

# Combinatorial Optimization in Computer Vision (IN2245)

Frank R. Schmidt  
Csaba Domokos

Winter Semester 2015/2016

21. Fast Trust Region for Object Tracking . . . . .	2
<b>Image Segmentation</b>	<b>3</b>
Submodular Image Segmentation . . . . .	4
Bayes Interpretation . . . . .	5
Hierarchical MRF . . . . .	6
<b>Regional Functionals</b>	<b>7</b>
Volume Constraint . . . . .	8
Holistic Histogram . . . . .	9
Holistic Distribution . . . . .	10
General Formulation . . . . .	11
Energy Approximation via Linearization . . . . .	12
Global Optimization of the Volume Constraint . . . . .	13
<b>Fast Trust Region</b>	<b>14</b>
Trust Region . . . . .	15
Regional Shape Distance . . . . .	16
Regional Shape Distance . . . . .	17

Volume Constraint . . . . .	18
Holistic Distribution Segmentation . . . . .	19
<b>Video Label Propagation</b>	<b>20</b>
Label Propagation . . . . .	21
Label Propagation . . . . .	22
Literature . . . . .	23

## 21. Fast Trust Region for Object Tracking

2 / 23

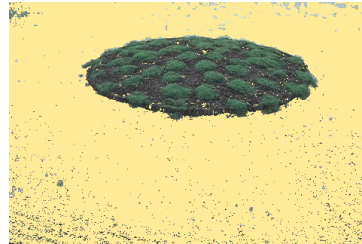
### Image Segmentation

3 / 23

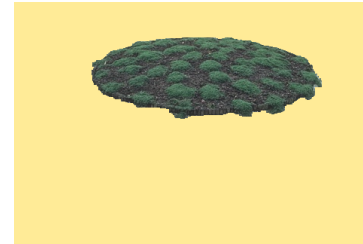
#### Submodular Image Segmentation



Given Image



Data Term



Data + Length Term

$$\begin{aligned}\operatorname{argmin}_{x \in \mathbb{B}^n} E(x) &= \operatorname{argmin}_{x \in \mathbb{B}^n} \sum_{i=1}^n f_i x_i + \sum_{i=1}^n \sum_{j \in \mathcal{N}(i)} f_{ij} x_i \bar{x}_j && f_{ij} \geq 0 \\ &= \operatorname{argmin}_{x \in \mathbb{B}^n} \sum_{i=1}^n f_i x_i + \operatorname{length}(x) \\ &= \operatorname{argmin}_{x \in \mathbb{B}^n} \langle f, x \rangle + \operatorname{length}(x)\end{aligned}$$

This can be efficiently minimized via graph cut.

## Bayes Interpretation

The above energy can be formulated by means of the Bayes' theorem.

