Combinatorial Optimization in Computer Vision (IN2245)

Frank R. Schmidt
Csaba Domokos

Winter Semester 2015/2016

19. Branch and Mincut

Introduction

We address the problem of **binary image segmentation**, where we also consider *non-local parameters* that are known *a priori*.

For example, one can assume prior knowledge about the **shape** of the foreground segment or the **color distribution** of the foreground and/or background.

Let us consider a set of pixels $\mathcal{V}$ and a set of edges $\mathcal{E}$ consisting of 8-connected pairs of pixels. We define an energy function for non-local parameter ω:

$$E(\mathbf{x}, \omega) = C(\omega) + \sum_{p \in \mathcal{V}} F^p(\omega) \cdot x_p + \sum_{p \in \mathcal{V}} B^p(\omega) \cdot (1 - x_p) + \sum_{(p,q) \in \mathcal{E}} P^{pq}(\omega) \cdot |x_p - x_q| \ ,$$

where $C(\omega)$ is a *constant potential*, and $F^p(\omega)$ and $B^p(\omega)$ are the *unary potentials* defining the cost for assigning the pixel p to the foreground and to the background, respectively. $P^{pq}(\omega)$ is the *pairwise potential* that is **non-negative** to ensure the tractability of $E(\mathbf{x}, \omega)$.

Globally optimal segmenation

The segmentation is given by binary labeling $\mathbf{x} \in 2^{\mathcal{V}}$, where individual pixel labels are denoted by $x_p \in \mathbb{B}$ (1:foreground, 0:background). We assume that non-local parameter $\omega \in \Omega$ are taken from a **discrete set**.

Shape priors will be encoded as a product space of various poses and deformations of the *template*, while color priors will correspond to the set of parametric color distributions.

The goal is to achieve a **globally optimal** segmentation under non-local priors. The applied optimization method relies on two techniques: **graph cuts** and **branch-and-bound**.

Although a global minimum can be achieved, the worst case complexity of the method is large (essentially, the same as the exhaustive search over the space of non-local parameters).

An alternative way to solve the problem is to apply *alternating minimization*.

Lower bound

$L(\Omega)$ denotes the lower bound for $E(\mathbf{x}, \omega)$ over $2^{\mathcal{V}} \times \Omega$:

$$\begin{aligned}
&\min_{\mathbf{x} \in 2^{\mathcal{V}}, \omega \in \Omega} E(\mathbf{x}, \omega) \\
&= \min_{\mathbf{x} \in 2^{\mathcal{V}}, \omega \in \Omega} \Big\{ C(\omega) + \sum_{p \in \mathcal{V}} F^p(\omega) \cdot x_p + \sum_{p \in \mathcal{V}} B^p(\omega) \cdot (1 - x_p) + \sum_{(p,q) \in \mathcal{E}} P^{pq}(\omega) \cdot |x_p - x_q| \Big\} \\
&\geqslant \min_{\mathbf{x} \in 2^{\mathcal{V}}} \Big\{ \min_{\omega \in \Omega} C(\omega) + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} F^p(\omega) \cdot x_p + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} B^p(\omega) \cdot (1 - x_p) + \\
&\quad \sum_{(p,q) \in \mathcal{E}} \min_{\omega \in \Omega} P^{pq}(\omega) \cdot |x_p - x_q| \Big\} \\
&= \min_{\mathbf{x} \in 2^{\mathcal{V}}} \Big\{ C_\Omega + \sum_{p \in \mathcal{V}} F^p_\Omega(\omega) \cdot x_p + \sum_{p \in \mathcal{V}} B^p_\Omega(\omega) \cdot (1 - x_p) + \sum_{(p,q) \in \mathcal{E}} P^{pq}_\Omega(\omega) \cdot |x_p - x_q| \Big\} \\
&= L(\Omega) \ .
\end{aligned}$$

C_Ω, F^p_Ω, B^p_Ω, P^{pq}_Ω denote the minima of $C(\omega)$, $F^p(\omega)$, $B^p(\omega)$, $P^{pq}(\omega)$ over $\omega \in \Omega$ referred as **aggregated potentials**.

Monotonicity

Suppose $\Omega_1 \subset \Omega_2$, then the inequality $L(\Omega_1) \geqslant L(\Omega_2)$ holds.

