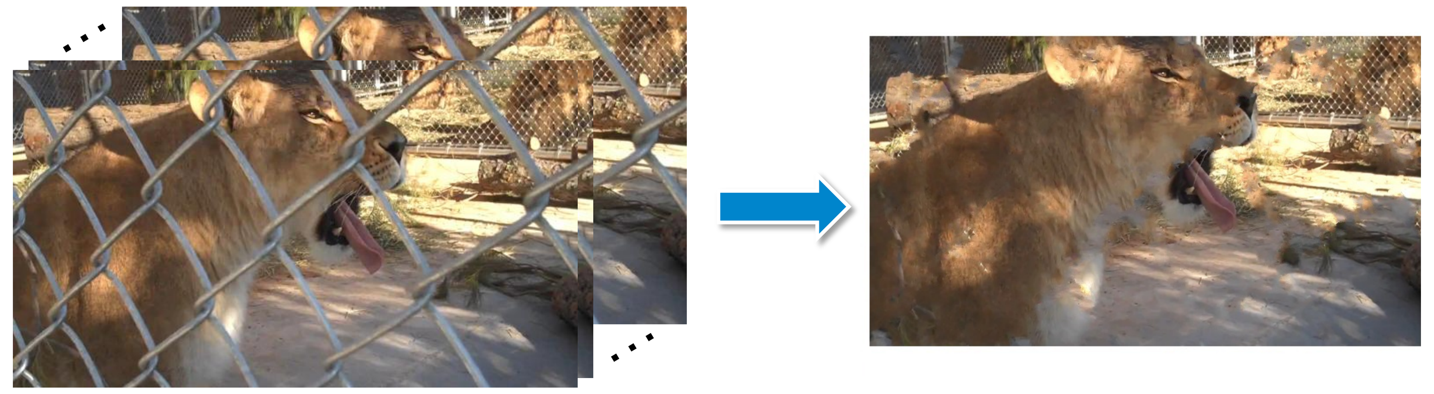
Automatic Fence Segmentation in Videos of Dynamic Scenes

Renjiao Yi[1,2], Jue Wang[3], Ping Tan[1],
[1]Simon Fraser University, [2]NUDT, [3] Adobe

IEEE 2016 Conference on Computer Vision and Pattern Recognition
CVPR 2016

Motivations



- Creating a fence-free image from a video clip

Challenges

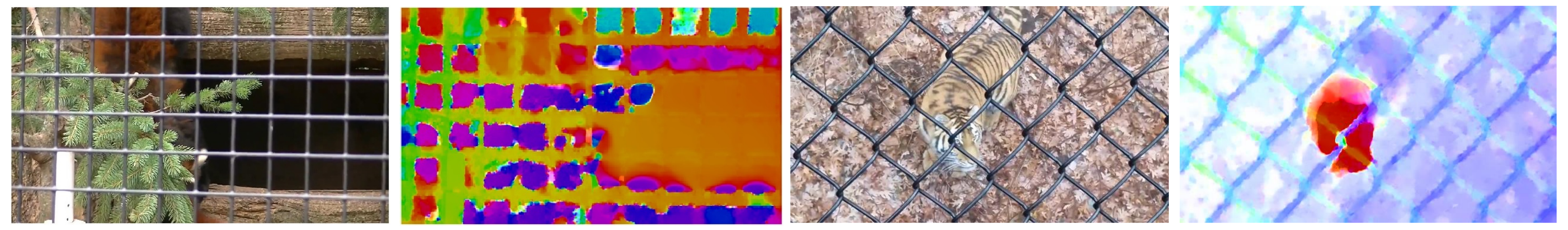
- Hard to detect & segment fence from the input video, due to

(moving backgrounds, distorted or irregular fence, thin and repetitive fence structures, changing illumination, etc)

- Recent works [1, 2] only deal with static scenes

Our observations

- Both background objects and the camera are often moving
- Motion contrast brings strong segmentation cues

