IUPUI

Automatic Heterogeneous Video Summarization in Temporal Profile

Hongyuan Cai, Jiang Yu Zheng
Indiana University Purdue University Indianapolis
Indiana, USA

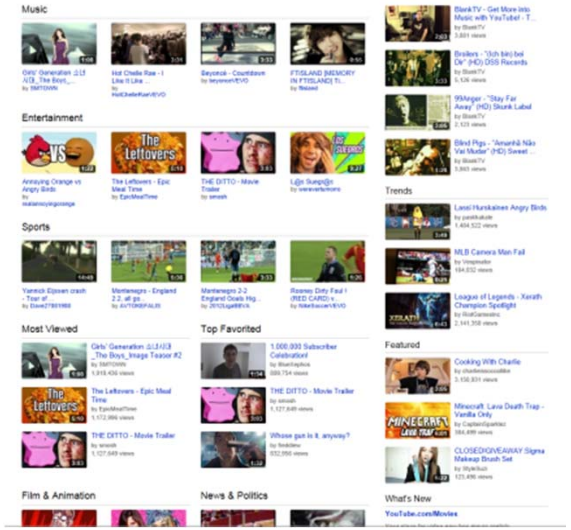
What is Temporal Profile of video

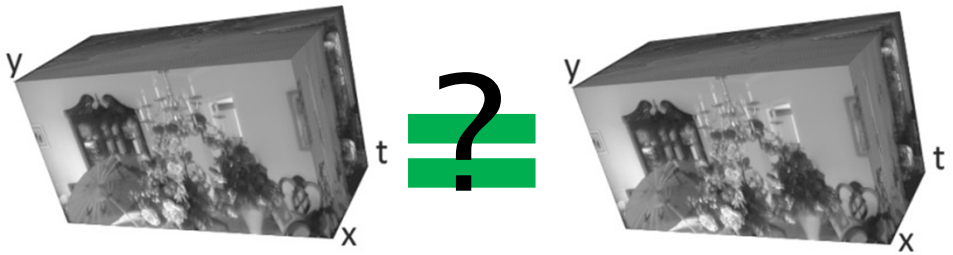
- Involving the time axis for selecting frames
- A space axis preserving shape and spatial layout of scenes in clips

- Different from spatial indexing of video so far
  - Key frames: find invariant video section and pick up representative frame
  - Stitching background and Onion-skinning foreground

Purposes – BigData

- Indexing large video DB for video sharing sites
- Filling in empty video track for video editing and retrieval
- Avoid expensive video matching for copyright
- Event description and detection
- Action and behavior discovery



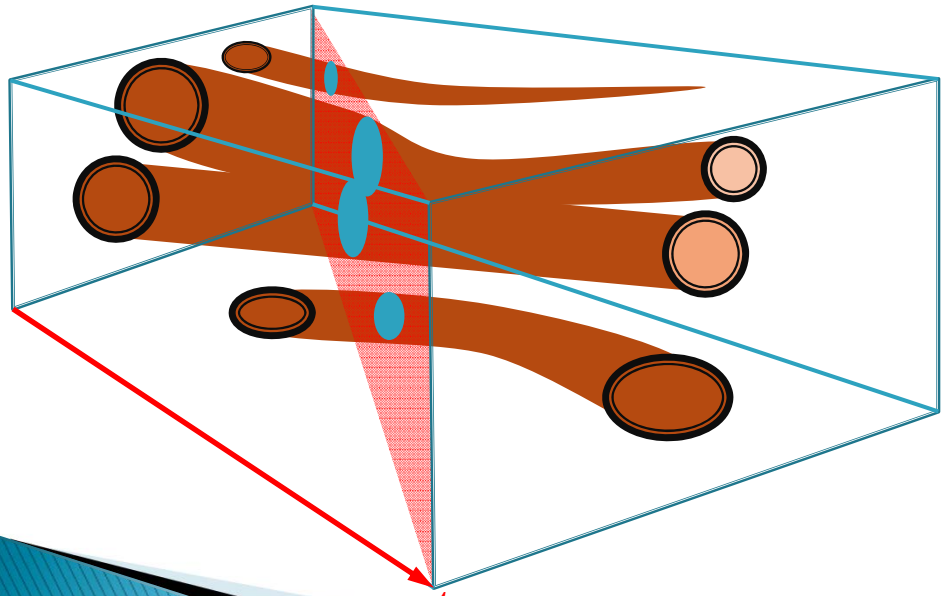


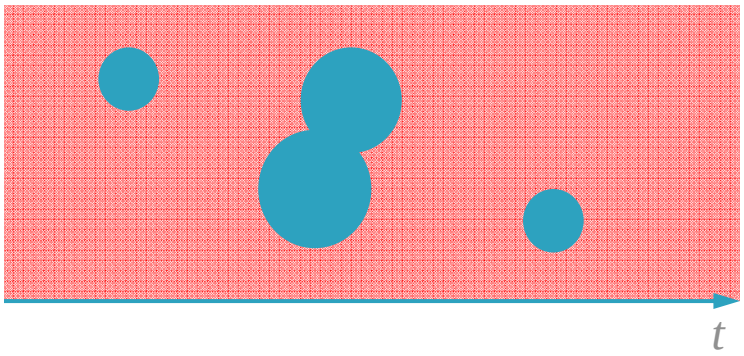
Why temporal digest?

