:

Supplemental video.

Sampling-based scene-space video processing

[Klose et al., SIGGRAPH 2015]

- Compute SfM and depth maps
- Project pixels (aka 'samples') into scene space
- Gather samples projecting into output pixel frustum
- Filter samples using their colour, position & timestamp
- Applications: denoising, deblurring, inpainting, virtual aperture/shutter effects

[Klose et al., SIGGRAPH 2015]

And most recently, on Tuesday [11 August 2015], Klose et al. presented their framework for sampling-based scene-space video processing.

They start by computing structure-from-motion and depth maps for each video frame using off-the-shelf tools.

Conceptually, they then project each pixel in every input video frame into the 3D scene space using its depth, and call them "samples".

For computing an output pixel colour, all samples that fall into its viewing frustum are collected, and filtered using their colour, position and timestamp relative to the same properties of the input pixel.

It is this step that makes this approach so robust to all the outliers in the computed depth maps.

By changing the filtering function, a large range of applications can be implemented, such as video denoising, deblurring, inpainting, but also virtual aperture and shutter effects, such as shown here.

The only apparent downside of this technique is that it stands and falls with structure-from-motion, so it fails if there is no camera motion.

Source:

Felix Klose, Oliver Wang, Jean-Charles Bazin, Marcus Magnor and Alexander Sorkine-Hornung
Sampling Based Scene-space Video Processing
ACM Transactions on Graphics (Proceedings of SIGGRAPH), **2015***, 34*(4), 67:1–11

DOI: http://dx.doi.org/10.1145/2766920

URL: http://www.disneyresearch.com/publication/scenespace/

Video source:
Supplemental video.

Summary

- Spatiotemporal video processing
  = exploit visual information in the same video frame + over time
- Requires robust temporal correspondences:
  - Assuming static camera or stabilised video
    [Liu et al. '05, Bennet & McMillan '07, Richardt et al. '12, Wadhwa et al. '13, Liao et al. '14]
  - Additional sensor data, e.g. gyroscope
    [Karpenko et al. '11, Instagram '14]
  - 2D feature tracking
    [Collomosse et al. '05, Rav-Acha et al. '08, Rüegg et al. '13, Joshi et al. '15]
  - Structure-from-motion / multi-view stereo
    [Bhat et al. '07, Lieng et al. '12, Tompkin et al. '13, Kopf et al. '14, Klose et al. '15]

In the last 16 minutes, we have seen a large variety of video processing effects enabled by spatiotemporal video processing.

They all rely on exploiting visual information from a video both within a video frame, and also over time.

This requires robust temporal correspondences for aligning corresponding pixels over time, for which the discussed techniques use different approaches.

Some techniques assume a static camera or at least a stabilised video.

Other techniques use external sensor information, for example from a gyroscope.

Most techniques use 2D feature tracking, and some feed these feature trajectories into structure-from-motion or multi-view stereo, to obtain 3D scene geometry and camera motion.

While structure-from-motion provides the most powerful information one can extract from videos, in the form of 3D geometric from a 2D video, it is also the most fragile approach as structure-from-motion has many failure modes.

And with this, I would like to take any questions you might have at this point, while Jiamin gets ready to talk about motion editing in videos.