KEY-SEGMENTS FOR VIDEO OBJECT SEGMENTATION

Yong Jae Lee, Jaechul Kim, and Kristen Grauman

BIL-722

Çağdaş Baş

WHAT IS OBJECT SEGMENTATION?

- Finding foreground objects.
- Finding interesting objects.
- Unsupervised segmentation and tracking has received little attention

RELATED WORK

- Bottom-up methods:
  - Most of the saliency detectors.
  - Exploring only image cues.
- Supervised methods:
  - Requires user annotation.
  - Tracking-based methods demand less user input (for only first frame)
- Semi-supervised methods:
  - Cannot generalize for all types of objects

APPROACH
A - Finding object like regions

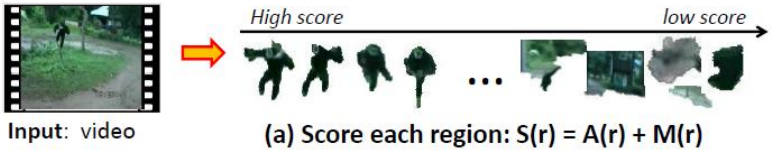
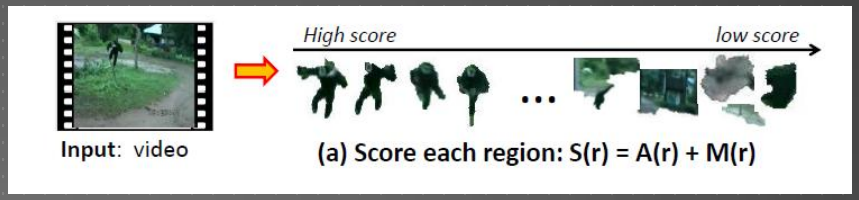


Figure Credit: Yong Jae Lee, Jaechul Kim, and Kristen Grauman

A - FINDING OBJECT LIKE REGIONS

- Appearance and motion cue is important to identfy object-like regions.
- 1000 regions are extracted for each frame
- $S(r) = A(r) + M(r)$



A - FINDING OBJECT LIKE REGIONS (CONT.)

- Use appearance cues to extract static object-like features.

- Use Histogram of Optical Flow Histograms on extracted regions to differ objects to background:

$$M(r) = 1 - \exp(-\chi^2_{flow}(r, \bar{r}))$$

- Large appearance change indicates background regions.

APPROACH -

B - DISCOVERING KEYSEGMENTS ACROSS FRAMES

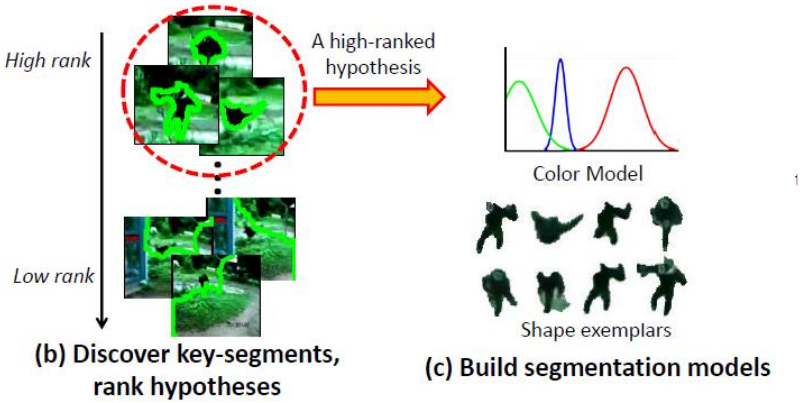


Figure Credit: Yong Jae Lee, Jaechul Kim, and Kristen Grauman

APPROACH -
B - DISCOVERING KEY SEGMENTS ACROSS FRAMES

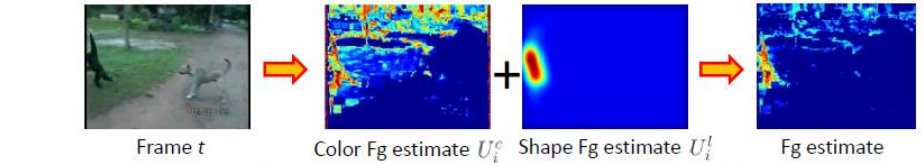
- Calculate affinity matrix with similarity measure:

