Video Tooning

Jue Wang, Yingqing Xu, Heung-Yeung Shum, Michael F. Cohen

Siggraph 2004

Presentation prepared by Kaleigh Smith, McGill University

Stylized Rending of Video

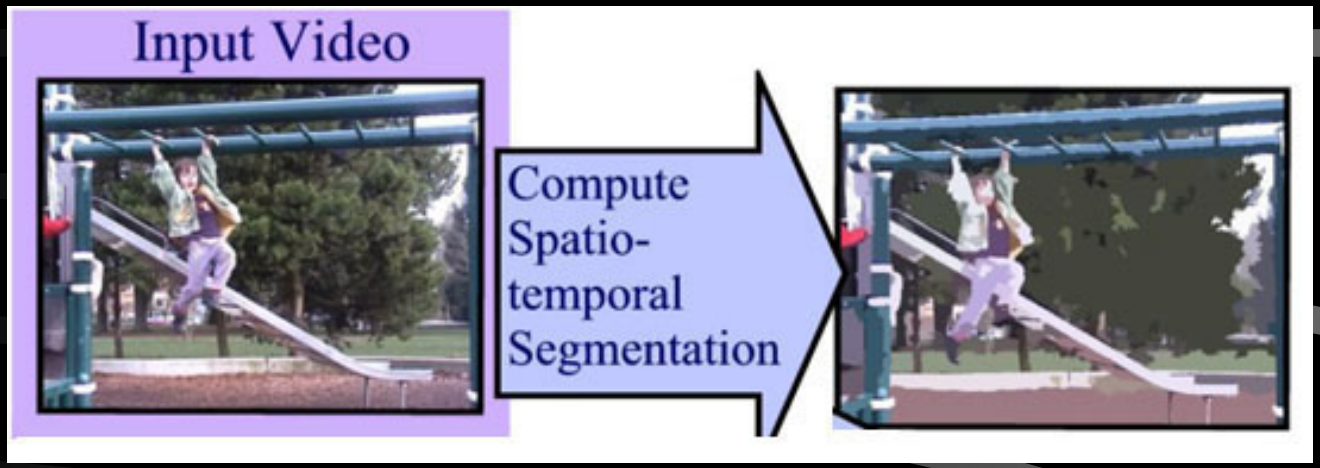
- Example and Motivation: Waking Life.
- Proprietary software by Bob Sabiston.

→ Show Waking Life clip.

