Fast Object Segmentation in Unconstrained Video

Anestis Papazoglou and Vittorio Ferrari

Outline

- Introduction
- Related Work
- Method
- Results
- References

Introduction

- Video object segmentation is the task of separating foreground objects from the background in a video

- Important for a wide range of applications, including providing spatial support for learning object class models, video summarization, and action recognition

Introduction

- There are two main model for segmentation:

  - *Require user annotation:* for example, user should annotate the object position

  - *Fully automatic:* the only input is the input video

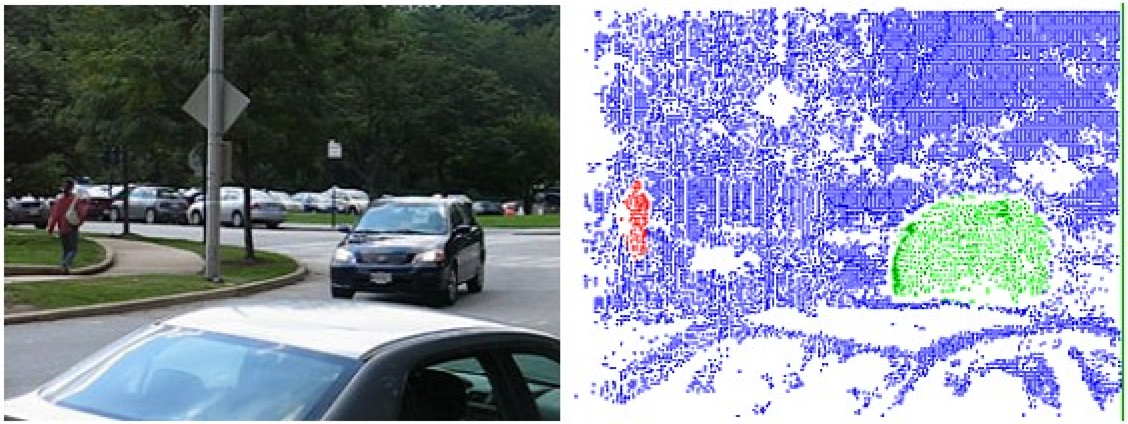
Introduction

- This paper proposes a technique for fully automatic video object segmentation in unconstrained settings

- It makes minimal assumptions about the video:the only requirement is for the object to move differently from its surrounding background in a good fraction of the video

