Structured-Cut: A Max-margin Feature Selection Framework for Video Segmentation

Nikhil Santosh Naikal

CS 294-69 Project Presentation.
December 5, 2011.

Motivation

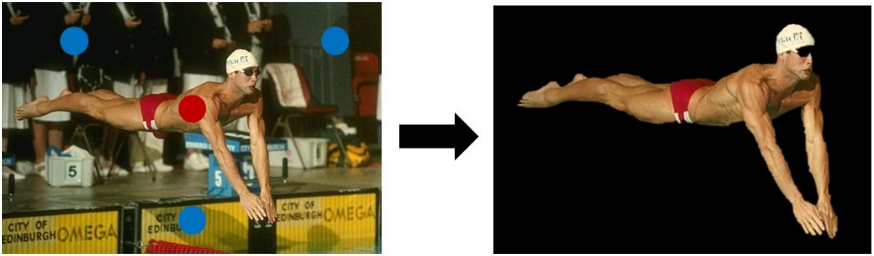
- Interactive video segmentation very useful for variety of applications.
  - Examples: Rotoscoping, Compositing, etc.
- State-of-the-art methods combine local and global classifiers [1] .

Goal:

Use multiple image features to improve foreground/background separation.

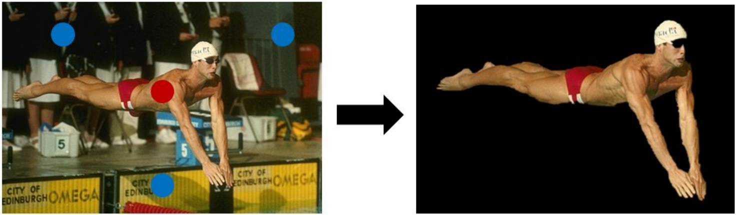
[1] Xue Bai, Jue Wang, David Simons, Guillermo Sapiro, Video SnapCut: Robust Video Object Cutout Using Localized Classifiers. ACM Transaction on Graphics (Proc. SIGGRAPH 2009).

Interactive Image Segmentation



Given marked Foreground/Background regions,

Assign binary label to all other pixels.



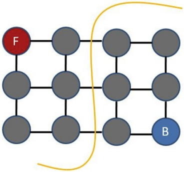
Markov Random Fields (MRF) formulation:

$$\mathbf{E}(\mathbf{s}) = \sum_{i} \Psi_u(y_i) + \lambda \sum_{i,j \in N} \Psi_p(y_i, y_j).$$

- $\mathbf{s}$: Image segmentation.
- $\mathbf{y}$: Image pixel nodes.
- Ψ_u: Unary potential.
- Ψ_p: Pairwise potential.

Can be minimized using Graph Cuts [1]

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} \mathbf{E}(\mathbf{s})$$



[1] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimisation via graph cuts", Tans. on PAMI, 2001.

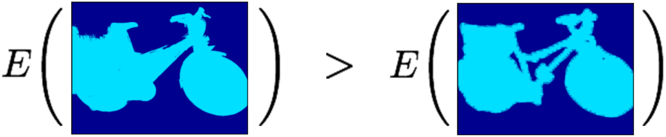
Using Multiple Features

- MRF framework general enough to incorporate multiple features (color, texture, shape, etc.)

$$\mathbf{E}(\mathbf{s}) = \sum_i \boldsymbol{\Psi}_u + \sum_{i,j \in N} \boldsymbol{\Psi}_p = \mathbf{w}^{\mathbf{T}} \boldsymbol{\Theta}(\mathbf{s})$$

- Composite unary potential $\boldsymbol{\Psi}_u = \lambda^{color} {\Psi_u}^{color} + \lambda^{texture} {\Psi_u}^{texture} + \ldots,$
- Composite pairwise potential $\boldsymbol{\Psi}_p = \mu^{color} {\Psi_p}^{color} + \mu^{texture} {\Psi_p}^{texture} + \ldots,$
- Feature weights $\mathbf{w} = [\lambda^{color} \lambda^{texture} \ldots \mu^{color} \mu^{texture} \ldots]^T.$

