Improved Image Boundaries for Better Video Segmentation

Anna Khoreva[1], Rodrigo Benenson[1], Fabio Galasso[2],
Matthias Hein[3], Bernt Schiele[1]

[1] Max Planck Institute for Informatics, Germany
[2] OSRAM Corporate Technology, Germany [3] Saarland University, Germany





Video Segmentation

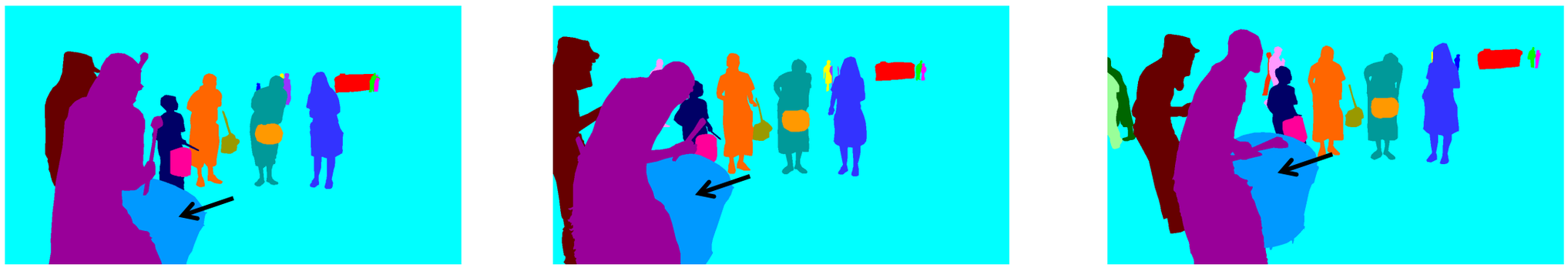
Goal: Find segments which describe the video in a way that mimics human annotations

Video

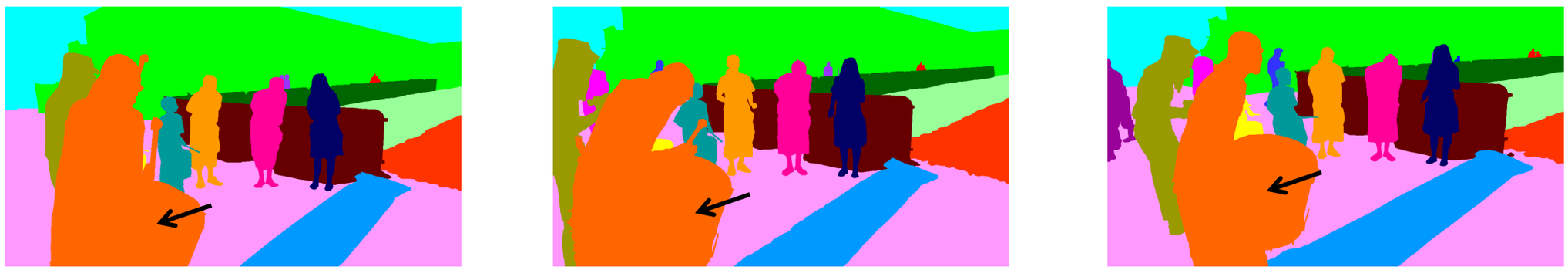
time

Human annotations:

Annotator 1



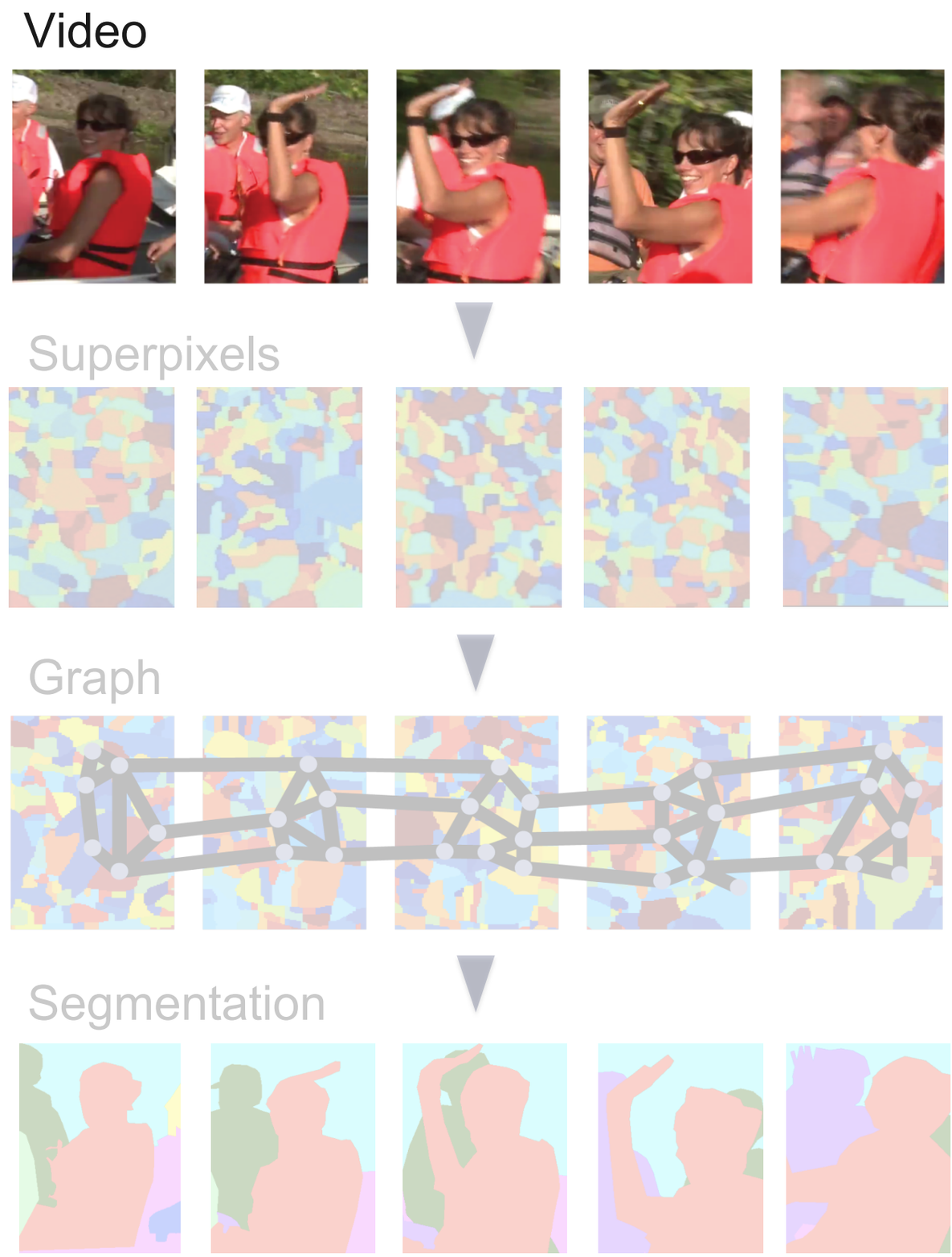
Annotator 2



Graph-Based Video Segmentation

[Fragkiadaki & Shi CVPR'12], [Ochs et al. PAMI'14], [Galasso et al. CVPR'14], [Yi & Pavlovic ICCV'15]

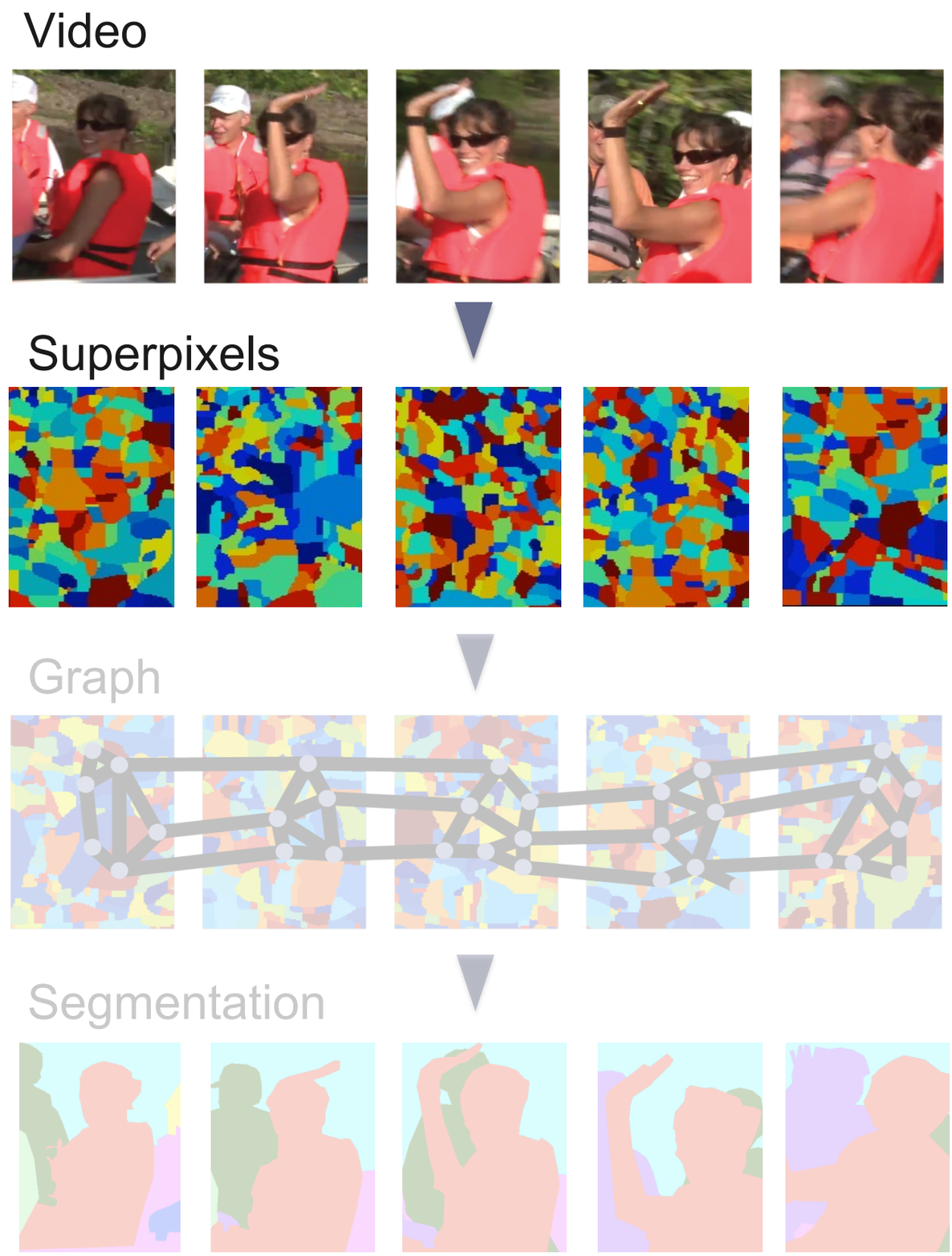
1. Superpixels per frame
   - Appearance and motion cues
2. Graph construction
   - Spatio-temporal neighborhood
3. Graph partitioning
   - Spectral clustering methods



Graph-Based Video Segmentation

[Fragkiadaki & Shi CVPR'12], [Ochs et al. PAMI'14], [Galasso et al. CVPR'14], [Yi & Pavlovic ICCV'15]

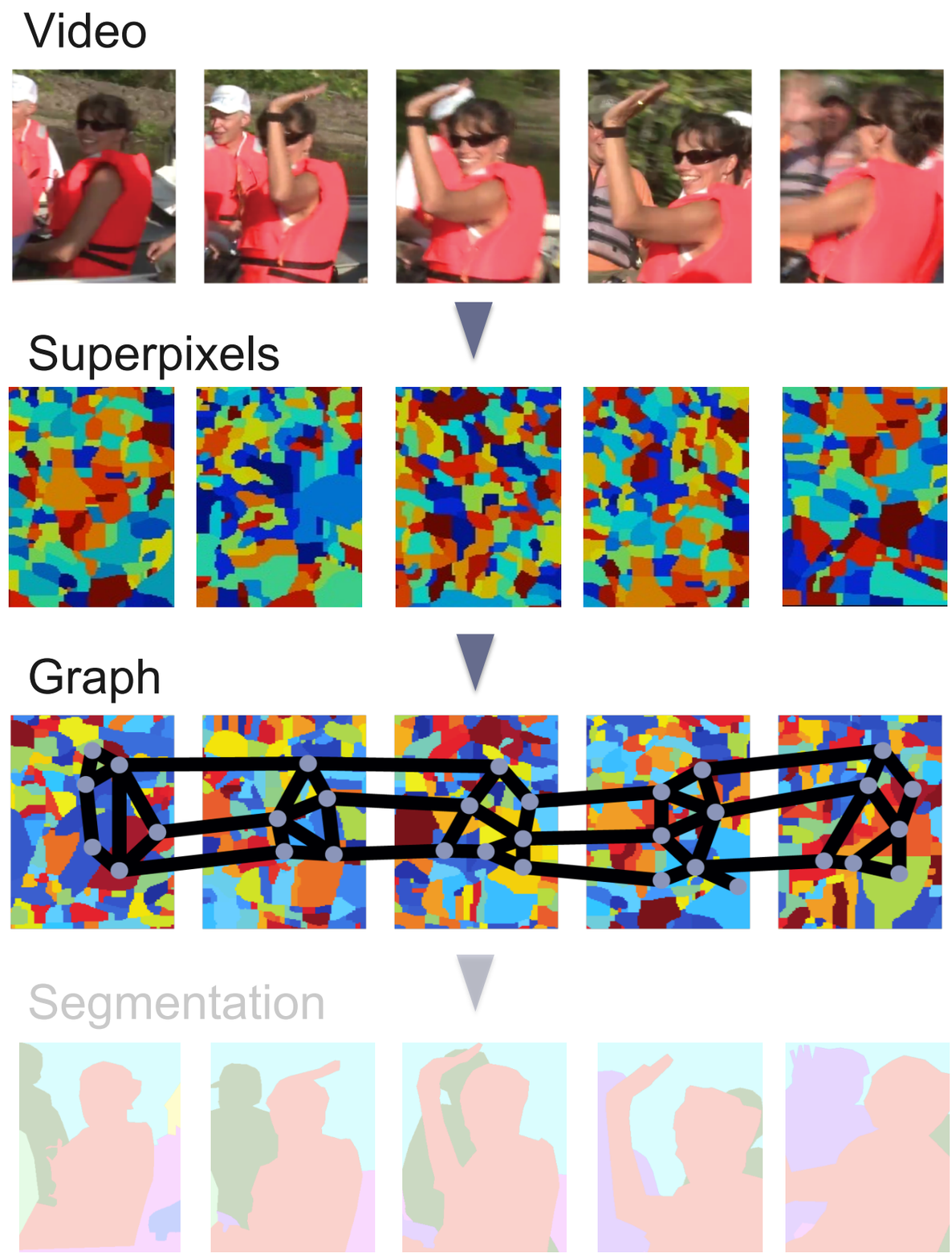
1. Superpixels per frame
   - Appearance and motion cues
2. Graph construction
   - Spatio-temporal neighborhood
3. Graph partitioning
   - Spectral clustering methods



Graph-Based Video Segmentation

[Fragkiadaki & Shi CVPR'12], [Ochs et al. PAMI'14], [Galasso et al. CVPR'14], [Yi & Pavlovic ICCV'15]

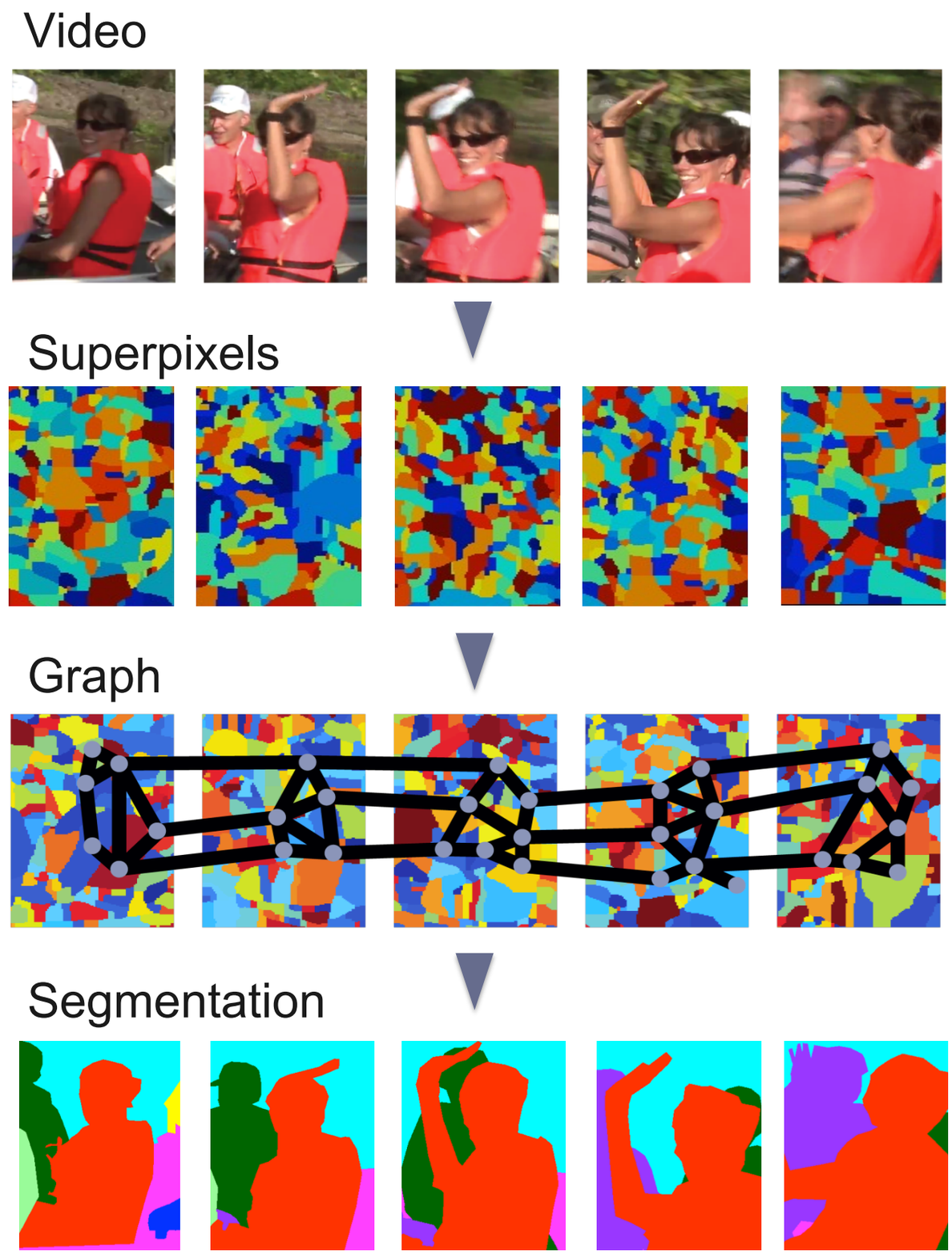
1. Superpixels per frame
   - Appearance and motion cues
2. Graph construction
   - Spatio-temporal neighborhood
3. Graph partitioning
   - Spectral clustering methods



Graph-Based Video Segmentation

[Fragkiadaki & Shi CVPR'12], [Ochs et al. PAMI'14], [Galasso et al. CVPR'14], [Yi & Pavlovic ICCV'15]

