Video Segmentation using Spectral Clustering on Superpixels

Asma H. Bhatti   Anis Rahman   Asad A. Butt

School of Electrical Engineering and Computer Science
National University of Sciences and Technology, Islamabad, Pakistan

IEEE International Conference on Image Processing 2016

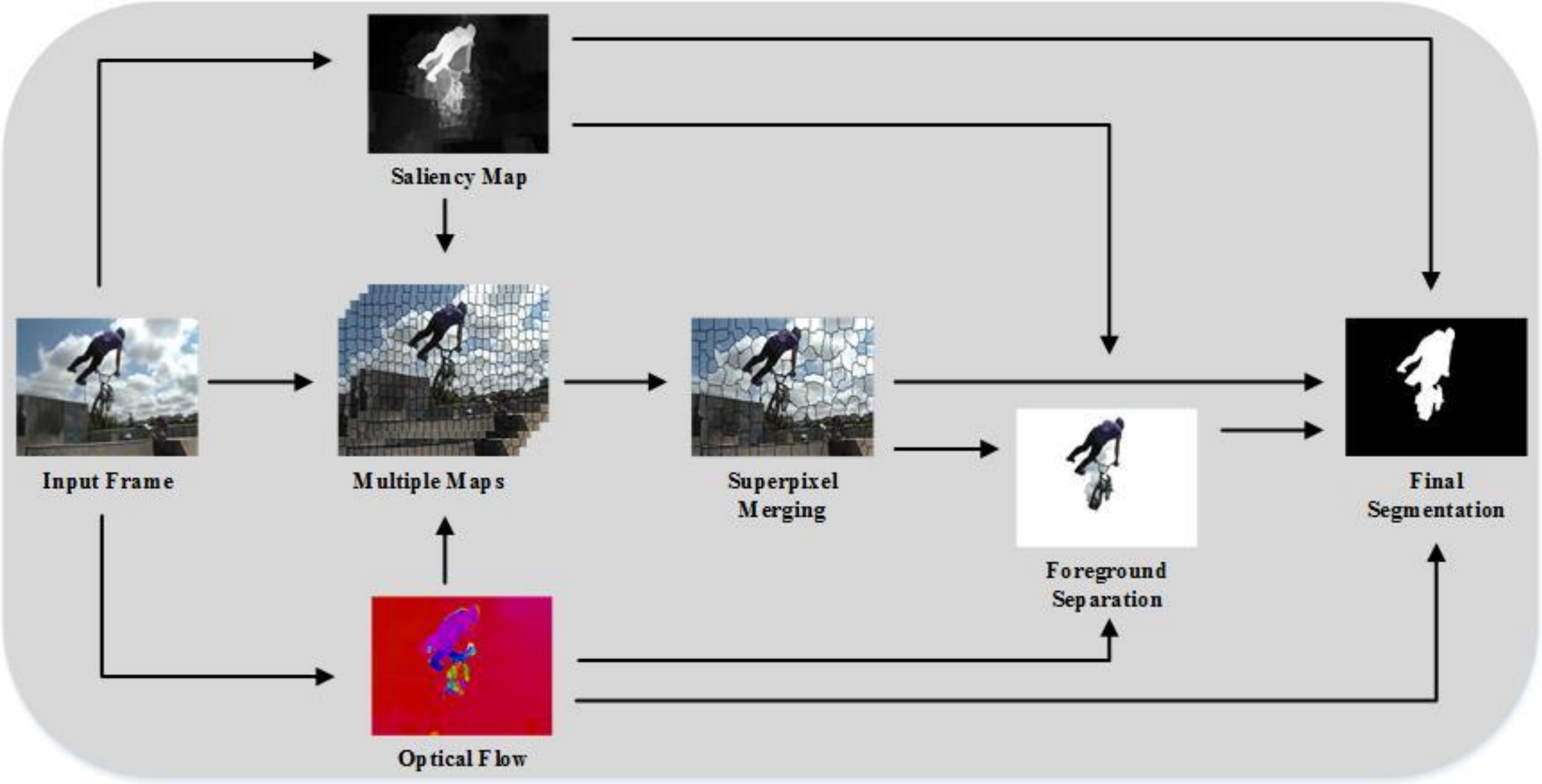
Video Object Segmentation

- Video object segmentation is the extraction of all objects from a given video sequence.

- A binary labelling problem where all the pixels belonging to an object are labelled as (1) while the background pixels are labelled as (0).

- Segmentation allows suppression of unnecessary/irrelevant information.

Proposed Framework



Feature Extraction

- We extracted multiple features for each frame of video sequence.

- Use of existing solutions to obtain:
  - CIELAB color representation.
  - CIELAB color representation of dense optical flow.
  - Spatio-temporal saliency score.

Superpixel Segmentation & Merging

- We extended existing segmentation technique to perform spatio-temporal superpixel segmentation of videos.