$$\mathrm{K}(r_m, r_n) = \exp(-\frac{1}{\Omega}\chi^2_{color}(r_m, r_n))$$

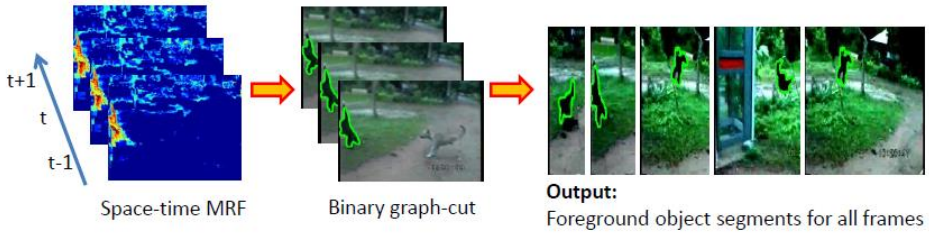
- Cluster regions via affinity matrix.
- Rank clusters based on avearage score S(r) .
- Cluster with highest score is primary foreground object.

APPROACH -

C - Foreground Object Segmentation



(d) Foreground likelihood estimation for each frame



(e) Space-time MRF for foreground object segmentation

Figure Credit: Yong Jae Lee, Jaechul Kim, and Kristen Grauman

APPROACH – C - FOREGROUND OBJECT SEGMENTATION

- Space-Time graph definition:
  - A pixel based graph is defined for whole video.
  - Then minimize the cost function

$$E(f,h) = \sum_{i \in \mathcal{S}} D_i^h(f_i) + \gamma \sum_{i,j \in \mathcal{N}} V_{i,j}(f_i, f_j),$$

  - $V_{i,j}$ is label smoothness parameter in space-time
  - D_i^h is cost of labeling pixel i with f_i, given key-segments in h (cluster).

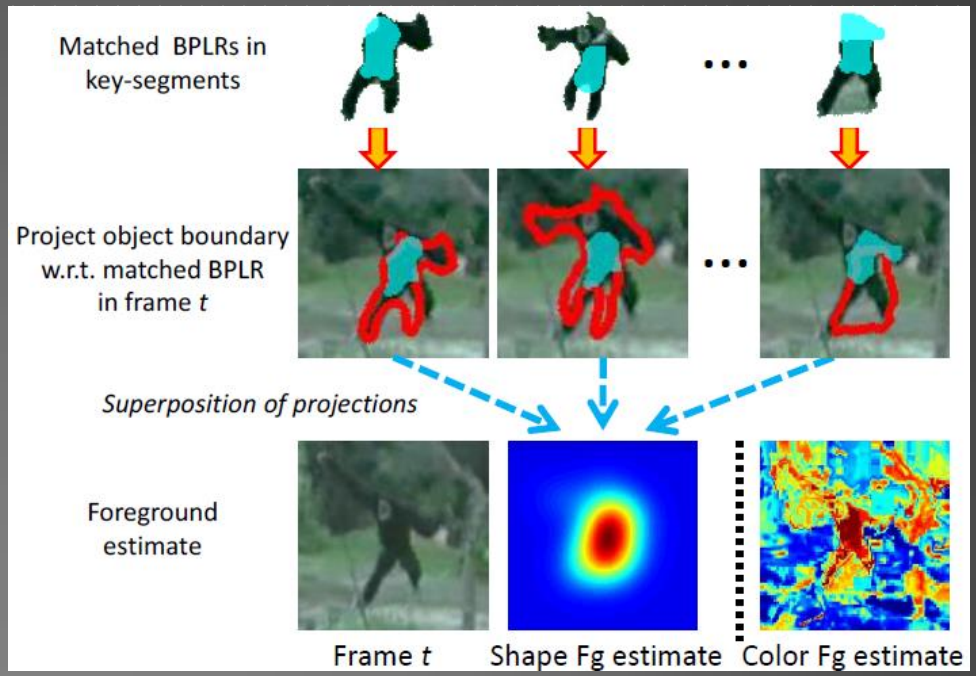
$$D_i^h(f_i) = -\log\left(\alpha \cdot U_i^c(f_i,h) + (1-\alpha)\cdot U_i^l(f_i,h)\right),$$

LABELLING COST EXPLAINED

- U_i^c is appearance based cost.
- Two Gaussian Mixture Model is estimated across a video: One for background, one for foreground.
  - fg^{color} for pixels in key-segments,
  - bg^{color} for pixels in the complement of key-segments, among all frames.
- Basically a pixel and its difference from bakcground have to be consistent across frames.

