Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation

Yuliya Tarabalka — Inria Sophia-Antipolis, AYIN team, Sophia Antipolis, France
yuliya.tarabalka@inria.fr

Guillaume Charpiat — Inria Sophia-Antipolis, STARS team
guillaume.charpiat@inria.fr

Ludovic Brucker — NASA GSFC & Universities Space Research Association, MD, USA
ludovic.brucker@nasa.gov

Bjoern H. Menze — Computer Vision Laboratory, ETH Zurich, Switzerland & Inria Sophia-Antipolis, Asclepios team
bjoern@ethz.ch

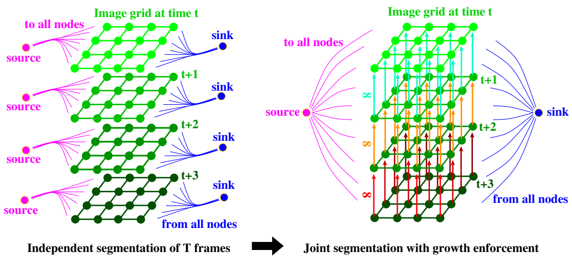

Figure 1: Enforcing shape growth in an image sequence.

One of the great challenges in computer vision is automatic segmentation of objects in videos. This task becomes more difficult when image sequences are subject to low signal-to-noise ratio or low contrast between intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing.

While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focus on segmenting shapes which only grow or shrink in time, from sequences of extremely noisy images. We consider image sequences, where both foreground and background intensity distributions can vary significantly over time, foreground can be heavily occluded or undistinguishable from a part of the background. Most of previously-proposed spatio-temporal methods rely on coherence of foreground/background intensity distributions in consecutive image frames, and therefore fail when segmenting such noisy data sets. Few approaches have been designed for spatio-temporal segmentation of shapes from magnetic resonance (MR) images with low signal-to-noise ratio [3, 4]. Applied to multi-temporal sequences that show a monotonously growing or shrinking structure, these smoothing methods bias results towards the mean shape obtained from averaging consecutive segmentations and, hence, underestimate rapid growth or shrinkage events.

To address this issue, we propose a new *segmentation framework* based on graph cuts for the joint segmentation of an image sequence. It *introduces growth or shrinkage constraint in graph cuts* by using directed infinite links, connecting pixels at the same spatial locations in successive image frames. By minimizing an energy computed on the resulting spatio-temporal graph of the image set, the proposed method yields a *globally optimal solution*. Differently from the state-of-the-art spatio-temporal techniques, it does not rely on the coherence of the intensity in time, but only on the coherence of the shape.

Graph cut is an optimization tool, which can be used to find the globally optimal binary segmentation of images [1], where the segmentation criterion E is related to a Markov Random Field with submodular interaction terms:

$$E(L) = \sum_{\text{pixels } i} V_i(L_i) + \sum_{i \sim j} W_{i,j}(L_i, L_j), \qquad (1)$$

where L is the binary labelling function to be found (L_i is the label of pixel i), individual potentials V_i are any binary real-valued functions measuring the disagreement between a prior probabilistic model and the observed data, $i \sim j$ denotes a pair of neighboring pixels, and $W_{i,j}$ are any real-valued interaction terms between neighboring pixels expressing spatial coherency of labels. A directed infinite link between two pixels expresses precisely the constraint that this pair of pixels cannot have the pair of labels (0,1). In the case of image binary segmentation, if 0 stands for the background and 1 for the foreground object, then this means that the second pixel may belong to the foreground only if the first one does.

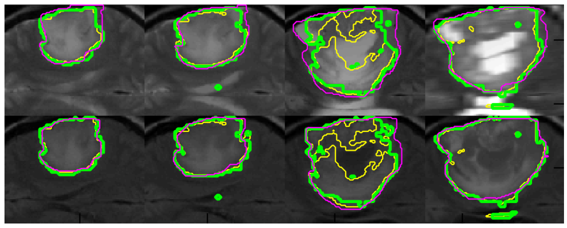

Figure 2: Time series of T2 and FLAIR MR image volumes, each of which acquired about 3-6 months apart. The rapidly growing tumor between the second and fourth scene leads to a suboptimal performance of the initial multimodal segmentation (yellow) [2]. The proposed segmentation with growth constraints (green) delineates areas similar to the manual evaluation (magenta). It does not smooth out outlines of rapidly growing tumors as conventional bi-directional temporal constraints would do.

Shape growth in a sequence of images $I(t)$ can be easily expressed as the property that the foreground object cannot lose any pixel when time advances. Otherwise said, if a pixel belongs to the foreground object at time t_1, then it belongs also to the foreground object for all times $t_2 > t_1$. Equivalently, and simpler: a pair of pixels $((x,y,t),(x,y,t+1))$, sharing the same location and immediately successive in time, cannot have the pair of labels (1,0), with the same binary segmentation notations as above. This can be enforced by setting directed infinite links from all pixels to their immediate predecessor in time.

Given T images $I(t), t \in [1,T]$, and as many associated submodular segmentation criteria E^t, we transform the problem of segmenting independently each image $I(t)$ according to its criterion E^t, into a joint segmentation of all images together, by enforcing the shape growth constraint with directed infinite links (Fig. 1). Thus, instead of applying graph cut T times independently to planar grids of the size of the images $W \times H$, we apply graph cut once to a 3D grid $W \times H \times T$, consisting of the same nodes and edges, but with additional directed infinite links in time. The criterion to be minimized is then $E = \sum_t E^t$ under the constraint of shape growth, and the solution found by graph cut is globally optimal, since the problem is binary and submodular.

One can enforce shape shrinkage instead of shape growth, by reversing the direction of infinite links. Another straightforward extension consists in applying this approach to the case of sequences of 3D images. The directed infinite links are then set for all pairs of voxels of the form $((x,y,z,t),(x,y,z,t-1))$ to enforce 3D shape growth.

We validated the performance of the proposed approach for segmentation of a shrinking ice floe and growing brain tumors from multimodal sequences of 2D satellite and 3D MR images (Fig. 2), respectively. In the latter application, we imposed an additional inter-sequences object inclusion constraint by adding directed infinite links to the joint graph associated to all sequences. The new method proved to be robust to very noisy or low-contrast images, and showed linear complexity.

[1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, 26(9):1124–1137, 2004.

[2] B. H. Menze et al. A generative model for brain tumor segmentation in multi-modal images. In *MICCAI*, pages 151–159, 2010.

[3] T. Riklin-Raviv et al. Segmentation of image ensembles via latent atlases. *Med. image analys.*, 14(5):654–665, Oct. 2010.

[4] R. Wolz et al. Measurement of hippocampal atrophy using 4d graph-cut segmentation: Application to ADNI. *NeuroImage*, 52(1):109–118, 2010.