- Spatial key frames are restrictive
  - Lack of temporal continuity
  - Mosaic of multiple poses of an action/events is confusing and cluttered
- Our temporal video profile
  - Continuous in time order and detailed to frame
  - Panorama property: Capable of having a glance at the whole video scene
  - Preserving shape to some extent

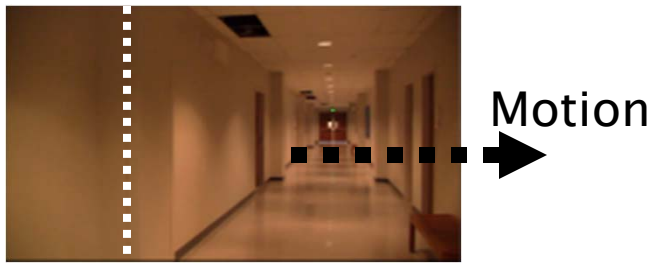
How to create a temporal profile?

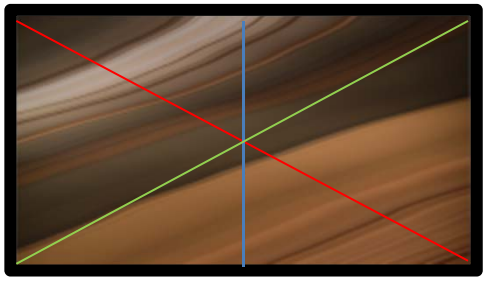
- Cutting a pixel slice across flow in video volume
- No stitching of regions from different frame based on matching – many camera motions do not allow frame stitching or guarantee correct matching





A slice crossing the major flow



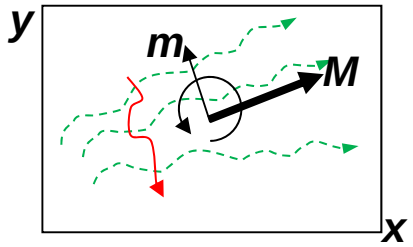

Flow trace

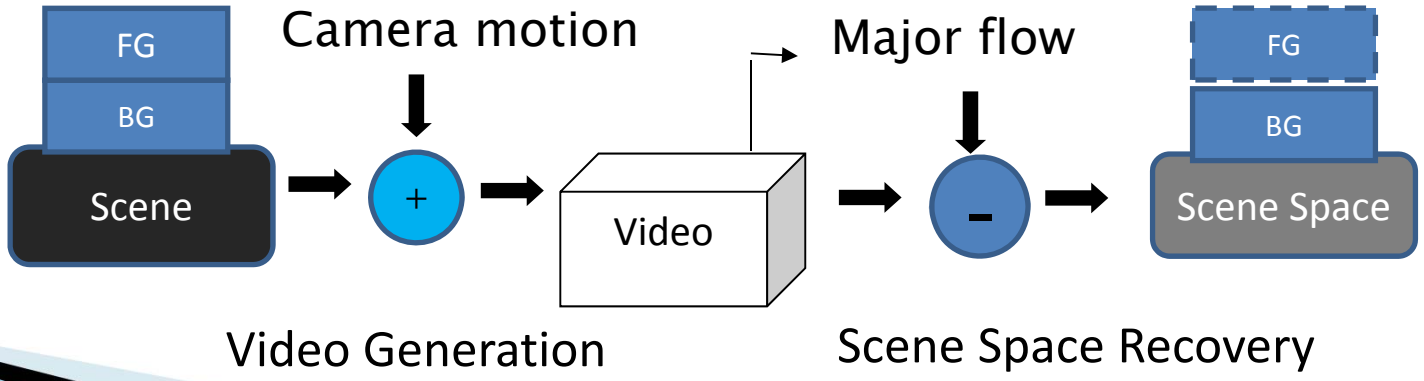
- The slice crossing the major flow for displaying the shape of structure
- Aligning a slice with the flow only shows motion traces in the slice

Shapes

Video from Motion and Scene

- Camera motion determines the major flow of background
- Dynamic objects determines the minor flow of foreground
- Separate the motion and scenes as flow and profile