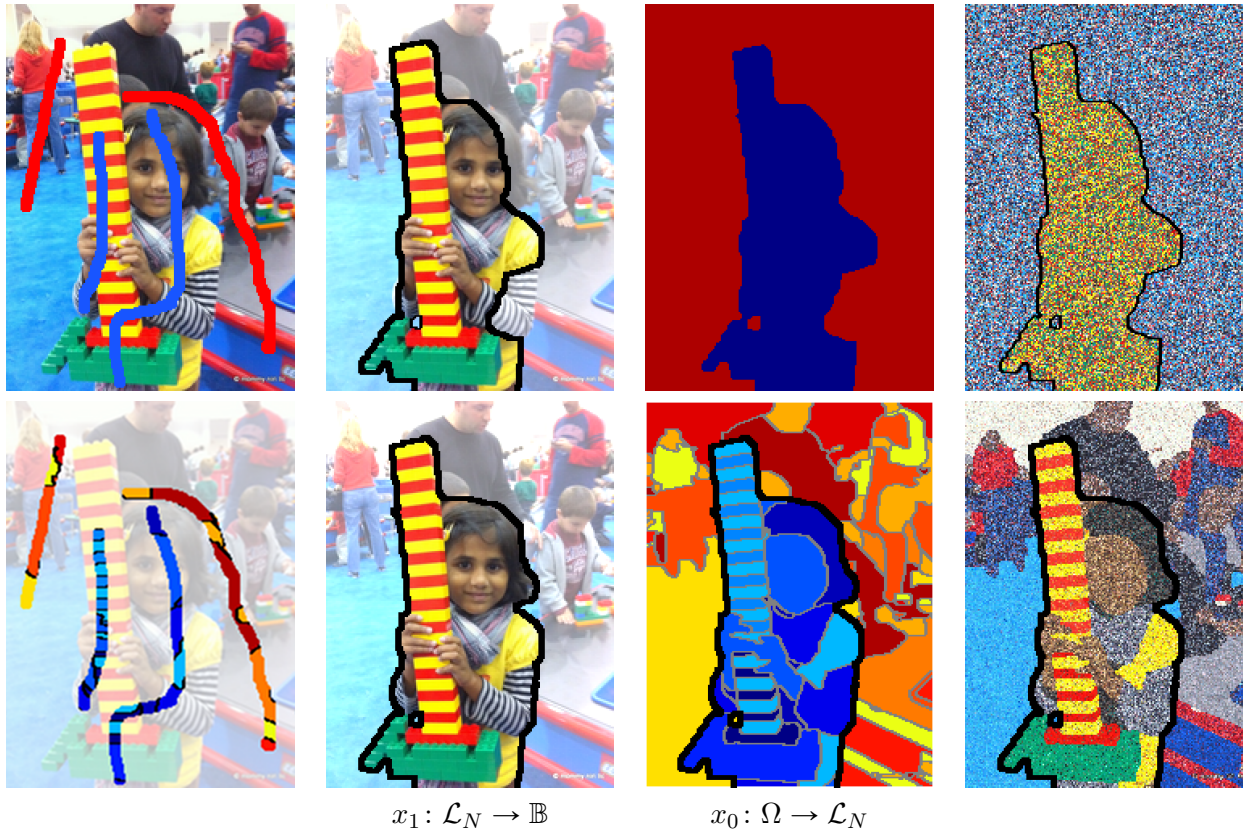
$$\max P(x|I) = \max \frac{P(I|x) \cdot P(x)}{P(I)}$$
$$E(x) = -\log(P(x|I)) = -\log(P(I|x)) - \log(P(x)) + \text{const}$$

Using

$$P(I|x) = \prod_{i \in \Omega} P_{\text{data}}(I(i)|x(i))$$
$$P_{\text{data}}(I(i)|0) = \text{pdf}^{(0)}(I(i)) = e^{-f^{(0)}(I(i))}$$
$$P_{\text{data}}(I(i)|1) = \text{pdf}^{(1)}(I(i)) = e^{-f^{(1)}(I(i))}$$
$$P(x) = \exp(-\text{length}(x))$$

we obtain  $\text{argmin}_x E(x) = \text{argmin}_x \langle f, x \rangle + \text{length}(x)$ .

## Hierarchical MRF



**Volume Constraint**

A **volume constraint** is very useful if we know the likely size of the observed image. A popular approach is to use the so called **ballooning term**.

The idea is to lower the data term in order to increase the size of the segmentation. The ballooning term is easy to implement

$$f(i) \rightsquigarrow f(i) - \lambda$$

and the resulting energy can be solved globally optimal via graph cut.

Whether this approach is robust depends highly on the used data term.

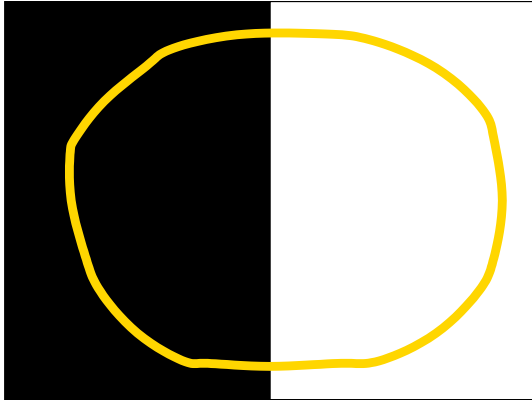
What we actually want to minimise is

$$\begin{aligned} E(x) &= E_0(x) + \lambda \cdot (\text{vol}(x) - V_0)^2 \\ &= E_0(x) + \lambda \cdot (\langle 1, x \rangle - V_0)^2 \end{aligned}$$

**The resulting energy is not submodular anymore.**

## Holistic Histogram

We can extend the volume constraints to multiple constraints that are applied to different colors or other appearances.