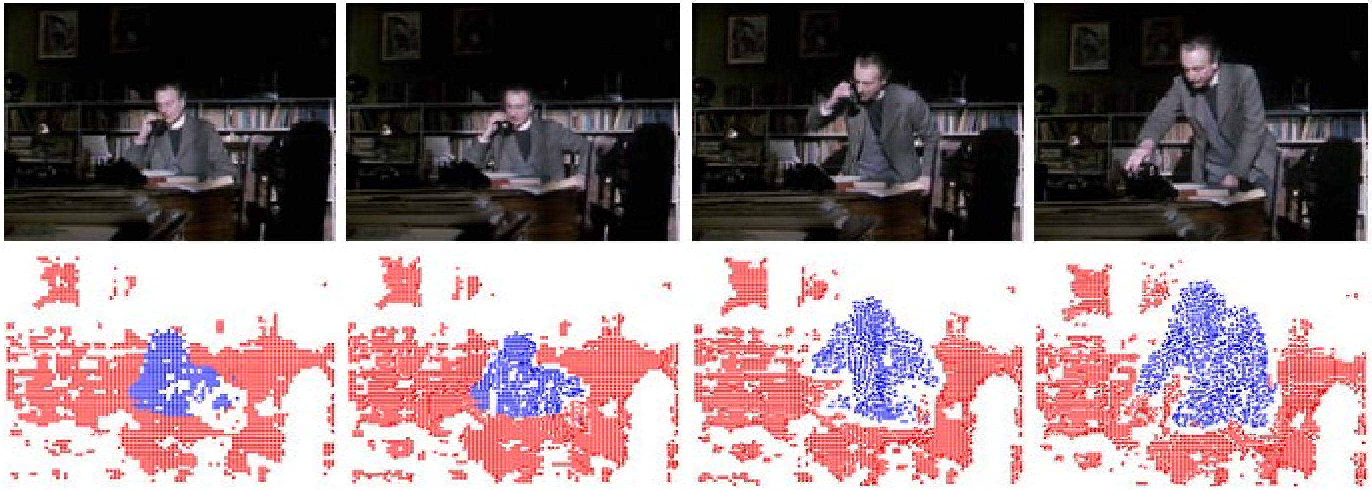
Related Work

- Object Segmentation by Long Term Analysis of Point Trajectories *(T. Brox, J. Malik), ECCV 2010.*
  - they describe a motion clustering method



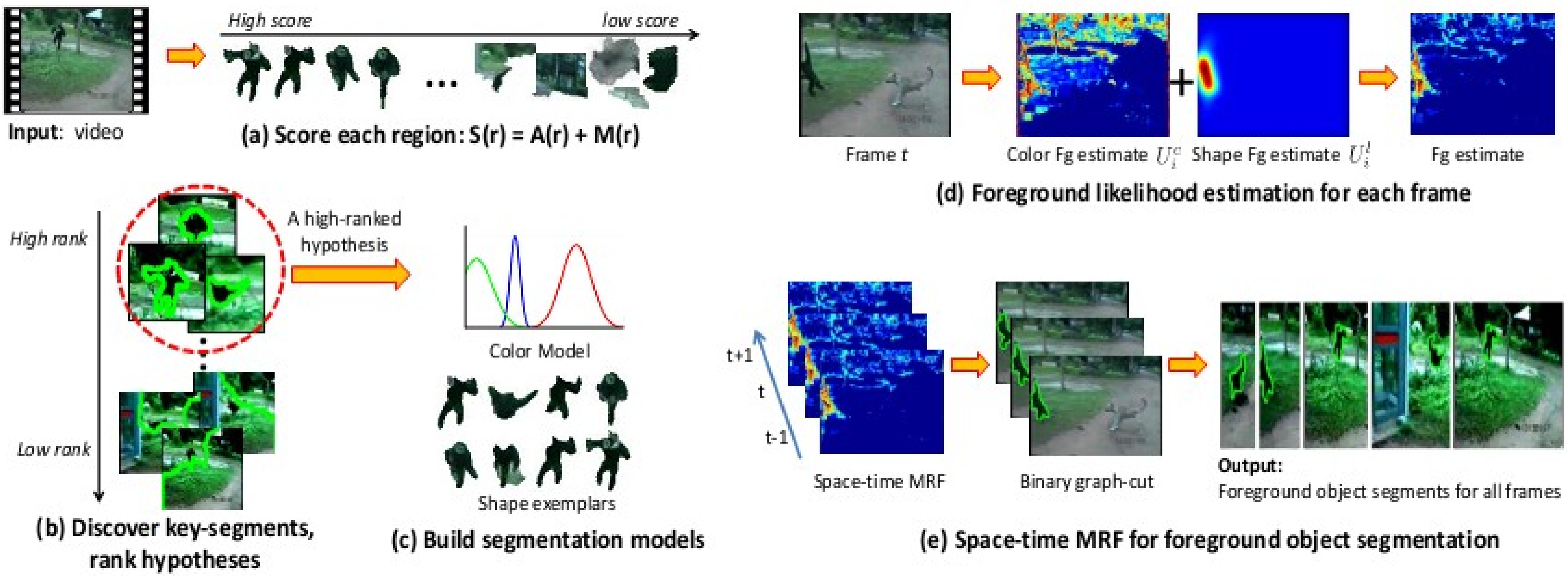
Related Work

- Object Segmentation by Long Term Analysis of Point Trajectories *(T. Brox, J. Malik), ECCV 2010.*
  - temporally consistent clusters over many frames can be obtained best by a nalyzing long term point trajectories rather than two-frame motion fields.



