



# Spatio-Temporal Video Segmentation with Shape Growth or Shrinkage Constraint

Yuliya Tarabalka, Guillaume Charpiat, Ludovic Brucker, Bjoern Menze

► **To cite this version:**

Yuliya Tarabalka, Guillaume Charpiat, Ludovic Brucker, Bjoern Menze. Spatio-Temporal Video Segmentation with Shape Growth or Shrinkage Constraint. *IEEE Transactions on Image Processing*, Institute of Electrical and Electronics Engineers (IEEE), 2014, 23 (9), pp.3829-3840. <hal-01052543>

**HAL Id: hal-01052543**

**<https://hal.inria.fr/hal-01052543>**

Submitted on 28 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatio-Temporal Video Segmentation With Shape Growth or Shrinkage Constraint

Yuliya Tarabalka, *Member, IEEE*, Guillaume Charpiat, Ludovic Brucker, and Bjoern H. Menze

**Abstract**—We propose a new method for joint segmentation of monotonously growing or shrinking shapes in a time sequence of noisy images. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or of shrinkage by introducing monodirectional infinite links connecting pixels at the same spatial locations in successive image frames. The globally optimal solution is computed with a graph cut. The performance of the proposed method is validated on three applications: segmentation of melting sea ice floes and of growing burned areas from time series of 2D satellite images, and segmentation of a growing brain tumor from sequences of 3D medical scans. In the latter application, we impose an additional intersequences inclusion constraint by adding directed infinite links between pixels of dependent image structures.

**Index Terms**—Video segmentation, shape growth, graph cut, energy minimization, infinite links.

## I. INTRODUCTION

**A**UTOMATIC segmentation of objects in videos is a difficult endeavor in computer vision [1]. This task becomes even more challenging for image sequences with low signal-to-noise ratio or low contrast between intensities of spatially adjacent objects in the image scene. Such challenging data are recorded frequently, for instance, in satellite remote sensing or medical imaging.

Image segmentation methods applied independently to each frame [2], [3] produce unstable results, while temporal coherence in video sequences yields a lot of information not available for a single image. There are two main categories of approaches for the spatio-temporal segmentation of image sequences. *Causal*, or *feedforward*, techniques consider only past data for segmenting each next frame [4], [5]. *Omniscient* approaches take advantage of both past and future data by analyzing the video as a  $2D+T = 3D$  spatio-temporal pixel

volume [6]–[8]. In this case, the segmentation of the entire image set supports each of the individual segmentations. Graph-based methods gained popularity among omniscient approaches, in particular those using hierarchical model [8], normalized cuts [6] or graph cuts [9]. These techniques do not impose any shape prior knowledge. It was proven that introducing shape priors into image segmentation, *i.e.* favoring segmentations similar in some sense to a given shape, allows to drastically improve segmentation of objects in the presence of strong noise and occlusions [10], [11]. However, imposing shape priors increases significantly both algorithmic and computational complexity of segmentation algorithms [12]. Schoenemann and Cremers proposed to compute minimal ratio cycles in a large product graph spanned by the image and the shape template for finding globally optimal segmentations, which are consistent both with edge information and with a shape prior [13], [14]. In order to speed up the algorithm, they implemented it in parallel on graphics hardware. This algorithm can be applied for segmenting shapes in time series in a causal way, so that the template determined for the last frame is matched to the next. However, it does not guarantee globally optimal solution over the whole temporal sequence.

In this paper, we focus on segmenting shapes which only grow or shrink in time, from sequences of extremely noisy images. Examples of growing shapes are forest fires or expanding cities in satellite images and organ development in medical imaging. In the image sequences we consider, both foreground and background intensity distributions can vary significantly over time: foreground can be heavily occluded or undistinguishable from a part of the background, and data for some pixels can be missing (see Figs. 2 and 7 for examples of such sequences). Most of previously-proposed spatio-temporal methods rely on coherence of foreground/background intensity distributions in successive image frames, and are therefore not suited for segmenting such noisy data sets. Few approaches have been specifically designed for spatio-temporal segmentation of magnetic resonance image (MRI) sequences with low signal-to-noise ratio [9], [15]. Applied to multi-temporal time series that show a monotonously growing or shrinking structure, however, 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 omniscient segmentation framework based on graph cuts for the joint segmentation of a multi-temporal image sequence. It *introduces growth or shrinkage constraint in graph cuts* by using directed

Manuscript received July 17, 2013; revised April 25, 2014; accepted July 2, 2014. Date of publication July 8, 2014; date of current version July 22, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xilin Chen.