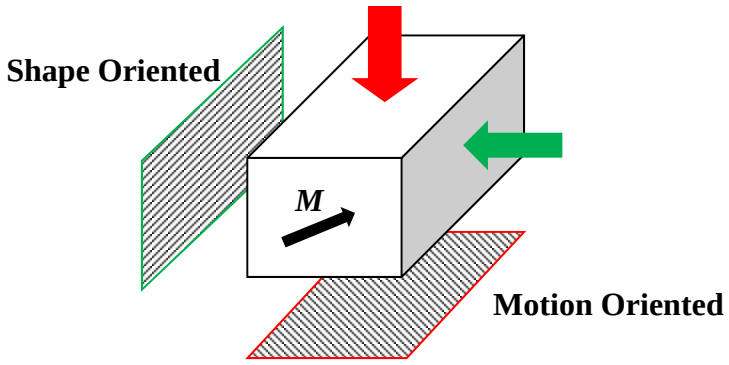
Profile and traces

- Show shapes without motion

- Show motion without shape

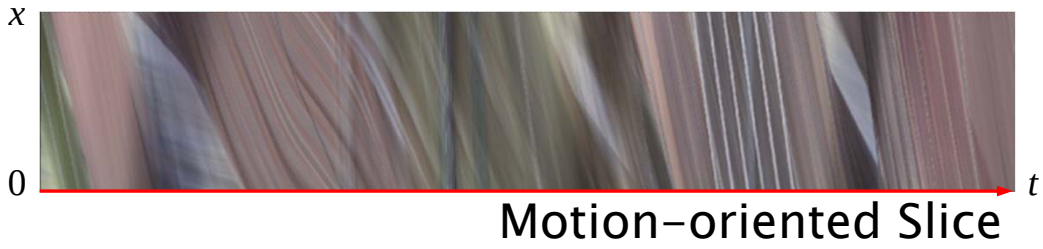
- To show both – watching video itself

Collecting Condensed Images



$$C_y(t,x) = \frac{1}{h}\sum_{y \in C} I(x,y,t)$$

$$C_x(t,y) = \frac{1}{w}\sum_{x \in C} I(x,y,t)$$



Motion-oriented Slice

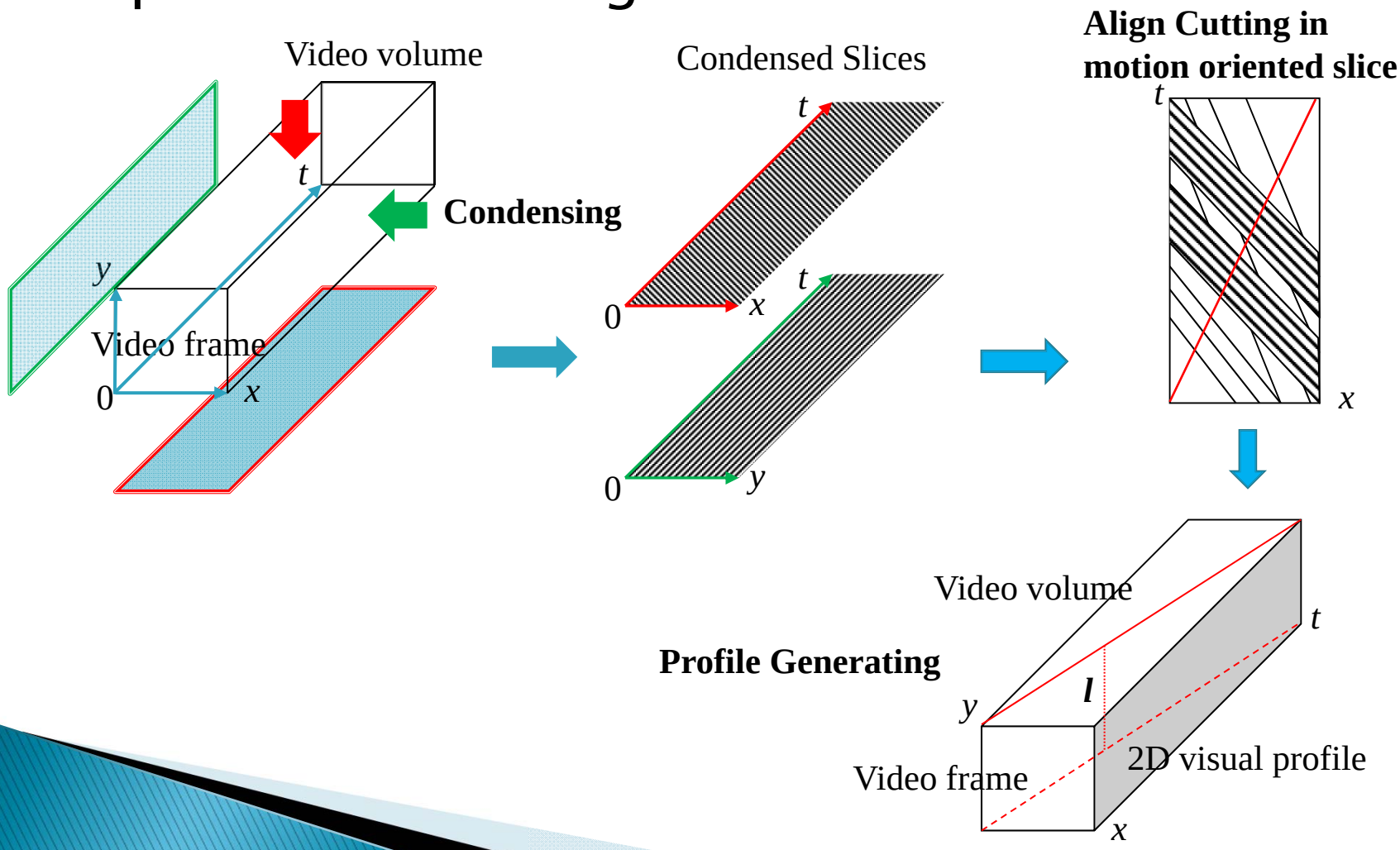
Shape-oriented Slice

Displays more traces from the linear features in the video

Shows more blurred shape than traces

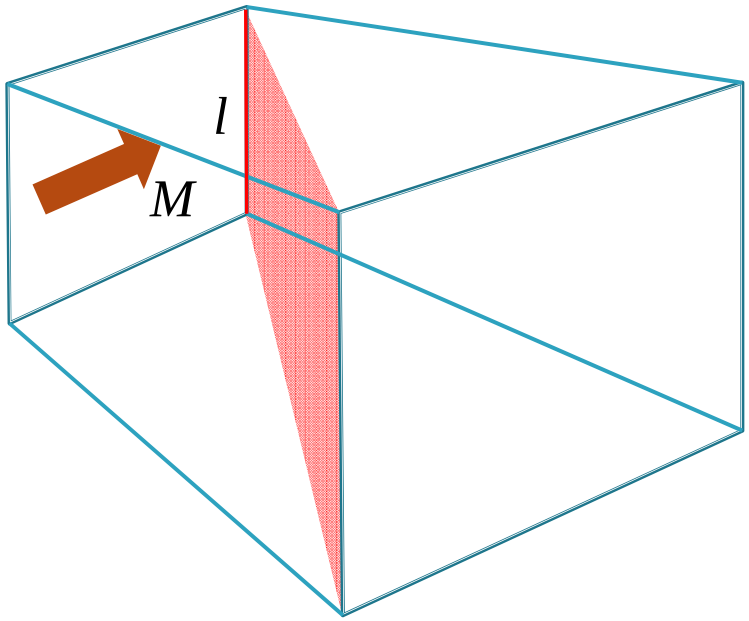
Camera Motion Understanding

- Spatial condensing of color



Aligning and moving a sampling line

