Fully Connected Object Proposals for Video Segmentation

Federico Perazzi[1,2], Oliver Wang[1], Markus Gross[1,2], Alexander Sorkine-Hornung[1]

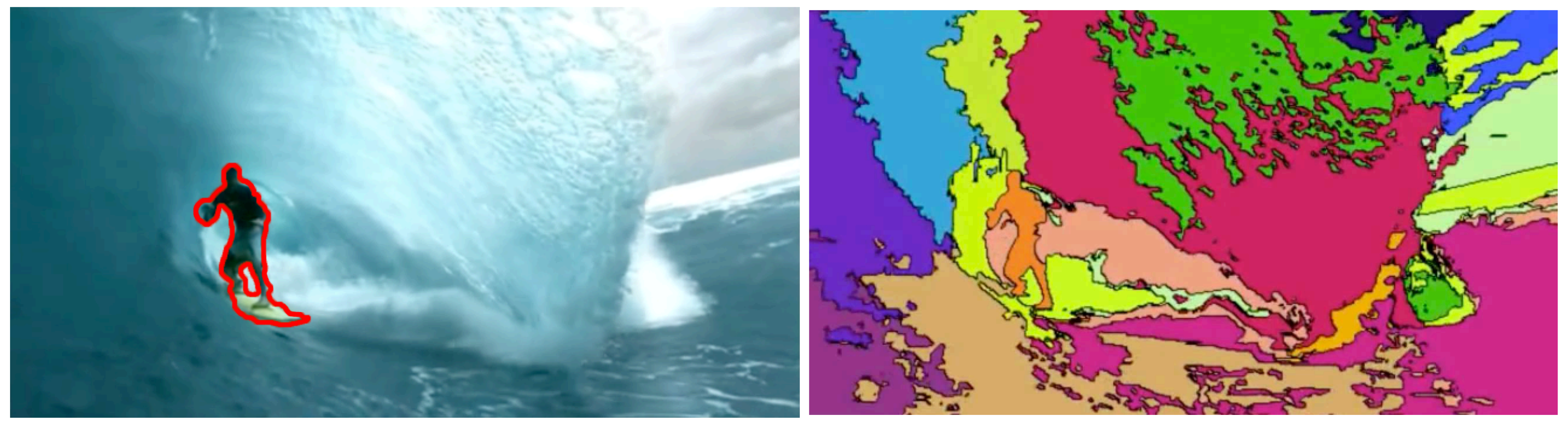
[1]Disney Research Zurich [2]ETH Zurich

Disney Research ETH zürich

SANTIAGO, CHILE DECEMBER 11-18, 2015 ICCV15 International Conference on Computer Vision