Y. Tarabalka and G. Charpiat are with Inria Sophia-Antipolis Méditerranée, Sophia Antipolis 06902, France (e-mail: yuliya.tarabalka@inria.fr; guillaume.charpiat@inria.fr).

L. Brucker is with the NASA Goddard Space Flight Center, Universities Space Research Association, Greenbelt, MD 20771 USA (e-mail: ludovic.brucker@nasa.gov).

B. H. Menze is with the Department of Computer Science, Institute for Advanced Study, Technische Universität München, München 80333, Germany (e-mail: bjoern.menze@tum.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2014.2336544

infinite links, which connect pixels at the same spatial locations in successive image frames. By minimizing a submodular energy computed on the resulting spatio-temporal graph of the image sequence, the proposed method yields a *globally optimal solution*. Differently from the state-of-the-art omniscient techniques, it does not rely on the coherence of the intensity in time, but only on the coherence of the shape. To summarize, the main contribution we make in this paper is:

- 1) a new framework for segmentation of 2D/3D image time series with the constraint of shape growth/shrinkage,
- 2) in order to be able to segment extremely noisy/low-contrast/incomplete data,
- 3) in a very low computational time (see Fig. 5: linear time in the number of frames in practice).

We validate the performance of the proposed framework on three applications with very noisy image sequences. The first one deals with the segmentation of multiyear sea ice floes in a set of satellite images acquired through different satellite sensors. The new method returns accurate melting profiles of sea ice, which is important for building climate models. The second application segments growing burned areas from time series of optical satellite images with missing data. The third application addresses the segmentation of brain tumors from longitudinal sets of multimodal MRI volumes, where we impose additional inter-modal inclusion constraints for the joint segmentation of different image structures (brain tissues).

Note that the preliminary results related to this research have been presented in the conference paper [16]. The rest of the paper is organized as follows. In the next section, the proposed graph-cut-based segmentation framework is presented. Experimental results are discussed in Sec. III and conclusions are drawn in Sec. IV.

## II. ENFORCING SHAPE GROWTH/SHRINKAGE IN GRAPH CUTS

Graph cut is an optimization tool coming from graph theory, based on the rewriting of image segmentation problems as (s,t)-min-cuts in graphs, on the equivalence of (s,t)-min-cut and max-flow problems, and on the existence of efficient algorithms to solve the latter ones [17]–[19]. In practice in computer vision, it can be used to find the globally optimal binary segmentation of images where the segmentation criterion is related to a Markov Random Field with submodular interaction terms, *i.e.* a criterion  $E$  of the form:

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

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 (any neighborhood system can be used), and  $W_{i,j}$  are any real-valued interaction terms between neighboring pixels expressing spatial coherency of labels, satisfying

$$W_{i,j}(0, 0) + W_{i,j}(1, 1) \leq W_{i,j}(0, 1) + W_{i,j}(1, 0). \quad (2)$$

A directed infinite link between two pixels expresses precisely the constraint that this pair of pixels cannot have the pair of labels (0,1), by assigning an infinite cost to such an interaction. The remaining possible pairs of labels are thus (0,0), (1,1) and (1,0), which means that either both pixels have the same label, or the order of labels is predefined (1 for first pixel and 0 for the second one). 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 already does.

### A. Related Works

Extensions of graph cuts to multi-label problems (*i.e.* multi-class segmentation) have been proposed but generally do not guarantee optimal solutions, except *e.g.* in the case of Ishikawa’s construction [20], which requires labels to be ordered and the interaction term to be a convex function of their differences, *i.e.*  $W_{i,j}(L_i, L_j) = g(L_i - L_j)$  with  $g$  convex. This graph construction makes intensive use of infinite links to constrain the min-cut solutions to satisfy desired properties required to interpret them as image segmentation solutions. This was the source of inspiration for our work.

A study related to shape constraints can be found in [21], where one image has to be segmented in several possibly-overlapping objects. Infinite links are used for imposing common boundaries, inclusion or exclusion conditions between objects in a same single image. A similar approach [22] segments jointly two surfaces in a same volumic image, under the constraint that they should be separated by a given minimal margin. There is however no work related to shape growth or shrinkage in time series.