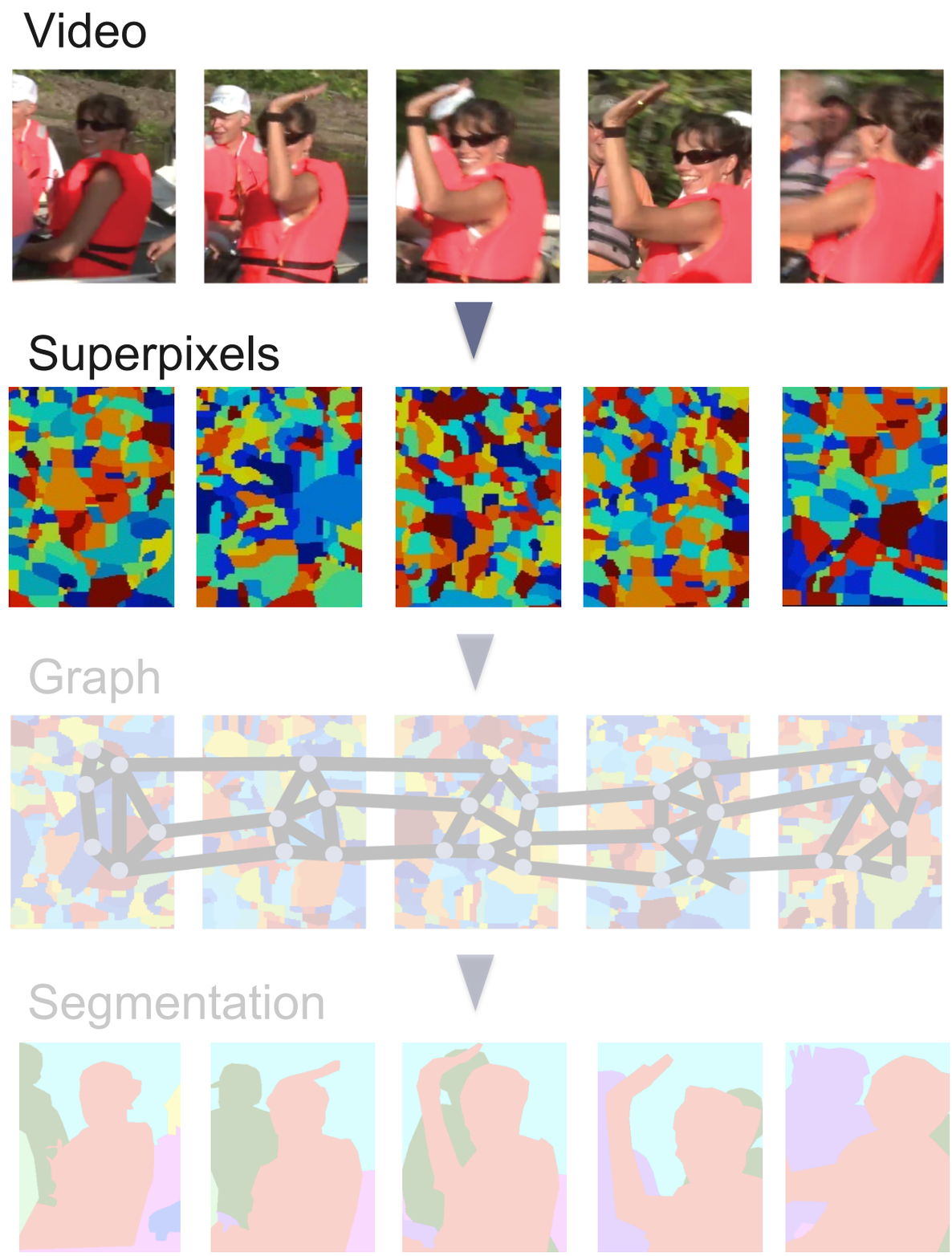
1. Superpixels per frame
   - Appearance and motion cues
2. Graph construction
   - Spatio-temporal neighborhood
3. Graph partitioning
   - Spectral clustering methods



Graph-Based Video Segmentation

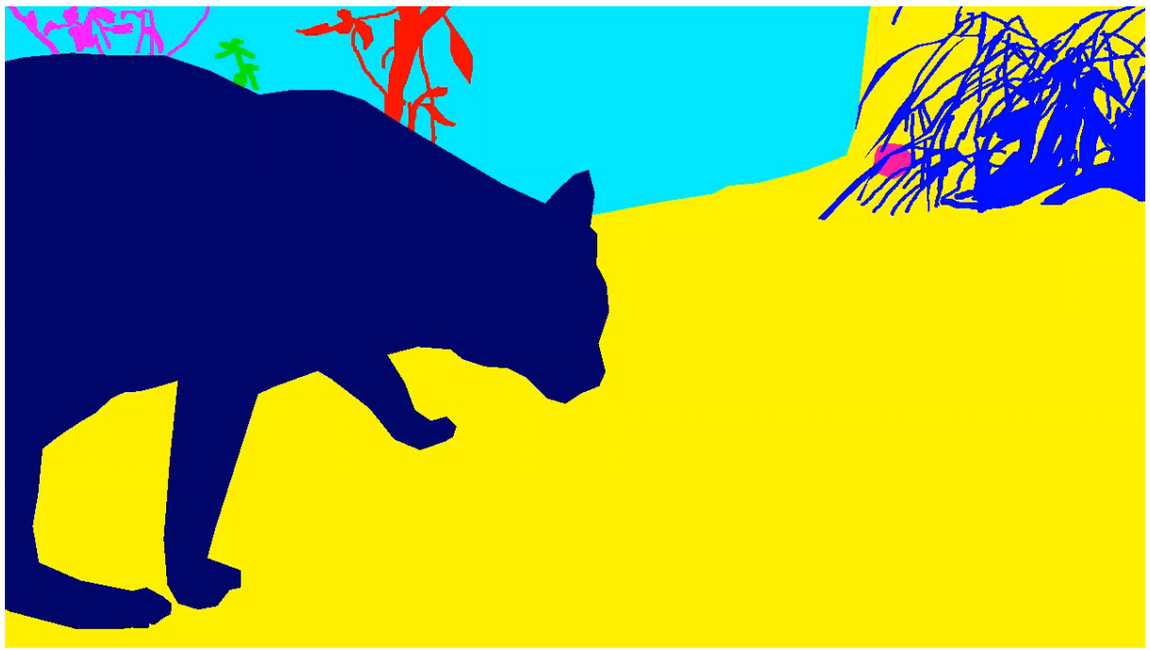
[Fragkiadaki & Shi CVPR'12], [Ochs et al. PAMI'14], [Galasso et al. CVPR'14], [Yi & Pavlovic ICCV'15]

1. Superpixels per frame
   - Appearance and motion cues
2. Graph construction
   - Spatio-temporal neighborhood
3. Graph partitioning
   - Spectral clustering methods

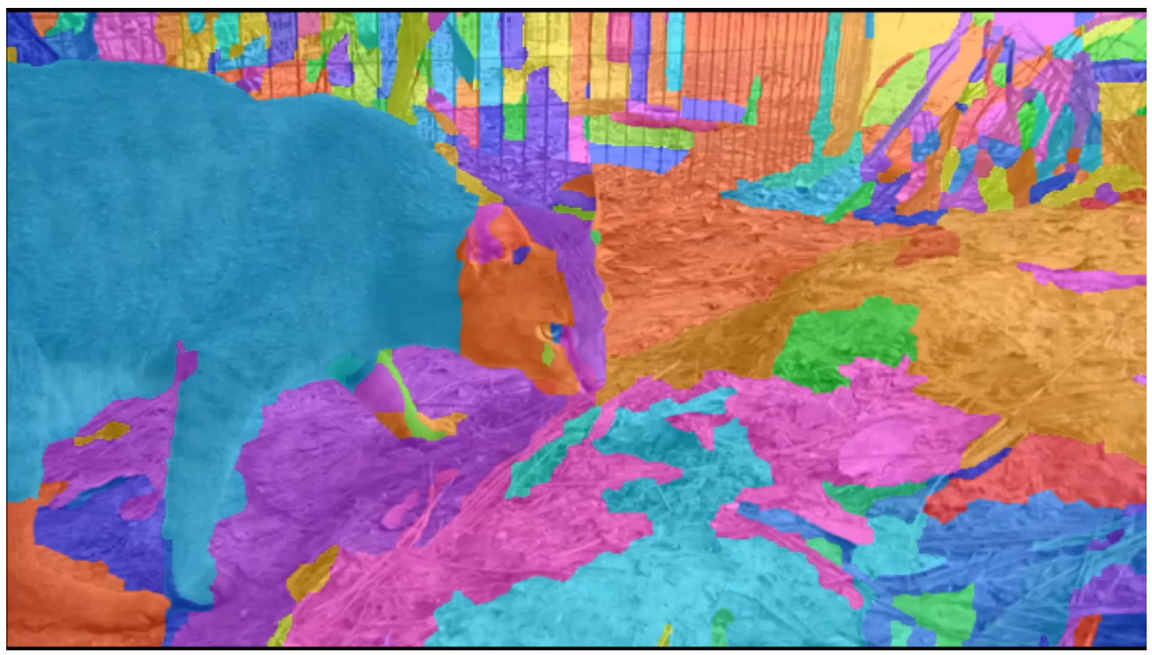


Motivation

Video



Ground truth

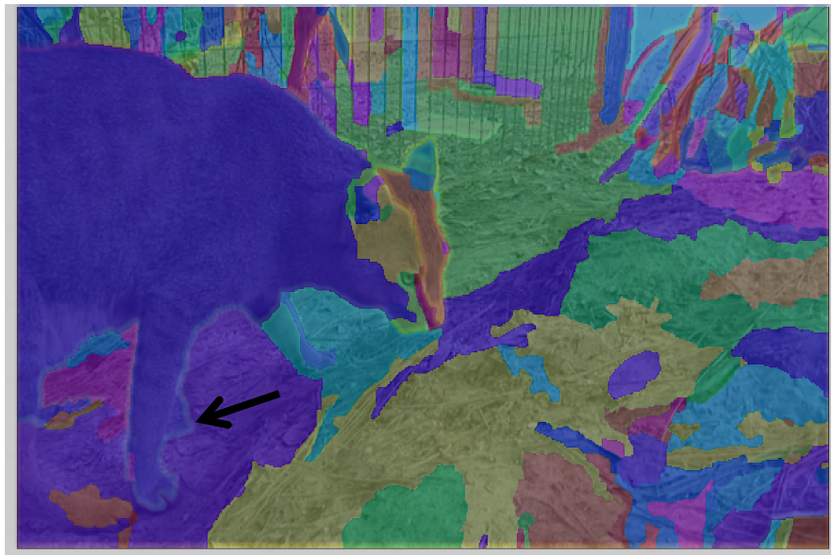


Superpixels
[gPb-owt-ucm, Arbelaez et al. PAMI'11]

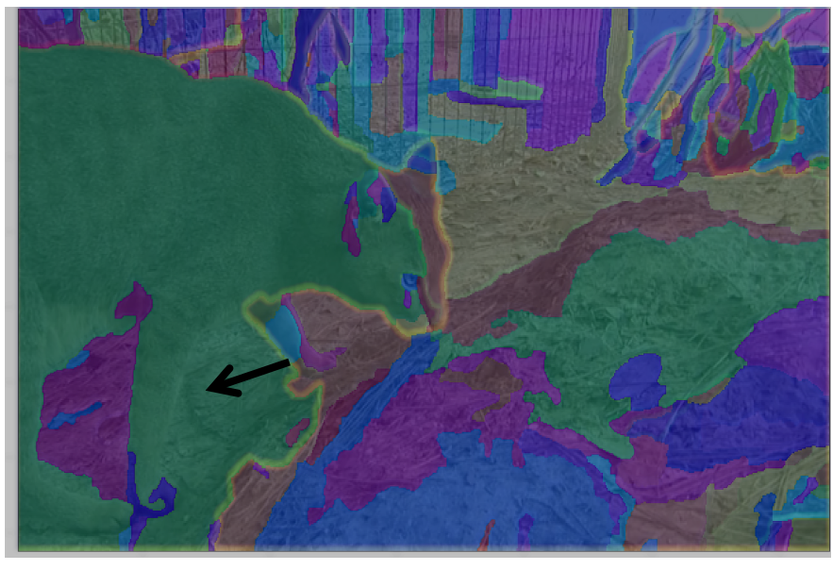
- Volatile and flickering superpixels
- Label leakage

Motivation

Video: frame *t*



Video: frame *t+1*



Superpixels

Motivation

Properties of good superpixels:

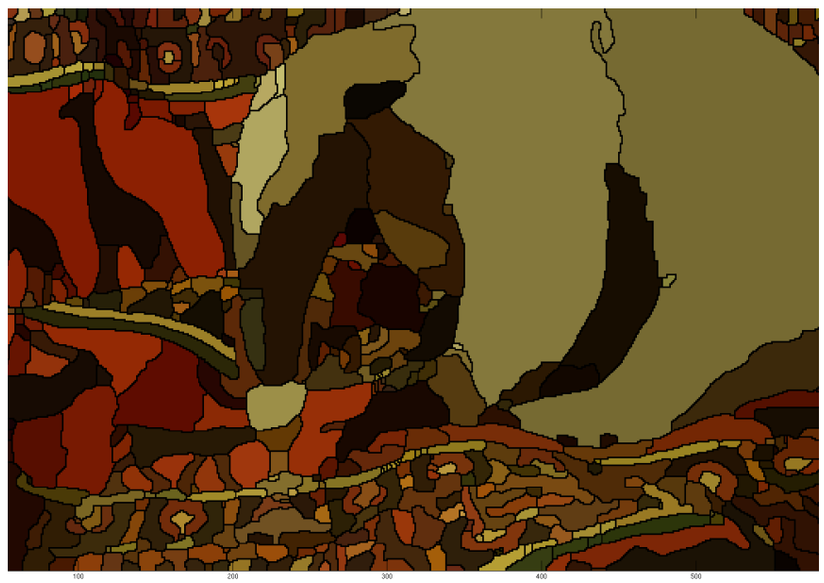
- high boundary recall
- good temporal consistency
- as few as possible

Video

Our superpixels

Better Superpixels for Video Segmentation

Video

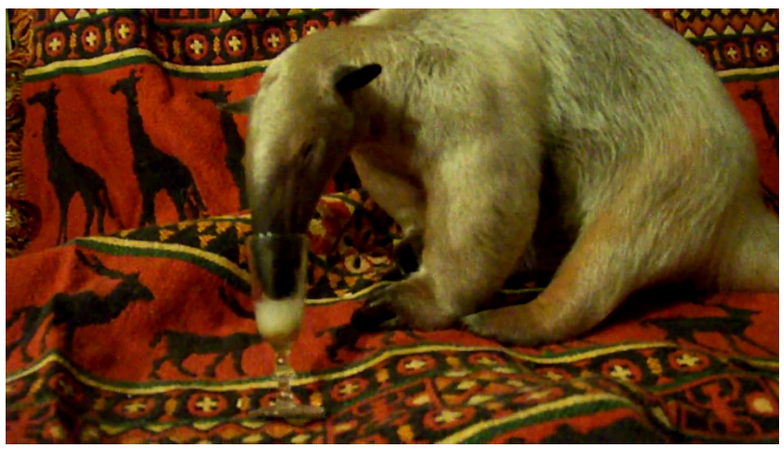


Superpixels

Better superpixels → Improved video segmentation

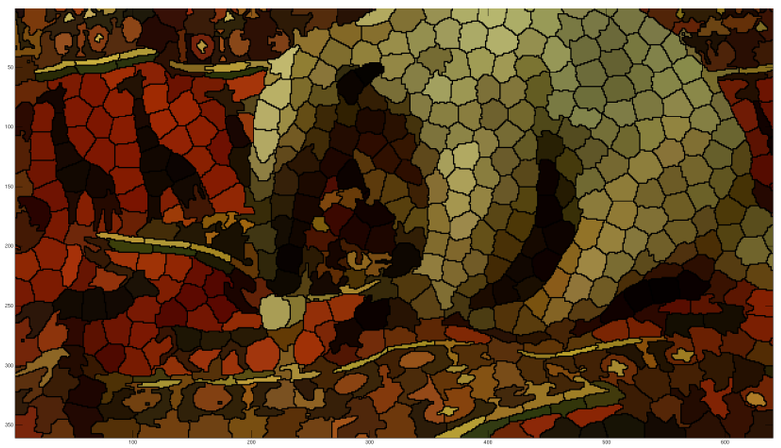
Superpixel/voxel Methods

1. Classical superpixel/voxel methods
    - homogeneous shape and size
    - regular topology

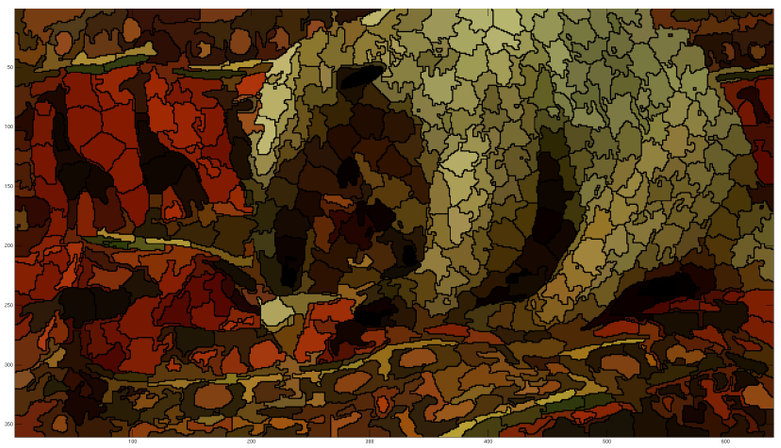


Video

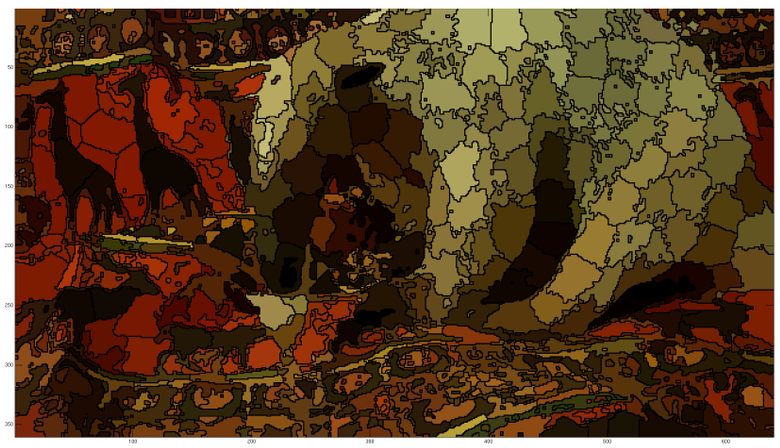
Sticky superpixels
[Dollar et al. PAMI'15]



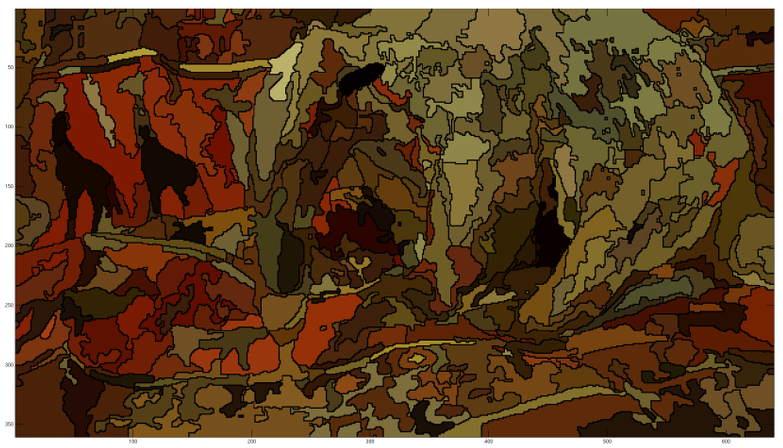
TSP
[Chang et al. CVPR'13]



