

SIGGRAPH 2015 Course

User-Centric Computational Videography

Christian Richardt

Intel Visual Computing Institute
MPI Informatik

James Tompkin

Harvard University

Jiamin Bai

Light, Co.

Christian Theobalt

MPI Informatik



Course website:

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

Course syllabus

- Introduction and welcome
[Richardt, 0:10]
- Background:
State-of-the-art video tools
[Richardt, 0:20]
- **Timeline editing**
[Bai, 0:20]
- Model-free video editing
[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]

Story Telling



2015-08-13

Jiamin Bai – User-Centric Computational Videography

3

Storytelling is essential for communicating ideas. When they are well told, stories help us make sense of information, appreciate cultural or societal differences, and imagine living in entirely different Worlds.

Story Telling

light reimagine
photography



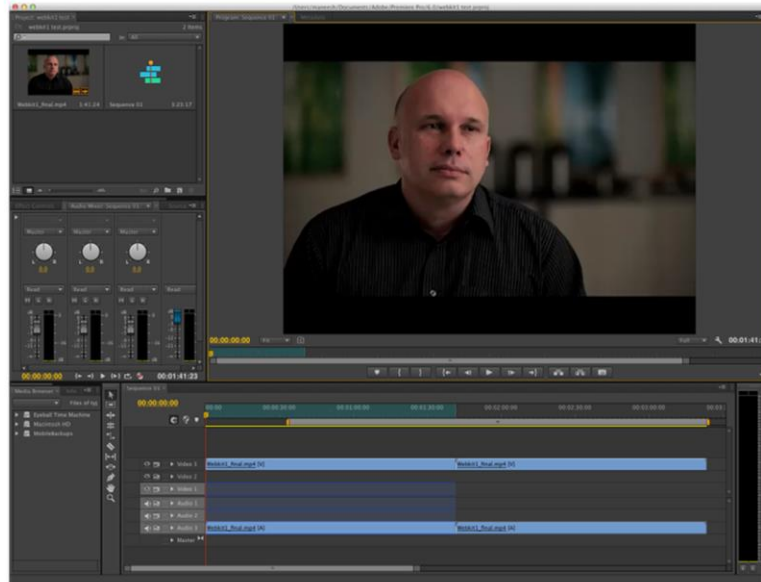
2015-08-13

Jiamin Bai – User-Centric Computational Videography

4

Audio/visual stories in the form of radio programs, books-on-tape, podcasts, television, movies and animations, are especially powerful because they provide a rich multisensory experience. Technological advances have made it easy to capture stories using the microphones and cameras that are readily available in our mobile devices, But, the raw media rarely tells a compelling story.

Story Telling



2015-08-13

Jiamin Bai – User-Centric Computational Videography

5

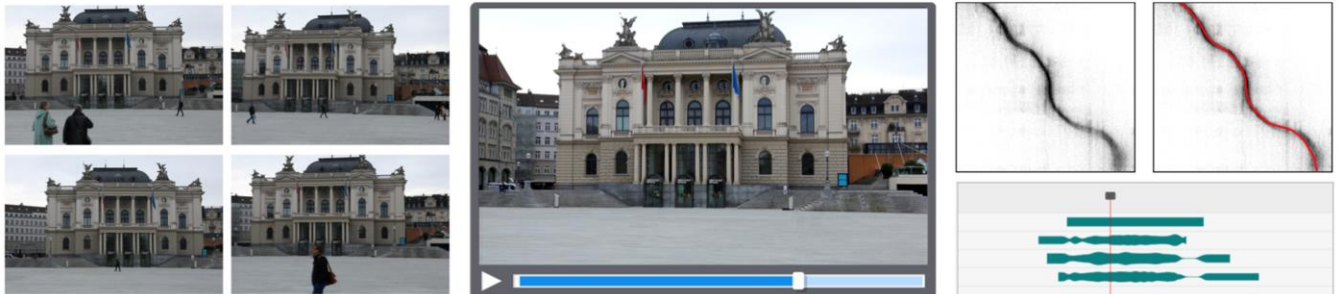
The best storytellers carefully compose, filter, edit and highlight the raw media to produce an engaging piece. Yet, the software tools they use to create and manipulate the raw audio/video media (e.g. Pro Tools, Premiere, Final Cut Pro, Maya etc.) force storytellers to work at a tediously low-level – selecting, filtering, cutting and transitioning between audio/video frames.