- 8-dimensional spatio-temporal feature vector used for superpixel segmentation:

$$p = (L, \alpha, \beta, x, y, u, v, s)$$

- Mapped to 16-dimensional feature vector:

$$l_1(i,j) = \cos\frac{\pi}{2} . l(i,j)$$
$$l_2(i,j) = \sin\frac{\pi}{2} . l(i,j)$$

- Spectral Clustering performed to get final superpixels.

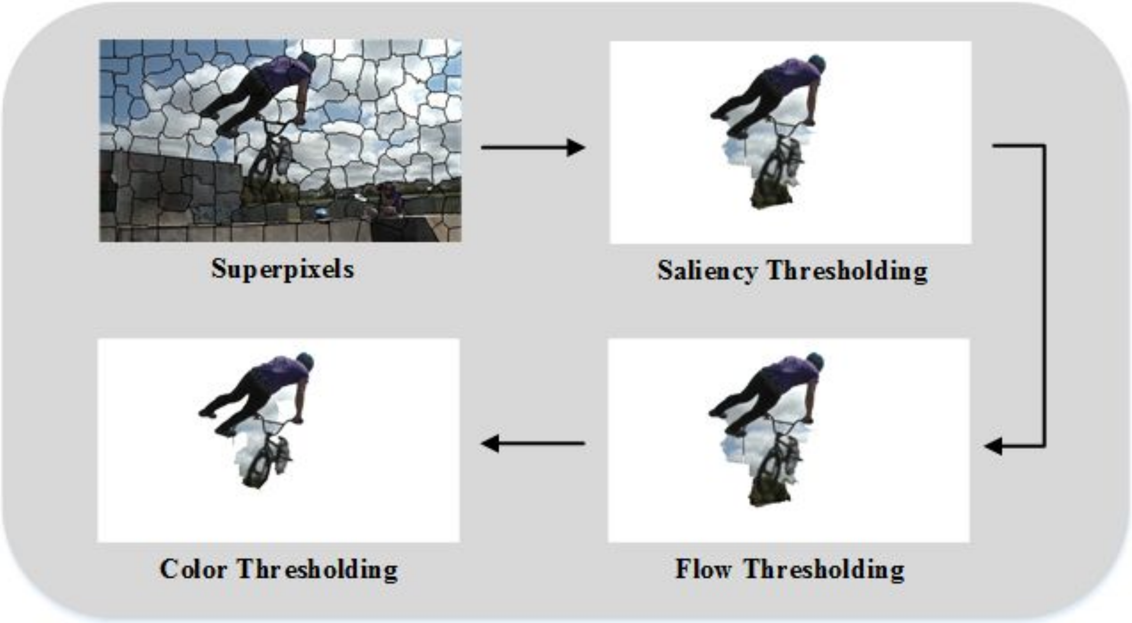
Superpixel Segmentation & Merging

- We obtained 5 different superpixel segmentations using different values for color, saliency and flow constants.

- Multiple segmentation merged into one using existing superpixel merging technique.

Foreground Separation Model

- Multiple thresholding steps performed on merged superpixel segmentation to get initial labelling for each frame.



Interframe Spectral Clustering

- Merged superpixel segmentation of each frame used.
- Distance between each feature (color, spatial, saliency and optical flow color) of every two segments calculated.
- Interframe graph constructed across all frames:

$$dist(i,j) = \sqrt{\lambda_1 C + \lambda_2 F + \lambda_3 P + \lambda_4 S + \lambda_5 L}$$

- Affinity matrix constructed using:

$$affinity(i,j) = \exp\left(\frac{-dist(i,j)}{2\sigma^2}\right)$$

- Spectral Clustering performed for final labelling.

Dataset

- We used SegTrack v2 dataset for experimentation.
- 14 video sequences with varying number of frames.
- Each video had different properties with respect to foreground, background and motion.
- For long videos, first 50 frames used for experimentation.

Experiments

- Evaluation measures:
  - Average per-frame pixel error rate
  - Precision, recall and F-measure.

- Comparison with state-of-the-art solutions:
  - Lee et al., 2011
  - Galasso et al., 2012
  - Zhang et al., 2013