LABELLING COST EXPLAINED (CONT.)

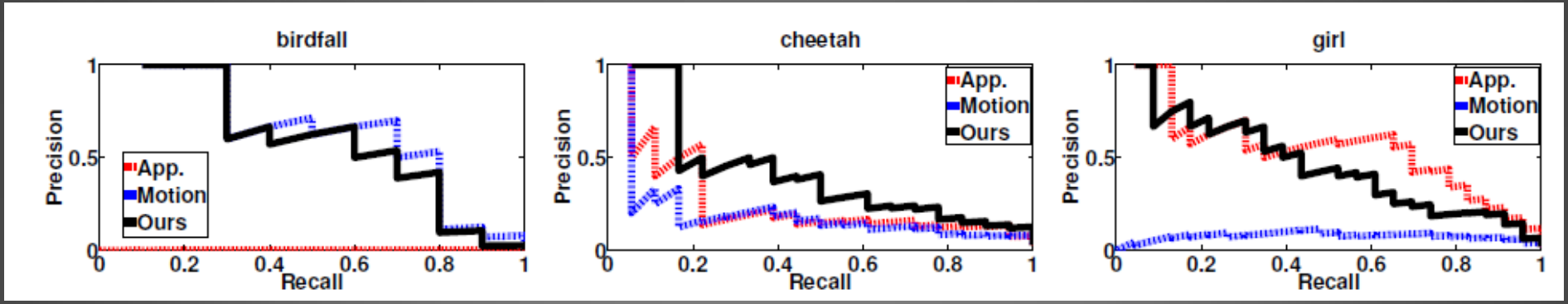
- Match each shape in a key segment across frames.
- Superposition all pairwise matches and create a Shape Fg mask for each frame.
- Each pixel will ve voted whether it moslty belongs to foreground or background.

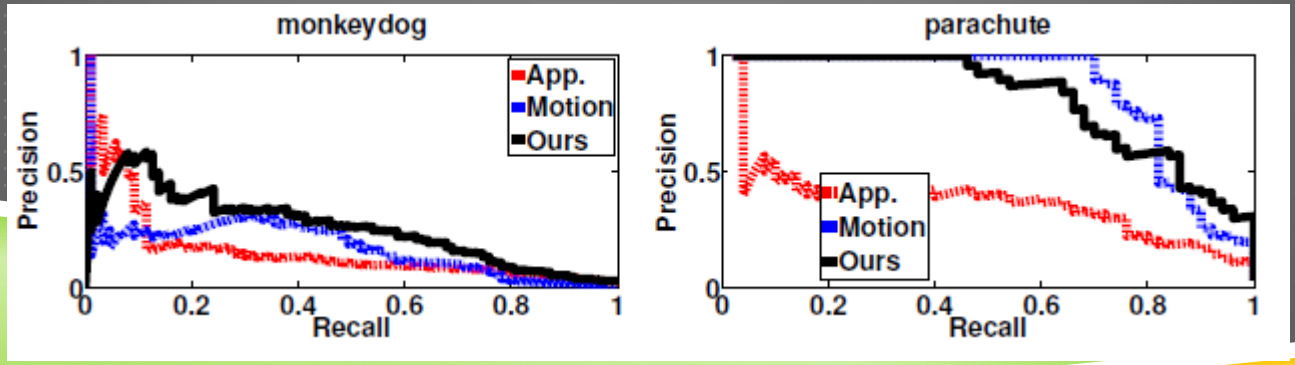


$$U_i^l(f_i) = \begin{cases} P(p_i|bg^{shape}(h)), & \text{if } f_i = 0; \\ P(p_i|fg^{shape}(h)), & \text{if } f_i = 1, \end{cases}$$