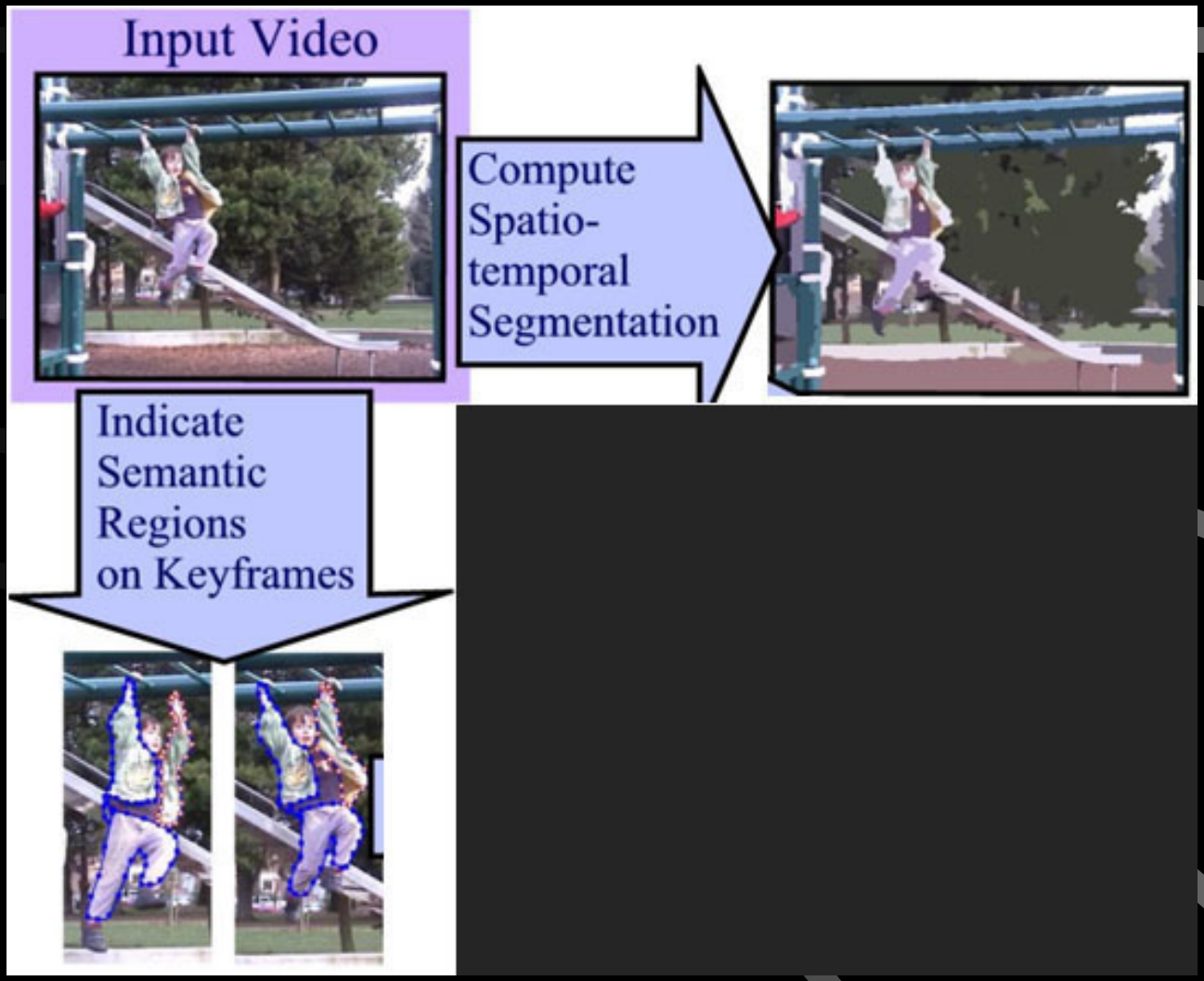
Outline

- System overview
- Related work
- Video segmentation
- User specification of video semantics
- Stylized rendering
- Summary and Discussion