Constraints on Energy

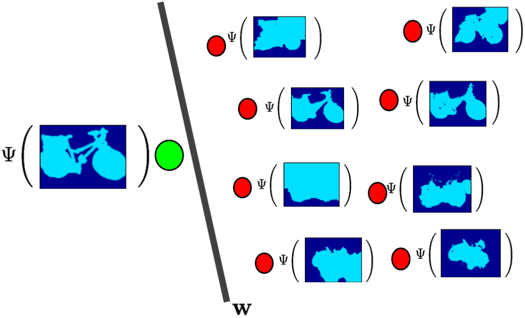


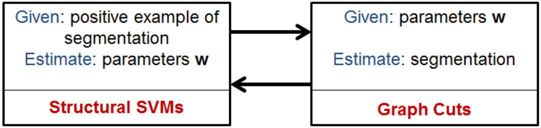
Incorrect segmentation   Ground truth segmentation

$$\mathbf{E}(\mathbf{s}_{incorrect}) > \mathbf{E}(\mathbf{s}_{gt}) \Rightarrow \mathbf{w}^{\mathsf{T}}\Theta(\mathbf{s}_{incorrect}) > \mathbf{w}^{\mathsf{T}}\Theta(\mathbf{s}_{gt}).$$

Exponentially large number of incorrect segmentations possible for an image.

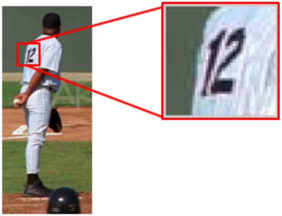
Learning Weights

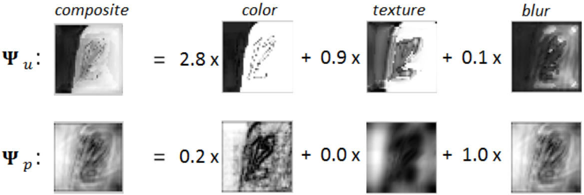




I. Tsochantaridis, *et. al.*, "Large Margin Methods for Structured and Interdependent Output Variables", JMLR, 2005.

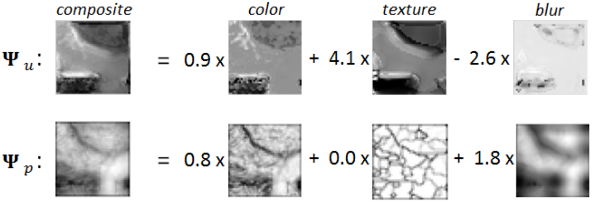
Example 1





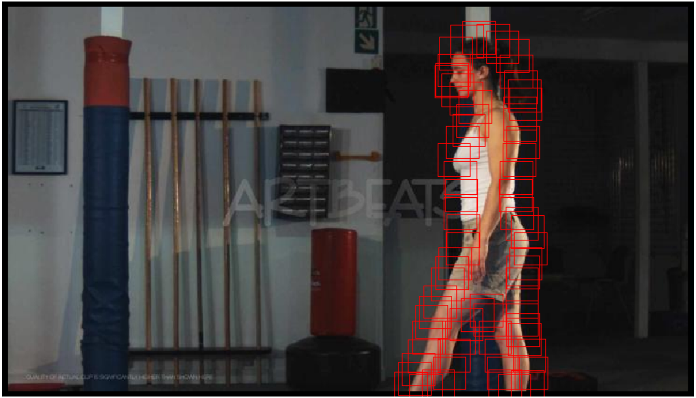
Minimizing composite energy using graph cuts

Example 2



Minimizing composite energy using graph cuts

Extension to Video



- Foreground/background models don't change drastically over subsequent frames.
- Local feature weights propagated to future frames.

Preliminary Results

Conclusion

- Color not always good discriminant for foreground/background segmentation.
- Incorporating multiple features (color, texture, blur, etc.) improves segmentation.
- Presented approach to weight multiple features using max-margin formulation.
- Demonstrated effectiveness of approach for image segmentation.
- Extending approach to video segmentation showed promise.

Questions?