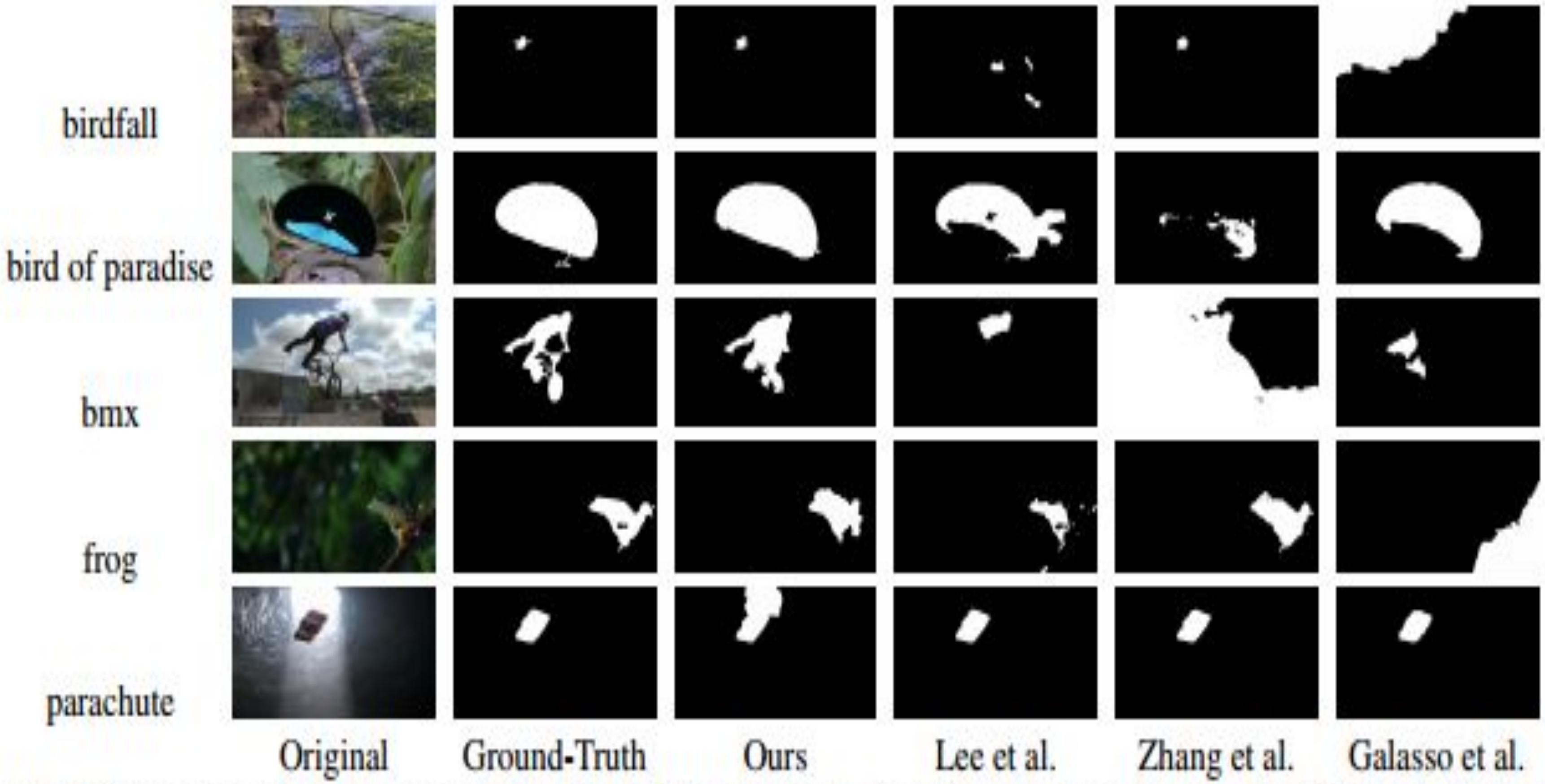
Results (Error Rate)

Sequence	Ours	(Lee et al.)	(Zhang et al.)	(Galasso et al.)
bird of paradise	**4845**	13041	35429	11832
birdfall	980	1091	**150**	18533
bmx	**10339**	12071	152527	17133
cheetah	2626	22977	**1914**	4532
drift	19614	**15503**	147480	42130
frog	2854	3387	**1767**	15216
girl	4802	2054	**1495**	4079
hummingbird	18237	18288	**13601**	21404
monkey	7230	**3205**	3430	18697
monkeydog	8065	9337	**2241**	3671
parachute	1104	**202**	219	382
penguin	36786	**24751**	30205	37022
soldier	6660	57586	24764	**5592**
worm	2839	**1470**	4147	17745
Mean	**9070**	13212	29963	15569

Results

	Ours	(Lee et al.)	(Zhang et al.)	(Galasso et al.)
Precision	0.612	**0.626**	0.617	0.430
Recall	0.597	0.571	**0.661**	0.306
F1-measure	**0.568**	0.521	0.535	0.284

Results


birdfall
bird of paradise
bmx
frog
parachute
Original
Ground-Truth
Ours
Lee et al.
Zhang et al.
Galasso et al.

Experiments

- We have proposed a Spectral Clustering based solution for video object segmentation.

- We extended existing segmentation technique to perform spatiotemporal superpixel segmentation of videos and introduced a foreground separation model to provide initial foreground labelling.

- Results show that our algorithm is comparable to other state-of-the-art solutions.