



Disney Research

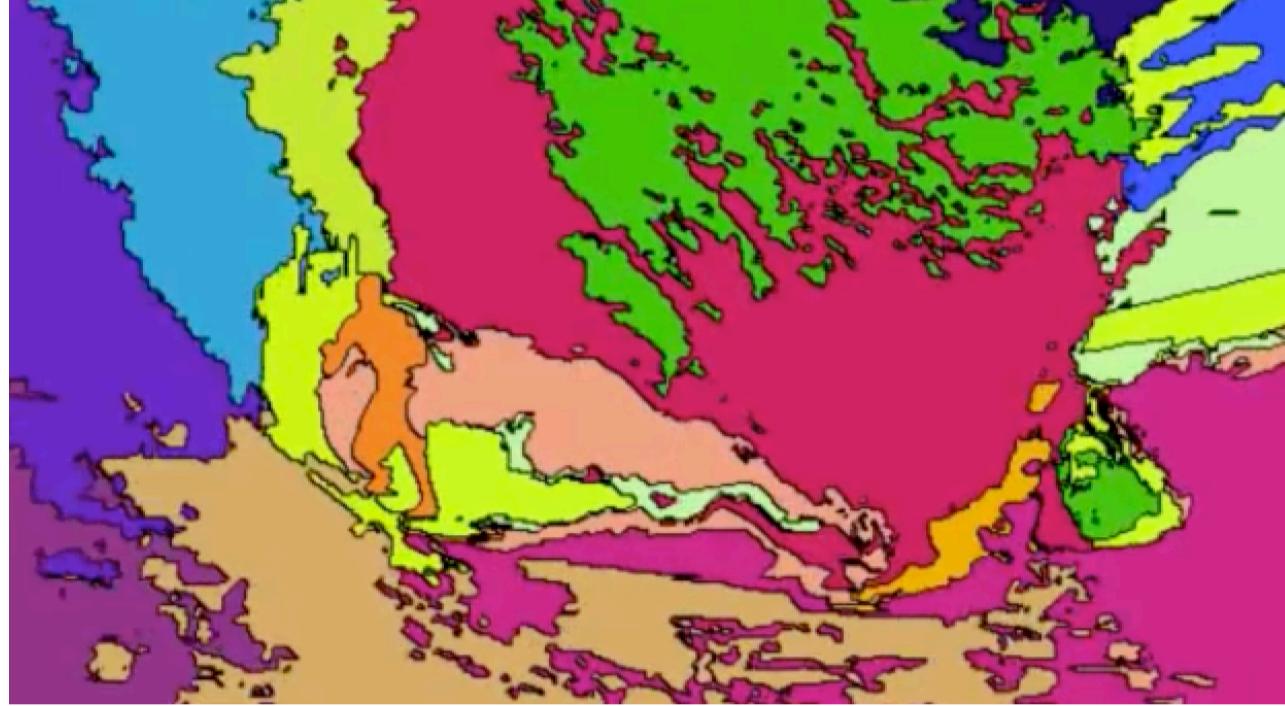
ETH zürich

Fully Connected Object Proposals for Video Segmentation

Federico Perazzi^{1,2}, Oliver Wang¹, Markus Gross^{1,2}, Alexander Sorkine-Hornung¹¹Disney Research Zurich²ETH Zurich

INTRODUCTION

Video Object Segmentation separates foreground object from background

Video Object SegmentationVideo Over-segmentation

EXISTING METHODS

Video Segmentation approaches based on object proposals have demonstrated promising results.

Algorithm

- Seek the best proposal per-frame
- Refine the segmentation on a locally connected graph .

Limitations

- Strongly rely on the quality of the generated proposals
- Suffer challenging scenarios such as fast motion and occlusions

OUR APPROACH



Key Idea

- Inference on a fully connected graph built over object proposals.
- Segmentation as grouping of multiple potentially imperfect object proposals

Contributions

- SVM classification and resampling to retain proposals with higher discriminative power
- Novel energy function combines appearance with long-range point tracks to ensure robustness with respect to fast motion and occlusions.