While these tools provide flexible and precise control over the look and sound of the final result, they are notoriously difficult to learn and are accessible primarily to experts.

In this talk I'll give an overview on recent projects that aim to significantly reduce the effort required to edit and produce high-quality audio/visual stories.

VideoSnapping

light reimagine
photography



VideoSnapping: Interactive Synchronization of Multiple Videos, Wang et al. 2014

2015-08-13

Jiamin Bai – User-Centric Computational Videography

6

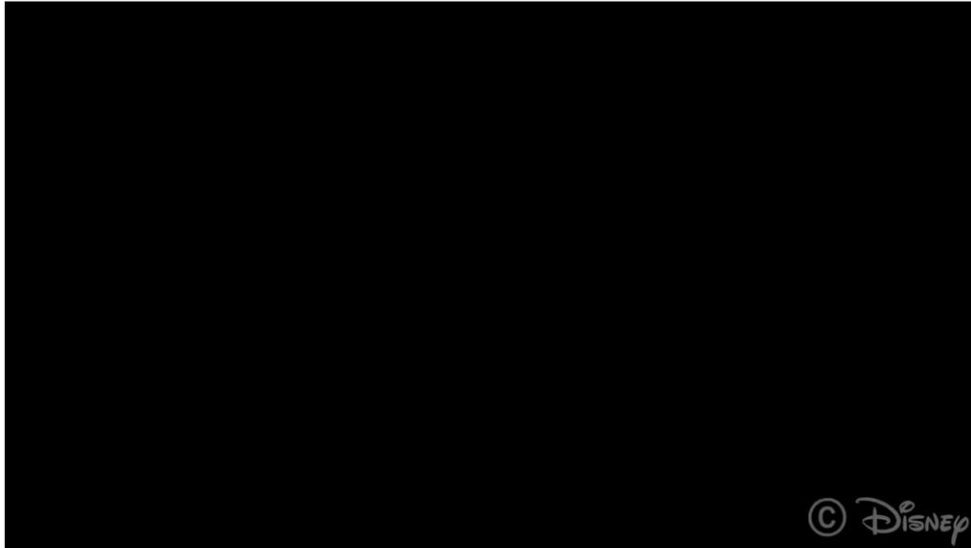
Trying to time align videos taken at different time points or from different cameras can be challenging.

One could try aligning them manually by comparing key frames and aligning them. However, if the cameras are not taken at the same time, one might have to stretch a video to match the other.

This work introduces a technique to snap 2 similar videos together. It computes features across videos and builds a distance matrix across frames. It then uses Dynamic Programming to find a non-linear mapping between the 2 videos. This allows the use to align videos together without manual effort.

VideoSnapping

light reimagine
photography



2015-08-13

Jiamin Bai – User-Centric Computational Videography

7

It can also be used to time align videos taken from multiple cameras. This way, the user can quickly time align the videos and focus his attention on telling the story, as oppose to figuring out the sequence in which they should come in.

Video Digest



Video Digests: A Browsable, Skimmable Format for Informational Lecture Videos, Pavel et al. 2014

2015-08-13

Jiamin Bai – User-Centric Computational Videography

8

There is an explosion of online lecture videos available on platforms such as youtube, coursea and EdX just to name a few.

However consumers of these platforms are largely using the same interface to browse through the video: a timeline.

There is no quick way to search or find sections that they are interested in.

Content creators can help by creating Video Digests. In this work, tools for creating Video digests are presented to allow content creators to quickly and efficiently create nice visual summaries of their lectures with section keyframes and section summaries.

Audio Stories

what is it about Detroit that launched all these musical careers, not just Motown, but so many others?

You know, so I just think it's a phenomenon. In New York, there's all the lights and the dances and the theater. In LA, there's the movies and whatever. But there was no reason for that to happen in Detroit. It, it's noted for cars and for factories. It's not necessarily just noted, even when, when the circuit or when the theater circuit was really big, while Detroit was big during those times, it wasn't a thing where people were leaving their homes trying to get to Detroit to be in show business. You know, so I just think it's a phenomenon.