SLIC 2D
[Achanta et al. PAMI'12]



SLIC 3D
[Achanta et al. PAMI'12]

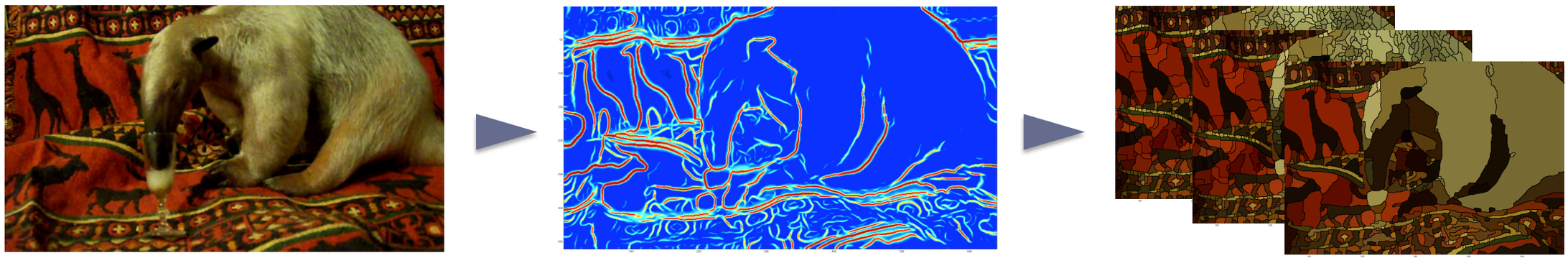


Video SEEDS
[Bergh et al. ICCV'13]

Superpixel/voxel Methods

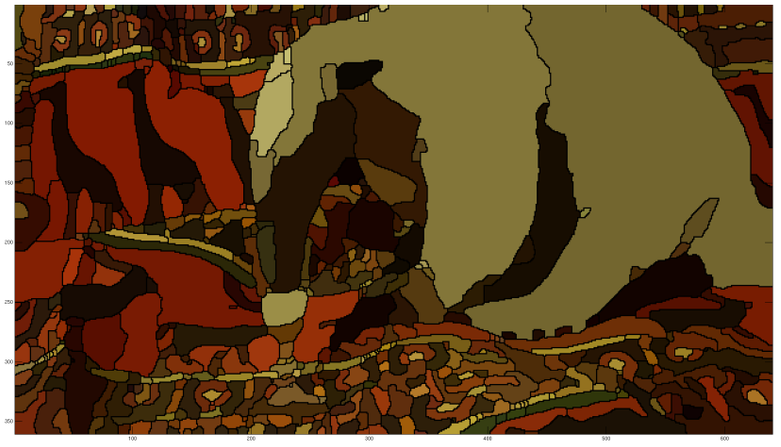
2. Boundary based superpixel/voxel methods
    - heterogeneous shape and size
    - semantic regions

gPb-owt-ucm [Arbelaez et al. PAMI'11]



Video | Boundaries | Hierarchical image segmentation

Video



gPb
[Arbelaez et al. PAMI'11]

PMI
[Isola et al. ECCV'14]

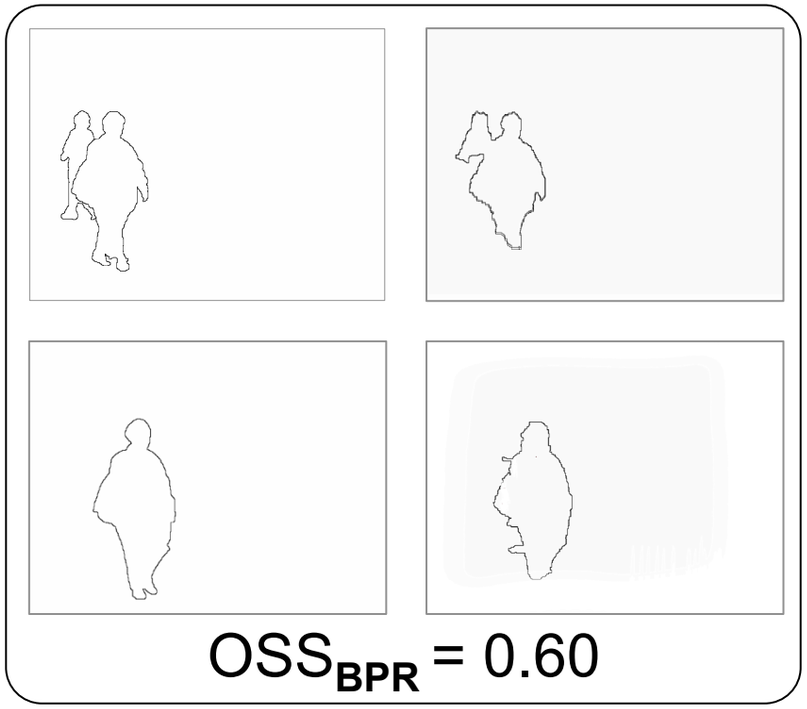
VSB100 [Galasso et al. ICCV'13]

- 100 HD-quality video sequences
- 4 sets of human annotations
- Training, validation and test sets [24 + 16 + 60]

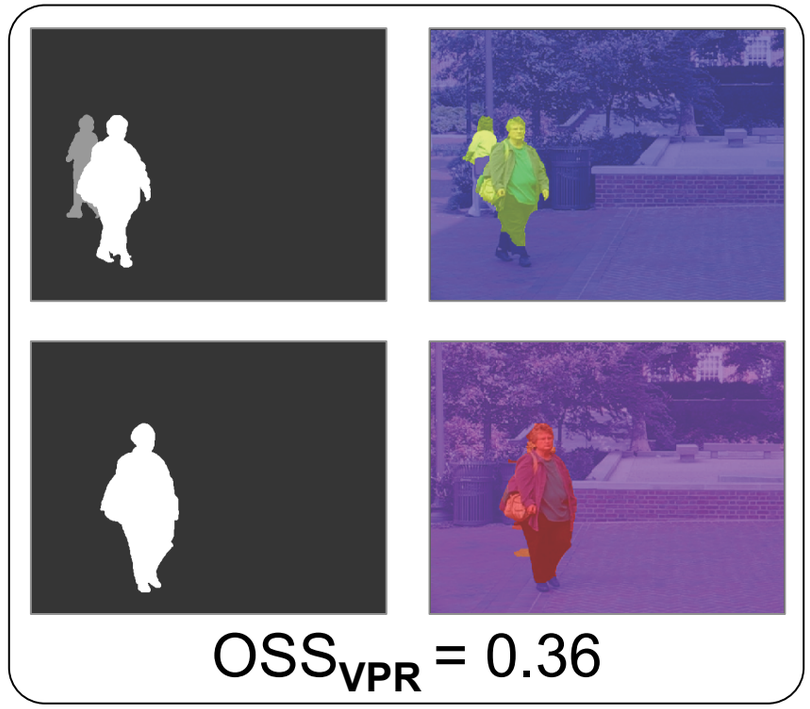
Benchmark metrics:

Boundary precision-recall (**BPR**)
[Martin et al. PAMI'04]

Volume precision-recall (**VPR**)
[Galasso et al. ICCV'13]



$OSS_{BPR} = 0.60$

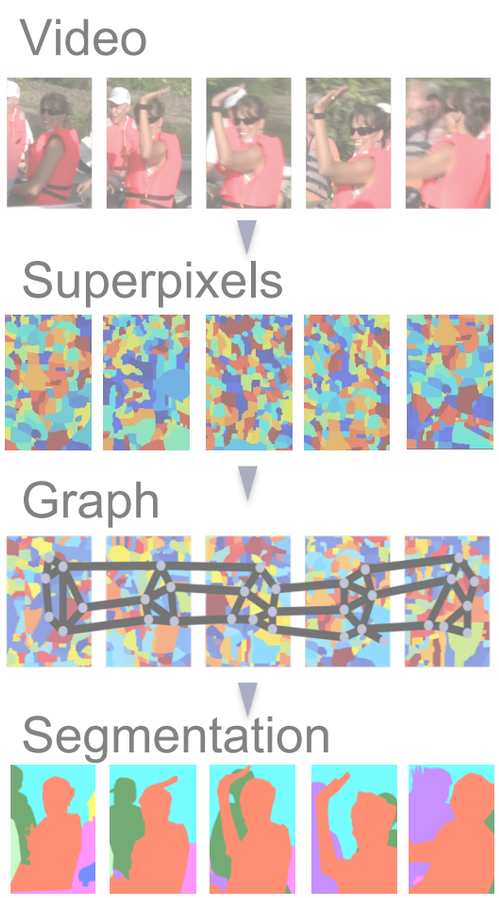


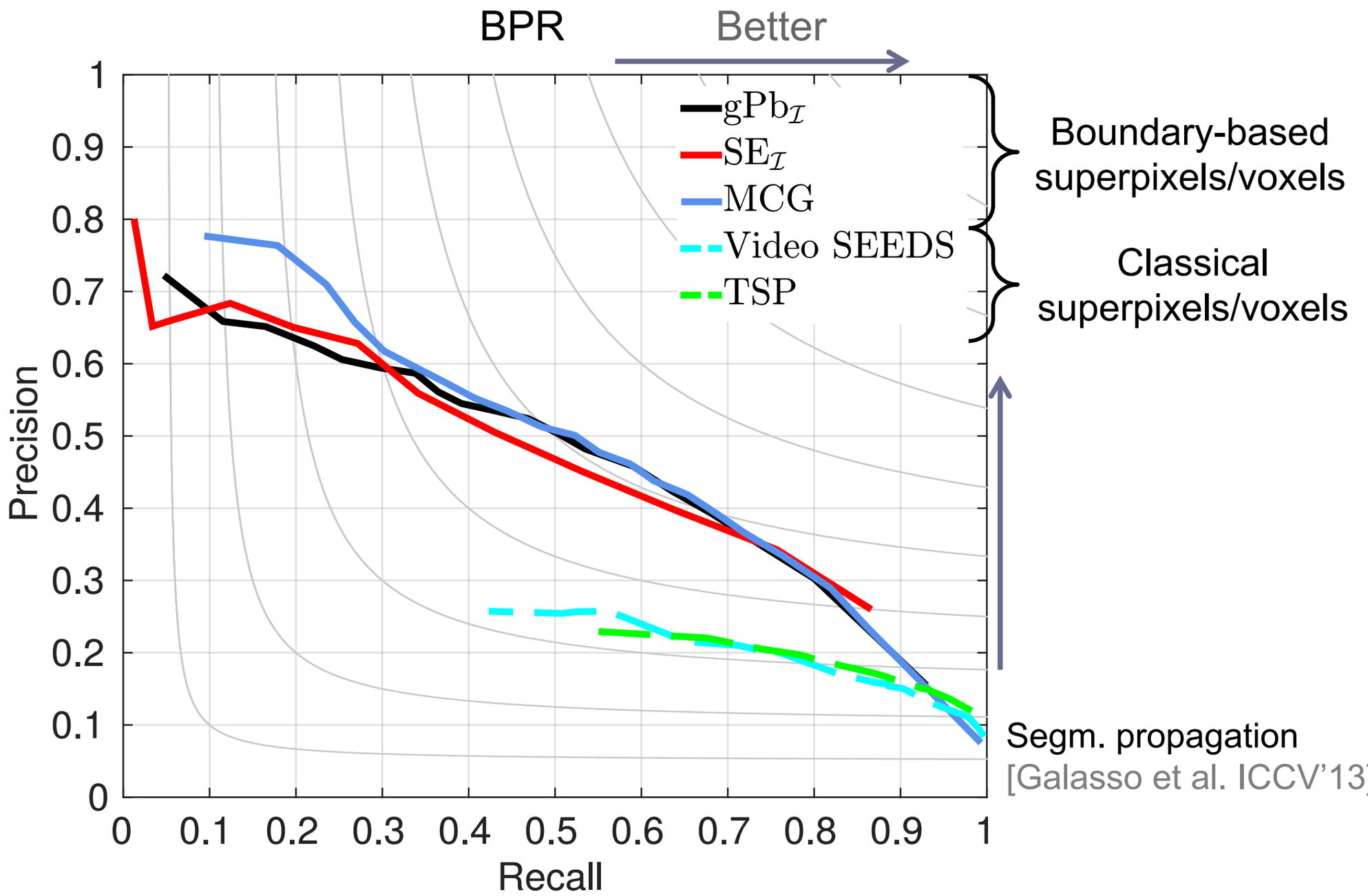
$OSS_{VPR} = 0.36$

OSS: F-measure at optimal segmentation scale

Superpixel/voxel Methods for Video Segmentation

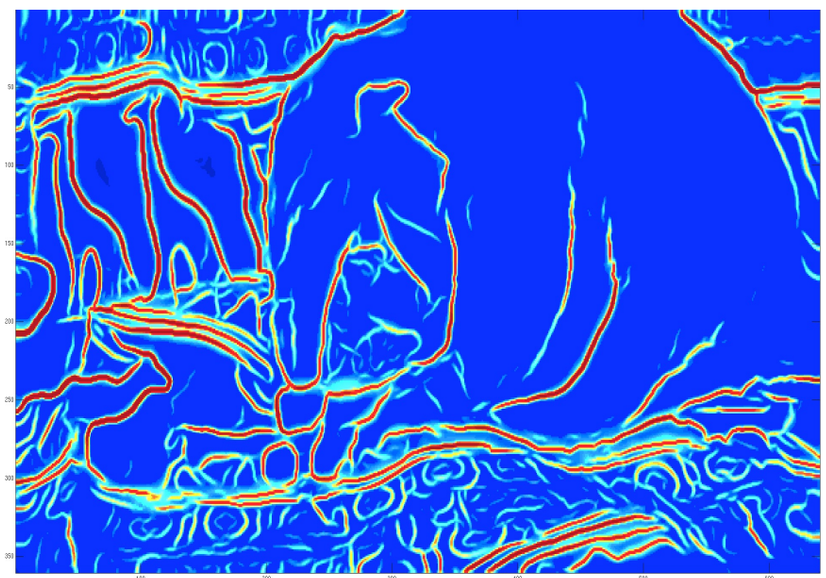
VSB100





Superpixels built from boundaries are more effective for graph-based video segmentation.

Better Boundaries for Superpixels



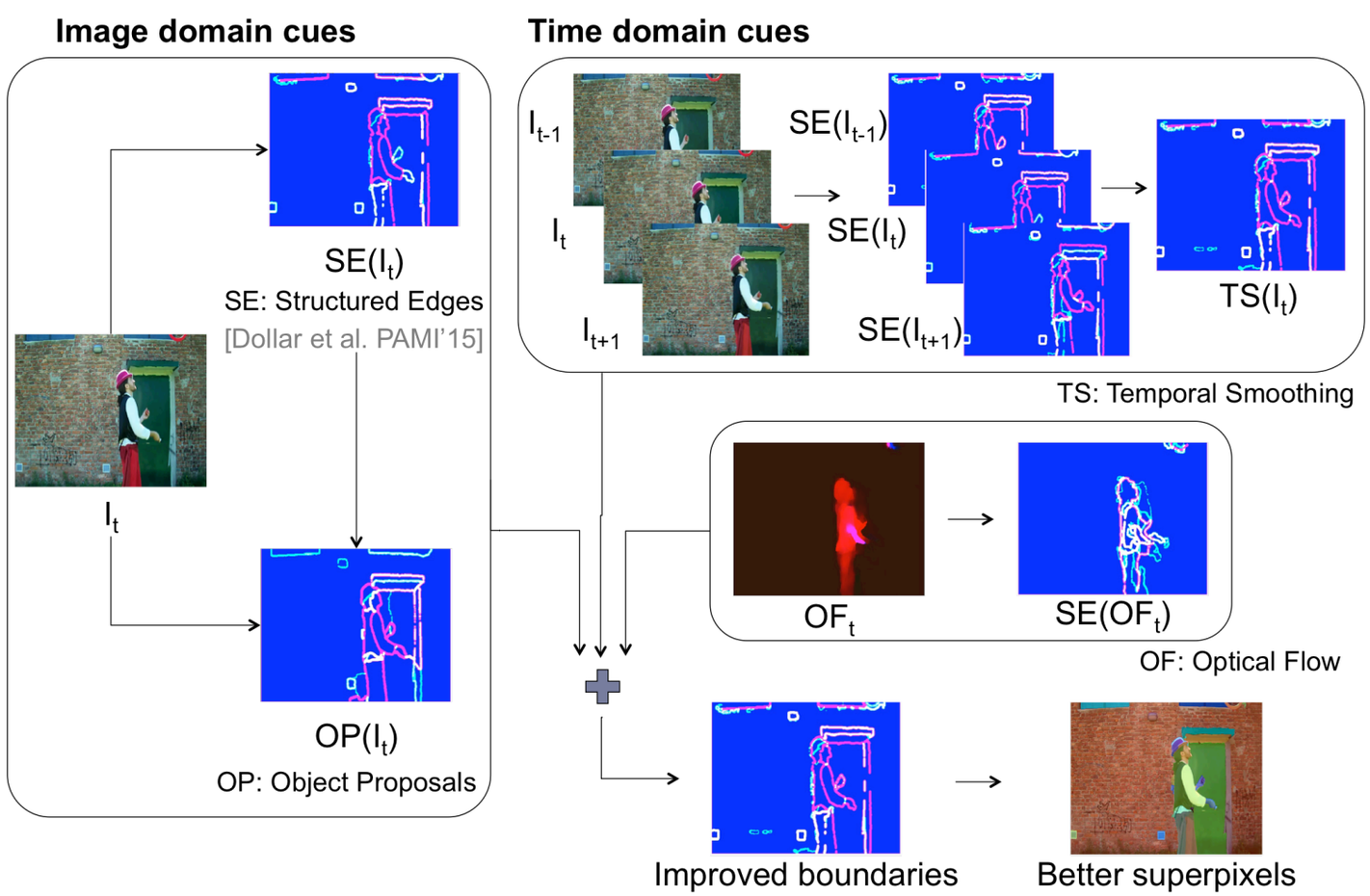
Boundaries

Superpixels

Better boundaries → Better superpixels → Improved video segmentation

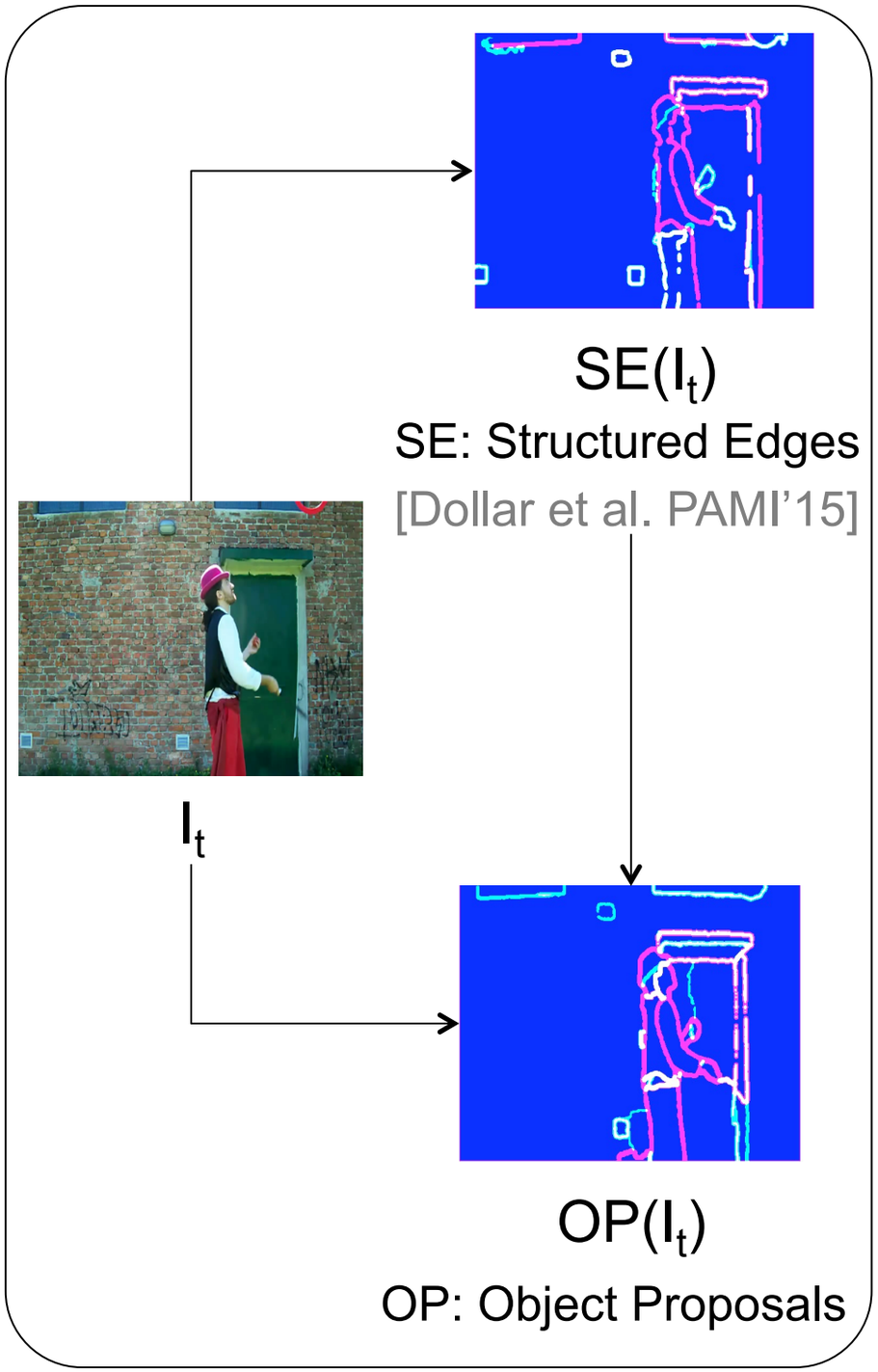
Improving Image Boundaries

- **Image domain cues**
  - Image boundaries
  - Higher-level cues via object proposals
- **Time domain cues**
  - Temporal smoothing
  - Motion boundaries