Project website

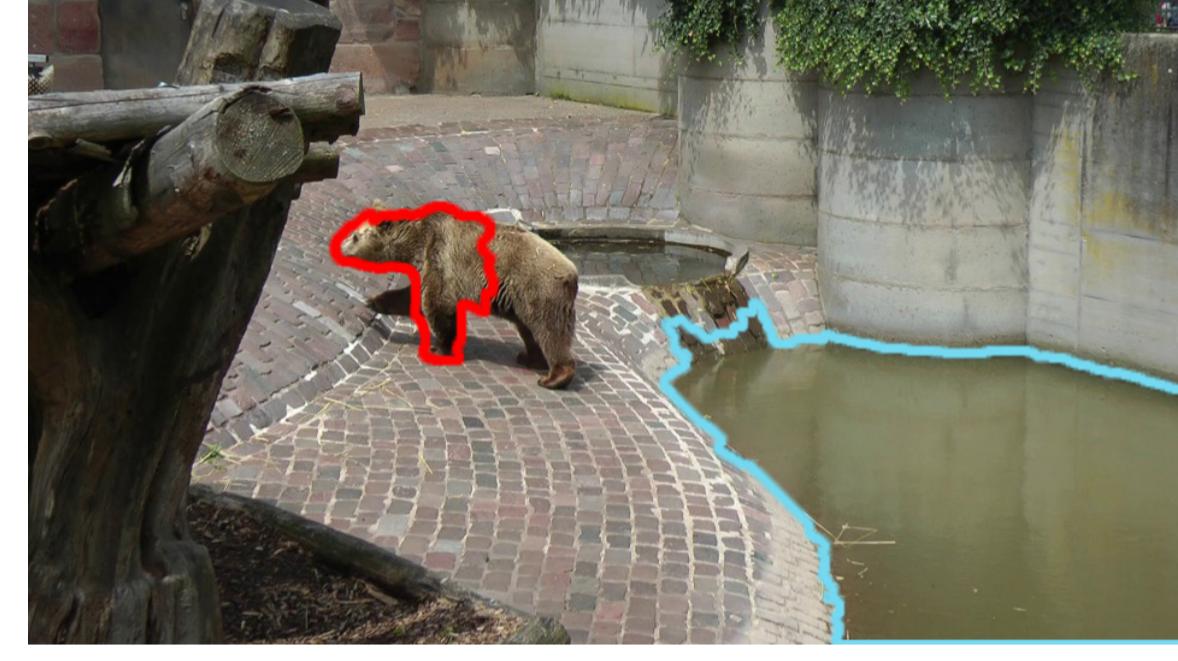
<https://graphics.ethz.ch/~perazzif/fcop>