A sweeping line for spatial integrity in the generated profile



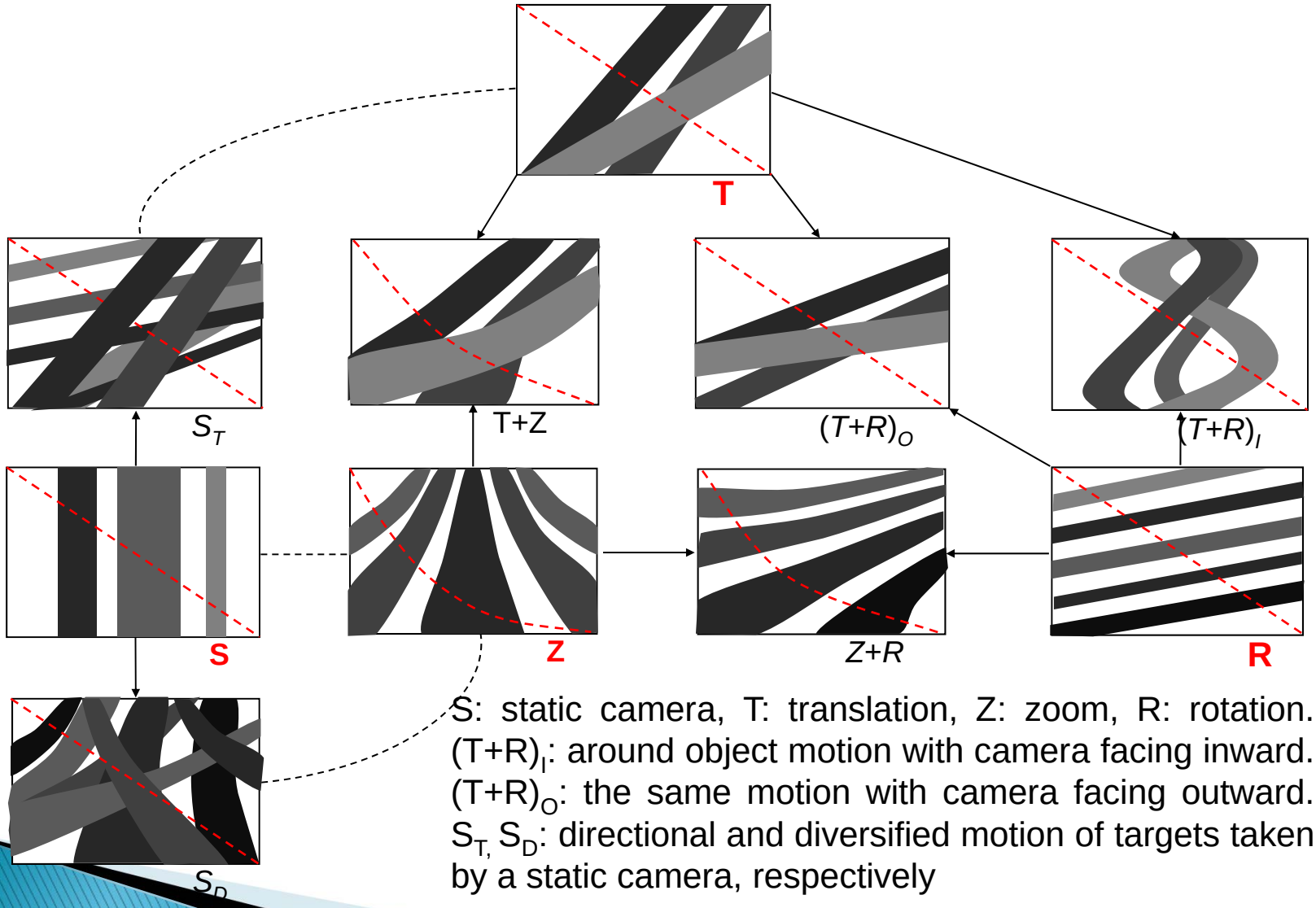
- Line alignment direction (vertical or horizontal slice for profile)
- Line moving direction in the volume in the volume – Diagonally across major flow
- Trajectory bending according to convergence factor

Is volume cutting possible for general videos ?

Camera motion	Static	Zoom	Rotation	Translation
Camera action Operation/work (combination of motion)	Crowd		Pan/tilt	Vehicle/rail
	Walking		Around-target	
			Crane	

Image flow	Flow without convergence, BG zero flow	BG flow converge	Equal velocity	Same direction, different velocity	Different moving directions
	Diversified flow		Directional flow		