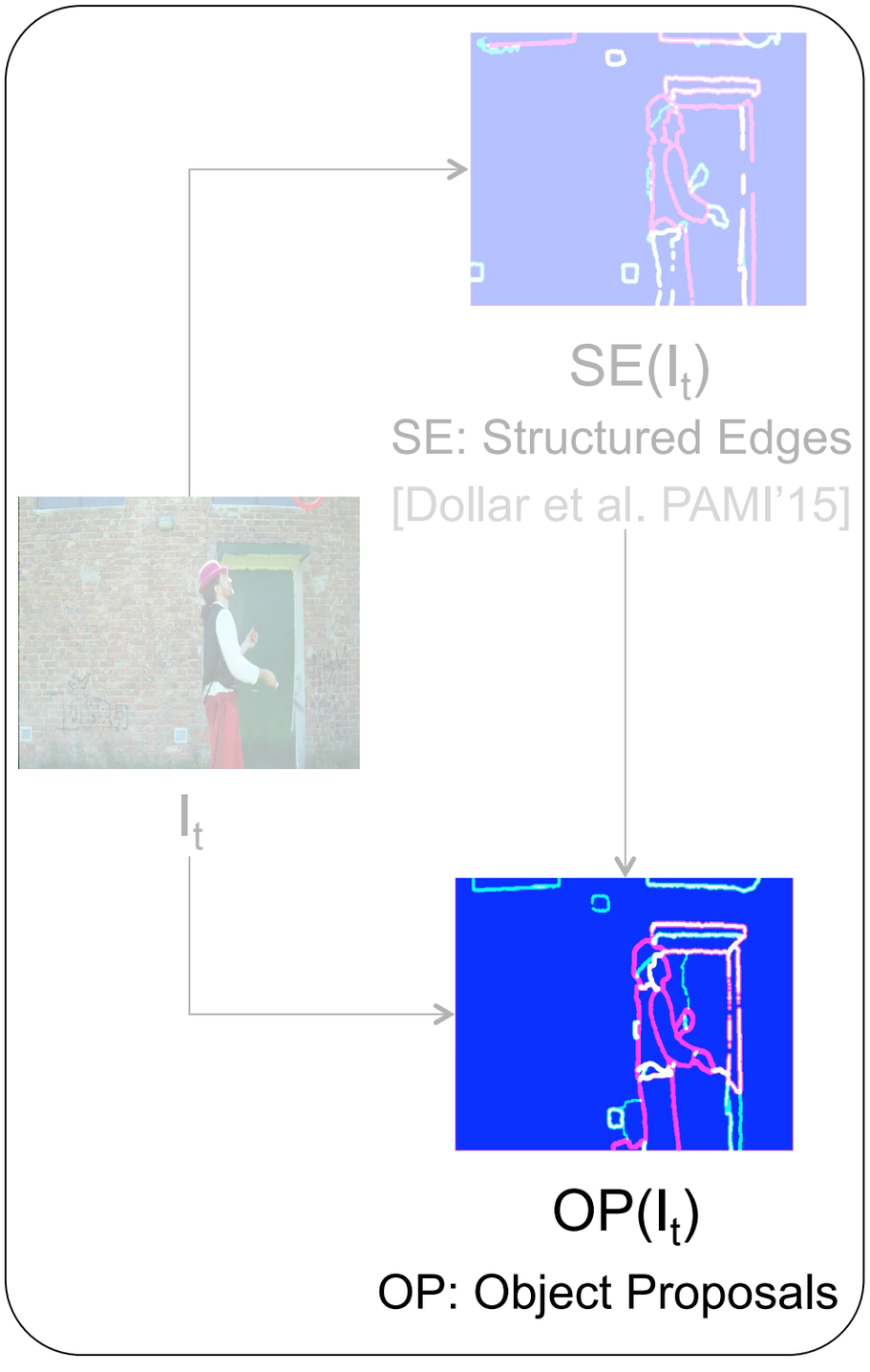
Improving Image Boundaries

Image domain cues



Improving Image Boundaries

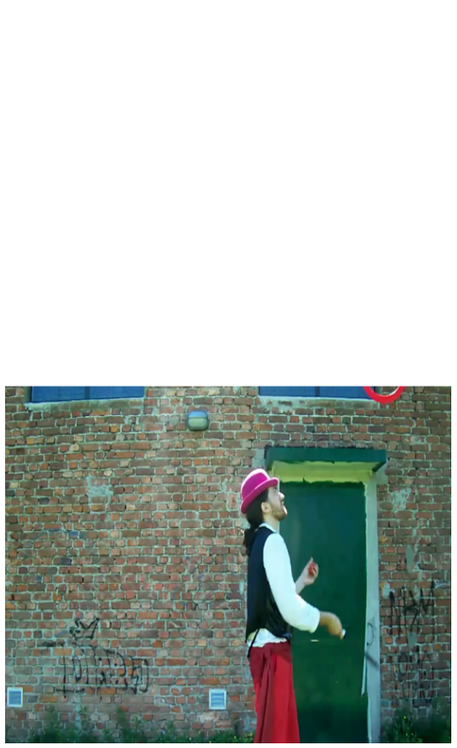
Image domain cues



Integration of Higher-Level Object Cues

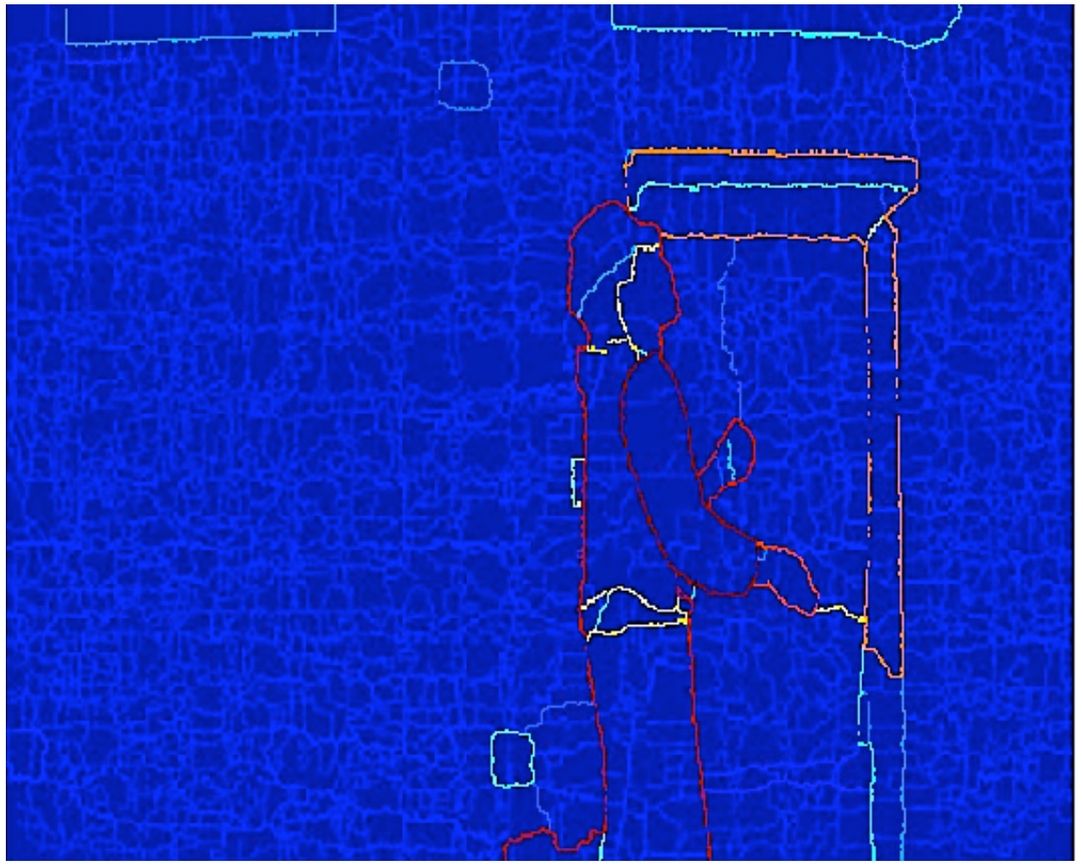
Object Proposals [Humayun et al. CVPR'14]