Wolz *et al.* [9] applied graph cuts for simultaneous segmentation of serially acquired MRI volumes. They defined temporal edge weights as the intensity differences of voxels at the same spatial locations. The same smoothness constraint was applied both in space and time, and the segmentations at different timepoints were forced to be consistent in areas where a small intensity difference between the images exist. This type of temporal constraint is suboptimal in image series where intensity distributions of foreground and background vary significantly over time. To the best of our knowledge, our work is the first to use infinite links to enforce a temporal growth constraint, and we illustrate in Sec. III the advantage of the new method over previous approaches such as [9].

### B. Growth/Shrinkage Constraint

Given a sequence of images  $I(t)$  preliminarily aligned, shape growth 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.

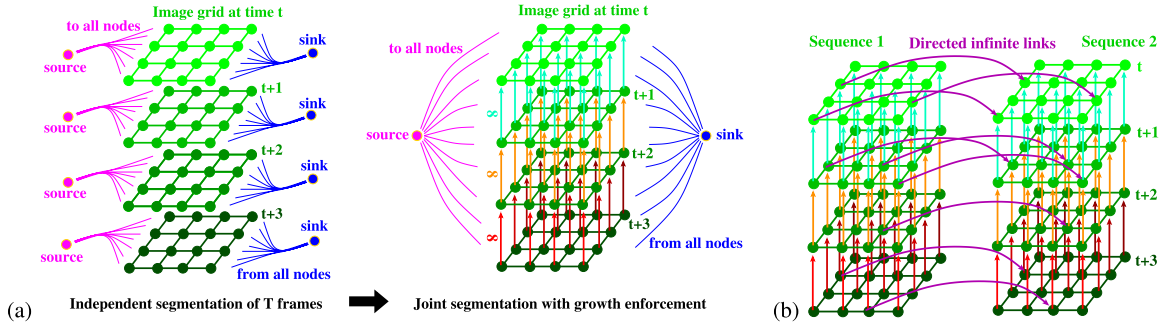


Fig. 1. (a) Enforcing shape growth in an image sequence. (b) Segmenting jointly two sequences  $S_1$  and  $S_2$ , by enforcing the foreground of  $S_1$  to contain the foreground of  $S_2$ , with directed infinite links from  $S_1$  towards  $S_2$ , between all pixels of coordinates  $(x, y, t, s_1)$  and  $(x, y, t, s_2)$ .

This can be enforced by setting monodirectional infinite links from all pixels to their immediate predecessor in time.

Given  $T$  images  $I(t)$ , with  $t \in [1, T]$ , 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 (see Fig. 1(a)). 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 monodirectional infinite links in time. The criterion to be minimized is then  $E = \sum_t E^t$  under the constraint of shape growth:

$$E(L) = \sum_{\text{pixels } i} V_i(L_i) + \sum_{i \sim j} W_{i,j}(L_i, L_j) + \infty \sum_t \delta_{L_t^i > L_{t+1}^i}. \quad (3)$$

Since the problem is binary and submodular, the solution found by graph cut is globally optimal.

Manifestly, one can enforce shape shrinkage instead of shape growth, by reversing the direction of the infinite links. Another straightforward extension, needed in Sec. III-C, 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.

In some applications, it may happen that growth (or shrinkage) is only very probable, but not with probability 1, *i.e.* growth should be considered as a probable hint but should not be enforced strictly at all locations at all times. In that case, one may replace directed infinite links by directed finite links: the weights of these links will encourage growth (more or less strongly depending on the weight), but sufficiently disagreeing potentials  $V_i$  may make the shape locally shrink instead. Thus, shrinkage would be discouraged but not forbidden.

### C. Inter-Sequences Inclusion Constraint

It is also possible to segment jointly several image sequences  $I(s)(t)$  with the constraint that the foreground object in some sequences should be included in the foreground

object of some other sequences. This can be done similarly by considering together the graphs associated to all sequences, and, for each inclusion constraint, by adding directed infinite links between pixels of the desired sequences  $s_1$  and  $s_2$ , sharing same location and time: such links from  $(s_1, x, y, t)$  to  $(s_2, x, y, t)$  for all  $x, y, t$  will force the foreground object in sequence  $s_1$  to contain the one of sequence  $s_2$  (see Fig. 1(b)). An example of such application is given in Sec. III-C, where image sequences correspond to four different MRI modalities, aligned both in space and time.

Naturally, instead of imposing an inclusion constraint over the whole time span and the whole image space, it is possible to specify spatio-temporal domains of constraints, for instance to express that the inclusion property between two sequences has to be satisfied inside a pre-defined region and/or during a pre-defined time span  $[t_1, t_2]$  only, by adding directed infinite links in these sets only.