ALGORITHM OVERVIEW

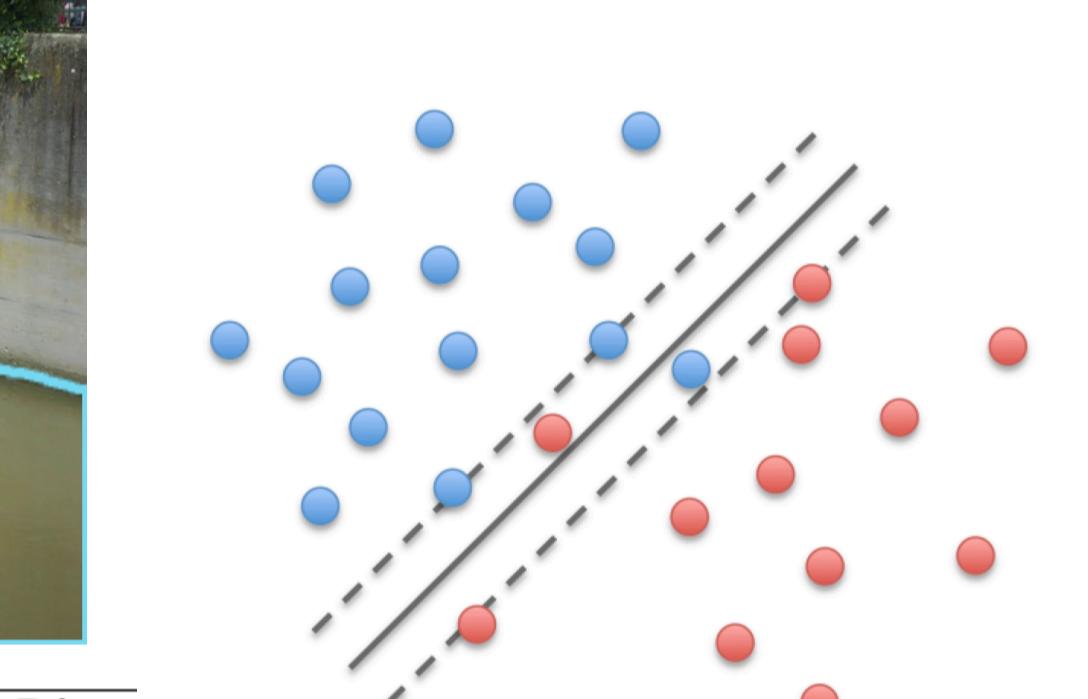


PROPOSAL GENERATION, FEATURES EXTRACTION AND RESAMPLING

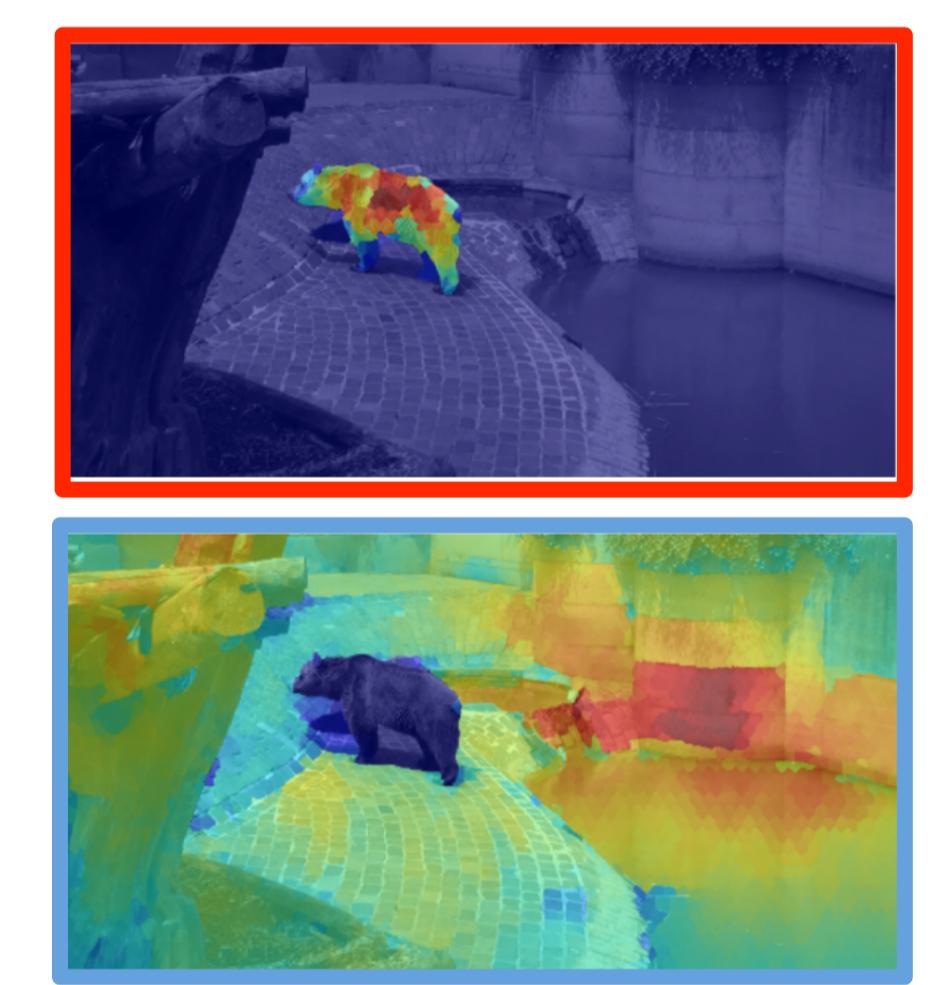
Proposals Generation and Features Extraction