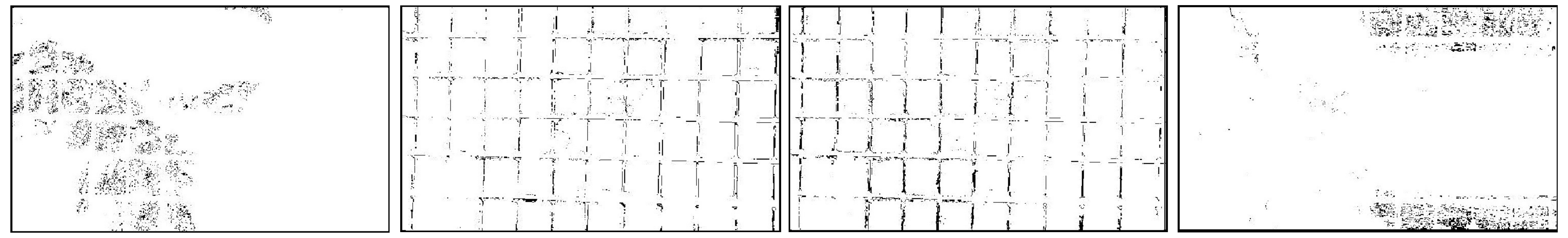


an input frame optical flow an input frame optical flow

Initial Segmentation

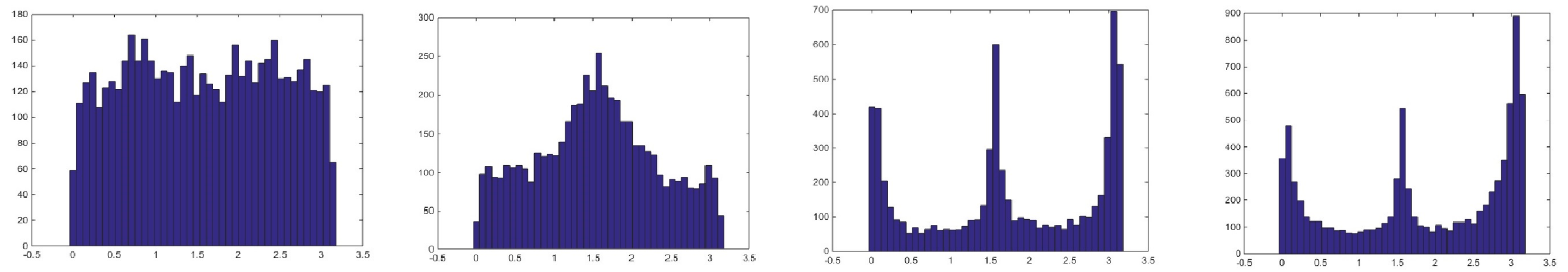
Selected frames from a input video

Group pixels according to color and motion



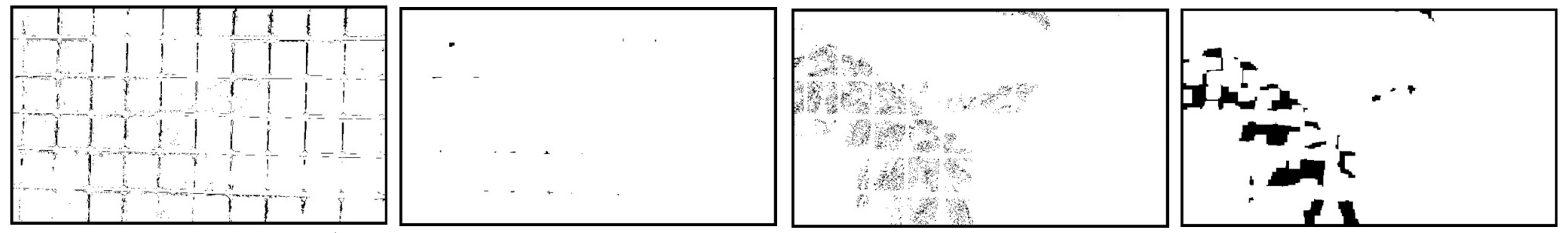
Fence groups have two distinctive features:

(a) bi-module gradient orientation histograms



non-fence groups fence groups

(b) thin structures (disappear after closed followed by erosion)



fence groups non-fence groups

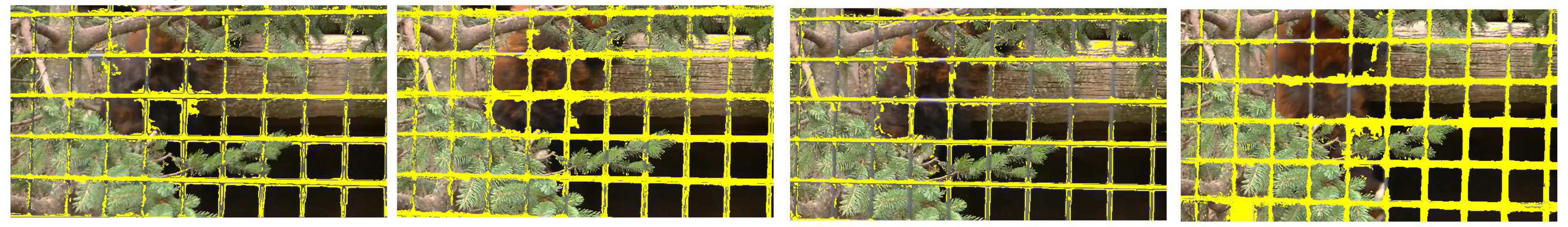
Graph-cut based segmentation (frame by frame):

- A fully-connected graph, each vertex is a pixel group.
- The data cost $D(c_i, l_i)=l_i(1-P(c_i))+(1-l_i)P(c_i)$

$P(c_i)$ is the probability of c_i being fence (according to the above two features).

- The smooth term $S(c_i, c_j, l_i, l_j)=\delta(l_i, l_j)(1-S_1(c_i, c_j))(1-S_2(c_i,c_j))(1-S_3(c_i,c_j))$

S_1, S_2, S_3 measures the similarity of c_i, c_j in color and gradient orientation distributions.



Segmentation Refinement

Dense CRF based segmentation (with all video frames):

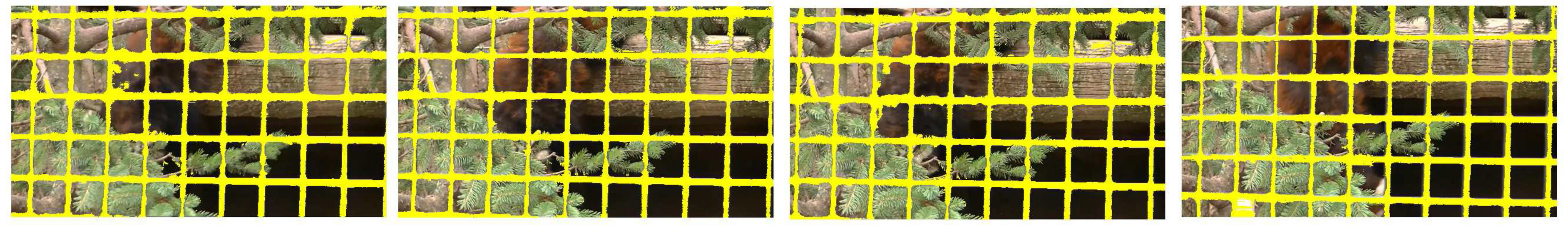
- A fully-connected graph, each vertex is a pixel.
- The data term $\mathbb{D}(x, l_x) = l_x \cdot (1-\mathbb{P}(x)) + (1-l_x)\cdot\mathbb{P}(x)$

$\mathbb{P}(x) = \mathbb{P}_1(x)\cdot\mathbb{P}_2(x)$ encodes the consistency with initialization, and the probability of a pixel being fence.

- The smooth term $\mathbb{S}(x, l_x; y, l_y)=\delta(l_x, l_y)k(x,y)$

$k(x,y)$ measures the similarity of x,y in both color and position

Segmentation results:



Results (fence removal results in the paper):

Reference

[1] T. Xue, M. Rubinstein, C. Liu, W. Freeman. A computational approach for obstruction-free photography. ACM Transactions on Graphics (TOG), 2015, 34(4): 79.

[2] Y. Mu, W. Liu, S. Yan, Video de-fencing. arXiv preprint, arXiv:1210.2388, 2012

[3] P. Krähenbühl, V. Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. NIPS 2011.

gruvi {graphics} ∪ {vision}