- average the contours of each object proposal segment



Video

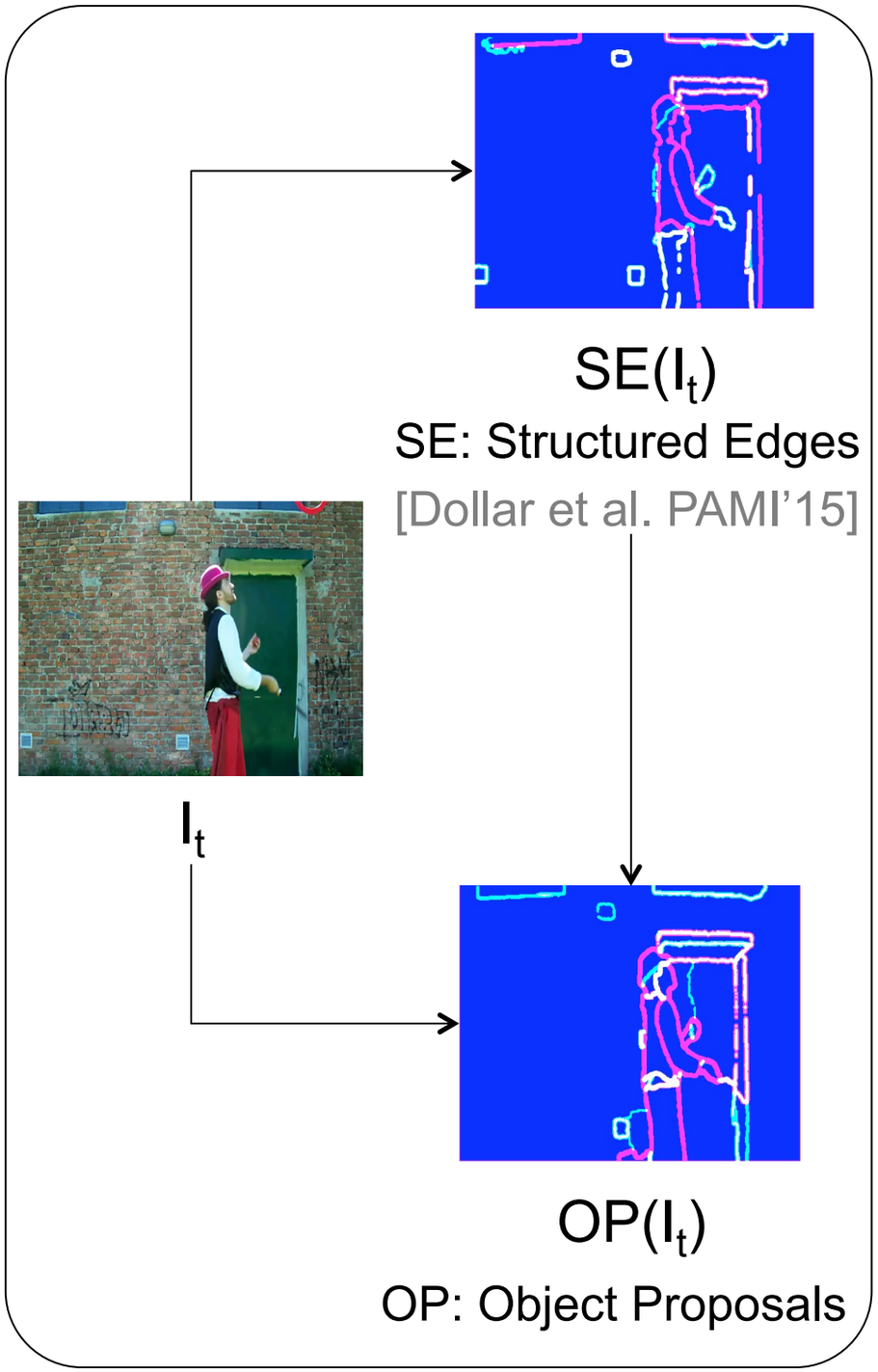
Object proposals



Boundary map

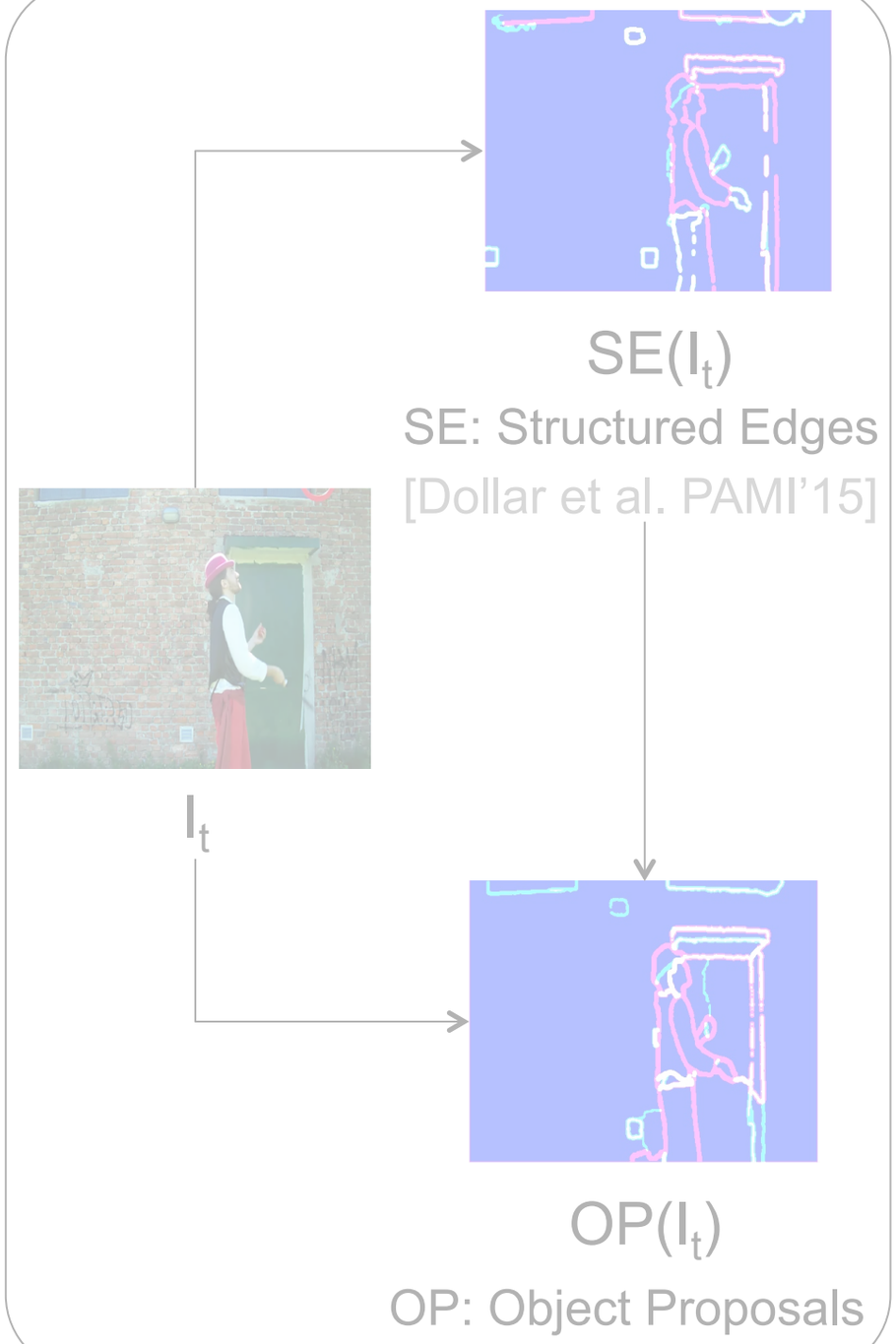
Improving Image Boundaries

Image domain cues

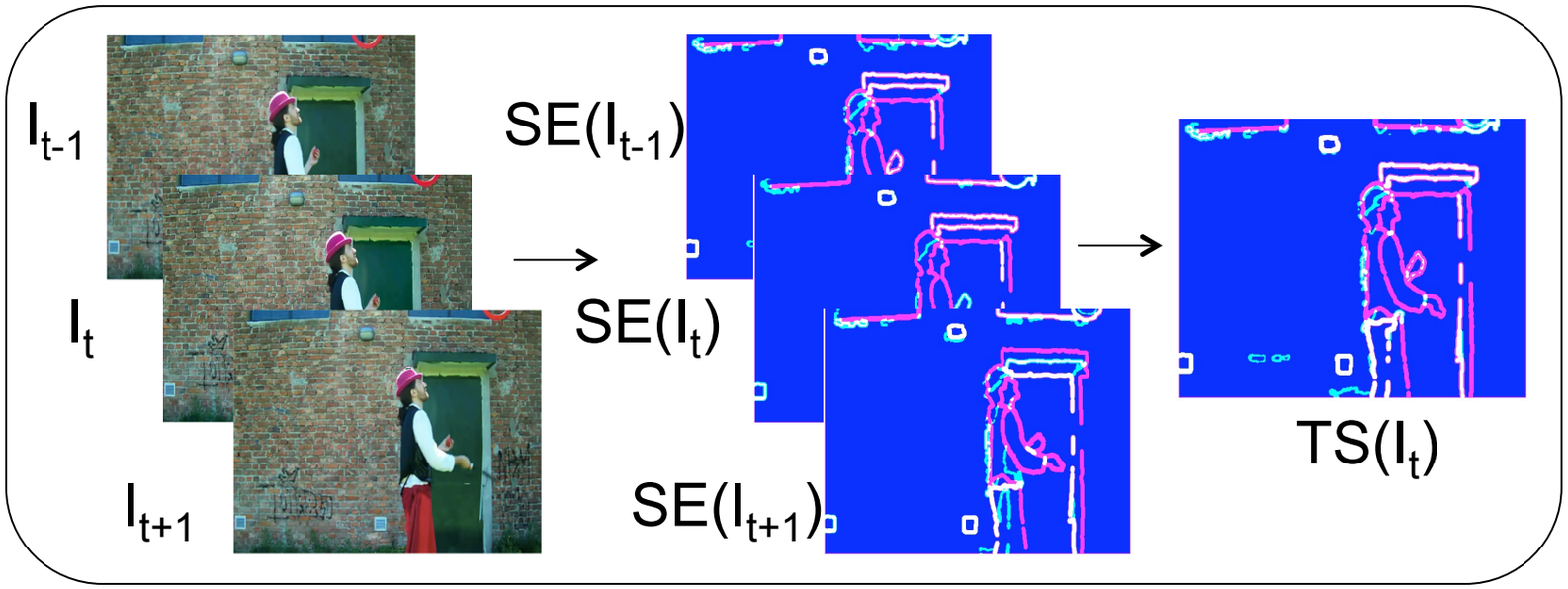


Improving Image Boundaries

Image domain cues



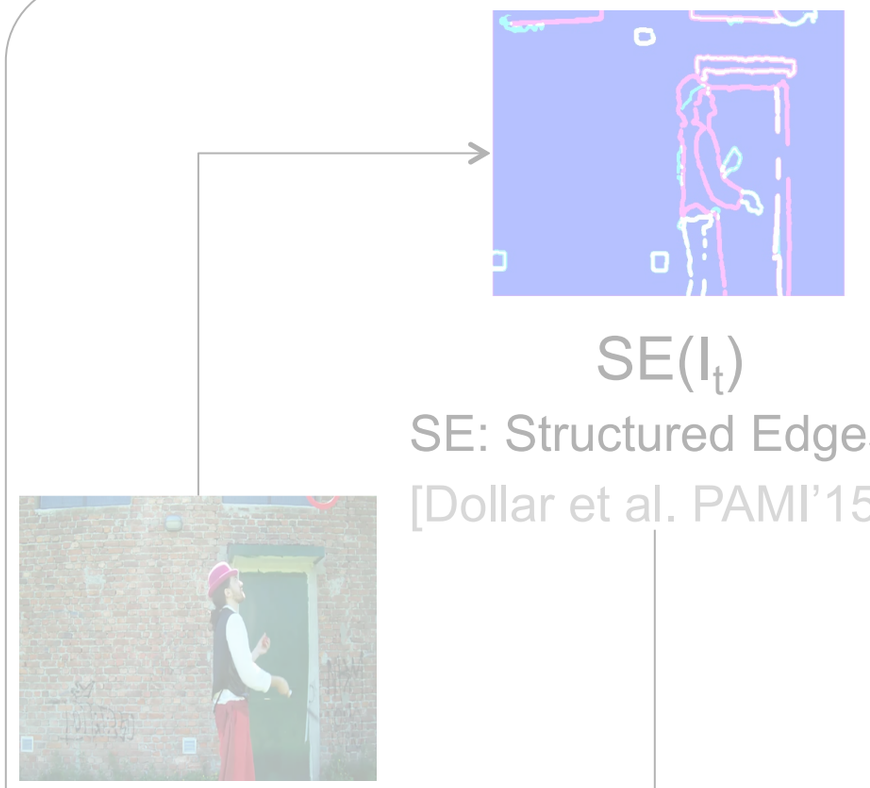
Time domain cues



TS: Temporal Smoothing

Improving Image Boundaries

Image domain cues



I_t

$SE(I_t)$

SE: Structured Edges
[Dollar et al. PAMI'15]

$OP(I_t)$

OP: Object Proposals

Time domain cues

I_{t-1}

I_t

I_{t+1}

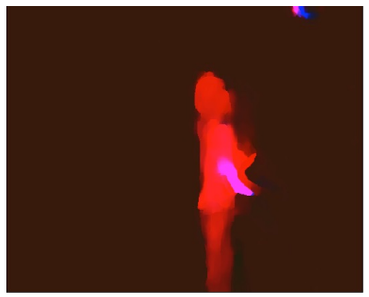
$SE(I_{t-1})$

$SE(I_t)$

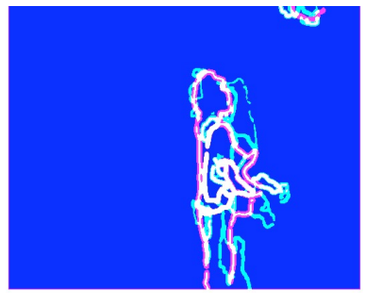
$SE(I_{t+1})$

$TS(I_t)$

TS: Temporal Smoothing



OF_t



$SE(OF_t)$

OF: Optical Flow

Improving Image Boundaries

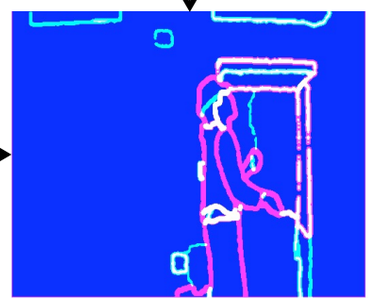
Image domain cues

I_t

$SE(I_t)$

SE: Structured Edges

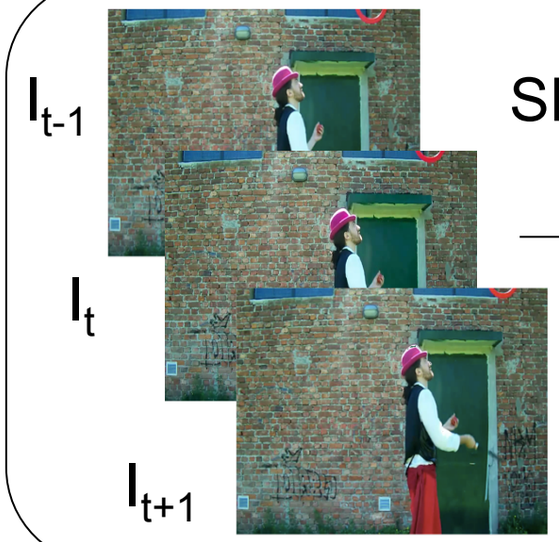
[Dollar et al. PAMI'15]



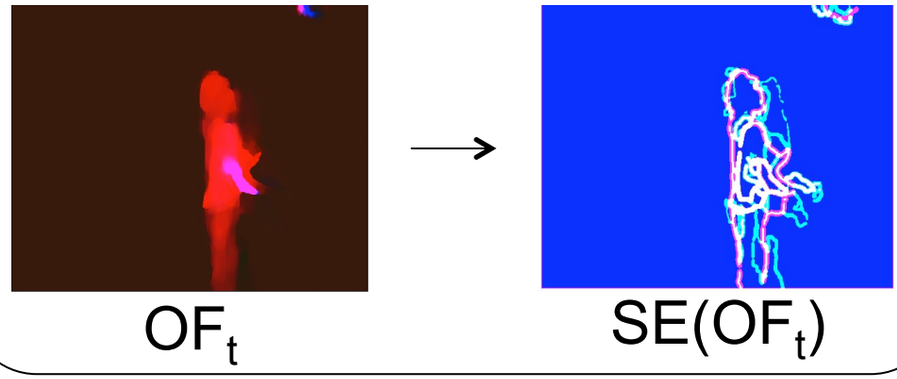
$OP(I_t)$

OP: Object Proposals

Time domain cues



TS: Temporal Smoothing



OF: Optical Flow

Improving Image Boundaries

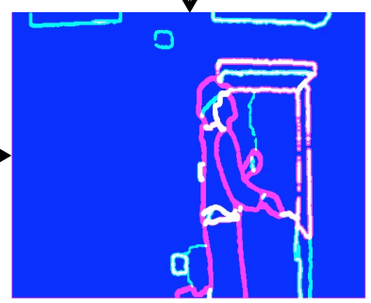
Image domain cues

$SE(I_t)$

SE: Structured Edges

[Dollar et al. PAMI'15]

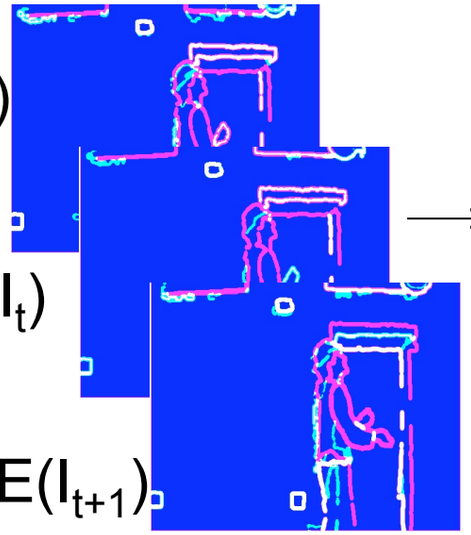
I_t



$OP(I_t)$

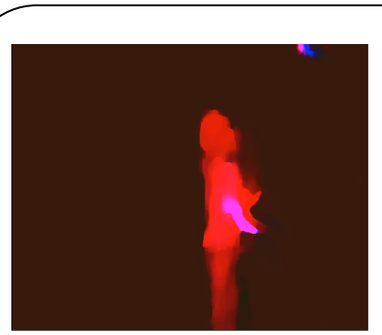
OP: Object Proposals

Time domain cues

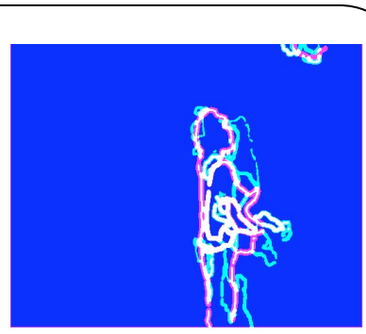


$TS(I_t)$

TS: Temporal Smoothing



OF_t



$SE(OF_t)$

OF: Optical Flow

Improving Image Boundaries

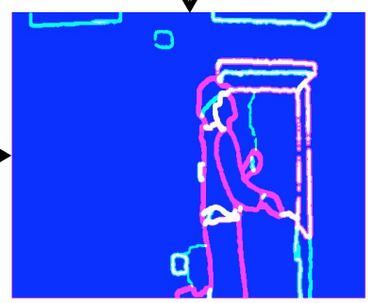
Image domain cues

$SE(I_t)$

SE: Structured Edges

[Dollar et al. PAMI'15]

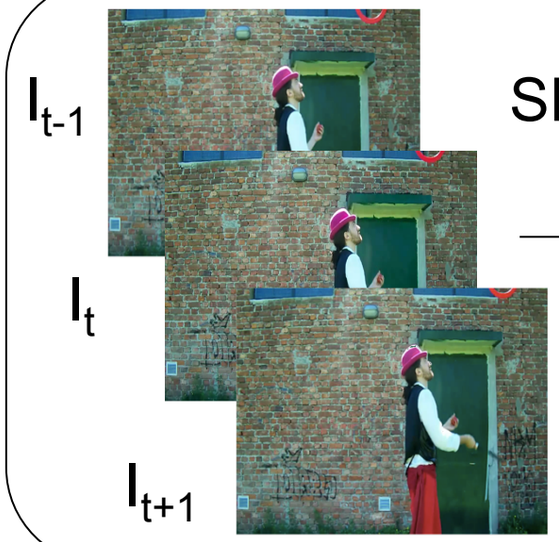
I_t

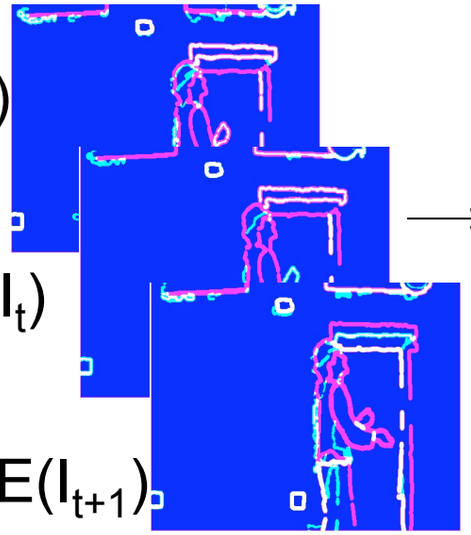


$OP(I_t)$

OP: Object Proposals

Time domain cues

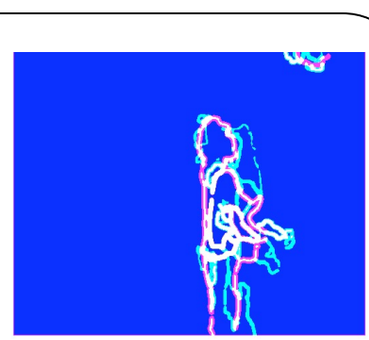




$TS(I_t)$

TS: Temporal Smoothing

OF_t



$SE(OF_t)$

OF: Optical Flow

Improved boundaries

Improving Image Boundaries

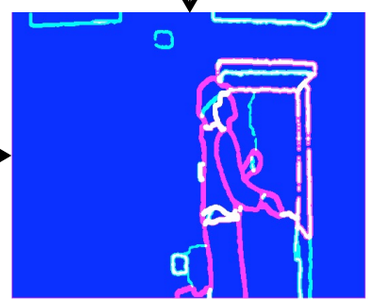
Image domain cues

$SE(I_t)$

SE: Structured Edges

[Dollar et al. PAMI'15]