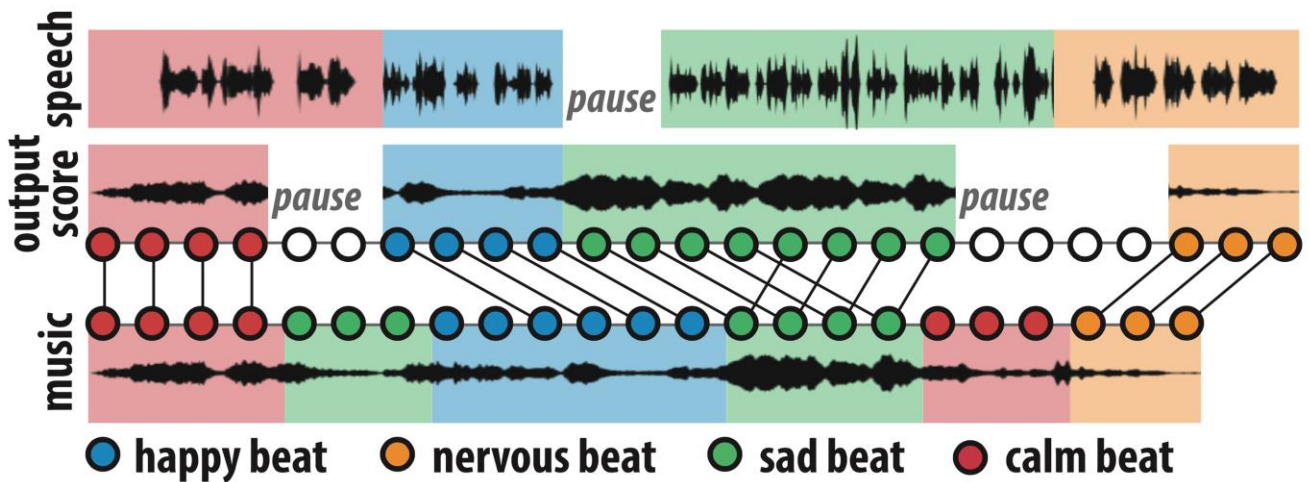
Music library	
+	Air <i>Hightschool Lover</i>
+	Jon Brion <i>You Learn</i>
+	Bill Frisell <i>Collier's Theme</i>
+	Liakt <i>Jaguar River</i>

2015-08-13

Jiamin Bai – User-Centric Computational Videography

10

Audio Stories



Generating Emotionally Relevant Musical Scores for Audio Stories, Rubin et al. 2014

2015-08-13

Jiamin Bai – User-Centric Computational Videography

11

Finding the right music to tell a story can be difficult.

This work helps to explore music collections through emotionally relevant tags such as happy, sad, calm.

The algorithm then helps match the right music to the right parts of the story line. It also stretches music and blends them seamlessly if needed.

Audio editing is an essential part of video editing.



AudeoSynth: Music-Driven Video Montage, Liao et al. 2015

2015-08-13

Jiamin Bai – User-Centric Computational Videography

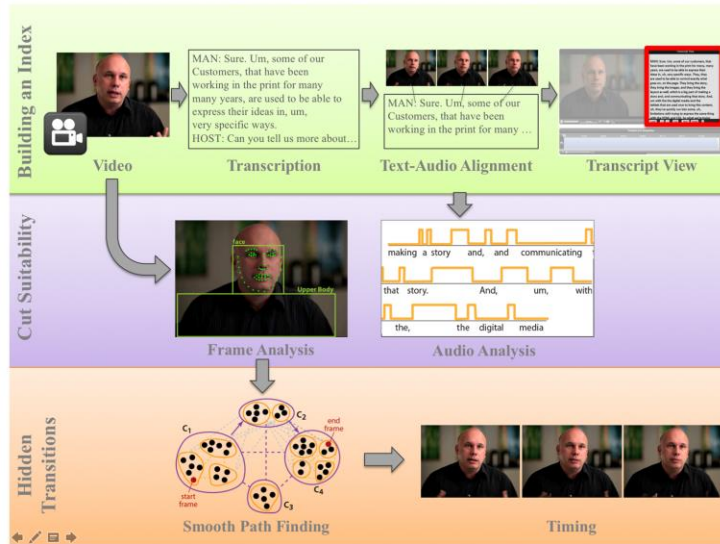
12

Let's say you have videos of an event you would like to capture and share with your friends. Chances are that you will have hours of footage, but your friends are not willing to spend more than a few minutes to watch your video.