Per-frame, SVM Classification



Object Proposals Pruning and Resampling



$$\mathcal{C}(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b$$

FULLY CONNECTED PROPOSAL LABELING

- We determine the fore- and background classification by solving for the maximum a posteriori of a fully connected conditional random field (CRF)
- Conditional Random Fields provide a natural framework to incorporate all mutual spatiotemporal relationships between proposals as well as our initial proposal confidences.

$$E(Y|\mathcal{X}, \mathcal{F}) = \sum_{i \in \mathcal{V}} \psi_u(y_i; \mathcal{X}) + \sum_{i,j \in \mathcal{E}} \psi_p(y_i, y_j; \mathcal{F})$$

$$e^{-\psi_u(y_i, \mathcal{X})} = \begin{cases} l_i + \hat{\epsilon}, & l_i \in \mathcal{L} \\ P(y_i | \mathbf{x}_i) & s_i \notin \tilde{\mathcal{S}} \end{cases}$$

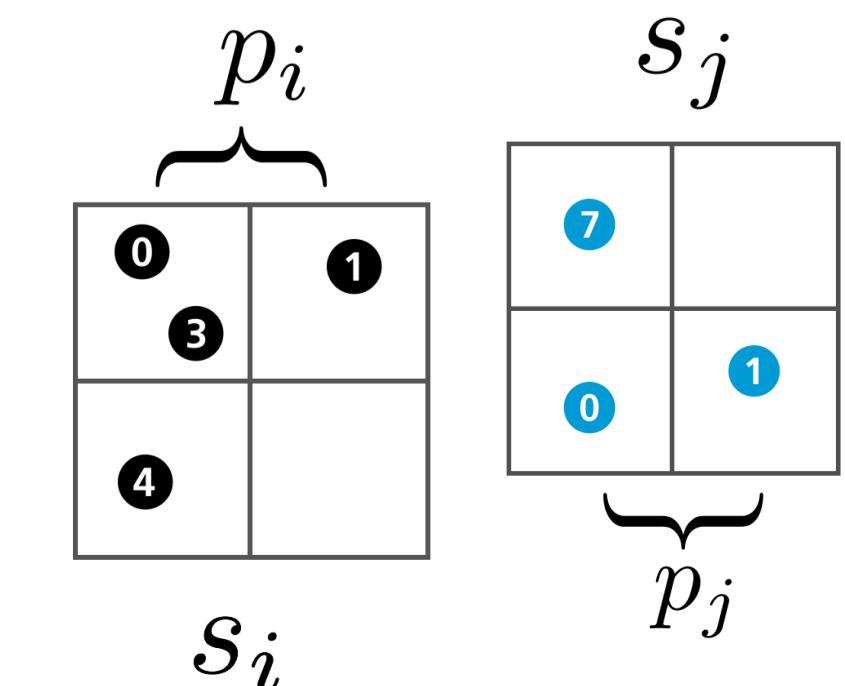
$$\psi_p(y_i, y_j; \mathcal{F}) = [y_i \neq y_j] \cdot (\underbrace{\omega_c k_c(\mathcal{D}_c(c_i, c_j))}_{\text{appearance kernel}} + \underbrace{\omega_s k_s(\mathcal{D}_s(s_i, s_j))}_{\text{spatial kernel}} + \underbrace{\omega_t k_t(\mathcal{D}_t(p_i, p_j))}_{\text{trajectory kernel}} + \underbrace{\omega_k k_k(|t_i - t_j|)}_{\text{temporal kernel}})$$

Appearance Kernel: $\chi^2(c_i, c_j)$

Temporal Kernel: $|t_i - t_j|$

Spatial Kernel: $\mathcal{D}_s(s_i, s_j) = 1 - \frac{|s_i \cap s_j|}{|s_i \cup s_j|}$

Trajectory Kernel: $\mathcal{D}_t(p_i, p_j) = 1 - \frac{|p_i \cup p_j|}{|s_i \cup s_j|}$