**Black:**  $V_0$  pixels.

**White:**  $V_1$  pixels.

The resulting energy is:

$$E(x) = (\langle f_0, x \rangle - V_0)^2 + (\langle f_1, x \rangle - V_1)^2 + \text{length}(x)$$

with indicator functions  $f_0, f_1: \Omega \rightarrow \{0, 1\}$  for black and white pixels.

## Holistic Distribution

Given are:

- $k$  appearance models and indicator functions  $f_i$  for  $i < k$ .
- Preferred distribution  $p_i$  for the models, i.e.,  $\sum_{i < k} p_i = 1$ .
- Histogram distance function, e.g., Bhattacharyya distance:

$$D(\{p_i\}, \{q_i\}) = -\log \left( \sum_{i < k} \sqrt{p_i q_i} \right)$$

Minimise the following energy functional:

$$E(x) = \text{length}(x) - \log \left( \sum_{i < k} \sqrt{p_i \cdot \frac{\langle f_i, x \rangle}{\langle 1, x \rangle}} \right)$$



## General Formulation

All these energies are instances of a special class of **higher-order pseudo-Boolean energies**

$$E(x) = E_0(x) + R_{\{f_i\}_{i < k}}^F(x)$$

- $E_0$  is a submodular function
- $R_{\{f_i\}_{i < k}}^F$  is a **regional function**, i.e.,

$$R_{\{f_i\}_{i < k}}^F(x) = F(\langle f_0, x \rangle, \dots, \langle f_{k-1}, x \rangle)$$

where

$$f_i : \Omega \rightarrow \mathbb{R}$$

“indicator” function

$$F : \mathbb{R}^k \rightarrow \mathbb{R}$$

smooth composition

## Energy Approximation via Linearization

Given the regional energy  $R_{\{f_i\}_{i < k}}^F(\cdot)$ , we receive

$$\begin{aligned} T_{x_0}^1 R_{\{f_i\}}^F(x) &= R_{\{f_i\}}^F(x_0) + \sum_{i < k} \frac{\partial F}{\partial v_i} \cdot \langle f_i, x - x_0 \rangle \\ &= \underbrace{R_{\{f_i\}}^F(x_0) - \sum_{i < k} \frac{\partial F}{\partial v_i} \cdot \langle f_i, x_0 \rangle}_{\text{depends only on } x_0} + \underbrace{\sum_{i < k} \frac{\partial F}{\partial v_i} \cdot \langle f_i, x \rangle}_{\text{linear in } x} \\ &= \text{const} + \left\langle \underbrace{\sum_{i < k} \frac{\partial F}{\partial v_i} \cdot f_i}_{\nabla R(x_0)}, x \right\rangle \end{aligned}$$

Note that the linearization results in a purely modular energy.

## Global Optimization of the Volume Constraint

Energy:

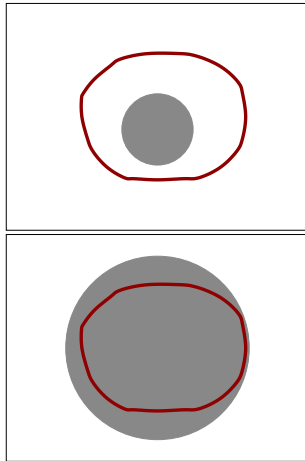
$$E(x) = \text{length}(x) + [\langle \mathbf{1}, x \rangle - V]^2$$

Approximation:

$$\tilde{E}(x) = \text{length}(x) + V^2 - \langle \mathbf{1}, x_0 \rangle^2 + 2\langle [\langle \mathbf{1}, \mathbf{x}_0 \rangle - \mathbf{V}], x \rangle$$

Minimize:

$$\text{length}(x) + 2\langle [\langle \mathbf{1}, \mathbf{x}_0 \rangle - \mathbf{V}], x \rangle$$



- Segmentation  $x_0$  too big
- $\langle \mathbf{1}, x_0 \rangle - V > 0$  (*shrinking bias*)
- $\arg \min \tilde{E}(x) = \emptyset$
  
- Segmentation  $x_0$  too small
- $\langle \mathbf{1}, x_0 \rangle - V < 0$  (*ballooning force*)
- $\arg \min \tilde{E}(x) = \Omega$

Global optimization fails to solve the problem. Optimise locally instead.