INTRODUCTION

Video Object Segmentation separates foreground object from background



Video **Object** Segmentation Video **Over**-segmentation

EXISTING METHODS

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

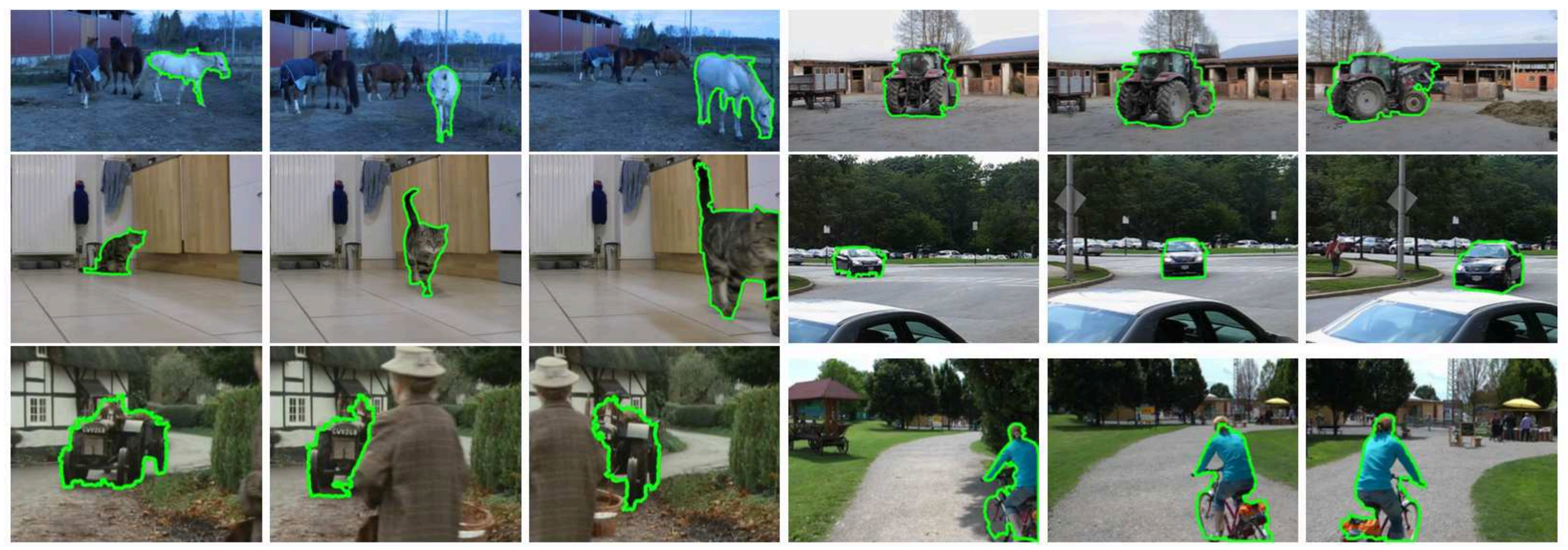
Algorithm

- 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

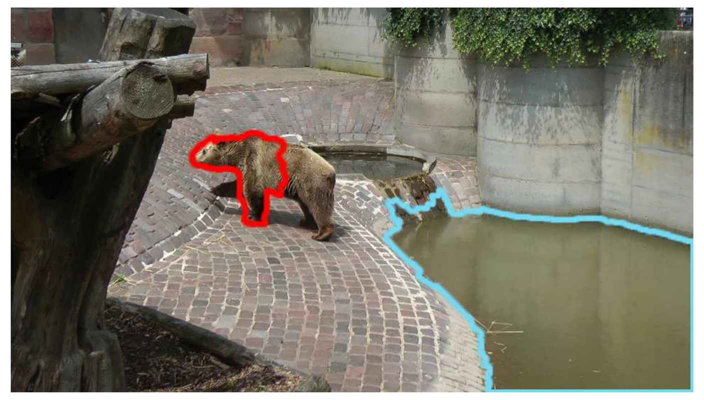
https://graphics.ethz.ch/~perazzif/fcop

ALGORITHM OVERVIEW

1. Object Proposals Generation and Features Extraction
2. SVM Classification and Resampling
3. Fully Connected Proposal Labeling
4. Grouping of multiple proposals into segmentation