### D. Weighting Frames by Reliability

As presented earlier, enforcing shape growth explicitly is particularly important when facing noisy sequences of images, as a joint segmentation incorporates information from different frames. With our approach, a fully noisy frame in the middle of several good-quality frames will be automatically ignored in practice, because statistically the (bad) potentials  $V_i$  at any pixel  $i$  in that frame will be neglectable with respect to all other (correct) potentials at the same pixel in other frames. Hence, the solution we bring will be strongly robust to any kind of noise, provided that the noise is not coherent in time during long time spans.

The quantity of noise that the method can handle depends on the number of neighboring frames with good-quality information at the same spatial location. One way to increase even more the trade-off between admissible level of noise and quantity of good information consists in estimating the reliability of the information given, in each image or even at each pixel, and in weighting accordingly the associated energies:  $E = \sum_t w_t E^t$ . For instance, if a strong level of noise is detected globally in one image  $I(t)$ , one may multiply the corresponding energy  $E^t$  by a small reliability factor  $w_t < 1$ , in order to make it less influential than other frames. Such a reliability factor could be computed



for instance from preliminary image processing steps (for example, correlation between an image  $I(t)$  and its neighboring ones  $\{I(t') ; 0 < |t - t'| \leq \delta t\}$ ).

### E. Complexity

The precise theoretical worst case complexity depends on the graph cut algorithm used, and is the same as for usual single-image graph cuts, but with  $T$  times more nodes and edges. Denoting by  $N$  the number of nodes and  $M$  the number of edges, this worst case complexity is  $O(NM^2)$  for the Edmonds-Karp algorithm,  $O(N^2M)$  for the Diniz blocking flow algorithm, which goes down to  $O(NM \log(N))$  using dynamic trees. However, the computational time observed in practice is known to be much faster on typical image segmentation problems. We applied the binary graph-cut algorithm of Boykov and Kolmogorov [17], and we report a linear observed complexity with the total number of pixels  $T \times W \times H$  (see for instance Fig. 5 for experiments of section III-A). As a consequence, enforcing shape growth on a long sequence, or on complementary shorter bits of the same sequence, will take approximately the same time.

In the case of long sequences of big images, the memory space required may exceed the capacities of a computer. This is however not an issue, as there exist graph cut implementations for massive grids [23] meant for such cases, where all information is not stored in the memory at all times. This was not required for the experiments presented in this paper though.

### F. Rewriting as a Multi-Label Problem

We show now another point of view on sequence segmentation with growth constraint. The successive labels  $L_i(t)$  of a given pixel  $i$  over time might change only once, and only from 0 (background) to 1 (foreground object). Hence, this vector of labels  $L_i(t)$  is of the form  $(0, 0, \dots, 0, 1, \dots, 1, 1)$  and can be represented by just the time index  $\tau_i$  of the first 1, *i.e.* the earliest time at which the pixel starts belonging to the object. This time  $\tau_i$  is in  $[1, T+1]$ , with  $T+1$  meaning “never.” We have thus transformed a binary optimization problem on a sequence of images with shape growth constraint into a multi-label problem defined on one single image, without any constraint. This new problem can be expressed in the Markov Random Field form (1) with

$$V_i(\tau_i) := \sum_{t < \tau_i} V_i^t(0) + \sum_{t \geq \tau_i} V_i^t(1)$$

and

$$W_{i,j}(\tau_i, \tau_j) := \sum_{t < \min(\tau_i, \tau_j)} W_{i,j}^t(0, 0) + \sum_{\tau_i \leq t < \tau_j} W_{i,j}^t(1, 0) \\ + \sum_{\tau_j \leq t < \tau_i} W_{i,j}^t(0, 1) + \sum_{t \geq \max(\tau_i, \tau_j)} W_{i,j}^t(1, 1)$$

where  $(V_i^t)$  and  $(W_{i,j}^t)$  define the energy  $E^t$  at time  $t$ , and where sets of summation may be empty (note that any time  $t$  appears in exactly one summation set only). The submodularity of the binary interaction terms  $W^t$  in each frame implies

the submodularity of the multilabel interaction term  $W$ , *i.e.*

$$W_{i,j}(\tau_1, \tau_2) + W_{i,j}(\tau'_1, \tau'_2) \leq W_{i,j}(\tau_1, \tau'_2) + W_{i,j}(\tau'_1, \tau_2)$$

for all labels satisfying  $\tau_1 \leq \tau'_1$  and  $\tau_2 \leq \tau'_2$  (proof in Appendix). Thus, this energy can be minimized globally efficiently with now standard techniques (see [24]). Note that in the particular case where interaction terms  $W^t$  do not depend on  $t$ , the interaction term  $W$  of this multi-label energy above can be rewritten as a convex function  $g$  of  $(\tau_i - \tau_j)$ , and then Ishikawa’s construction [20] can be applied. It turns out that the graph built this way is precisely the graph that we built in our initial binary multi-frame problem. Our initial formulation is however more flexible, in that interaction terms can depend on  $t$ , and more natural, in that inclusion constraints can easily be enforced in spatial or/and time subregions only, while this would not be expressible with the multi-label formulation.

