Combinatorial Optimization in Computer Vision (IN2245)

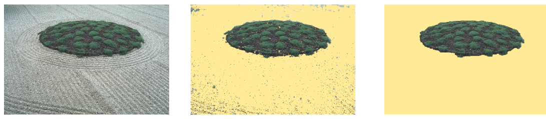
Frank R. Schmidt
Csaba Domokos

Winter Semester 2015/2016

21. Fast Trust Region for Object Tracking

Image Segmentation

Submodular Image Segmentation



Given Image | Data Term | Data + Length Term

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

This can be efficiently minimized via graph cut.

Bayes Interpretation

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

$$\max P(x|I) = \max \frac{P(I|x)\cdot P(x)}{P(I)}$$

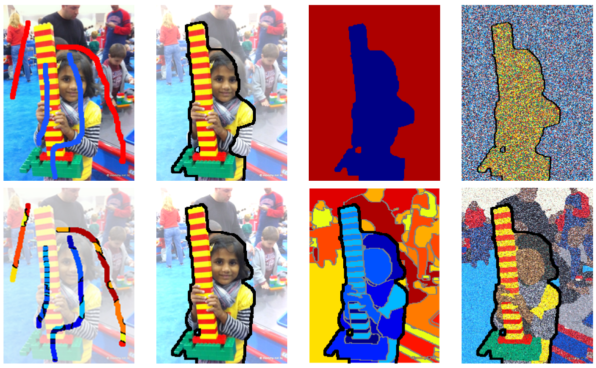
$$E(x) = -\log(P(x|I)) = -\log(P(I|x)) - \log(P(x)) + \text{const}$$

Using

$$\begin{aligned}P(I|x) &= \prod_{i\in\Omega} P_{\mathsf{data}}(I(i)|x(i))\\ P_{\mathsf{data}}(I(i)|0) &= \operatorname{pdf}^{(0)}(I(i)) = e^{-f^{(0)}(I(i))}\\ P_{\mathsf{data}}(I(i)|1) &= \operatorname{pdf}^{(1)}(I(i)) = e^{-f^{(1)}(I(i))}\\ P(x) &= \exp(-\operatorname{length}(x))\end{aligned}$$

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

Hierarchical MRF



$x_1: \mathcal{L}_N \to \mathbb{B}$ $\qquad x_0: \Omega \to \mathcal{L}_N$

Regional Functionals

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 **balooning 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.

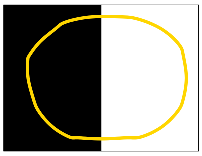
Whether this approach is robust depends highly on the used data term. What we actually want to minimise is

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

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 $f_0, f_1 \colon \Omega \to \{0,1\}$ 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**

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

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

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

where

$f_i : \Omega \to \mathbb{R}$	"indicator" function
$F : \mathbb{R}^k \to \mathbb{R}$	smooth composition

Energy Approximation via Linearization

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

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

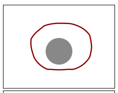
Note that the linearization results in a purely modular energy.

Global Optimization of the Volume Constraint

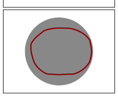
Energy: $E(x) = \text{length}(x) + [\langle 1, x\rangle - V]^2$

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

Minimize: $\text{length}(x) + 2\langle [\langle \mathbf{1}, \mathbf{x_0}\rangle - \mathbf{V}], x\rangle$



- Segmentation x_0 too big
- $\langle 1, x_0\rangle - V > 0$ (*shrinking bias*)
- $\arg\min \widetilde{E}(x) = \varnothing$

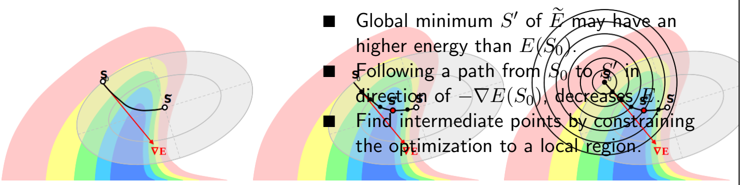


- Segmentation x_0 too small
- $\langle 1, x_0\rangle - V < 0$ (*ballooning force*)
- $\arg\min \widetilde{E}(x) = \Omega$

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

Fast Trust Region

Trust Region



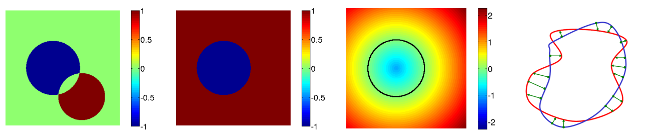
- Global minimum S' of $\widetilde{E}$ may have an higher energy than $E(S_0)$.
- Following a path from S_0 to S' in direction of $-\nabla E(S_0)$, decreases E.
- Find intermediate points by constraining the optimization to a local region.