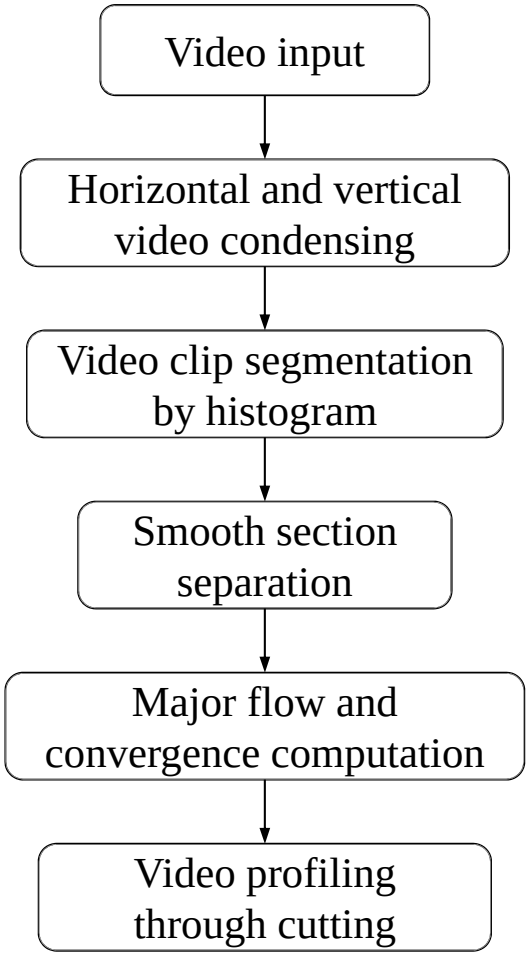
Major flow for typical camera actions



S: static camera, T: translation, Z: zoom, R: rotation.
$(T+R)_I$: around object motion with camera facing inward.
$(T+R)_O$: the same motion with camera facing outward.
S_T, S_D: directional and diversified motion of targets taken by a static camera, respectively

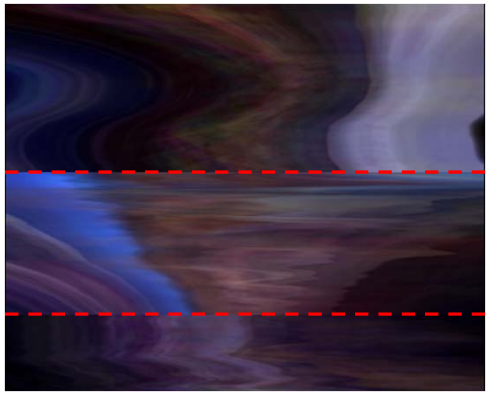
Approach to obtain temporal profile

- Video volume segmentation
- Automatic detection of motion
- Design a *cutting* in the video volume for each type of video clip
- Sweep a sampling line across the video volume



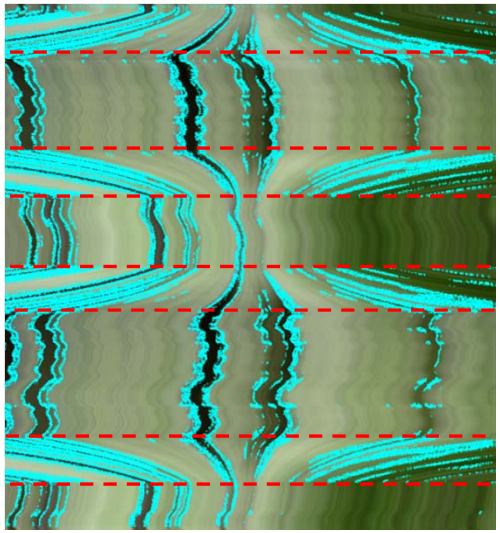
Video clip segmentation

- Camera transitions
  - Histogram Differentiation



- Smooth Camera Operations

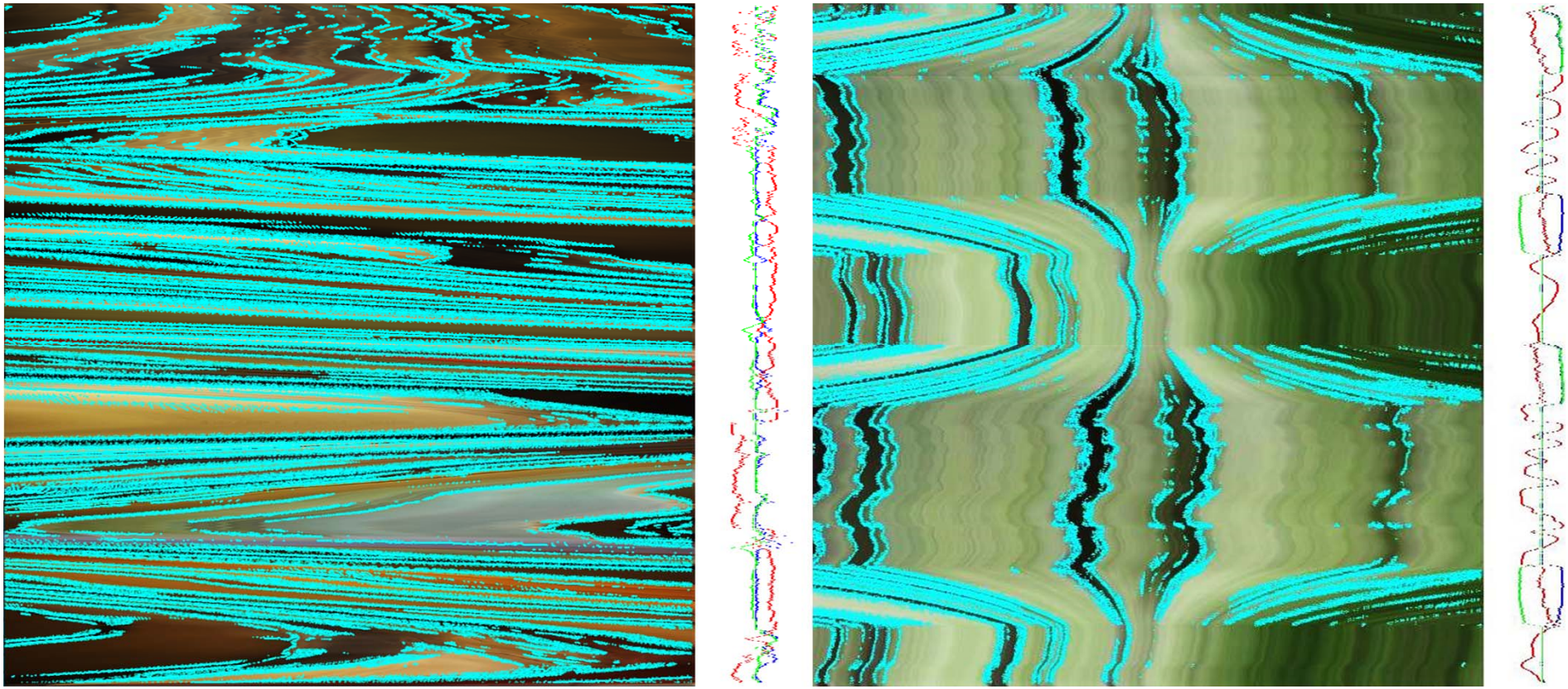
	small $\sigma_x(t)$	large $\sigma_x(t)$
$\lvert v_x(t)\rvert$ small	Static camera and scene	Zooming
$\lvert v_x(t)\rvert$ large	Camera pan	Pan+zoom, translation