## III. EXPERIMENTAL RESULTS

We applied the proposed method to three different applications, in order to validate its use to enforce shape shrinkage or growth, for 2D or 3D image sequences:

(Application 1) We impose shrinkage constraint for a 2D sequence, to segment a melting multiyear ice from a time series of satellite measurements.

(Application 2) We enforce shape growth for a 2D sequence, to segment burned areas from optical satellite images.

(Application 3) We use growth constraint for 3D image sequences, to segment growing brain tumors from multimodal MRI volumes.

The performance of our framework with monodirectional links, [*Mono=const*], is compared with other graph-cut-based methods:

- [*w/o*]: Graph cut with no temporal links, *i.e.* independent segmentation of each frame.
- [*Feedforward*]: After segmenting the first frame with graph cut approach, foreground/background pixels are marked as seeds with infinite unary costs in the next frame for enforcing shape growth/shrinkage.
- [*Bi=const*]: Smoothing by introducing bidirectional temporal links, *i.e.* links in both directions (from  $t$  to  $t+1$  and from  $t+1$  to  $t$ ), with a constant weight (finite or infinite).
- [*Bi=variable*]: As proposed in [9], bidirectional temporal links are computed based on intensity differences between pixels in successive image frames, *i.e.* in the same way as spatial links.

For comparison between the methods we used the Dice score [25],  $D = (2|\hat{M} \cap M|)/(|\hat{M}| + |M|)$ , where  $\hat{M}$  and  $M$  are manually and automatically segmented foreground regions, respectively.

### A. Application 1: Melting Sea Ice in Satellite Images

The melting of sea ice is correlated to the increases in sea surface temperature and to associated climatic changes. Thus it is very important to monitor sea ice evolution and to develop methods for automated analysis of satellite measurements. Previous works on ice floe segmentation attempted to extract

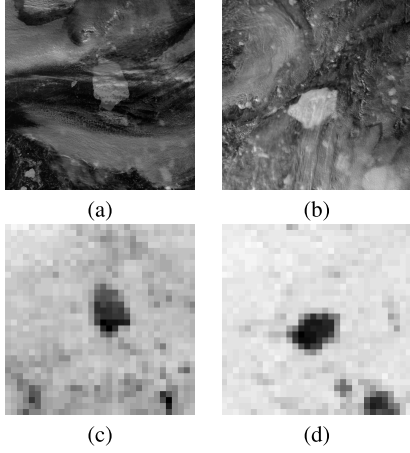


Fig. 2. (a, b) Reprojected MODIS images captured on the 244th and 252th days, respectively. (c, d) Corresponding AMSR-E images.

temporal information by using ice percentages, area and shape parameters of the ice floes at the previous time moments as prior information to segment the floe at the next time moment [5], [26]. These feed-forward approaches were unable to accurately estimate melting ice profiles because of the low signal-to-noise ratio and of the lack of contrast in satellite data, producing area estimations that are noisy in time, while a multiyear ice floe can actually only melt in the summer period.

We aimed at segmenting a multiyear ice floe from a 45-day sequence (summer period from the 227th to 271th day of 2008) of measurements over the polar regions by two Aqua satellite sensors: Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E, 89 GHz, 6.25 km spatial resolution) and Moderate-Resolution Imaging Spectroradiometer (MODIS, band 1, 0.620–0.670  $\mu\text{m}$ , 250 m spatial resolution). Fig. 2 shows MODIS and AMSR-E images at two time moments from the considered data set. The floe was tracked from the AMSR-E data, where multiyear ice has a low microwave emissivity (dark area in Fig. 2), and is in this way distinguishable from clouds and younger ice which has a higher emissivity (white area in Fig. 2). However, the low spatial resolution of these data does not allow to quantify the ice floes areas accurately. In accordance with the tracking measurements, a time series of  $T = 75$  MODIS and upscaled AMSR-E images with the ice floe was built, with spatial dimensions of  $800 \times 800$  pixels. We denote each MODIS image by  $I^t$ , and each upscaled AMSR-E image smoothed by Gaussian filter by  $A^t$ ,  $t = 1, \dots, T$ .