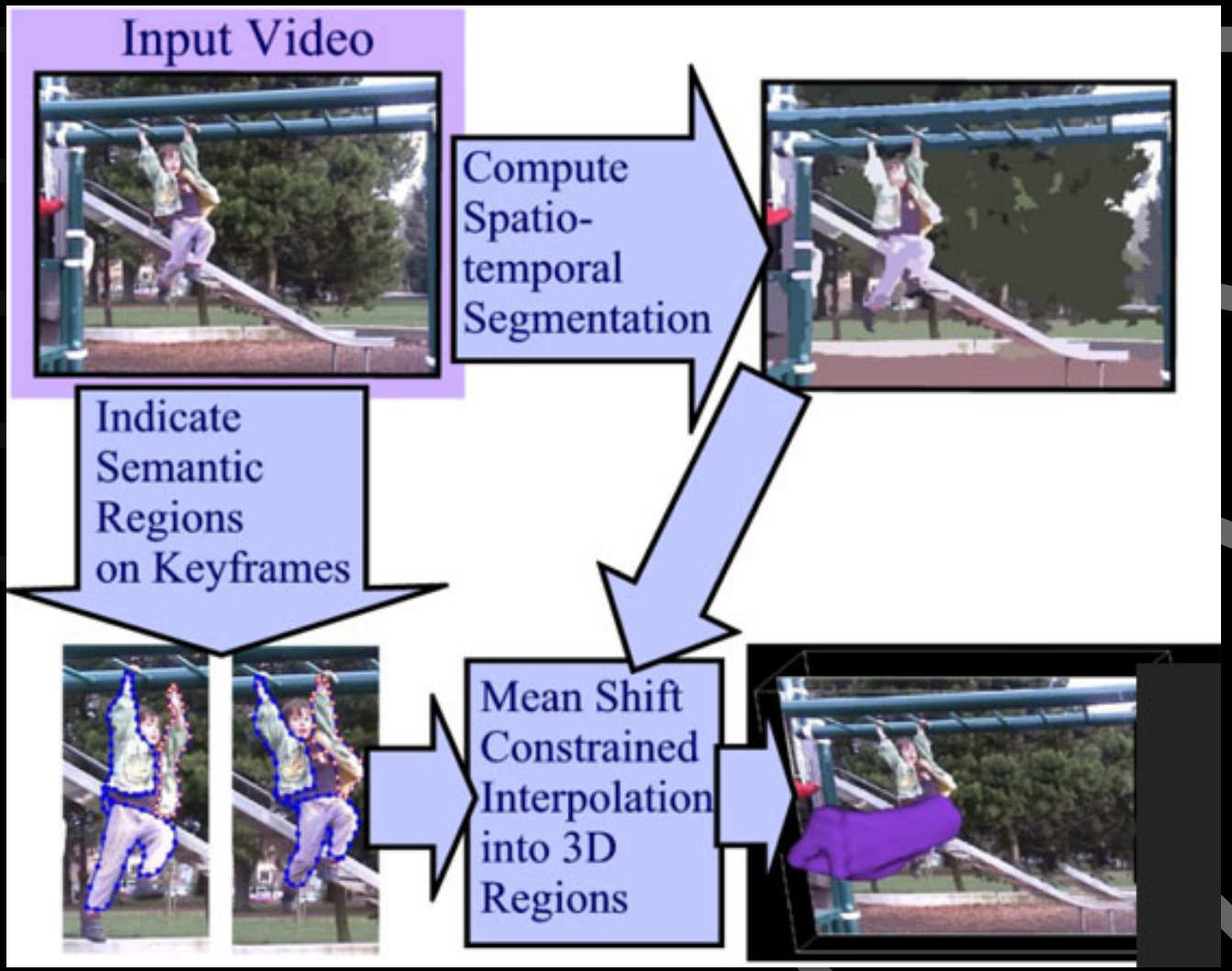
System Overview



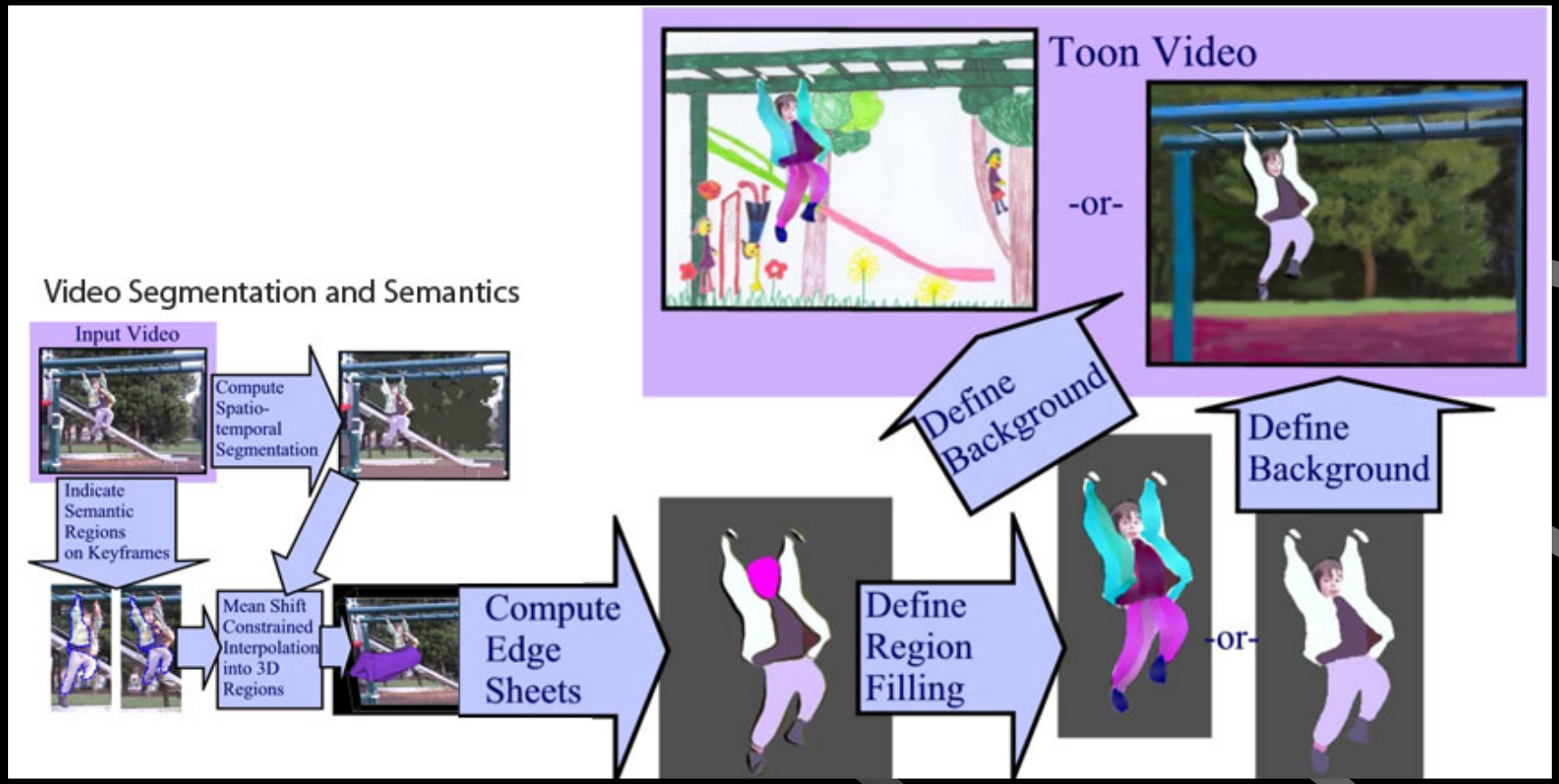
System Overview



System Overview



System Overview



System Overview

Criteria of successful system:

- Spatio-temporal consistency.
- Higher level semantic representation.
- Control over the style of the result.

→ Show Video Tooning clip.