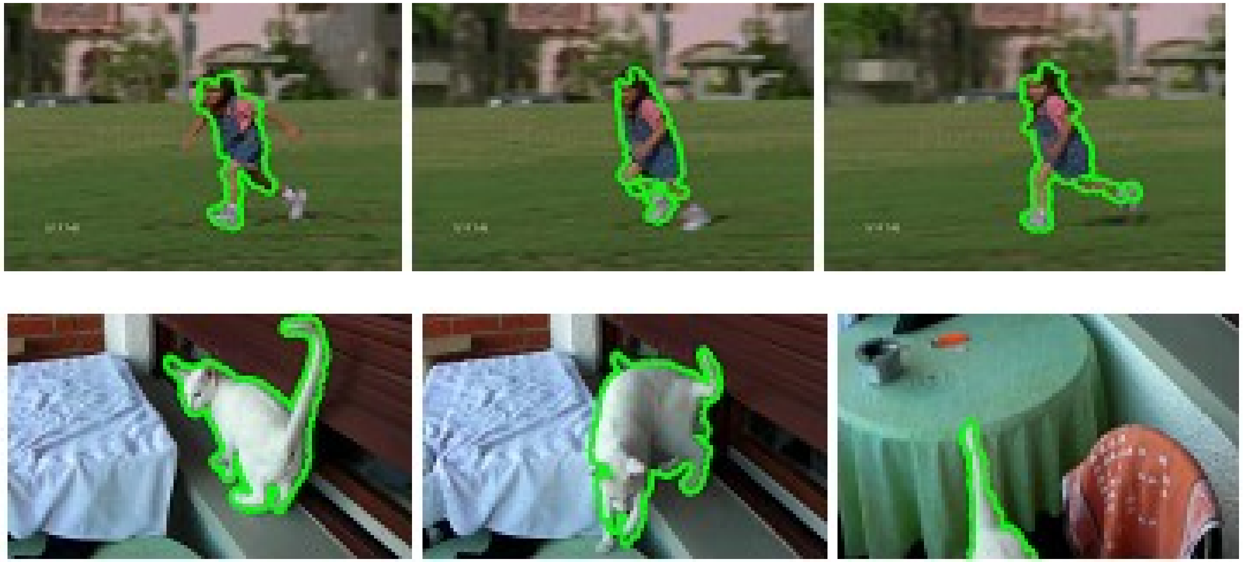
Related Work

- Key-Segments for Video Object Segmentation *(Y.J. Lee, J. Kim, K. Grauman), ICCV 2011.*



Method

- The method aims to segment objects that move differently than their surroundings.



Method

- The method consists of two steps:

  I. Initial foreground estimation

  III. Foreground-background labelling refinement

Method

I. Initial foreground estimation

- The goal of the first stage is to rapidly produce an initial estimate of which pixels might be inside the object based purely on motion.