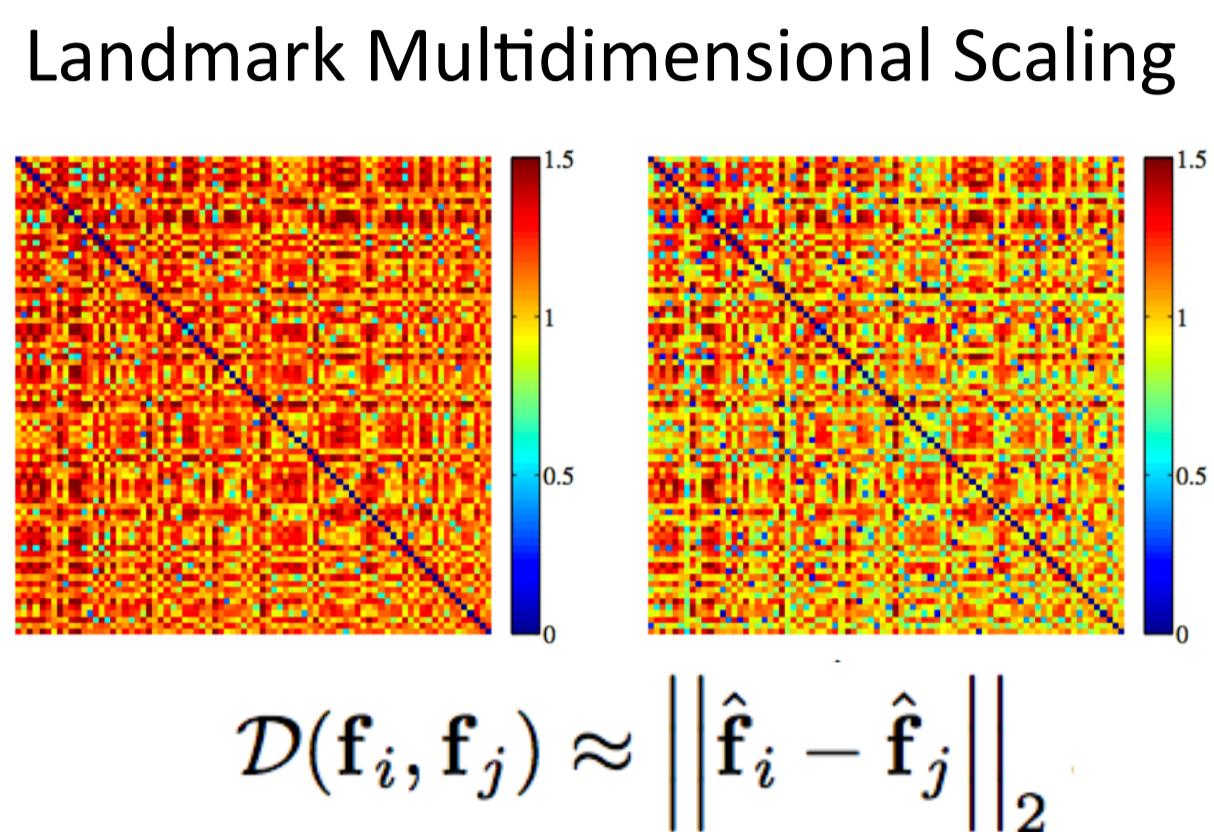
EUCLIDEAN EMBEDDING

Pairwise potentials: linear combination of Gaussian Kernels

$$\psi_p(y_i, y_j, \mathcal{F}) = \mu(y_i, y_j) \sum_{m=1}^K w_m k_m(\mathbf{f}_i, \mathbf{f}_j)$$

$$k_m(\mathbf{f}_i, \mathbf{f}_j) = \exp\left(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)^T \Lambda_m(\mathbf{f}_i - \mathbf{f}_j)\right)$$