Outline

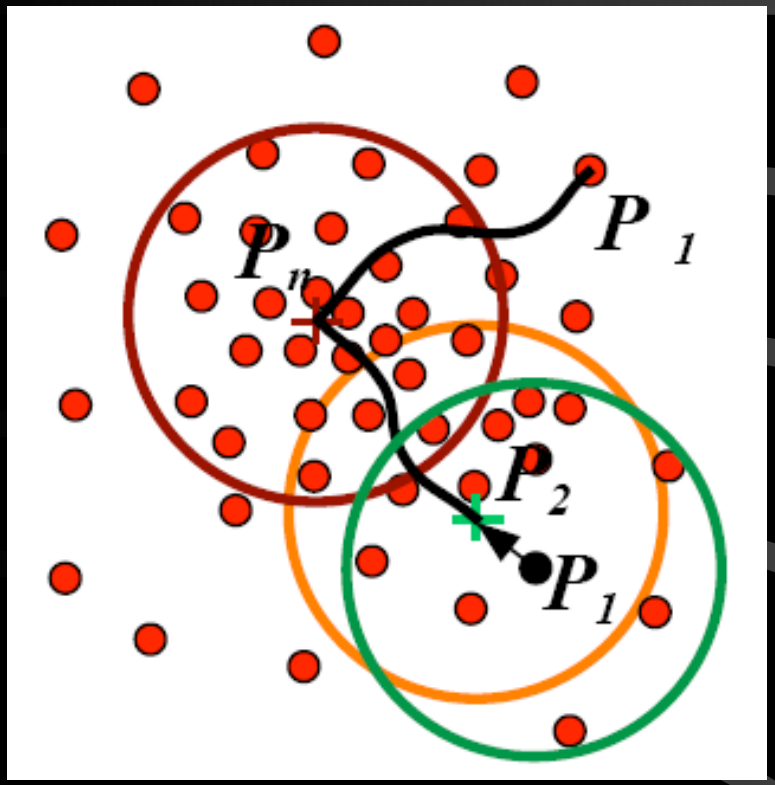
- System overview
- Related work
- Video segmentation
- User specification of video semantics
- Stylized rendering
- Summary and Discussion

Related Work

Video Segmentation and Spatio-temporal Coherence

- *Stylized Video Cubes*, A.W. Klein et al.
- *Spatio-temporal segmentation of video by hierarchical mean shift analysis*, D. DeMenthon, R. Megret.

Related Work: Mean Shift

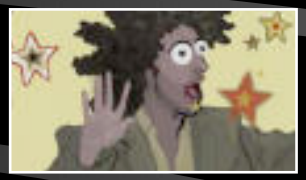


DeMenthon

- The mean shift point (cluster center) for P1is found by repeatedly re-centering the kernel at the centroid of points inside the kernel of P1.
- Kernel of influence gives spatial and colour distance measure.
- Pixels with same mean shift point belong to same segment.

Related Work

Stylized Rendering



- Waking Life

- *Stylization and Abstraction of Photographs,* D. DeCarlo and A. Santella.

Variable level of detail is a common goal of NPR styles.

Very Related Work

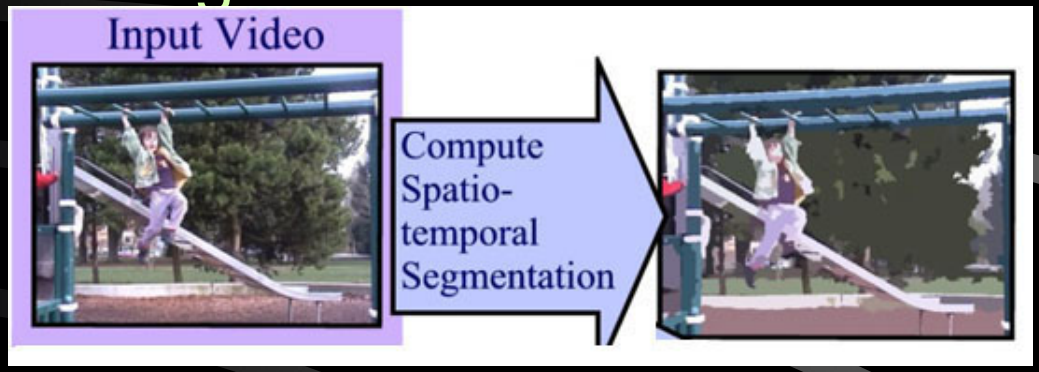
Video Paintbox Project, J.P Collomosse, D. Rowntree and P.M. Hall.

- Difference is that mean-shift segmentation is done in 2D on keyframes and associations are created between segments according to segment colour, shape and location.

→ Show Video Painbox clip.

Outline

- System overview
- Related work
- Video segmentation



- User specification of video semantics
- Stylized rendering
- Summary and Discussion

Video Segmentation

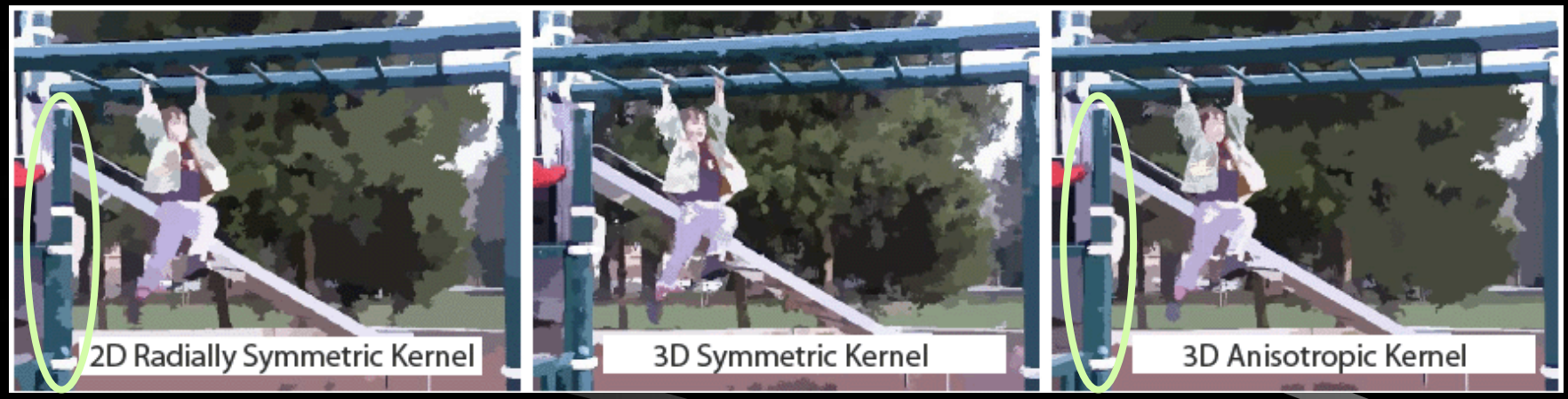
- Goal: create a representation that will provide a basis for spatio-temporal coherence.
- Application of standard mean shift image segmentation to 3D (x,y,t) video space.