- The motion boundaries detected by optical flow

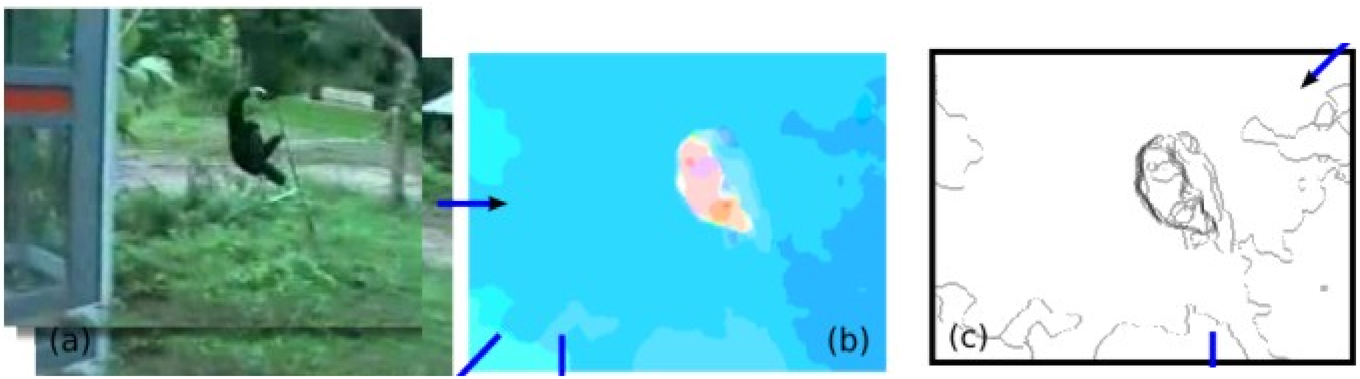
Initial foreground estimation

i. Optical flow estimation

Initial foreground estimation

ii. Motion Boundaries

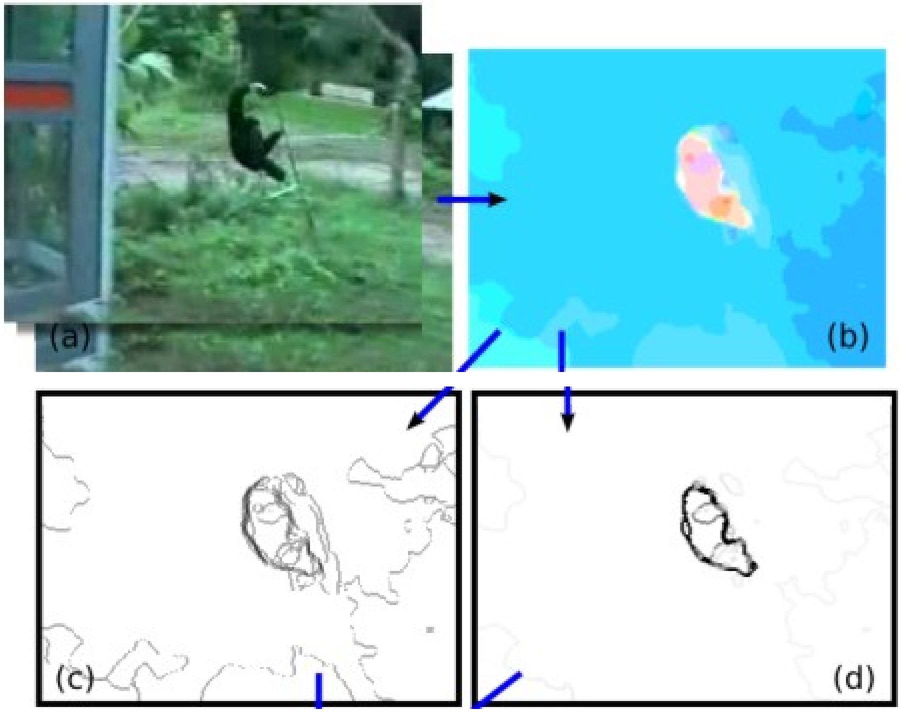
$$b_p^m = 1 - \exp\left(-\lambda \|\nabla \vec{f}_p\|\right)$$