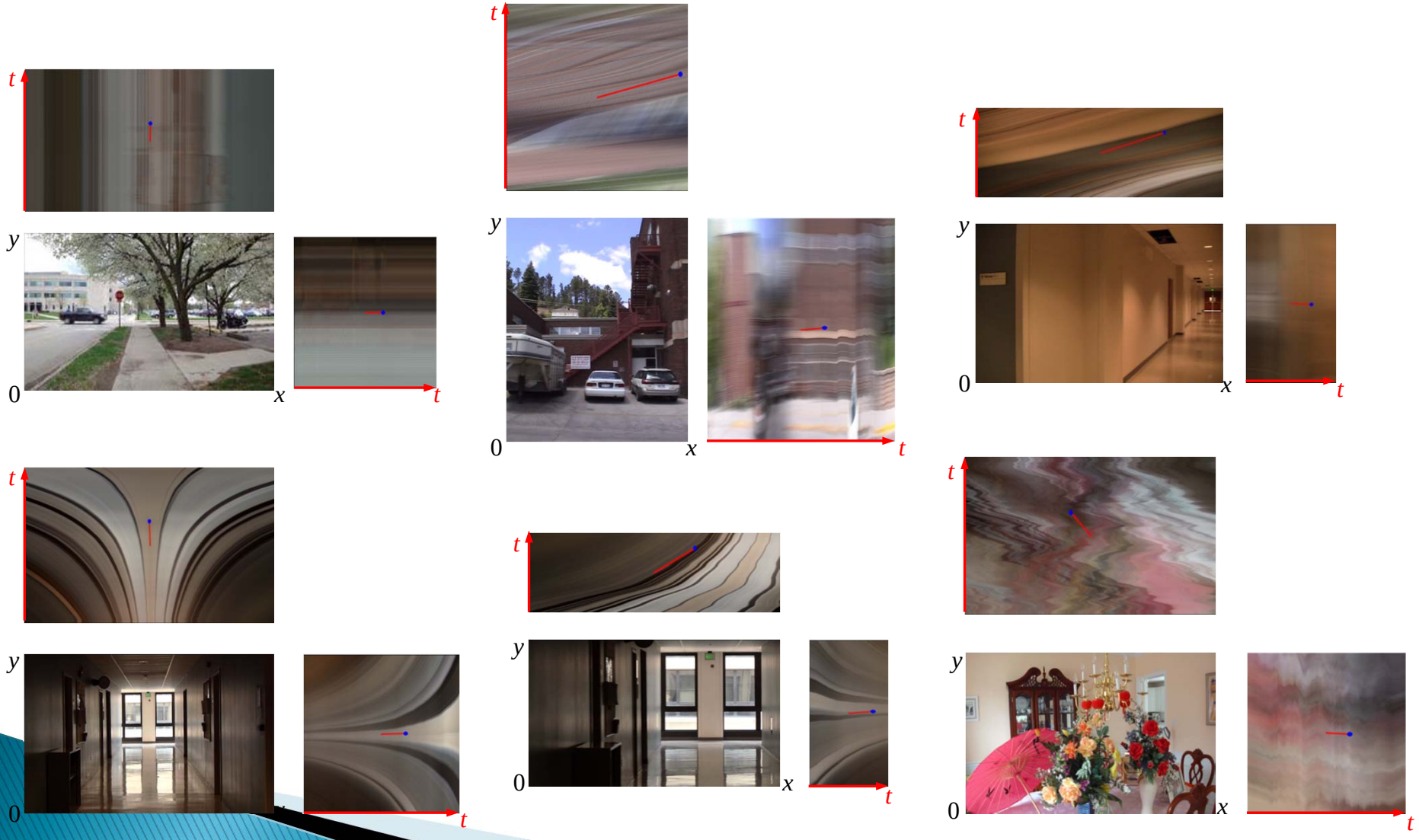
$v_x(t)$: major flow component
$\sigma_x(t)$: variance of major flow component

Flow properties: flow direction and divergence

- Average trace direction in each frame $v(t)$
- Variance of the trace direction in each frame $\sigma_v(t)$
- Convergence factor $\kappa(t)$