Anisotropic Mean Shift Video Segmentation

- Main contribution in *Image and Video Segmentation by Anisotropic Kernel Mean Shift* (2004).

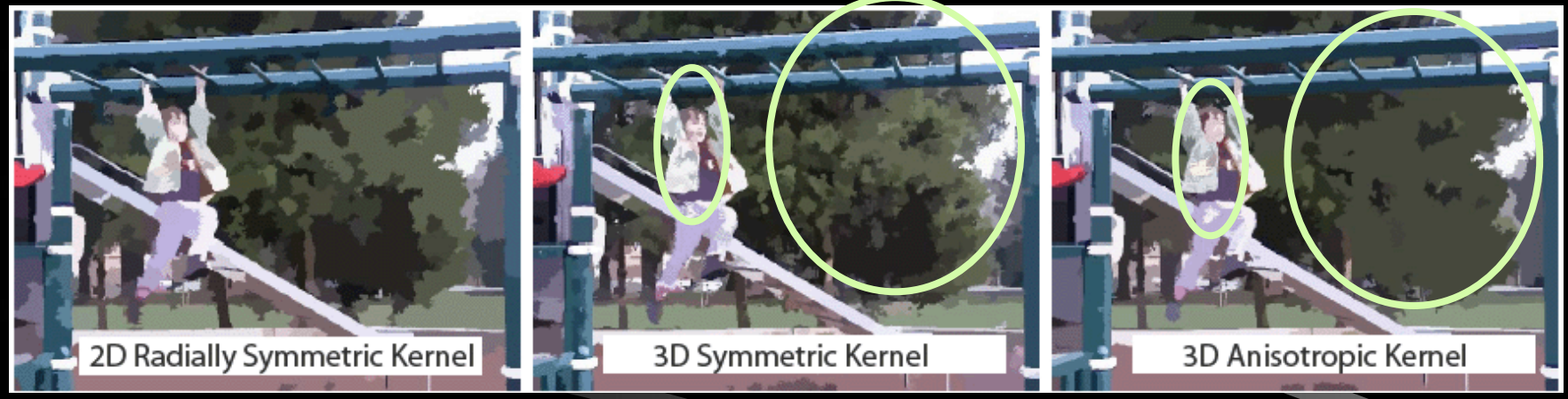
- Kernel bandwidth and shape for each pixel adapts to local video structure (density of similar pixels).

Anisotropic Mean Shift Video Segmentation



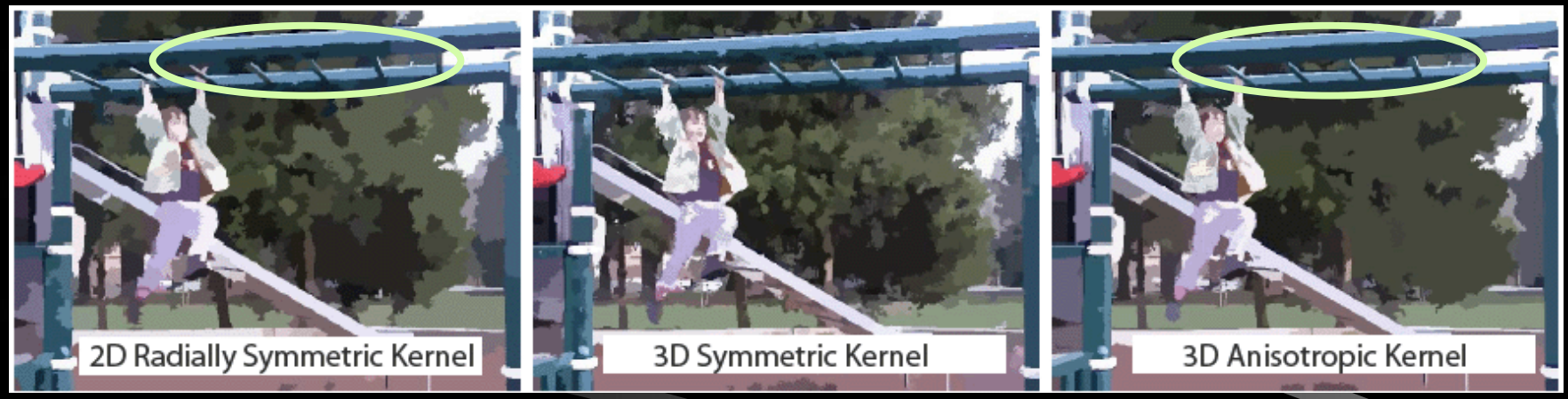
- Segments more smooth
- Variable coarseness (level of detail in foreground vs. background)
- Thin objects better segmented