Initial foreground estimation

ii. Motion Boundaries

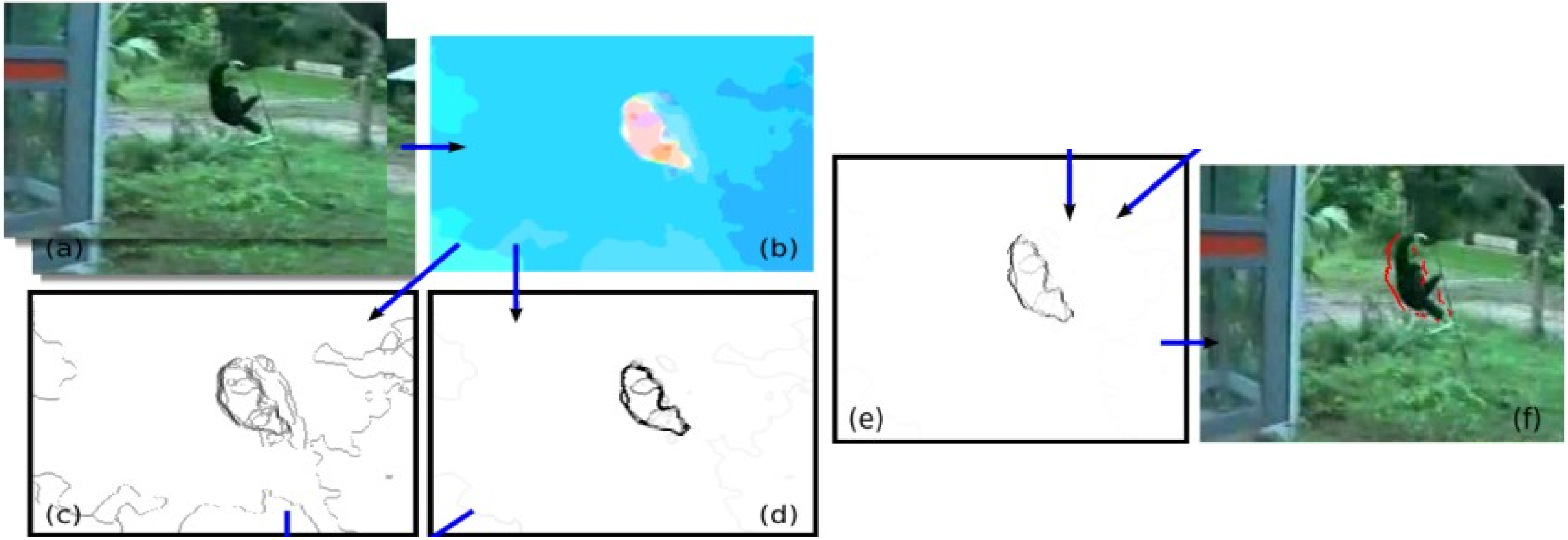
$$b_p^{\theta} = 1 - \exp\left(-\lambda^{\theta} max_{q \in N}\left(\delta \theta_{p,q}^2\right)\right)$$



Initial foreground estimation

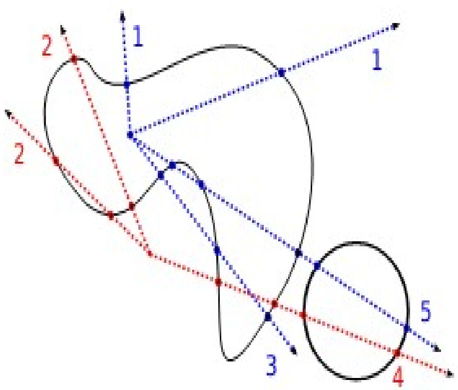
ii. Motion Boundaries

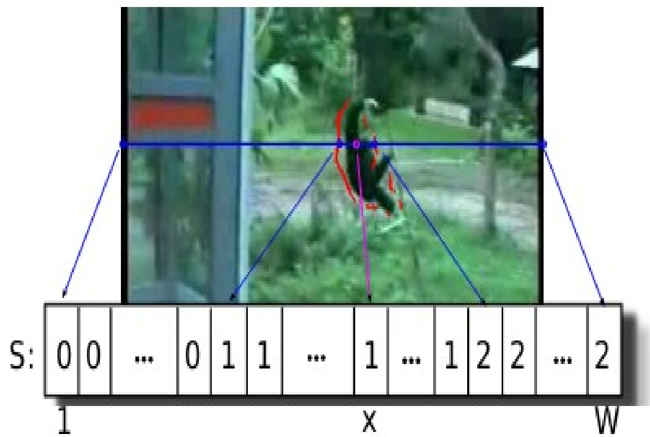
$$b_p = \begin{cases} b_p^m & \text{if } b_p^m > T \\ b_p^m \cdot b_p^\theta & \text{if } b_p^m \leq T \end{cases}$$



Initial foreground estimation

iii. Inside-outside maps





Method

II. Foreground-background labelling refinement

- They formulate video segmentation as a pixel labelling problem with two labels (foreground and background)

$$E(\mathcal{L}) = \sum_{t,i} A_i^t(l_i^t) + \alpha_1 \sum_{t,i} L_i^t(l_i^t) + \alpha_2 \sum_{(i,j,t)\in\mathcal{E}_s} V_{ij}^t(l_i^t, l_j^t) + \alpha_3 \sum_{(i,j,t)\in\mathcal{E}_t} W_{ij}^t(l_i^t, l_j^{t+1})$$