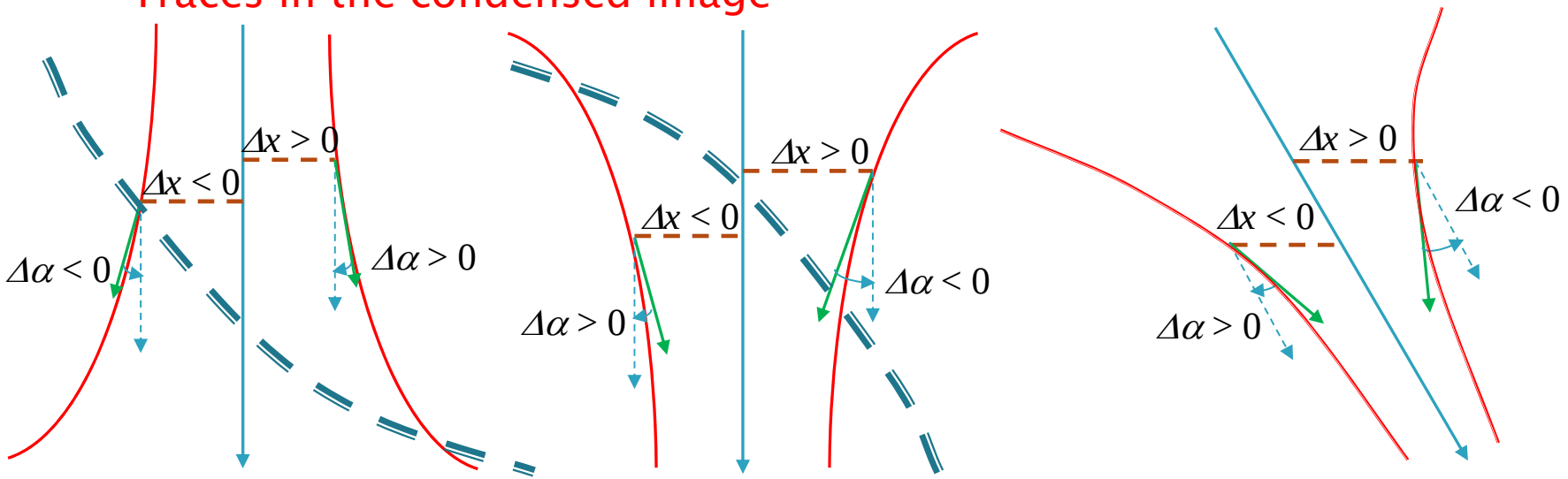
Extracting major flow direction



Convergence Factor – zooming

$$\kappa(t)=\sum_i (e_i(t,x_i)-v_x)\times sign\left[x_i-(x_m+v_x(t-t_m))\right]=\begin{cases}<0 & converge: zoom-out\\ >0 & diverge\ \ : zoom-in\end{cases}$$

Traces in the condensed image



Zoom In

$\Delta x \cdot \Delta\alpha > 0$

Zoom Out

$\Delta x \cdot \Delta\alpha < 0$

Zoom Out + Pan

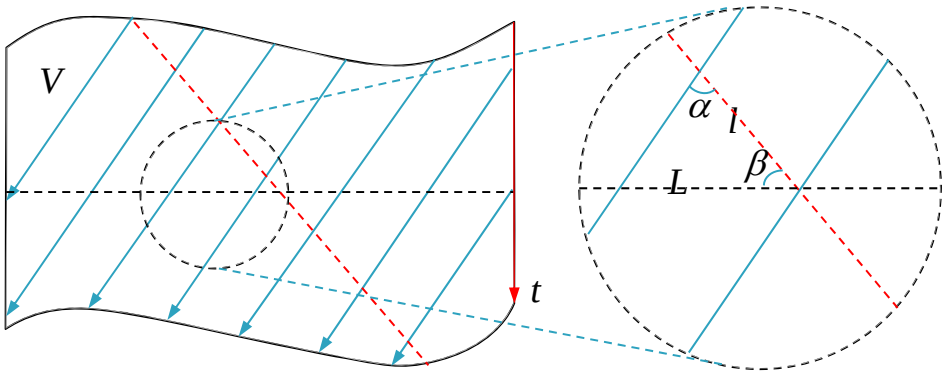
$\Delta x \cdot \Delta\alpha < 0$

Temporal Mode and Shape Mode

- The slice cutting is determined by section length
  - Profile different from the perspective projection
  - Shape distortions

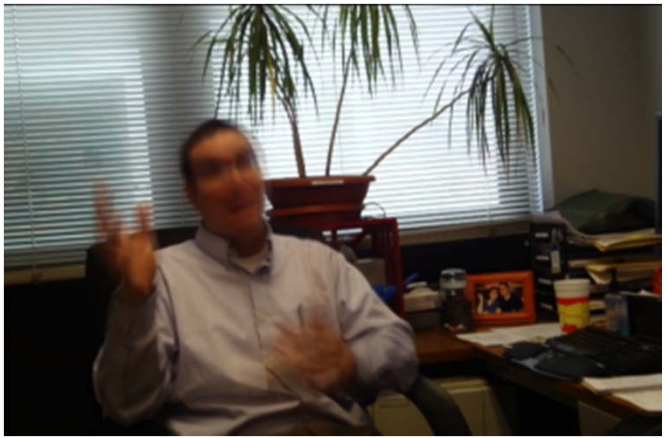
- Shape Mode is developed to solve this problem
  - Work only on profiles with camera pan/translation