Proof. Let us define $A(\mathbf{x}, \Omega)$ as

$$\begin{aligned}
A(\mathbf{x}, \Omega) \triangleq \min_{\omega \in \Omega} C(\omega) + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} F^p(\omega) \cdot x_p + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} B^p(\omega) \cdot (1 - x_p) \\
+ \sum_{(p,q) \in \mathcal{E}} \min_{\omega \in \Omega} P^{pq}(\omega) \cdot |x_p - x_q| \ .
\end{aligned}$$

Assume $\Omega_1 \subset \Omega_2$. Then, for any $\mathbf{x}$

$$\begin{aligned}
&A(\mathbf{x}, \Omega_1) \\
&= \min_{\omega \in \Omega_1} C(\omega) + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega_1} F^p(\omega) x_p + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega_1} B^p(\omega)(1 - x_p) + \sum_{(p,q) \in \mathcal{E}} \min_{\omega \in \Omega_1} P^{pq}(\omega) |x_p - x_q| \\
&\geqslant \min_{\omega \in \Omega_2} C(\omega) + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega_2} F^p(\omega) x_p + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega_2} B^p(\omega)(1 - x_p) + \sum_{(p,q) \in \mathcal{E}} \min_{\omega \in \Omega_2} P^{pq}(\omega) |x_p - x_q| \\
&= A(\mathbf{x}, \Omega_2) \ .
\end{aligned}$$

Monotonicity

Proof. Continued

Note that $L(\Omega) = \min_{\mathbf{x} \in 2^{\mathcal{V}}} A(\mathbf{x}, \Omega)$.

Let $\mathbf{x}_1 \in \operatorname{argmin}_{\mathbf{x} \in 2^{\mathcal{V}}} A(\mathbf{x}, \Omega_1)$ and $\mathbf{x}_2 \in \operatorname{argmin}_{\mathbf{x} \in 2^{\mathcal{V}}} A(\mathbf{x}, \Omega_2)$, then from the monotonicity, one gets:

$$L(\Omega_1) = A(\mathbf{x}_1, \Omega_1) \geqslant A(\mathbf{x}_1, \Omega_2) \geqslant A(\mathbf{x}_2, \Omega_2) = L(\Omega_2) \ .$$

$\square$

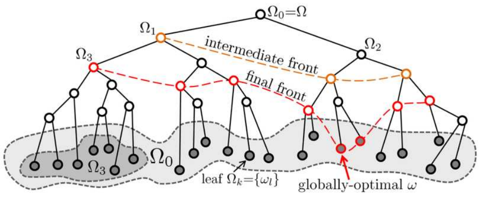
Computability and tightness

Computability: the lower bound $L(\Omega)$ equals the minimum of a submodular quadratic pseudo-boolean function, which can be globally minimized via graph-cut.

Tightness: for a singleton $\Omega = \omega$ (i.e. $|\Omega| = 1$) the bound $L(\Omega)$ is **tight**, that is

$$L(\{\omega\}) = \min_{\mathbf{x} \in 2^{\mathcal{V}}} E(\mathbf{x}, \omega) \ .$$

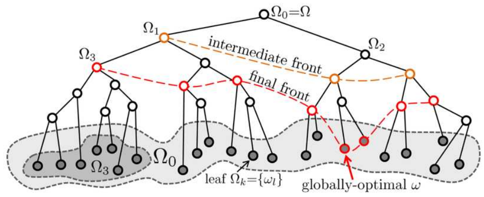
Best-first branch-and-bound optimization



The discrete domain Ω can be hierarchically clustered and the binary tree of its subregions can be considered.

At each step the *active node* with the **smallest** lower bound is removed from the *active front*, while two of its *children* are added to the *active front* (due to *monotonicity property* they have higher or equal lower bounds).

Best-first branch-and-bound optimization



If the *active node* with the smallest lower bound turns out to be a **leaf** $\{\omega'\}$ and $\mathbf{x}'$ is the corresponding optimal segmentation, then $E(\mathbf{x}', \omega') = L(\omega')$ due to the *tightness property*. Consequently, $(\mathbf{x}', \omega')$ is a **global minimum**.

Remark that in worst-case any optimization has to search exhaustively over Ω.

Pseudo code of Branch-And-Mincut *

Front $\leftarrow \varnothing$ ▷ initializing the priority queue
$[C_0, \{F_0^p\}, \{B_0^p\}, \{P_0^{pq}\}] \leftarrow$`GetAggregPotentials`$(\Omega_0)$
$\mathrm{LB}_0 \leftarrow$`GetMaxFlowValue`$(\{F_0^p\}, \{B_0^p\}, \{P_0^{pq}\}) + C_0$
Front.`InsertWithPriority`$(\Omega_0, -\mathrm{LB}_0)$
while true **do** ▷ advancing front
  $\Omega \leftarrow$ Front.`PullHighestPriorityElement()`
  if `IsSingleton`(Ω) `then` ▷ `global minimum found`
    $\omega \leftarrow \Omega$
    $[C, \{F^p\}, \{B^p\}, \{P^{pq}\}] \leftarrow$`GetAggregPotentials`$(\omega)$
    $\mathbf{x} \leftarrow$`FindMinimumViaMincut`$(\{F^p\}, \{B^p\}, \{P^{pq}\})$
    `return` $(\mathbf{x}, \omega)$
  end if
  $[\Omega_1, \Omega_2] \leftarrow$`GetChildrenSubdomains`$(\Omega)$
  $[C_1, \{F_1^p\}, \{B_1^p\}, \{P_1^{pq}\}] \leftarrow$`GetAggregPotentials`$(\Omega_1)$
  $\mathrm{LB}_1 \leftarrow$`GetMaxFlowValue`$(\{F_1^p\}, \{B_1^p\}, \{P_1^{pq}\}) + C_1$
  Front.`InsertWithPriority`$(\Omega_1, -\mathrm{LB}_1)$
  $[C_2, \{F_2^p\}, \{B_2^p\}, \{P_2^{pq}\}] \leftarrow$`GetAggregPotentials`$(\Omega_2)$
  $\mathrm{LB}_2 \leftarrow$`GetMaxFlowValue`$(\{F_2^p\}, \{B_2^p\}, \{P_2^{pq}\}) + C_2$
  Front.`InsertWithPriority`$(\Omega_2, -\mathrm{LB}_2)$
end while