$$\mathcal{L}^* = \operatorname*{argmin}_{\mathcal{L}} E(\mathcal{L})$$

Method

II. Foreground-background labelling refinement

- Appearance Model (A^t)
  - The appearance model consists of two GMM over RGB colour values,one for the foreground and one for the background.
  - They are estimated automatically based on the inside-outside maps M^t
  - Weight of each superpixel $s_i^{t'}$ in frame t'
    - *foreground:* $\exp\left(-\lambda^{A.}(t-t')^2\right).r_i^{t'}$

      background: $\exp\left(-\lambda^{A.}(t-t')^2\right).\left(1-r_i^{t'}\right)$

Method

II. Foreground-background labelling refinement

- Location Model (L^t)
  - inside-outside maps can provide a valuable location prior to anchor the segmentation to image areas likely to contain the object, as they move differently from the surrounding region

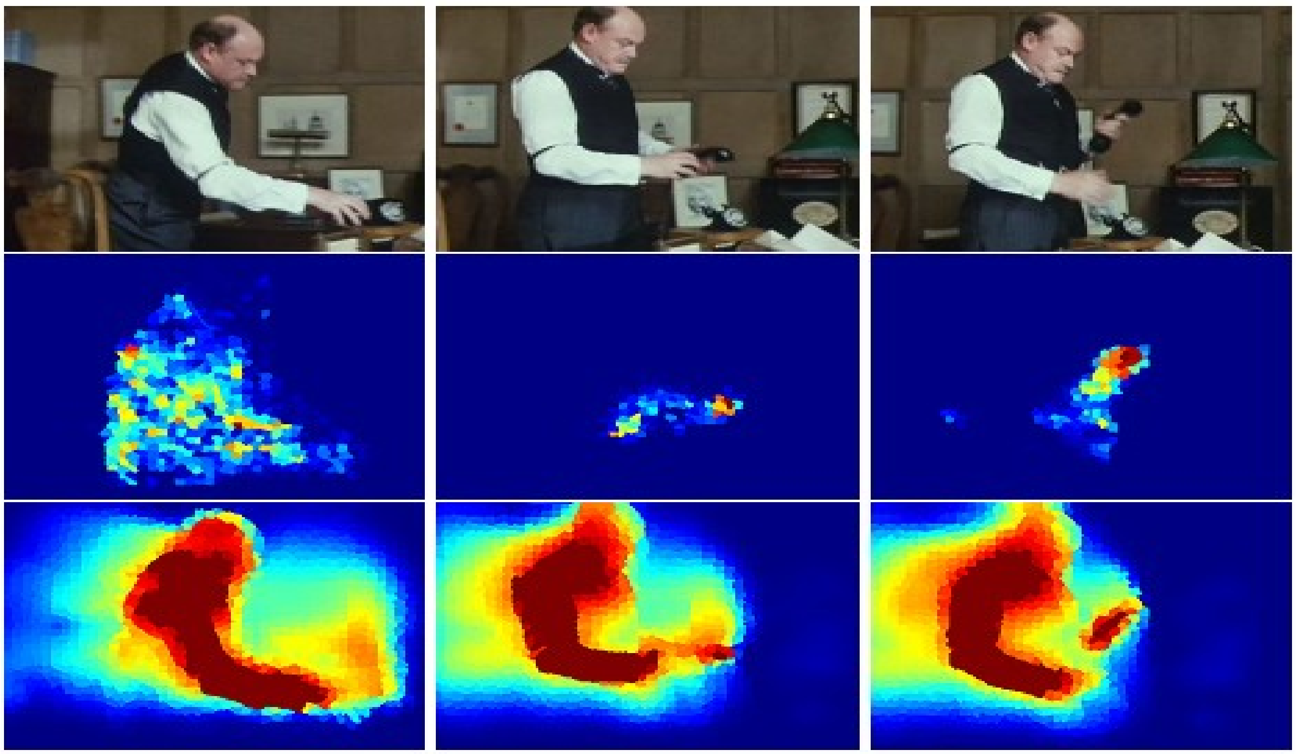
$$L_j^{t+1} := L_j^{t+1} + \gamma \frac{\sum_i \phi(s_i^t, s_j^{t+1}) \cdot \psi(s_i^t) \cdot L_i^t}{\sum_i \phi(s_i^t, s_j^{t+1})}$$

$$\psi(s_i^t) = \exp(-\lambda^{\psi} \sum_{p \in s_i^t} ||\nabla \vec{f}_p||)$$

Method

II. Foreground-background labelling refinement

- Location Model (L^t)