**Trust Region**

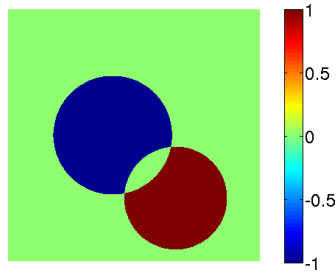
- Global minimum  $S'$  of  $\tilde{E}$  may have an higher energy than  $E(S_0)$ .
- Following a path from  $S_0$  to  $S'$  in direction of  $-\nabla E(S_0)$ , decreases  $E$ .
- Find intermediate points by constraining the optimization to a local region.

$$\min_{\text{dist}(S_0, S) < \delta} E_0(S) + \langle \nabla R(S_0), S \rangle$$

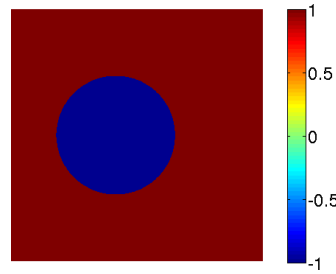
using shape distance  $\text{dist}(S_0, S)$ . With Lagrangian multiplier  $\lambda$ :

$$\min_S E_0(S) + \langle \nabla R(S_0), S \rangle + \lambda \cdot \text{dist}(S_0, S)$$

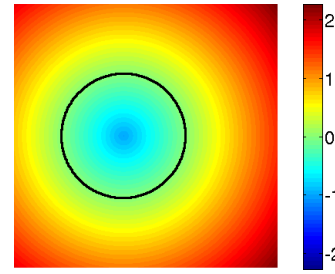
## Regional Shape Distance



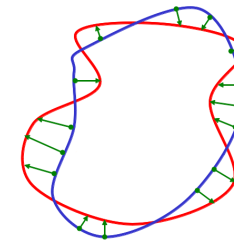
Hamming



Indicator  $\phi_{S_0}$



Signed Distance



Shape Distance

$$\text{dist}_1(S_0, S) = \int_S \phi_{S_0}(x) dx - \int_{S_0} \phi_S(x) dx$$

(Hamming Distance)

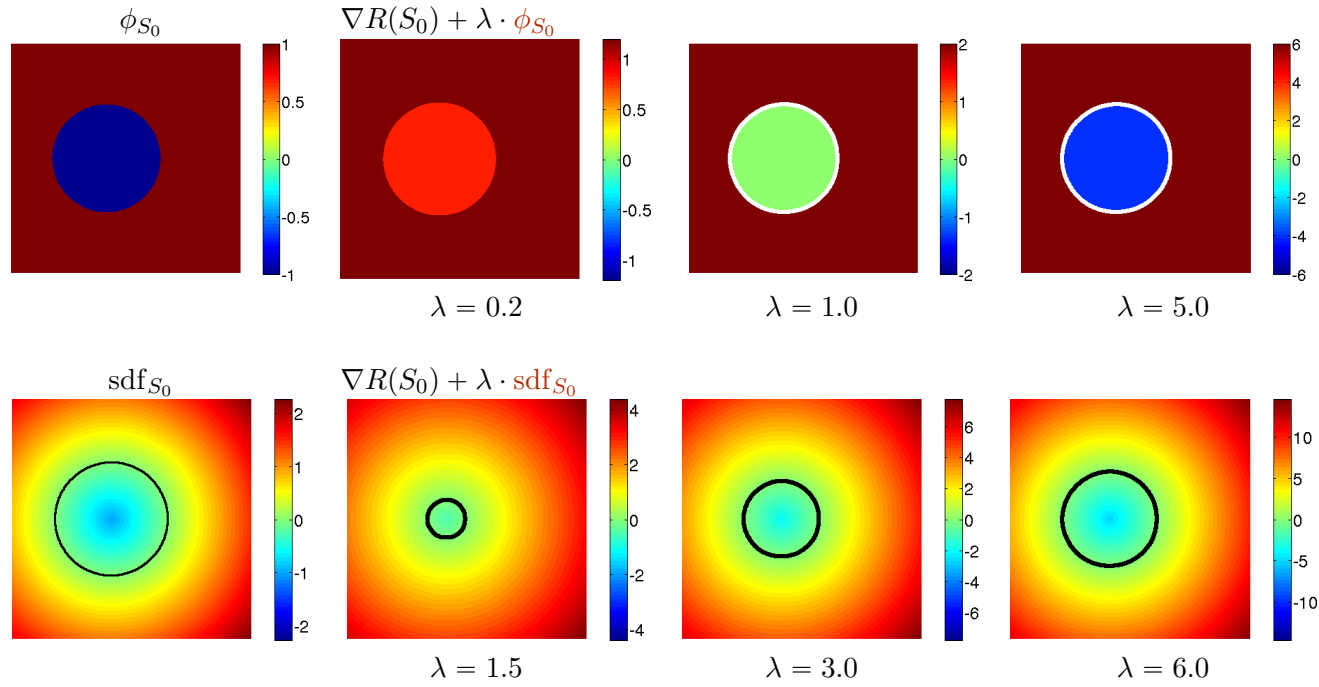
$$\text{dist}_2(S_0, S) = \int_S \text{sdf}_{S_0}(x) dx - \int_{S_0} \text{sdf}_S(x) dx$$