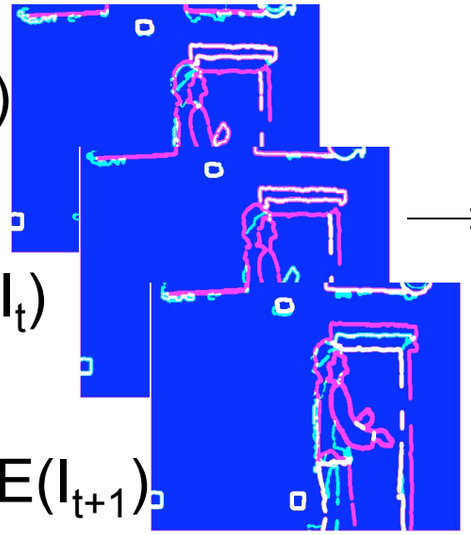
I_t



$OP(I_t)$

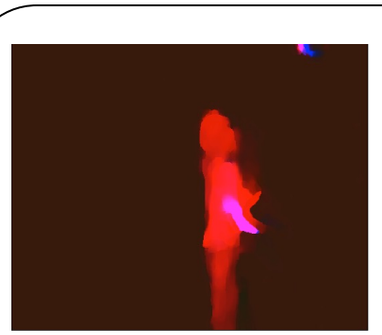
OP: Object Proposals

Time domain cues

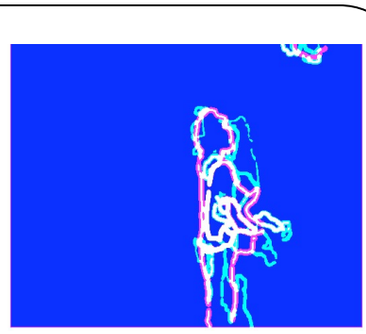


$TS(I_t)$

TS: Temporal Smoothing



OF_t



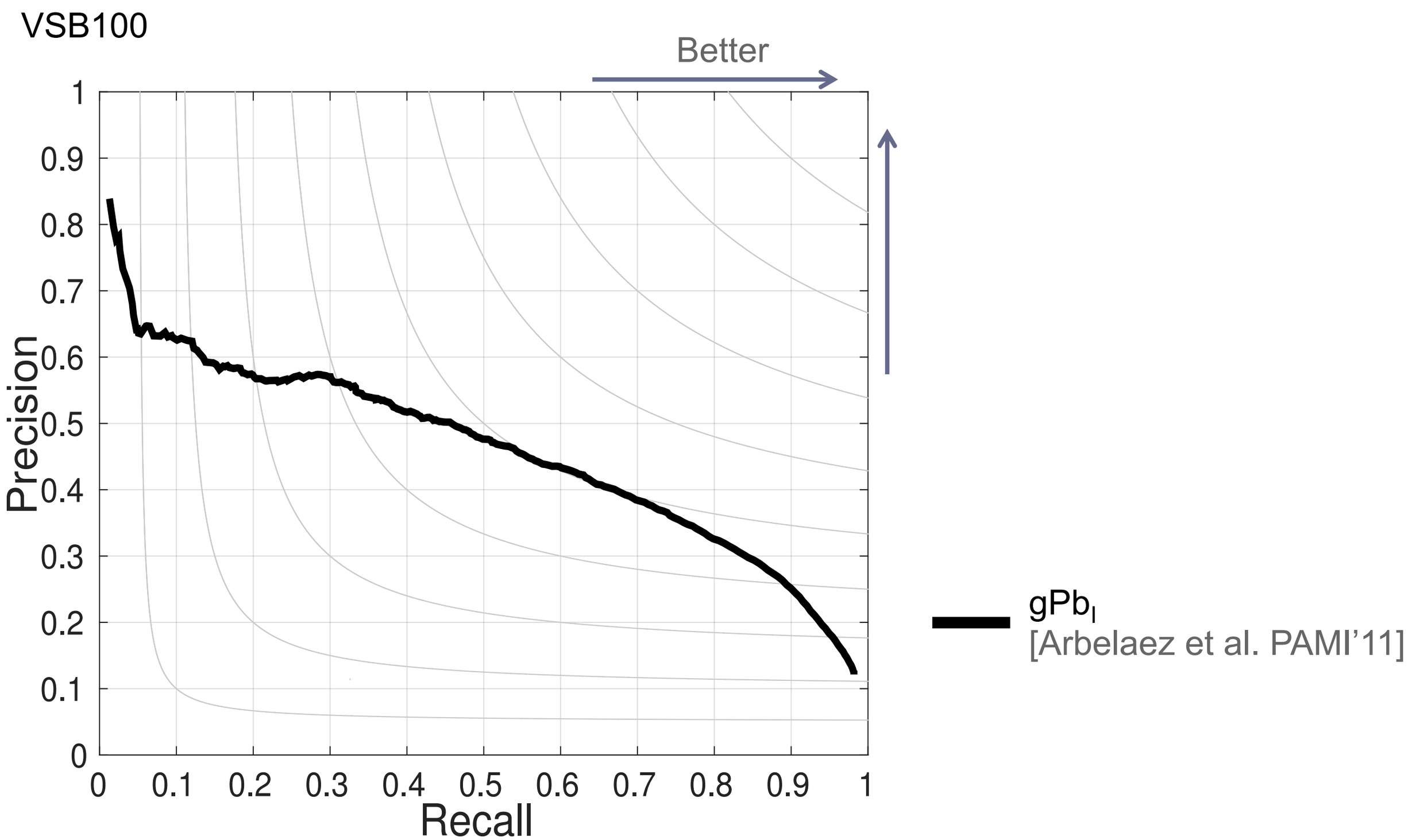
$SE(OF_t)$

OF: Optical Flow

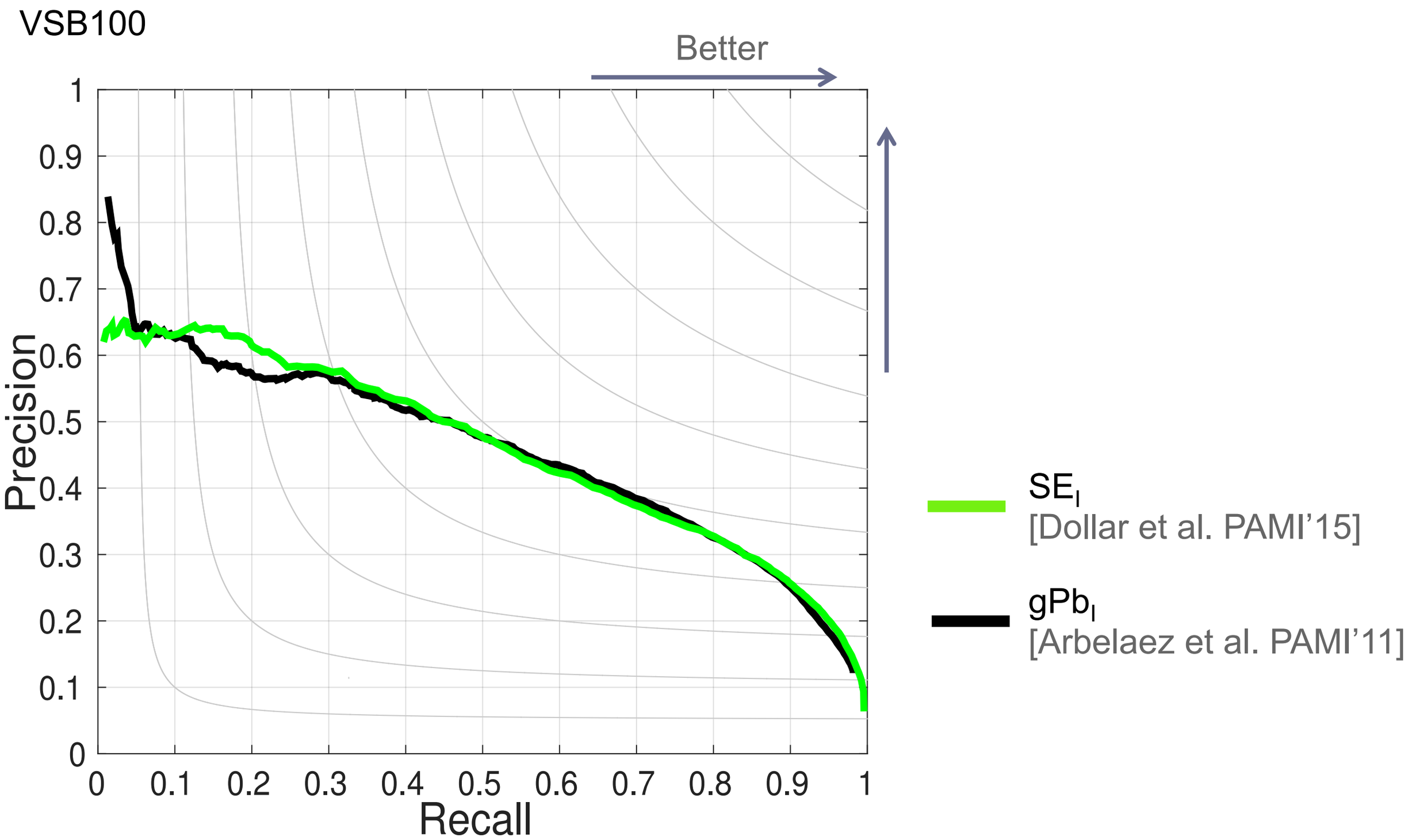
Improved boundaries

Better superpixels

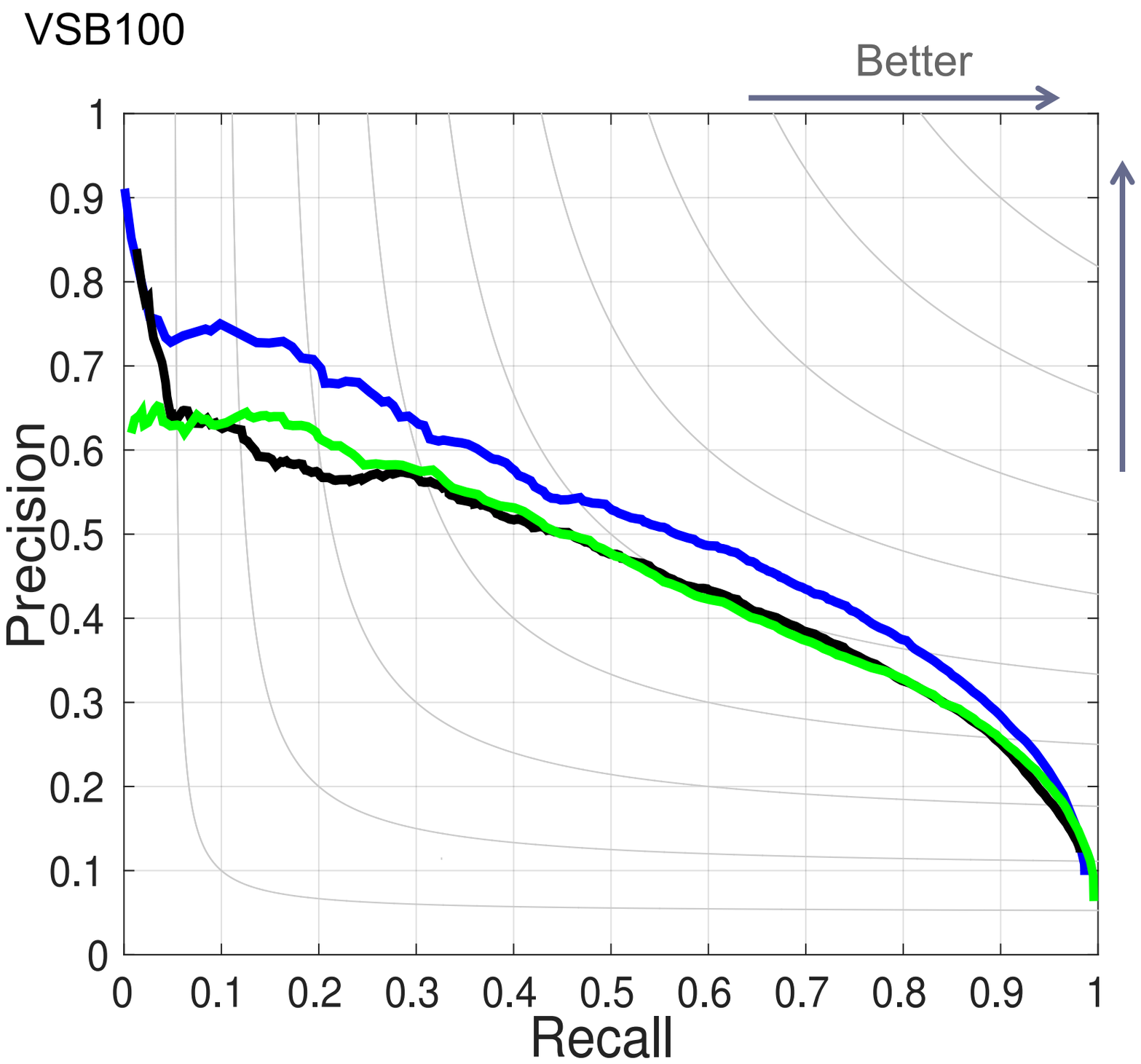
Evaluation of Image Boundaries on VSB100



Evaluation of Image Boundaries on VSB100



Evaluation of Image Boundaries on VSB100

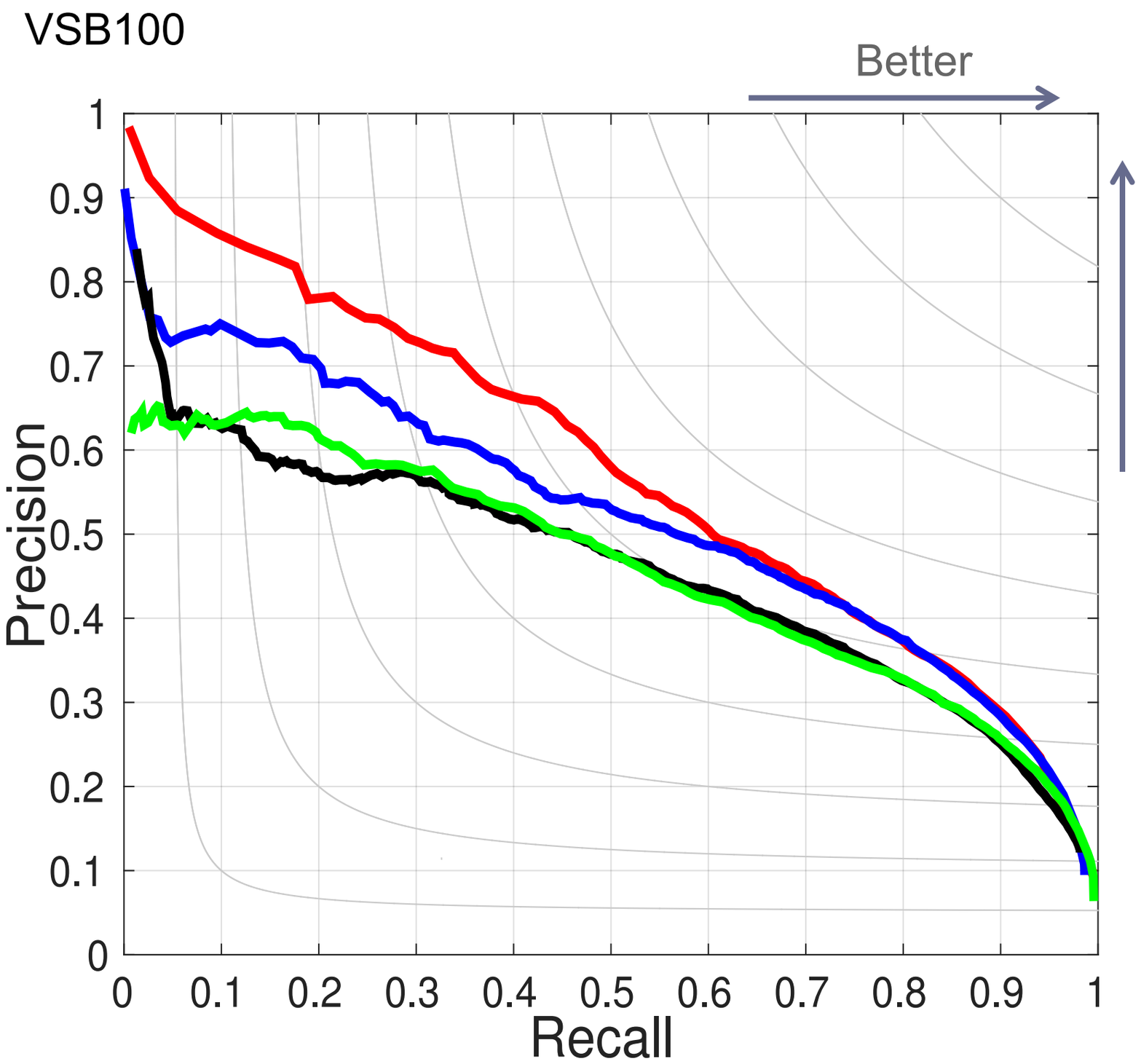


$SE_l + OP_l = \alpha$
[Humayun et al. CVPR'14]

SE_l
[Dollar et al. PAMI'15]

gPb_l
[Arbelaez et al. PAMI'11]

Evaluation of Image Boundaries on VSB100



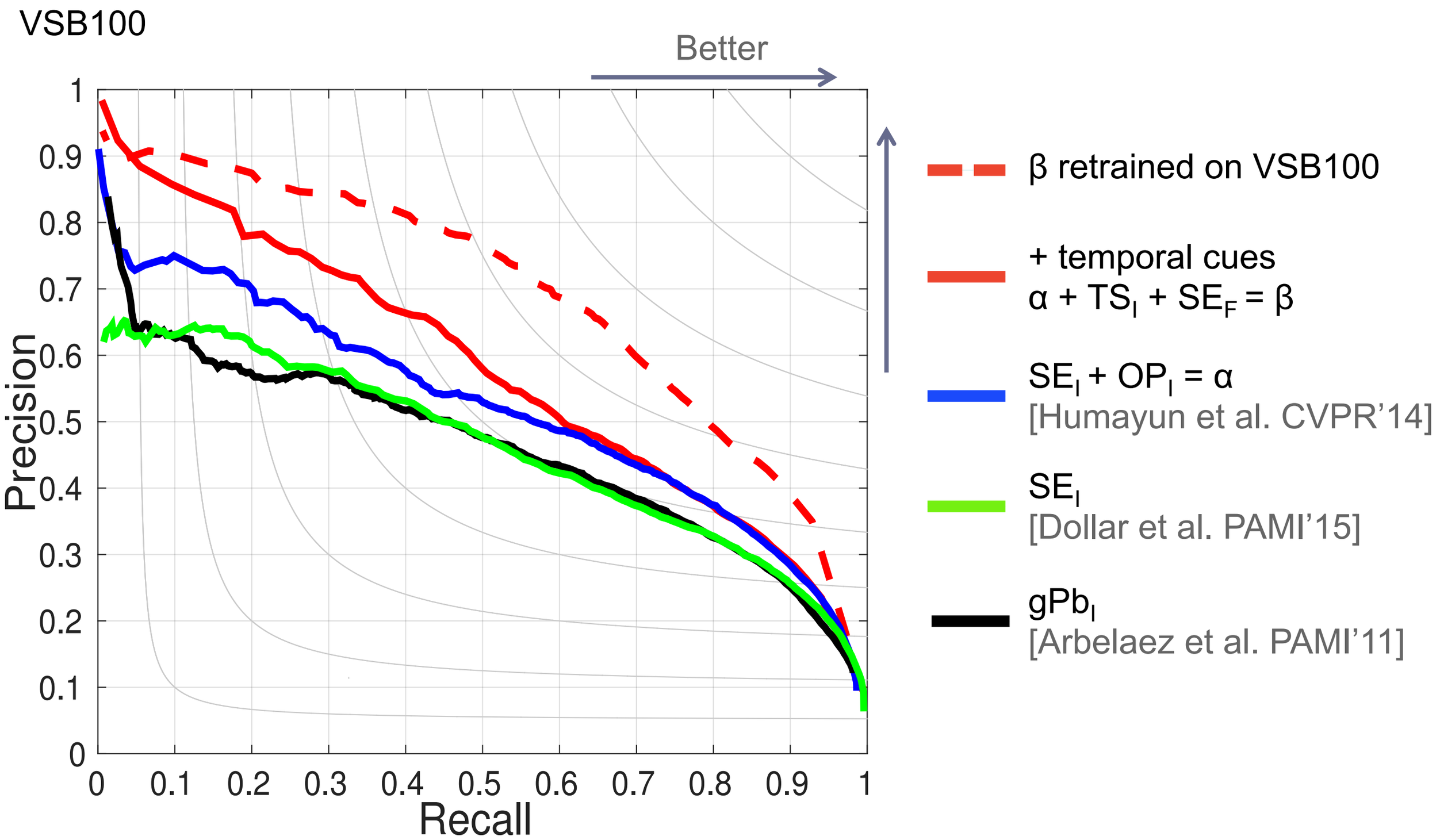
+ temporal cues
$\alpha + TS_I + SE_F = \beta$

$SE_I + OP_I = \alpha$
[Humayun et al. CVPR'14]

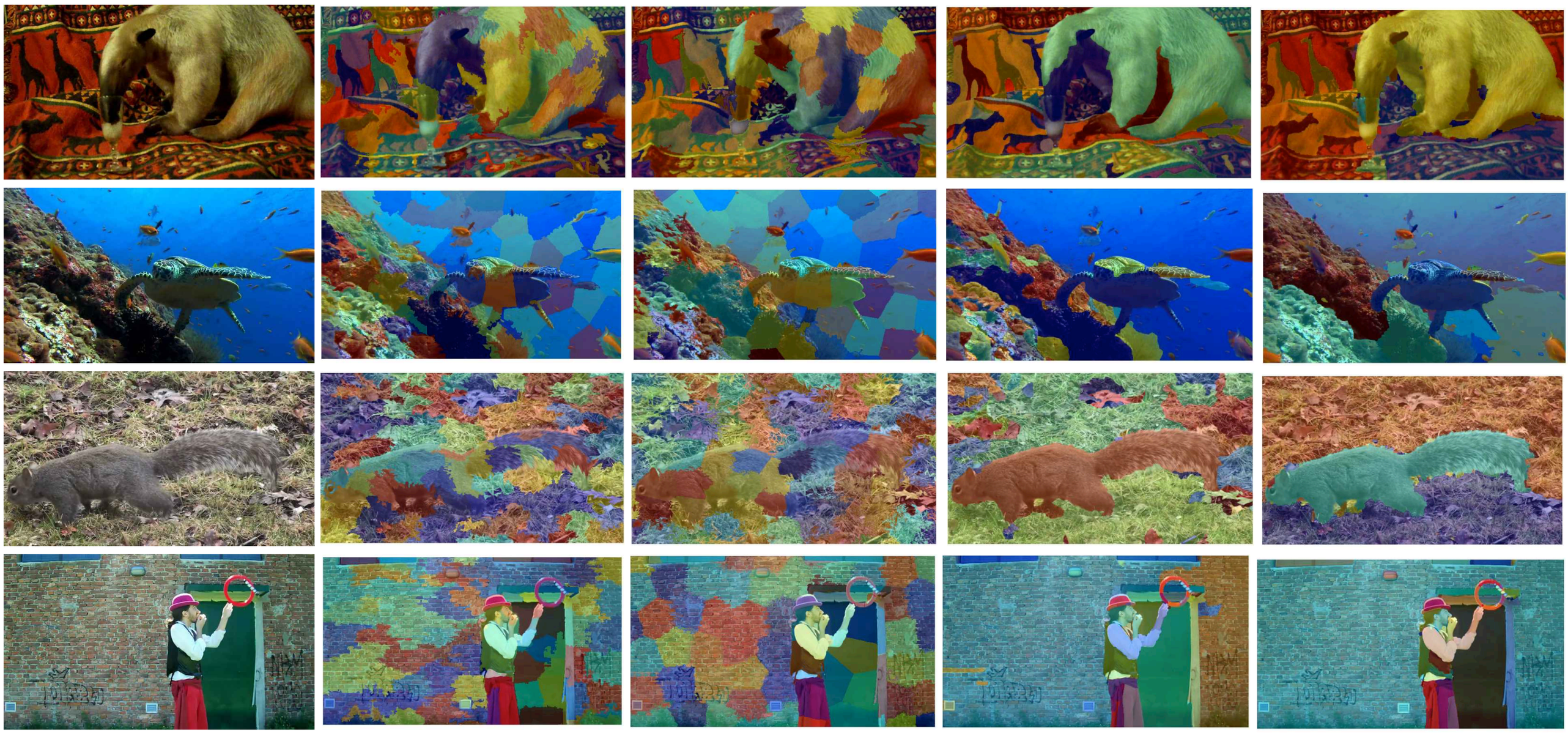
SE_I
[Dollar et al. PAMI'15]

gPb_I
[Arbelaez et al. PAMI'11]