Segmentation with shape priors

The prior is defined by the set of exemplar binary segmentations $\{\mathbf{y}^\omega \mid \omega \in \Omega\}$, where Ω is a discrete set indexing the exemplar segmentations.

We define a joint prior over the segmentation and the non-local parameter:

$$E_{\text{prior}}(\mathbf{x}, \omega) = \sum_{p \in \mathcal{V}} (1 - y_p^\omega) \cdot x_p + \sum_{p \in \mathcal{V}} y_p^\omega \cdot (1 - x_p) .$$

This encourages the segmentation $\mathbf{x}$ to be close in the Hamming-distance to one of the exemplar shapes.

The segmentation energy may be defined by adding a standard *contrast-sensitive term* for $\lambda, \sigma > 0$:

$$E_{\text{shape}}(\mathbf{x}, \omega) = E_{\text{prior}}(\mathbf{x}, \omega) + \lambda \sum_{(p,q) \in \mathcal{E}} \frac{e^{-\frac{\|K_p - K_q\|}{\sigma}}}{|p - q|} \cdot |x_p - x_q| ,$$

where K_p denotes RGB colors of the pixel p.

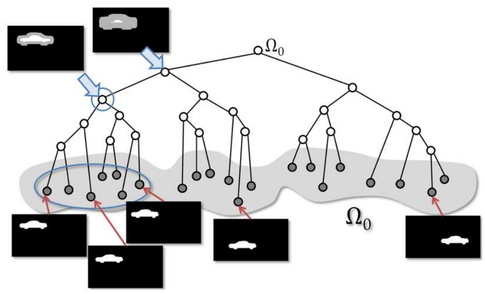
Multiple templates $\times$ translations

The shape prior is given by a set of templates, whereas each template can be located anywhere within the image.

$\Omega_{\text{shape}} = \Delta \times \Theta$, where the set Δ indexes the set of all exemplar segmentations y_δ and Θ corresponds to translations.

Any exemplar segmentation y^ω for $\omega = (\delta, \theta)$ is then defined as some exemplar segmentation y_δ centered at the origin and then translated by the shift θ.

Clustering tree



For Δ we use agglomerative bottom-up clustering resulting clustering tree $T_\Delta = \{\Delta = \Delta_0, \Delta_1, \dots, \Delta_N\}$.

To build a clustering tree for Θ, we recursively split along the "longer" dimension. This leads to a tree $T_\Theta = \{\Theta = \Theta_0, \Theta_1, \dots, \Theta_N\}$.

Branch operation

Each *nodeset* Ω_t in the combined tree is defined by a pair $\Delta_{p(t)} \times \Theta_{q(t)}$

The **looseness** of a nodeset Ω_t is defined as the number of pixels that change their mask value under different shapes in Ω_t (i.e. neither background nor foreground):

$$\Lambda(\Omega_t) = |\{p \mid \exists \omega_1, \omega_2 : y_p^{\omega_1} = 0 \text{ and } y_p^{\omega_2} = 1\}| .$$

The tree is built in a recursive top-down fashion as follows.

We start by creating a root nodeset $\Omega_0 = \Delta_0 \times \Theta_0$. Given a nodeset $\Omega_t = \Delta_t \times \Theta_t$ we consider (recursively) two possible splits: 1) split along the shape dimension or 2) split along the shift dimension. The split that minimizes the sum of loosenesses is preferred.

The recursion stops when the leaf level is reached within both the shape and the shift trees.

Results *

Yellow: global minimum of E_{shape}; Blue: feature-based car detector; Red: global minimum of the combination of E_{shape} with detection results (detection is included as constant potential)

The prior set Δ was built by manual segmentation of 60 training images coming with the dataset.

Victor Lempitsky, Andrew Blake, and Carsten Rother. **Branch-and-Mincut: Global Optimization for Image Segmentation with High-Level Priors**. *Journal of Mathematical Imaging and Vision*, vol. 44(3), pp. 315–329, March, 2012.