Anisotropic Mean Shift Video Segmentation



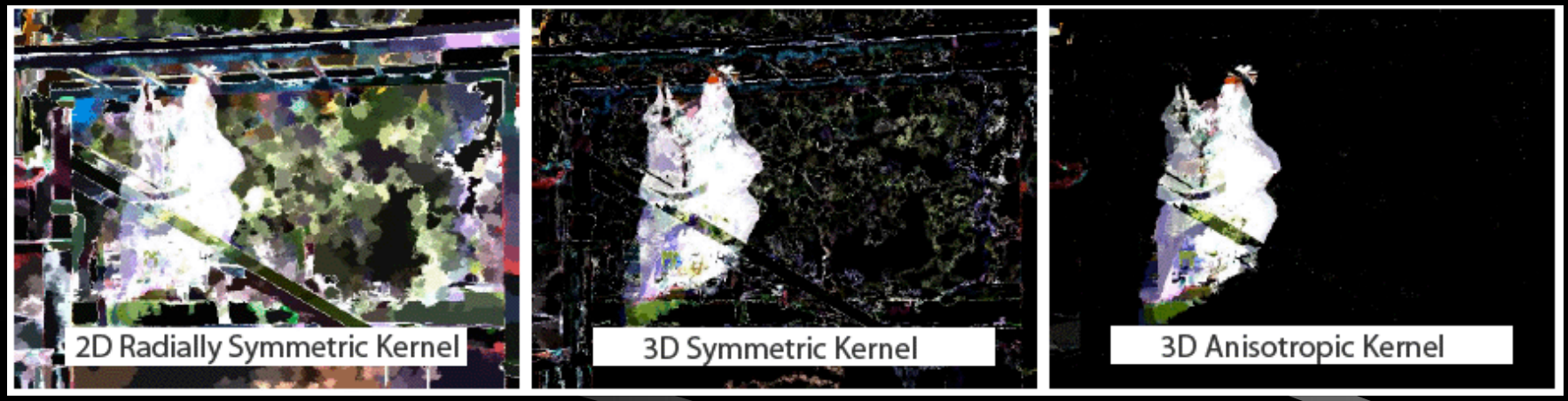
- Segments more smooth
- Variable coarseness (level of detail in foreground vs. background)
- Thin objects better segmented

Anisotropic Mean Shift Video Segmentation



- Segments more smooth
- Variable coarseness (level of detail in foreground vs. background)
- Thin objects better segmented

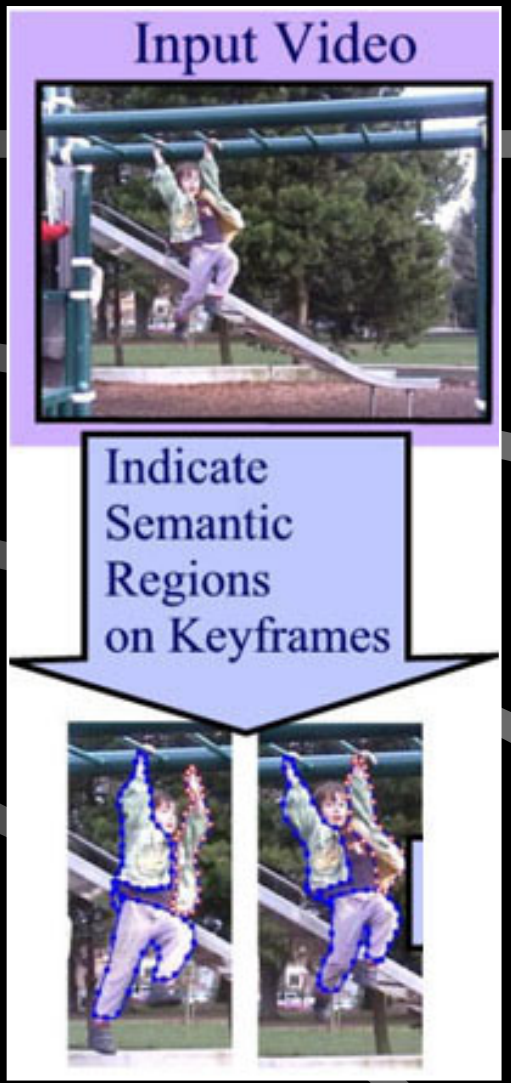
Anisotropic Mean Shift Video Segmentation



- Total absolute differences across 9 frames in a 10 frame sequence.
- Clean segmentation of moving girl.

Outline

- System overview
- Related work
- Video segmentation
- User specification of video semantics
- Stylized rendering
- Summary and Discussion



User specification of video semantics

- Create high level semantic REGIONS from low level segments.
- User inputs boundaries and keypoints.
  - These boundaries used to create 3D unions of mean-shift segments that belong to a region.