Evaluation of Image Boundaries on VSB100

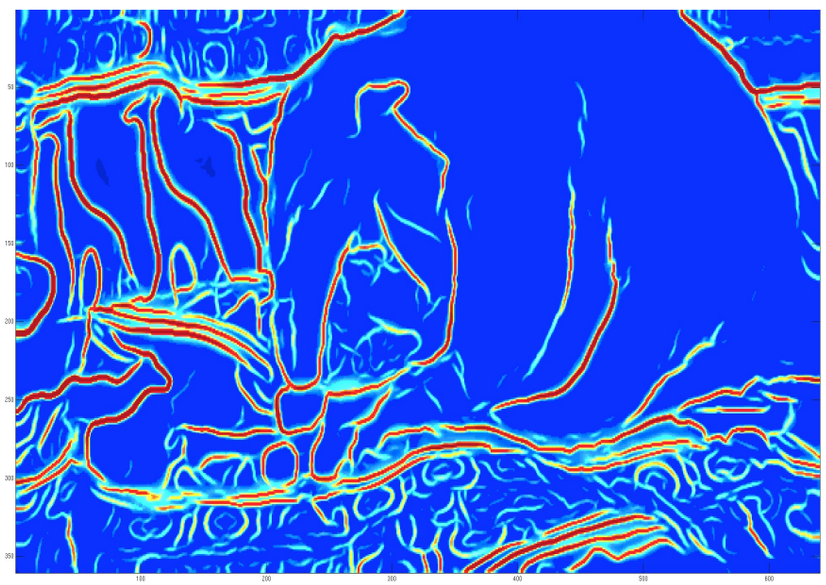


Comparison of Superpixel/voxel Methods

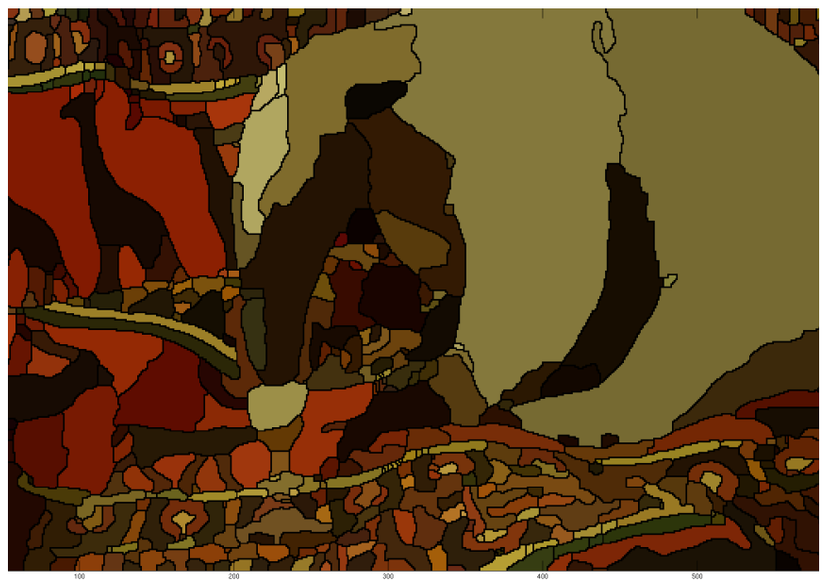


Video | SLIC 2D [Achanta et al. PAMI'12] | TSP [Chang et al. CVPR'13] | gPb [Arbelaez et al. PAMI'11] | Our superpixels