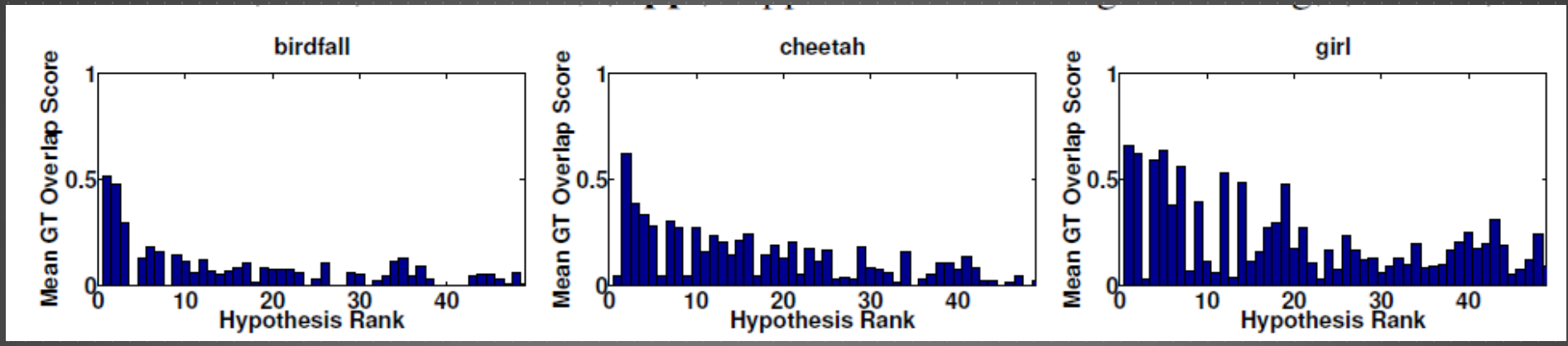
RESULTS

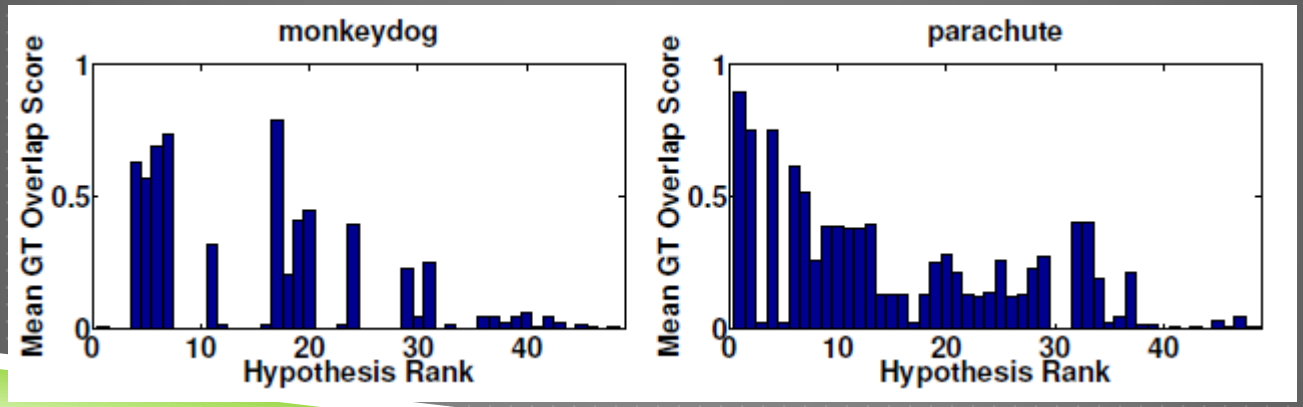
- SegTrack dataset is used [29].





RESULTS (CONT.)





RESULTS (CONT.)

	Ours	[29]	[7]	Top $A(r)$ region	Bg Sub
birdfall	288	**252**	454	26156	7435
cheetah	**905**	1142	1217	27728	28763
girl	1785	**1304**	1755	10236	45019
monkeydog	**521**	563	683	38083	31099
parachute	**201**	235	502	75168	27242
penguin	136285(*)	**1705**	6627	147686	61089
Manual seg?	No	Yes	Yes	No	No

- Segmentation errors.

	Ours	Ours w/o partial shape match
birdfall	**288**	414
cheetah	**905**	1024
girl	1785	**1534**
monkeydog	**521**	1261
parachute	201	**188**

CONCLUSION AND THANKS

- Only motion or appearance models are inadequate to extract foreground object properly.
- Top-down information can be used to improve existing methods.

Questions?