$$L(t) = l(t)\frac{\sin\alpha(t)}{\sin(\alpha(t)+\beta)}$$



Static BG and dynamic FG – slow cutting

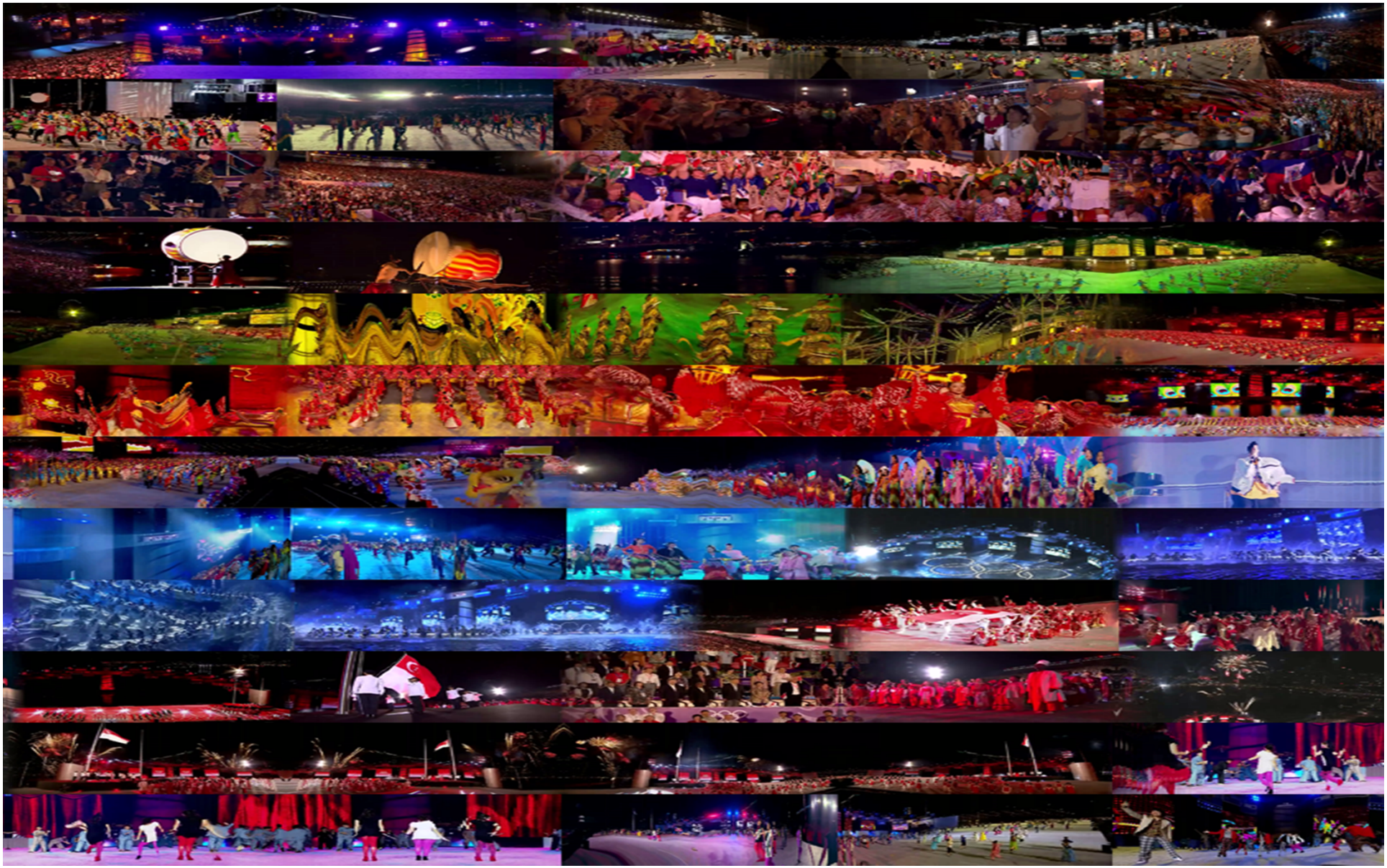
- Motion blur along the major flow direction to show FG motion



Consecutive pans with dynamic FG – fast cutting

Forward motion with pan

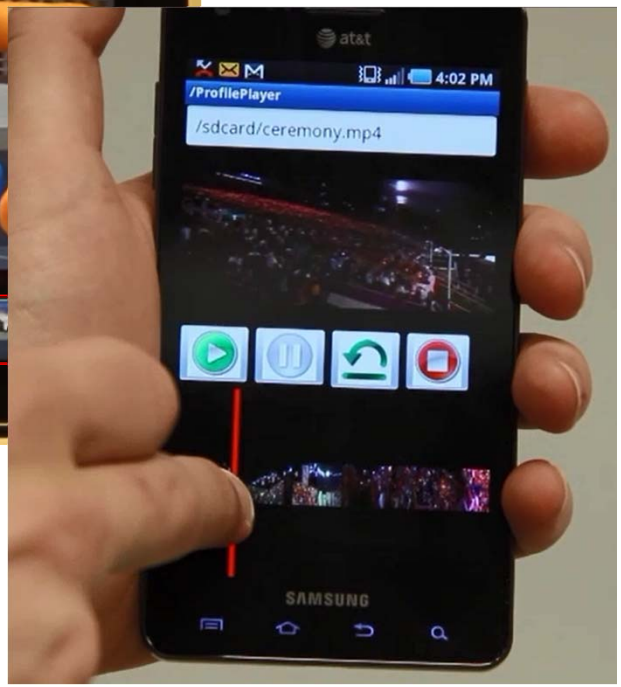
Video wall



Applications – mobile video browsing

PC

Pad



Smart Phone

Applications – surveillance

- Count passing people, which is easier than counting in overlapped frames because the data in profile is non-redundant