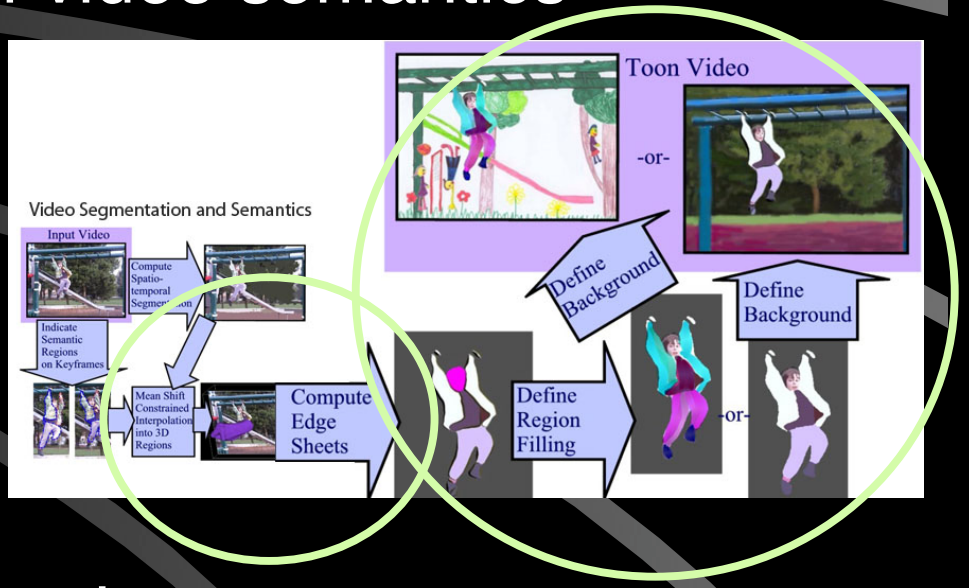
User specification of video semantics: Create 3D regions

- Interpolate user-defined boundary points and points on mean-shift segments to determine region at intermediate frames.

- Goals:
  - Smooth region shape (user points).
  - Minimize temporal difference (user points and points on mean-shift segments).

Outline

- System overview
- Related work
- Video segmentation
- User specification of video semantics
- Stylized rendering



- Summary and Discussion

Stylized Rendering

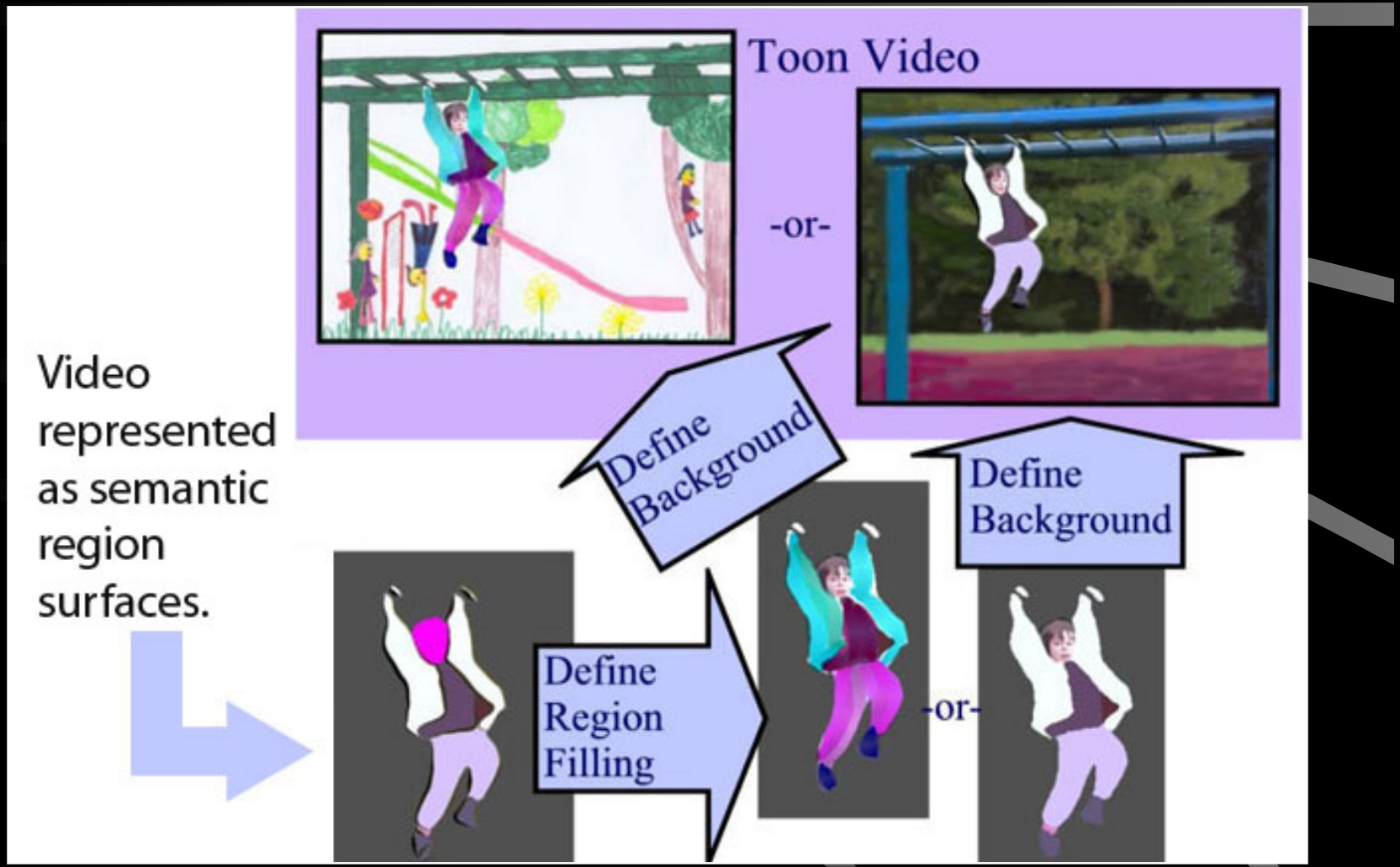
- Video is now represented as a set of pixelized 3D regions.