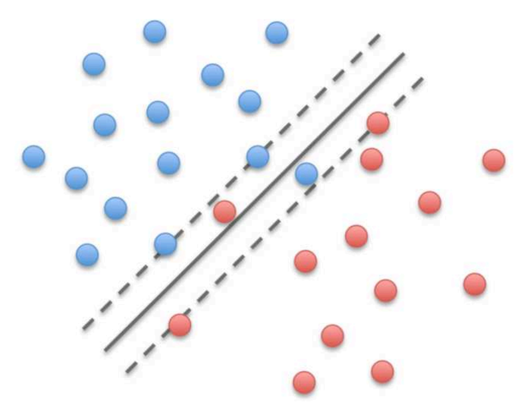
PROPOSAL GENERATION, FEATURES EXTRACTION AND RESAMPLING

Proposals Generation and Features Extraction



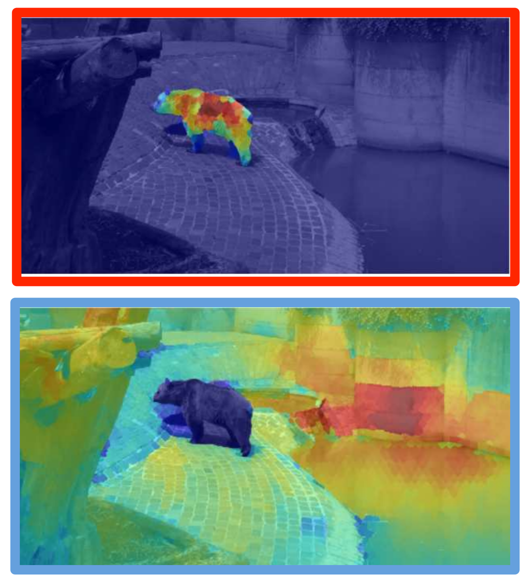
Feature	Description	Dim
(ACC)	Area, centroid, average color	6
(HOOF)	Histogram of Oriented Optical Flow	32
(NG)	Objectness via normalized gradients	64
(HOG)	Histogram of oriented gradients	1296

Per-frame, SVM Classification



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

Object Proposals Pruning and Resampling



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.

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

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

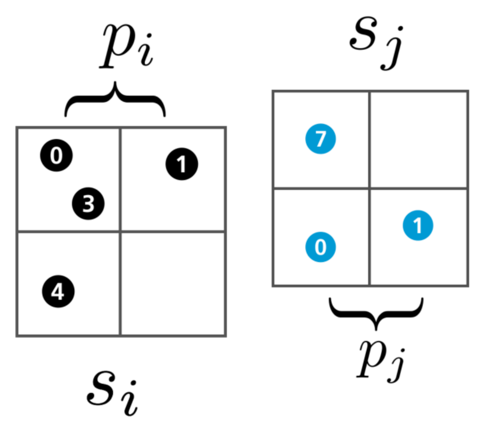
$$\psi_p(y_i,y_j;\mathcal{F}) = [y_i\neq y_j]\cdot(\underbrace{\omega_c k_c(\mathcal{D}_c(c_i,c_j))}_{\text{appearance kernel}} + \underbrace{\omega_s k_s(\mathcal{D}_s(s_i,s_j))}_{\text{spatial kernel}} + \underbrace{\omega_p k_p(\mathcal{D}_p(p_i,p_j))}_{\text{trajectory kernel}} + \underbrace{\omega_t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

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



EUCLIDEAN EMBEDDING

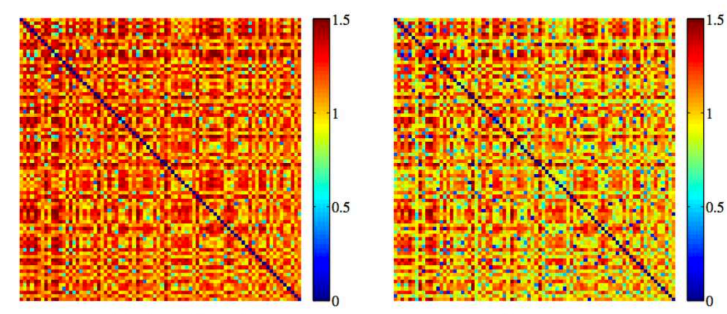
Pairwise potentials: linear combination of Gaussian Kernels