(Shape Distance)

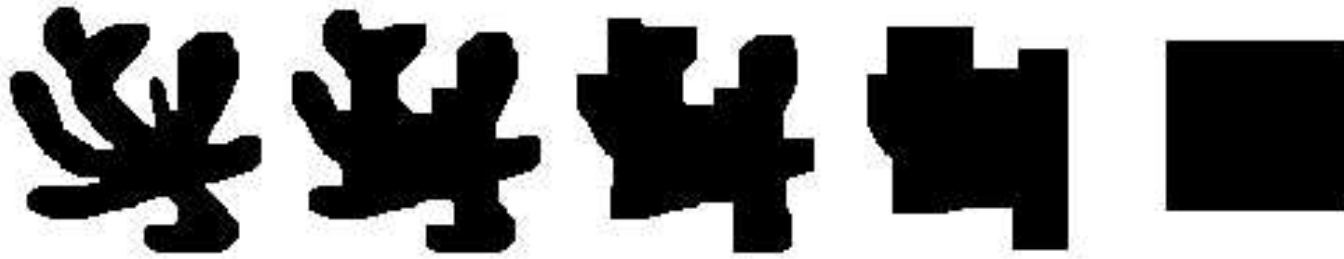
$$\min_S E_0(S) + \langle \nabla R(S_0) + \lambda \cdot \phi_{S_0}, S \rangle$$

$$\min_S E_0(S) + \langle \nabla R(S_0) + \lambda \cdot \text{sdf}_{S_0}, S \rangle$$