Method

II. Foreground-background labelling refinement

- Smoothness Terms
  - Spatial smoothness potential

$$V_{ij}^{t}(l_i^t, l_j^t) = \mathrm{dis}(s_i^t, s_j^t)^{-1}[l_i^t \neq l_j^t]\exp(-\beta \mathrm{col}(s_i^t, s_j^t)^2)$$

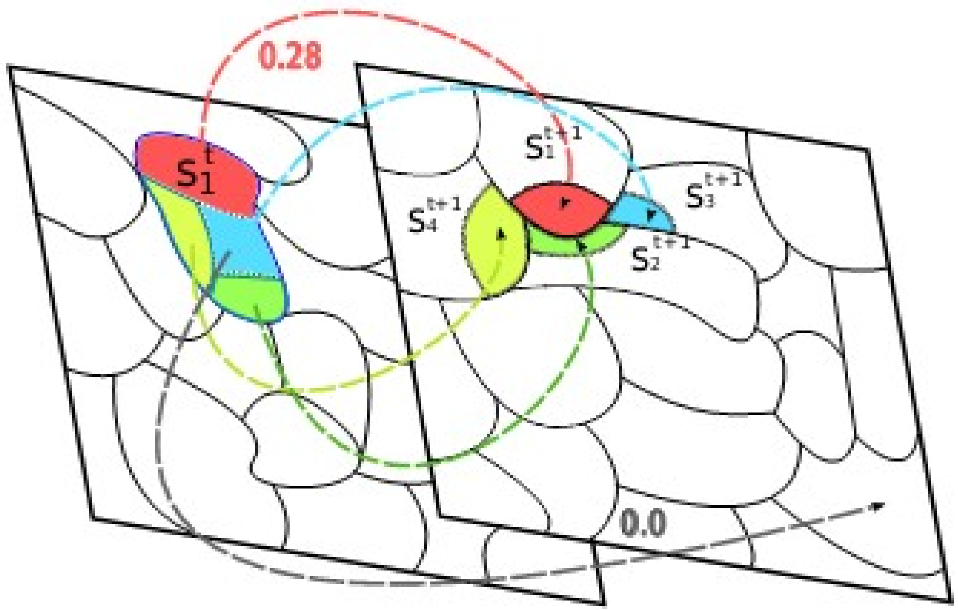
  - Temporal smoothness potential

$$W_{ij}^{t}(l_i^t, l_j^{t+1}) = \phi(s_i^t, s_j^{t+1})[l_i^t \neq l_j^t]\exp(-\beta \mathrm{col}(s_i^t, s_j^{t+1})^2)$$

Method

II. Foreground-background labelling refinement

- Smoothness Terms



Results

1) SegTrack

precision	ours	[14]	[16]	[27]	[6]	[18]	[4]
birdfall	217	288	189	155	468	468	606
cheetah	890	905 (34228)	806	633	1968	1175	11210
girl	3859	1785	1698	1488	7595	5683	26409
monkey	284	521 (64339)	472	365	1434	1434	12662
parachute	855	201	221	220	1113	1595	40251

Results

1) SegTrack

Results

2) Youtube Objects

	aero	bird	boat	car	cat	cow	dog	horse	mbike	train	avg
Clustering tracks [6]	53.9	19.6	38.2	37.8	32.2	21.8	27.0	34.7	45.4	37.5	34.8
Automatic segment selection [19]	51.7	17.5	34.4	34.7	22.3	17.9	13.5	26.7	41.2	25.0	28.5
ours	65.4	67.3	38.9	65.2	46.3	40.2	65.3	48.4	39.0	25.0	50.1

Results

2) Youtube Objects

Thanks