Requires Embedding of features in Euclidean Space



GROUPING INTO SEGMENTATION

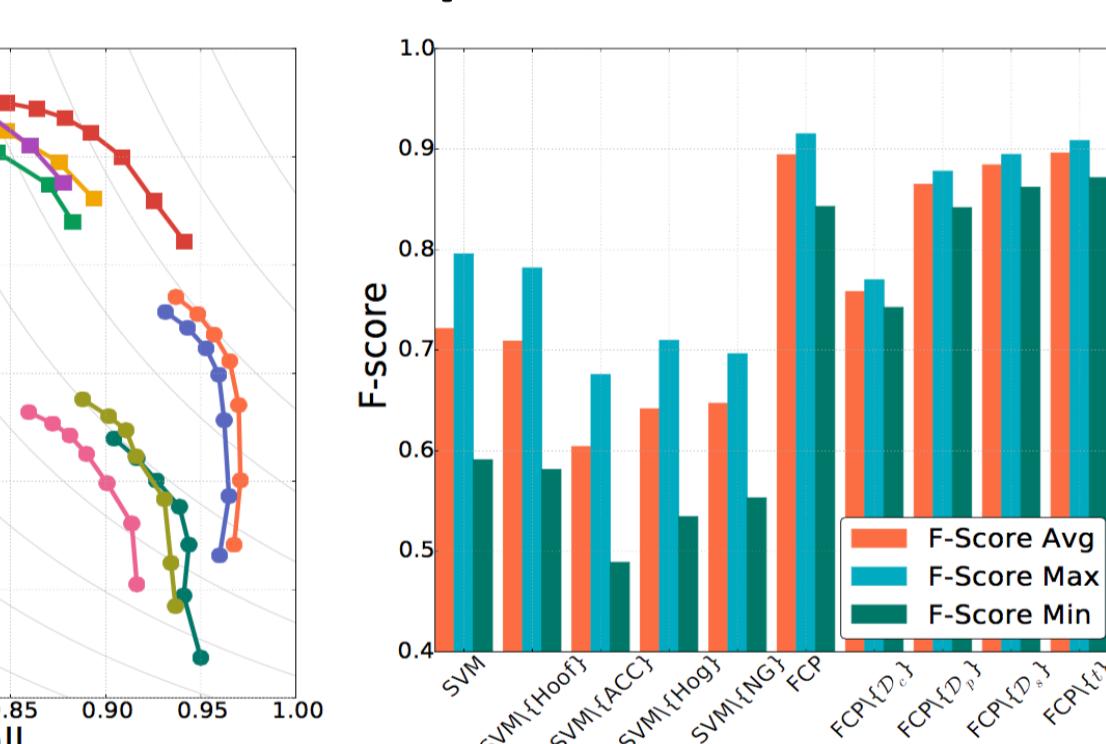


QUANTITATIVE EVALUATION

FBMS – Intersection-over-Union

Sequence	FCP	CRF	SVM	SEA	FST	DAG	NLC
cars1	0.69	0.80	0.68	(0.83)	0.82	0.10	0.27
cats01	(0.83)	0.68	0.76	0.62	0.80	0.34	0.71
cats03	0.39	0.00	0.11	0.17	(0.53)	0.32	0.12
dogs01	0.55	0.22	0.39	0.38	0.53	(0.56)	0.54
goats01	0.54	(0.84)	0.47	0.34	0.44	0.38	0.58
horses05	(0.77)	0.47	0.69	0.34	0.44	0.38	0.58
lions01	(0.84)	0.74	0.80	0.73	0.80	0.77	0.67
marple2	0.59	0.57	0.71	(0.78)	0.65	0.56	0.60
marple4	(0.88)	0.73	0.87	0.69	0.15	0.45	0.19
marple6	0.77	0.64	0.77	(0.86)	0.24	0.18	0.48
people1	0.68	0.64	0.22	0.58	0.54	0.69	(0.85)
people2	0.81	0.78	0.76	0.77	(0.92)	0.48	0.77
rabbit02	0.65	0.11	0.33	0.27	0.21	0.21	(0.71)
rabbit03	0.43	0.40	0.28	0.42	0.41	0.22	(0.44)
rabbit04	0.29	0.00	0.12	0.23	(0.38)	0.12	0.20
tennis	0.48	0.27	0.41	0.55	0.30	0.51	(0.64)
Avg. Test	(0.65)	0.51	0.53	0.58	0.56	0.43	0.51
Avg. Training	(0.77)	0.62	0.61	0.71	0.68	0.60	0.56

Features Importance



RESULTS