Better Boundaries for Superpixels



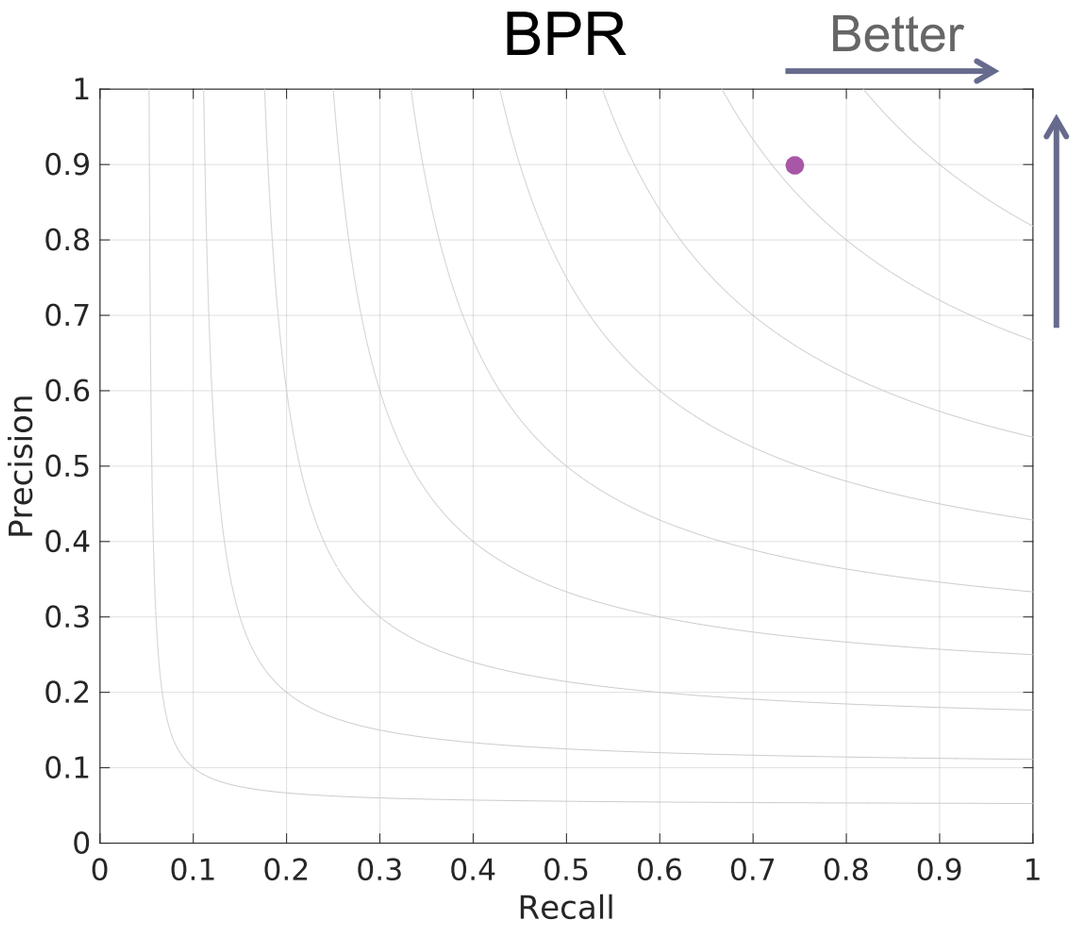
Boundaries

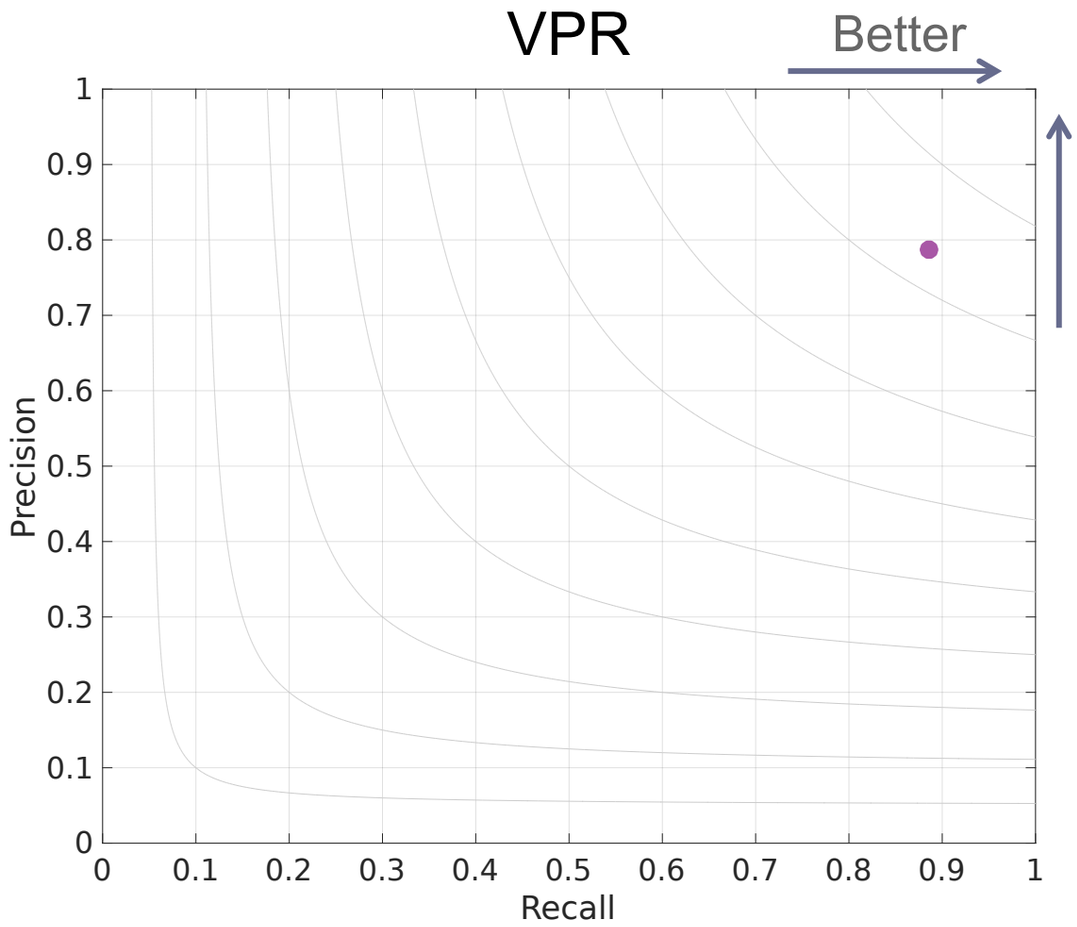


Superpixels

Better boundaries → Better superpixels → Improved video segmentation

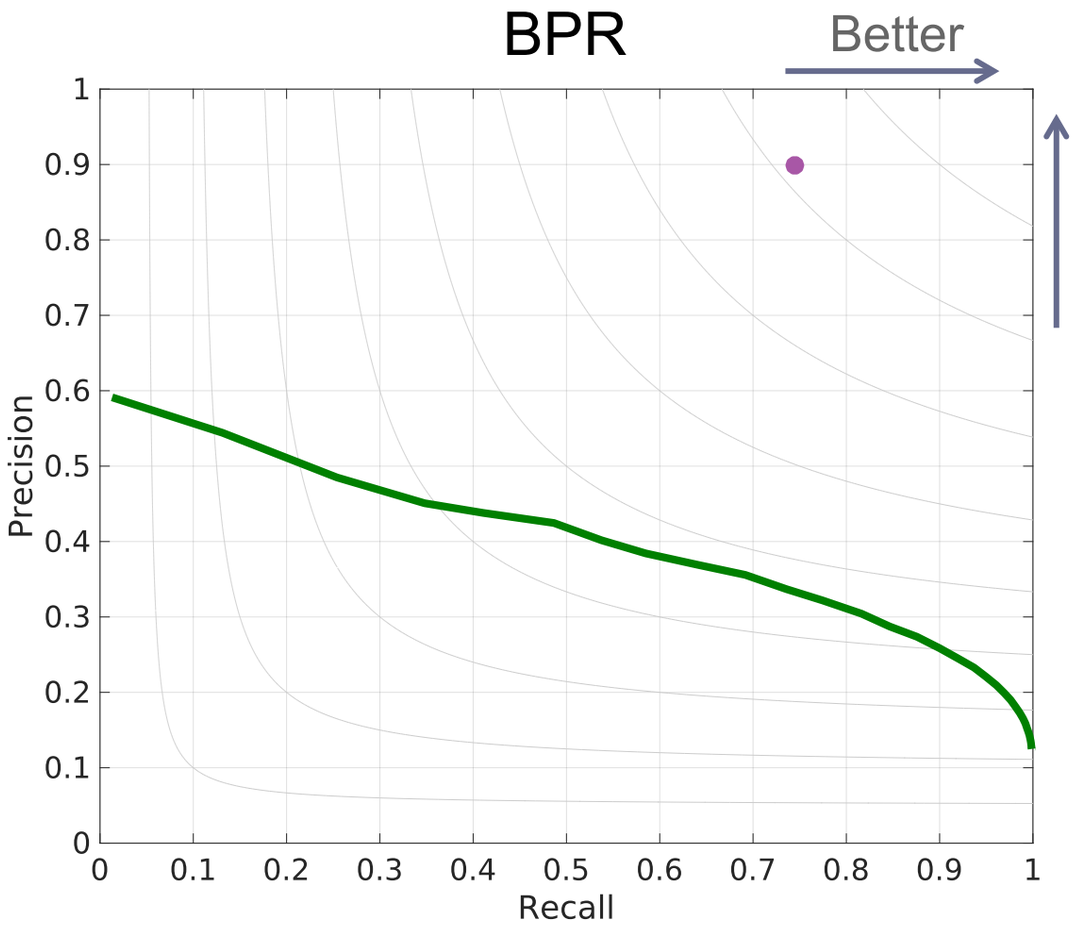
Video Segmentation Results on VSB100

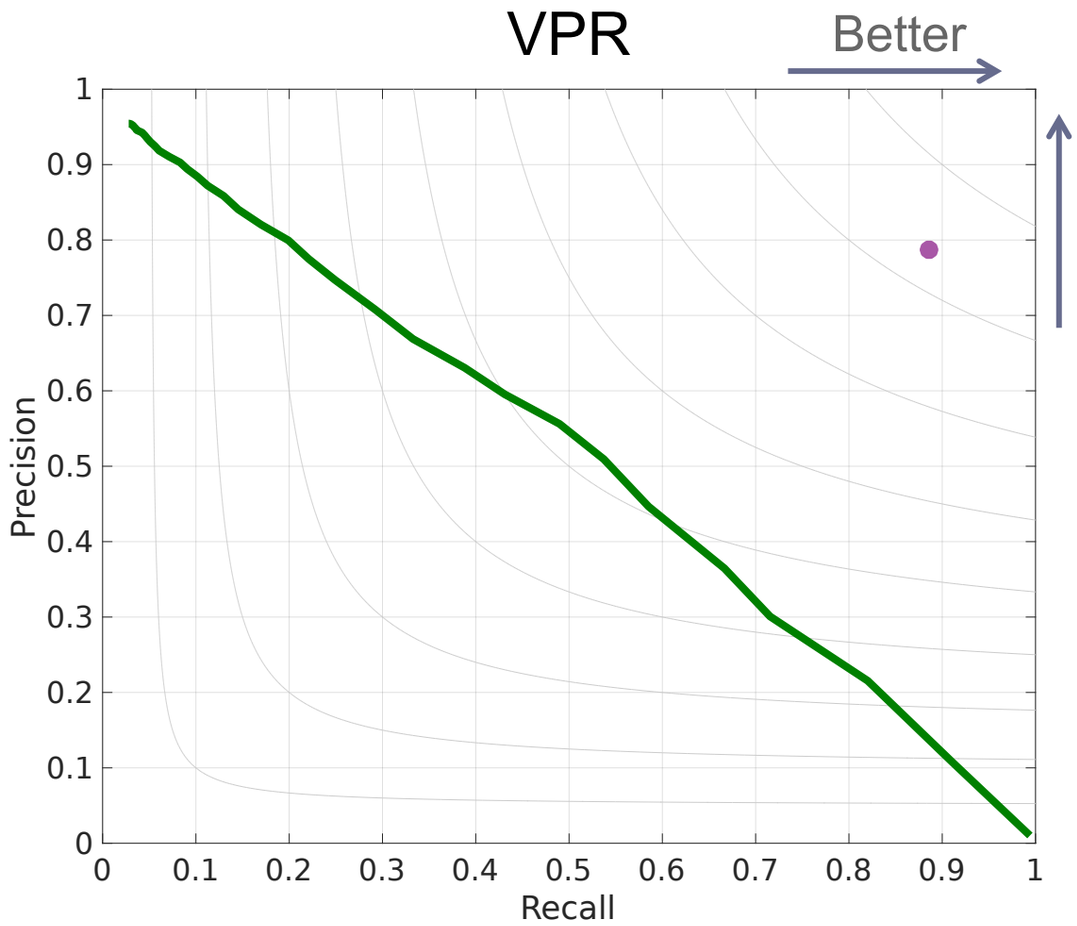




- Human

Video Segmentation Results on VSB100

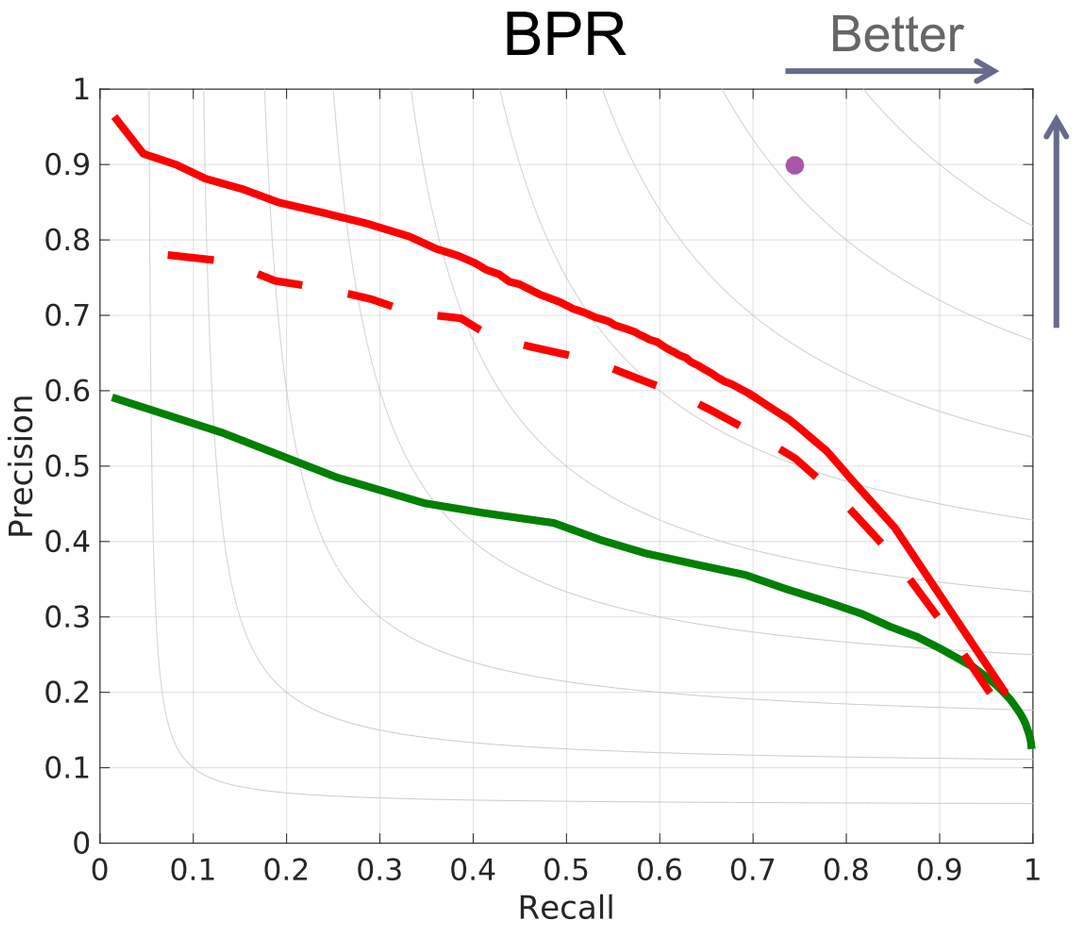


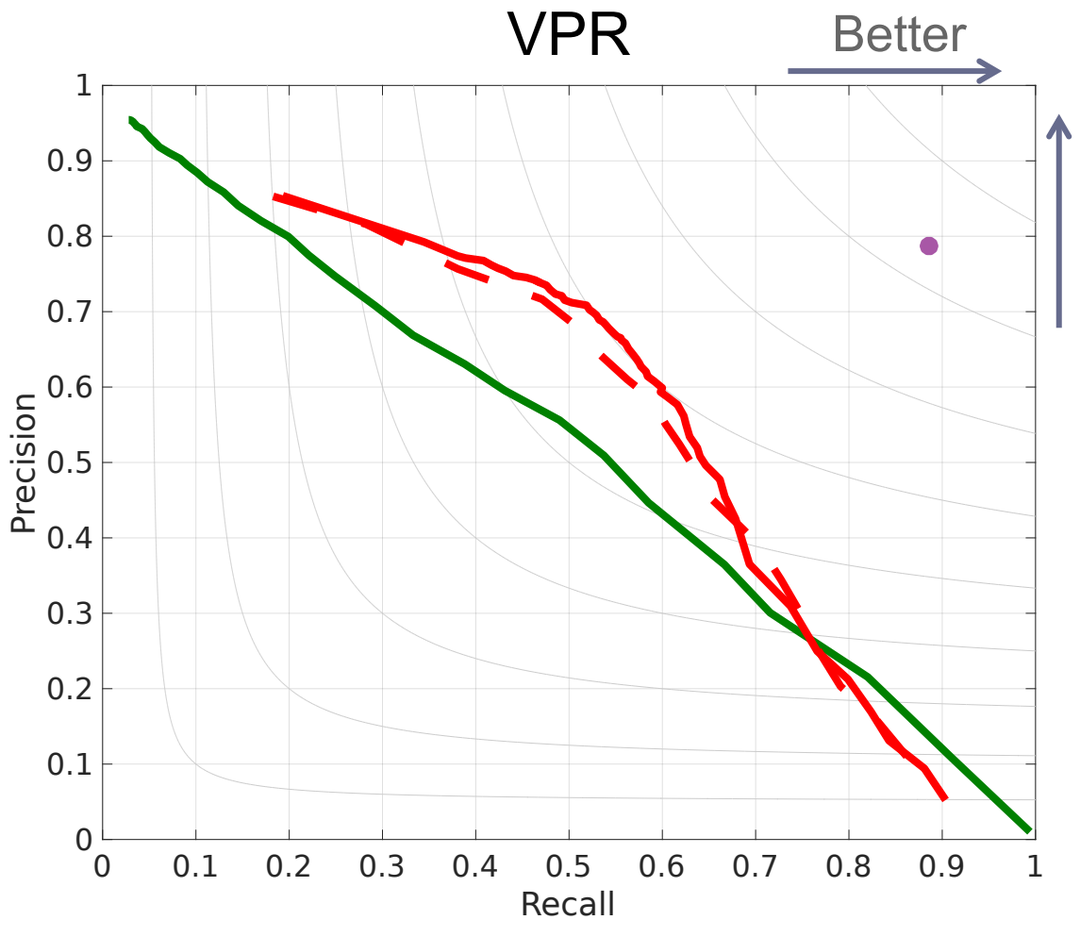


Human

Grundmann et al. CVPR'10

Video Segmentation Results on VSB100





Human

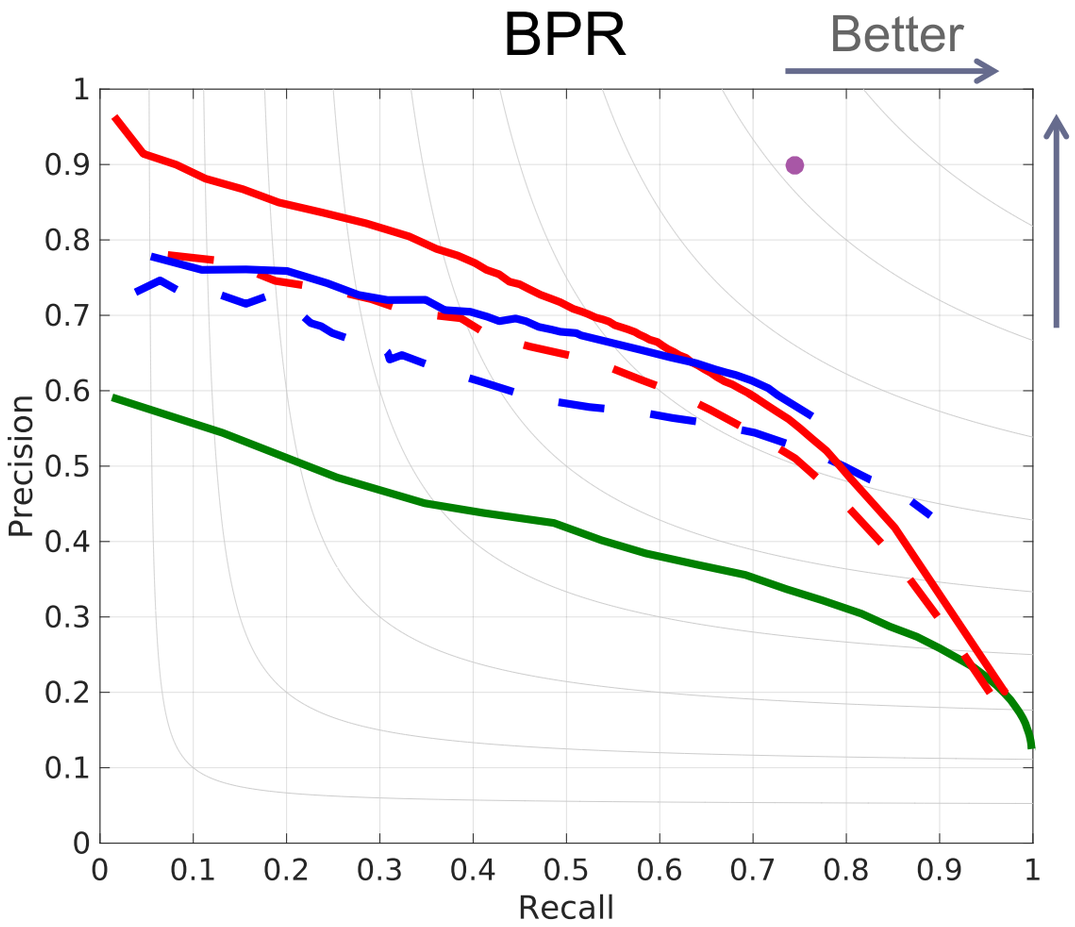
Grundmann et al. CVPR'10

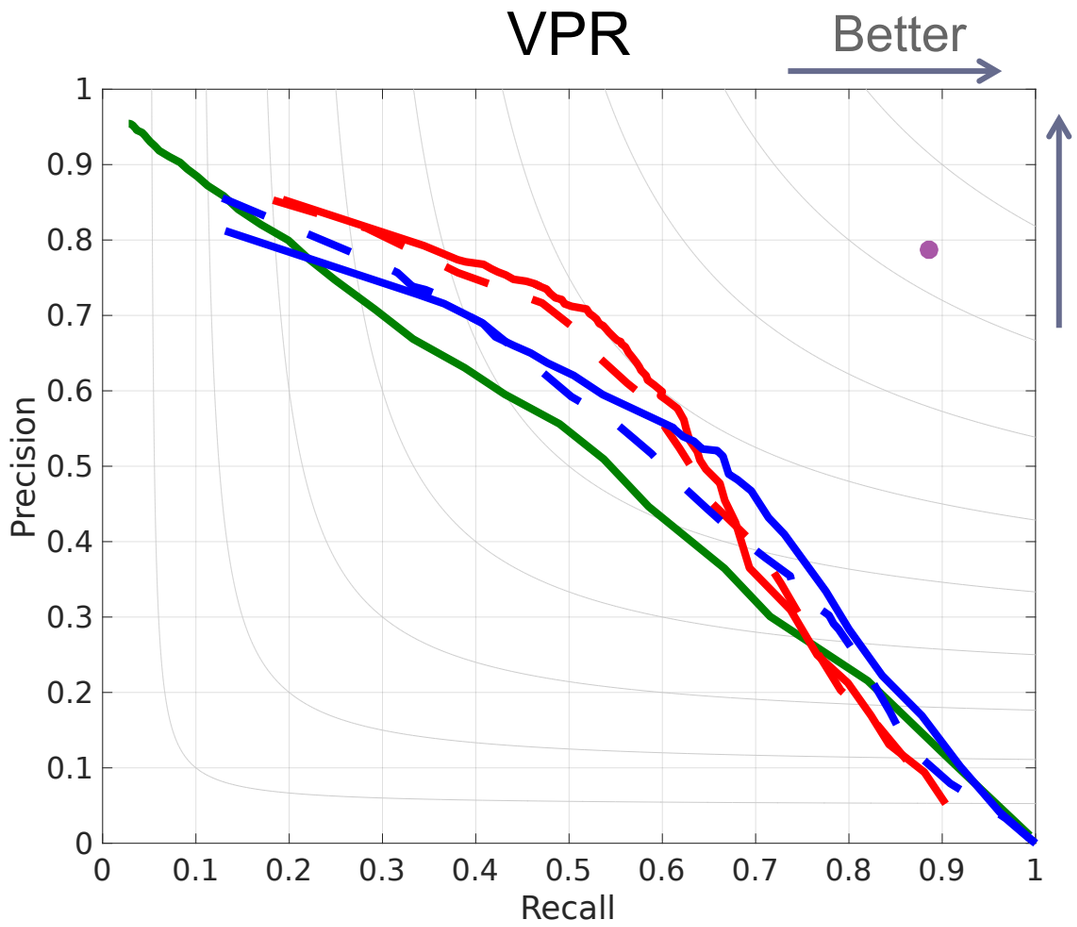
gPb

Our SPX

} + Segm. propagation ICCV'13

Video Segmentation Results on VSB100





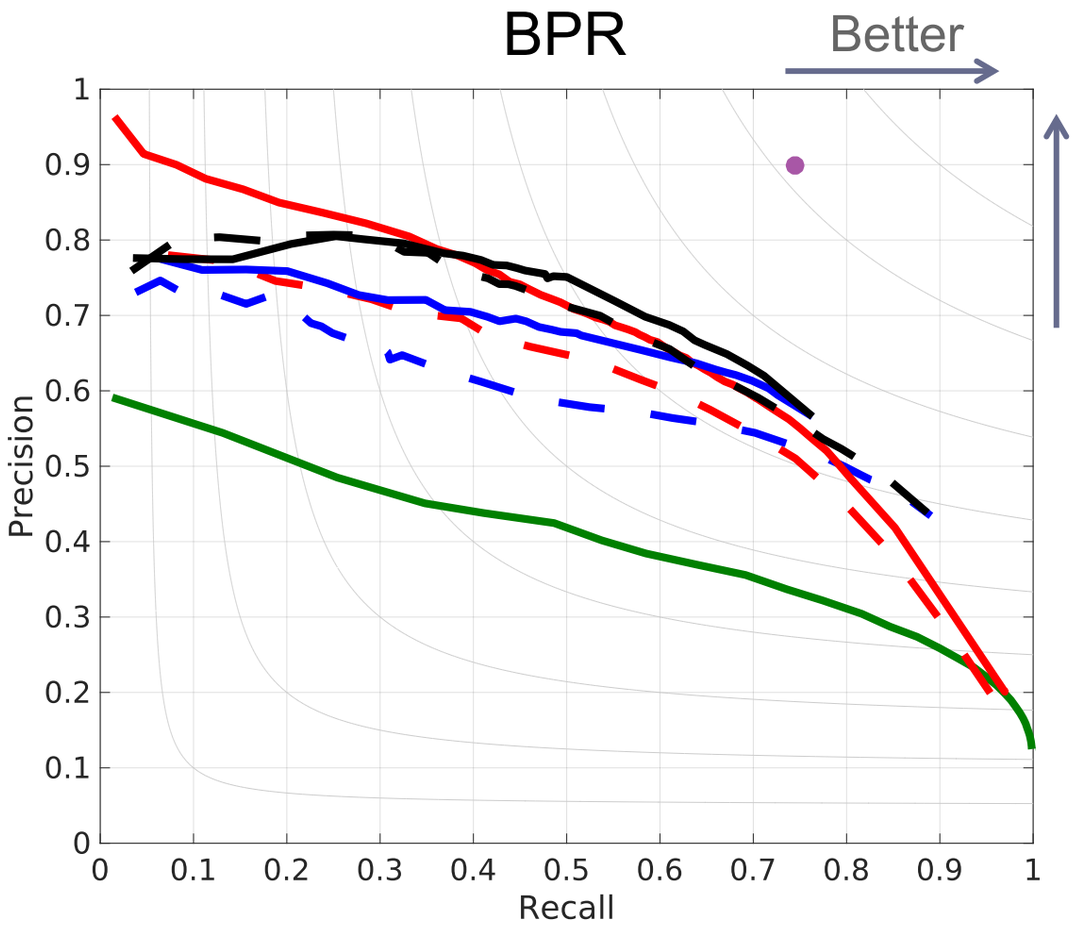
Human

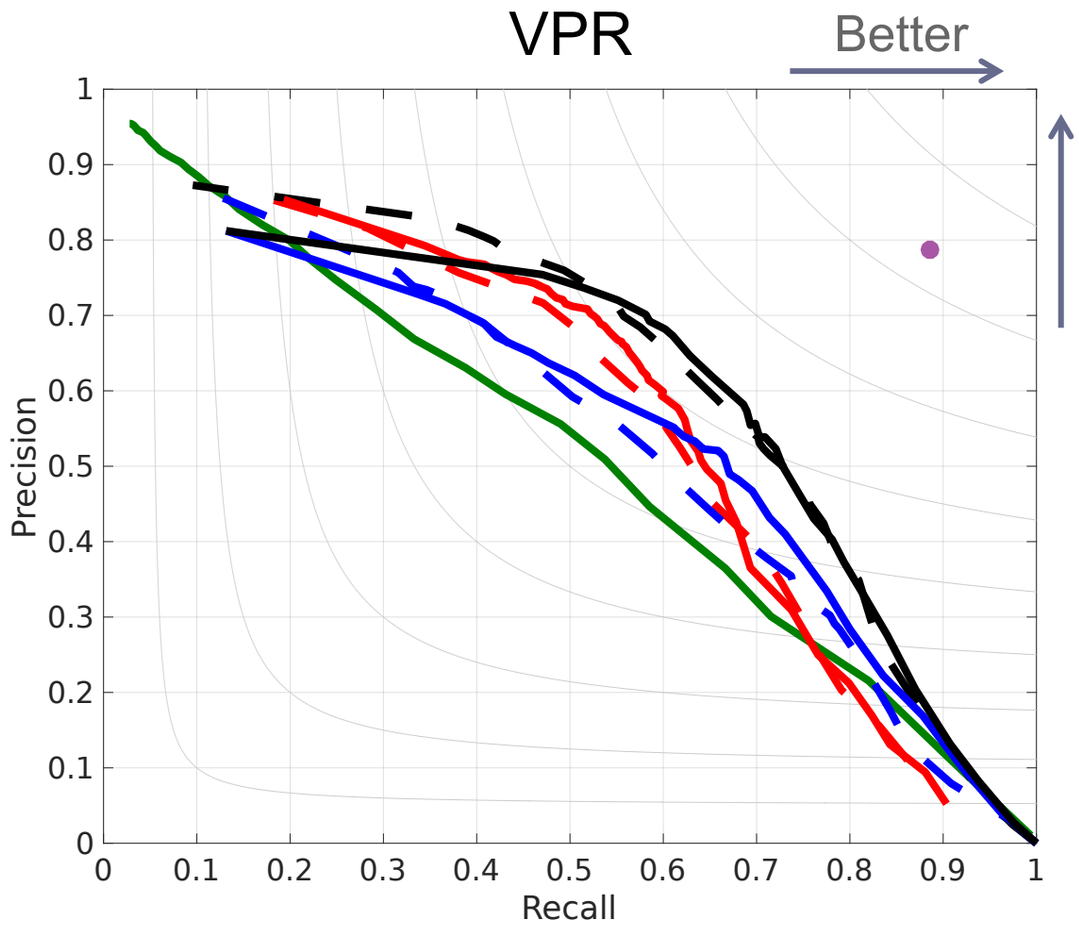
Grundmann et al. CVPR'10

gPb
Our SPX
} + Segm. propagation ICCV'13

gPb
Our SPX
} + Galasso et al. CVPR'14

Video Segmentation Results on VSB100





Human

Grundmann et al. CVPR'10

gPb
Our SPX
} + Segm. propagation ICCV'13

gPb
Our SPX
} + Galasso et al. CVPR'14

gPb
Our SPX
} + Khoreva et al. CVPR'15

buffalos

Video

Segm. propagation

Galasso et al.'14

Ground truth

Our SPX + Segm. propagation

Our SPX + Galasso et al.'14

kim yu na

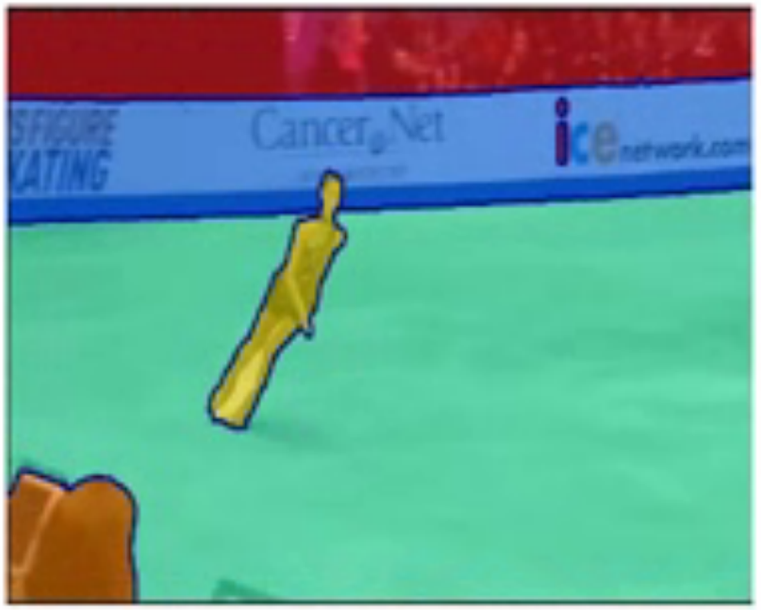
Video

Segm. propagation

Galasso et al.'14

Ground truth

Our SPX + Segm. propagation



Our SPX + Galasso et al.'14

Contributions

- a comparative evaluation of the superpixels/voxels methods
- improved boundary estimates (and thus superpixels) by the fusion of image and time domain cues
- integration of high-level object-related cues into the local image segmentation
- state-of-the-art video segmentation results on the VSB100 and BMDS datasets

Video

TSP

gPb

Our superpixels