Manually creating such a video montage to represent the event is non trivial and requires hours of work.

This work seems to automate the entire process. It takes a MIDI file and computes features of both the audio and video. It looks for pace as well as transition points. It then finds suitable parts of the video to use it to create the final output.

Interview Video



Tools for Placing Cuts and Transitions in Interview Video, Berthouzoz et al. 2012

2015-08-13

Jiamin Bai – User-Centric Computational Videography

13

Let me talk about a specialization of Video editing: Interview footage. We would have seen this often, in newsreports, interviews, and apple commercials.

It is not uncommon to shoot up to hours of footage and only use a few minutes worth in the end. Manually finding and editing it together is tedious. On top of that, cutting together separate pieces often will result in jump cuts which are not pleasant if the subject suddenly jerks.

This works address the problem by providing a UI which allows the editor to create the final out by manipulating a script instead. They use a transcription service to get accurate time stamped script. The editor edits directly on the script and ththe video and audio data is edited indirectly.

Suitable areas to cut are recommended by the system as it analyses facial motion as well as audio content.

To remove the unpleasant jump cuts, the system uses a hierarchal path finding to search for other frames that can smoothly interpolate between the cuts. It also uses optical flow to interpolate cuts if necessary.

Interview Video



2015-08-13

Jiamin Bai – User-Centric Computational Videography

14

Here the editor decides to remove a phrase

Interview Video



2015-08-13

Jiamin Bai – User-Centric Computational Videography

16

In comparison, here is what a jump-cut would look like. – Watch the hands

Interview Video



2015-08-13

Jiamin Bai – User-Centric Computational Videography

17

And here is the resulting smooth transition generated by our system.

Interview Video



2015-08-13

Jiamin Bai – User-Centric Computational Videography

18

Interview Video

Our result hidden transitions in blue
pauses in green

Comp. time:	
clusters	52m
hidden	8s
pauses	16s

2015-08-13

Jiamin Bai – User-Centric Computational Videography

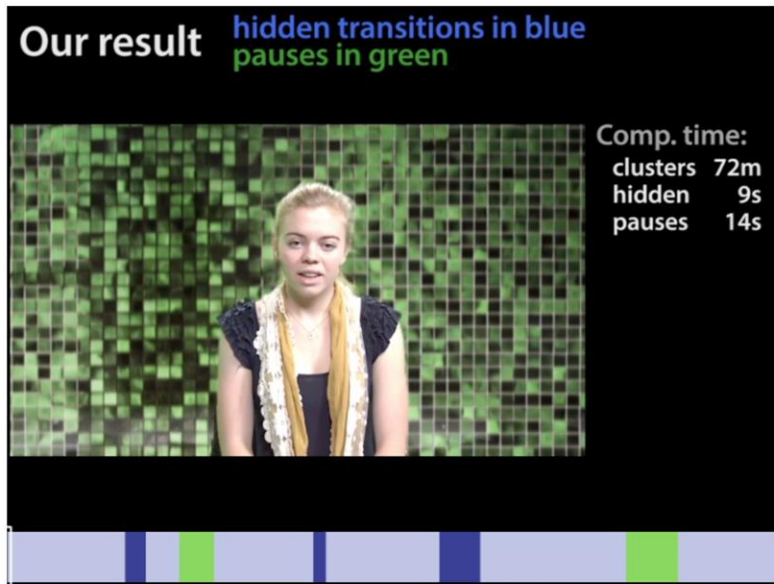
19

Interview Video



Another example with jumpcuts

Interview Video



Another example with jumpcuts


Interview Video



And one final example with jumpcuts

Interview Video

Our result hidden transitions in blue
pauses in green



Comp. time:
clusters 22m
hidden 5s
pauses 9s

And our result

Conclusion

- Use the right representation for editing content
 - Structures of external representation should match structure of mental representation
- Computation can alleviate tedious tasks
- Empower novice users to create high quality content