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

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

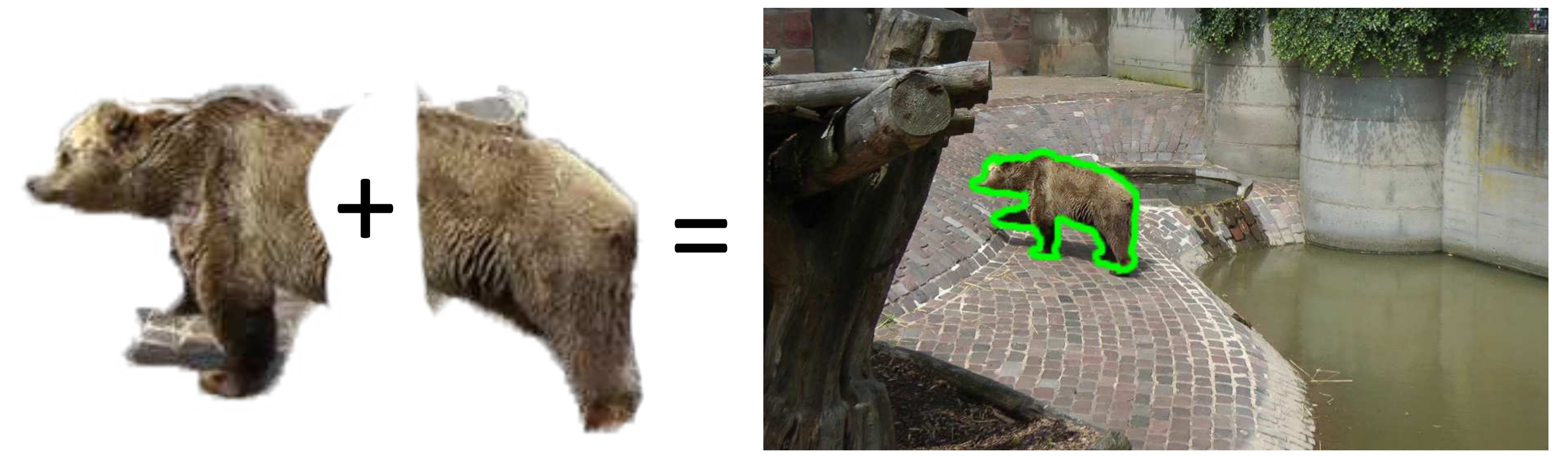
Requires Embedding of features in Euclidean Space

Landmark Multidimensional Scaling



$$\mathcal{D}(\mathbf{f}_i,\mathbf{f}_j) \approx \left|\left|\hat{\mathbf{f}}_i - \hat{\mathbf{f}}_j\right|\right|_2$$

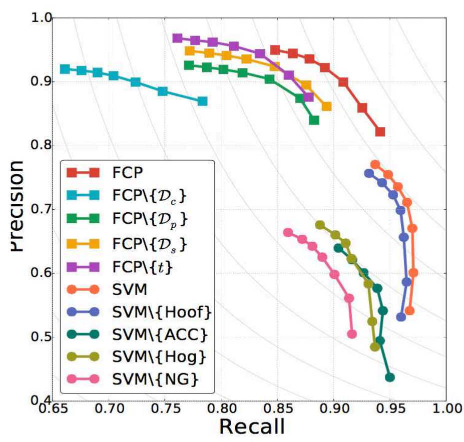
GROUPING INTO SEGMENTATION



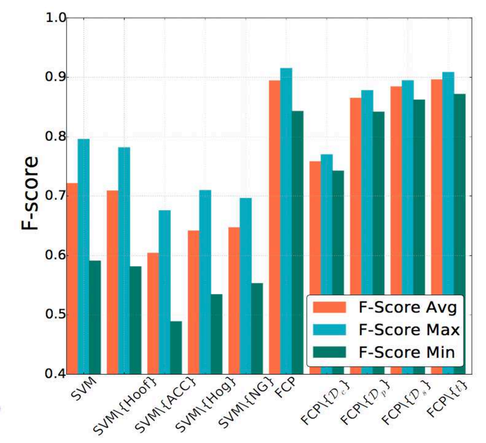
QUANTITATIVE EVALUATION

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	0.82	**(0.84)**	0.78	0.53	**(0.84)**	0.79	0.58
horses05	**(0.77)**	0.66	0.47	0.69	0.34	0.44	0.38
lion01	**(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
rabbits02	0.66	0.11	0.33	0.42	0.65	0.32	**(0.71)**
rabbits03	0.43	0.40	0.23	0.42	0.41	0.22	**(0.44)**
rabbits04	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