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

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

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

Regional Shape Distance



Hamming | Indicator ϕ_{S_0} | Signed Distance | Shape Distance

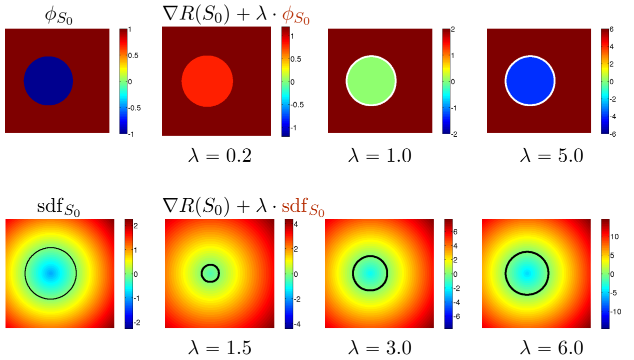
$$\text{dist}_1(S_0, S) = \int_S \phi_{S_0}(x)\,\mathrm{d}x - \int_{S_0} \phi_{S_0}(x)\,\mathrm{d}x \qquad \text{(Hamming Distance)}$$

$$\text{dist}_2(S_0, S) = \int_S \text{sdf}_{S_0}(x)\,\mathrm{d}x - \int_{S_0} \text{sdf}_{S_0}(x)\,\mathrm{d}x \qquad \text{(Shape Distance)}$$

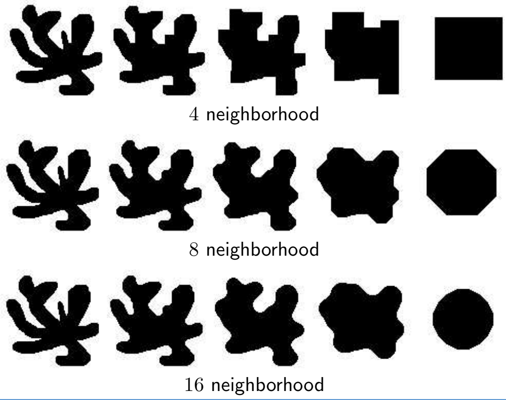
$$\min_S E_0(S) + \langle \nabla R(S_0) + \lambda \cdot \phi_{S_0}, S\rangle$$

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

Regional Shape Distance



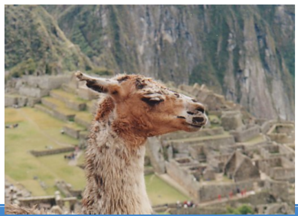
Volume Constraint



Holistic Distribution Segmentation

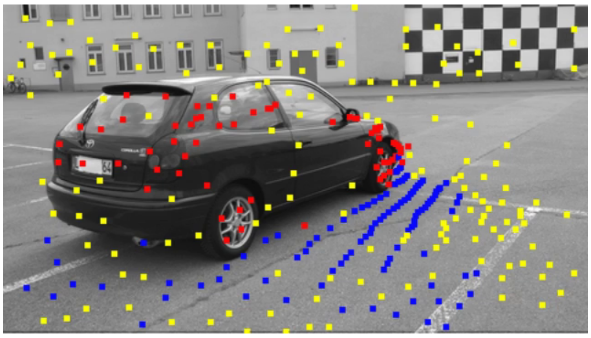
$$E(S) = \text{Hist}(S) + \text{length}(S)$$

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

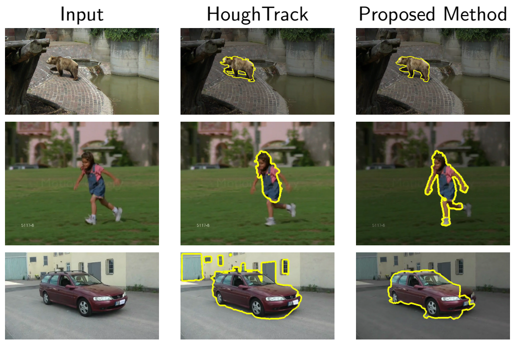


Video Label Propagation

Label Propagation



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.