- Two major steps in rendering:
  1. Construct and smooth semantic region surfaces.
  2. Specify styles for rendered video.

Region surface reconstruction

- Pixelized semantic regions converted into 3D polyhedral surfaces using the marching cubes algorithm.
- Smoothing of semantic regions.

Stylized Rendering



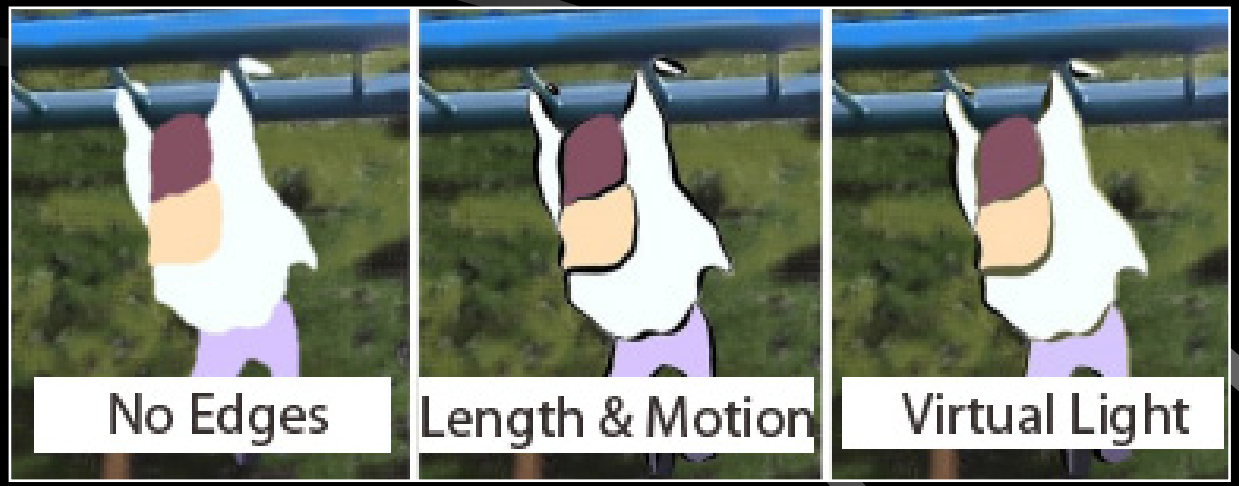
Edge sheets

- Used to add solid strokes in final rendering.
- For temporal coherence of edges, construct 2D edge sheets (region wrappers).

DeCarlo & Santella edge style

Edge sheets

- Render edges in different styles.
- Edge style affected by:
  - edge importance weight
  - edge length, motion & direction
  - virtual light direction



Region colouring

- Semantic regions filled with weighted combination of colours:
  - user defined colour
  - original pixel or average pixel colour
  - define subregions with distinct colours

3D Paint Strokes

- User adds paint strokes within regions at keyframes – creates ‘Stroke Sheets’.
- Authors provide a method to ‘flow’ paint strokes between keyframes.

Outline

- System overview
- Related work
- Video segmentation
- User specification of video semantics
- Stylized rendering
- Summary and Discussion

Stylized Results

- Show Video Tooning full video.

Performance

- Automated mean-shift runs overnight for 300 frame video.
- User interaction lengthy (2 hours) for short video example.
- Optimization of regions and edge and stroke sheets an additional ½ hour.
- No measure on perceived temporal coherence of stylized video.

Future Work

- Other stylistic choices for region filling, edges and paint strokes.
- Enhance interface to allow for segments to be cut (currently only segment joining).
- Video representation is resolution independent in space and time: working on vectorized encoding as a compression technique.

Discussion

- Very similar to the framework presented in Video Paintbox.
- The main contribution seems to be the use of a 3D video segmentation technique before user interaction and stylized rendering.

Discussion

- Video results are temporally coherent (at the expense of speedy video segmentation?)
- Enables users to specify semantics (easily? complicated regions that split into two?)
- Video results show a wide variety of styles (which styles can not be achieved?)

Questions?