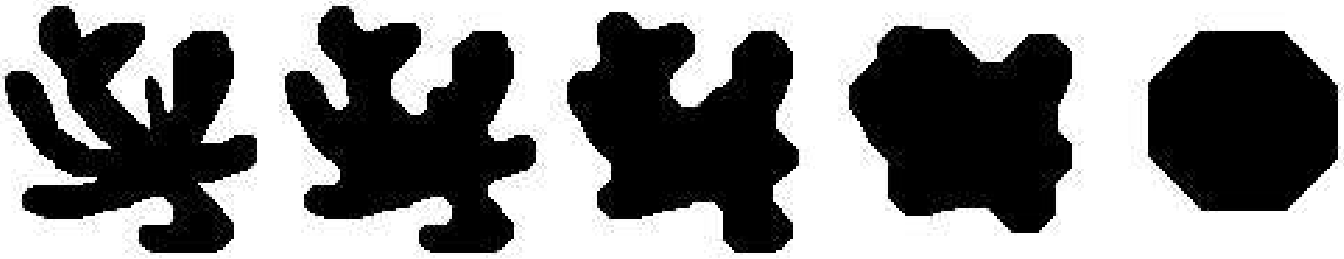
## Regional Shape Distance



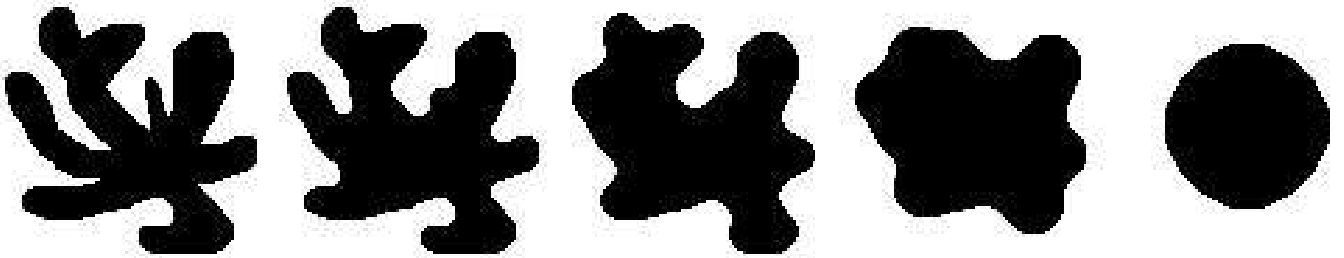
Volume Constraint



4 neighborhood



8 neighborhood



16 neighborhood

## Holistic Distribution Segmentation

$$E(S) = \text{Hist}(S) + \text{length}(S)$$



Level Set (t=1)  
after 1 hour



Level Set (t=50)  
after 1 hour



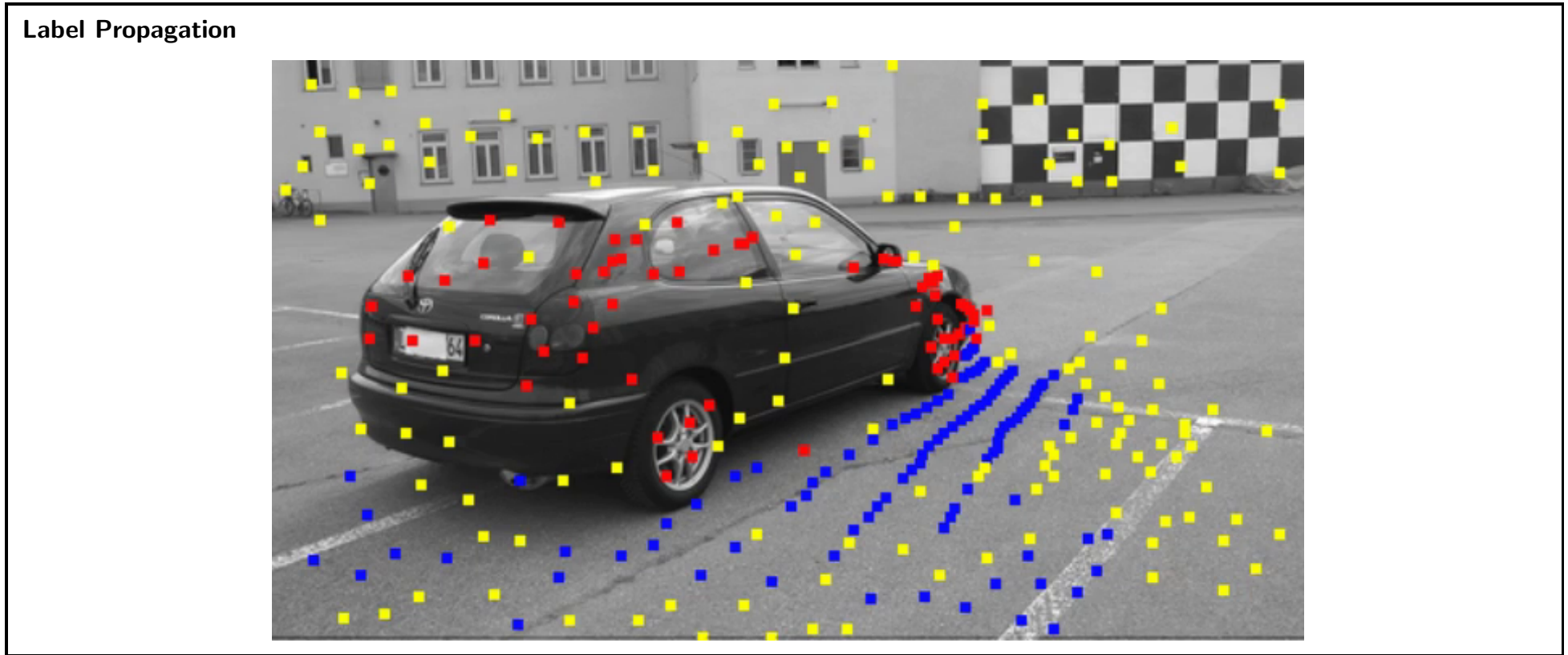
Level Set (t=1000)  
after 1 hour



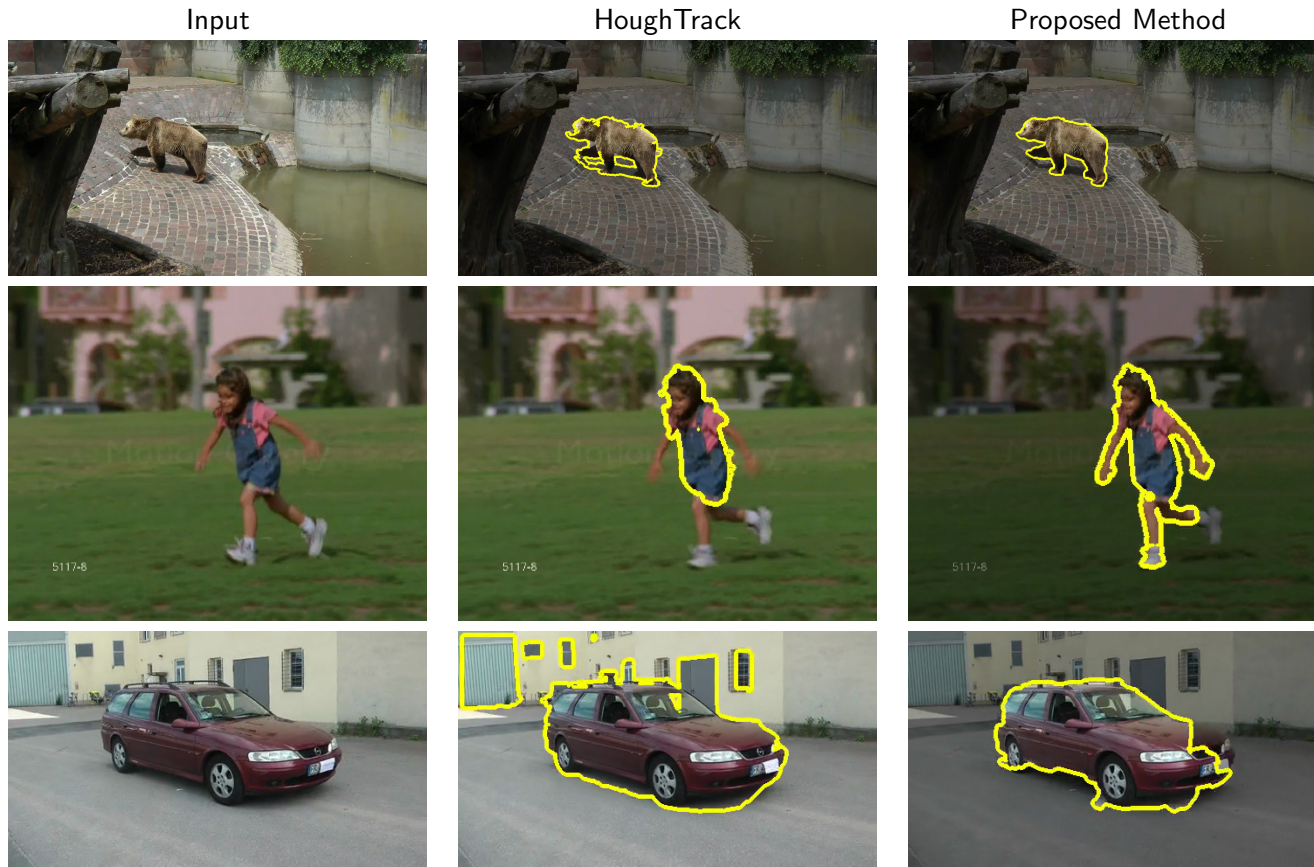
Trust Region  
64 seconds







## Label Propagation



## Literature

### Fast Trust Region in Computer Vision

- Gorelick, Schmidt, Boykov, DeLong, Ward. *Segmentation with non-linear regional constraints via line-search cuts*, 2012, ECCV, 583–597.
- Gorelick, Schmidt, Boykov. *Fast Trust Region for Segmentation*, 2013, IEEE CVPR.
- Gorelick, Boykov, Veksler, Ben Ayed, DeLong. *Submodularization for Binary Pairwise Energies*, 2014, IEEE CVPR.
- Nagaraja, Schmidt, Brox. *Video Segmentation with Just a Few Strokes*, 2015, IEEE ICCV.