The objective is to compute  $T$  segmentation maps  $L^t$ , where each pixel  $(x, y)$  has label  $L^t_{(x,y)} = 1$  if it belongs to the floe at time  $t$ , and 0 otherwise. In order to apply the proposed method with a shrinkage constraint to the selected time series, the images must be aligned, so that the property that the floe in the image  $I^{t+1}$  is included in the floe of the image at the previous time moment  $I^t$  can be expressed directly in terms of pixel locations.

$$\forall t, x, y, \quad L^t_{(x,y)} = 1 \implies L^{t+1}_{(x,y)} = 1.$$

For this purpose, we estimated a reliable region of the foreground (*i.e.*, a region which can be considered with high probability as a part of the floe),  $R_F$ , and a reliable region of the background,  $R_B$ , from the AMSR-E images, where a multiyear ice floe is darker than water, young ice and clouds. From the AMSR-E tracking measurements, we derived a foreground seed point and an approximate area of the floe for each time moment  $t$ . We then grew a reliable region of the floe  $R_F$  in each  $A^t$  from the foreground seed point, until it reached approximately half of the size of the floe. Another region,  $\overline{R}_B$ , was grown in  $A^t$  from the same foreground seed point, but until an area half larger than the approximate floe size. The complementary region  $R_B$ , of pixels not in  $\overline{R}_B$ , was considered with high probability as part of the background.

We then computed the histograms of the intensities  $I^t$  of the floe,  $p^t(I|F)$ , and of the background,  $p^t(I|B)$ , respectively, and a map of floe probabilities as

$$p^t(F|I) = \frac{p^t(I|F)P^t(F)}{p^t(I|F)P^t(F) + p^t(I|B)P^t(B)}, \quad (4)$$

$$P^t(B) = \frac{A^t - \min_{x,y} A^t_{(x,y)}}{\max_{x,y} A^t_{(x,y)} - \min_{x,y} A^t_{(x,y)}}, \quad P^t(F) = 1 - P^t(B).$$

The quantities above depend on the pixel location. The images  $I^t$  were aligned by exhaustive searching over rigid motions (rotations and translations) to maximize the correlation between maps of foreground probabilities at the current and previous moments. This search is largely affordable given the low frame rate. We computed potentials and interaction terms between neighboring pixels as:

$$V_i^t(1) = -\ln[p^t(F|I)], \quad V_i^t(0) = -\ln[p^t(B|I)], \quad (5)$$

$$W_{i,j}^t = \delta_{L_i \neq L_j} \beta \exp \left[ -\frac{(I_i^t - I_j^t)^2}{2\sigma^2} \right], \quad (6)$$

where  $\sigma$  is a standard deviation of  $I^t$ ,  $\beta$  is a parameter that controls the importance of the spatial interaction energy term. We found experimentally that setting  $\beta = 2$  yields robust results. The proposed method was applied with monodirectional temporal links to enforce floe shrinkage, as described in Sec. II. We performed several experiments, with different values of the constant  $w$  standing for the temporal link weight, from 0.25 to  $\infty$ . The results [*Mono* = 0.25... $\infty$ ] are compared with those obtained with other graph-cut-based approaches (listed in the beginning of Sec. III) in Table I and in Fig. 3–4. Both graph-cut with no temporal links and feedforward approaches show the worst performances, and prove to be not well suited for segmenting such noisy data sets. When a feedforward method encounters a frame with the part of the floe obscured by clouds and thus undistinguishable from the background, it segments only the visible part of the foreground, and then is trapped in a non-sense segmentation for the rest of future times. The method using gradient-based temporal links [*Bi* = *variable*] [9] also yields poor segmentation accuracies, because it is sensitive to both noise and variation of foreground/background intensities in consecutive frames.

TABLE I  
RESULTS FOR APPLICATION 1. MEAN AND STANDARD DEVIATIONS OF THE DICE SCORES FOR THE PROPOSED METHOD  
[ $Mono = \infty$ ] AND GRAPH-CUT-BASED APPROACHES USED FOR COMPARISON

Method	Feedforward	w/o	Bi = variable	Bi = 16	Mono = $\infty$
Dice score	.554 $\pm$ .128	.933 $\pm$ .099	.958 $\pm$ .048	.978 $\pm$ .007	.980 $\pm$ .007

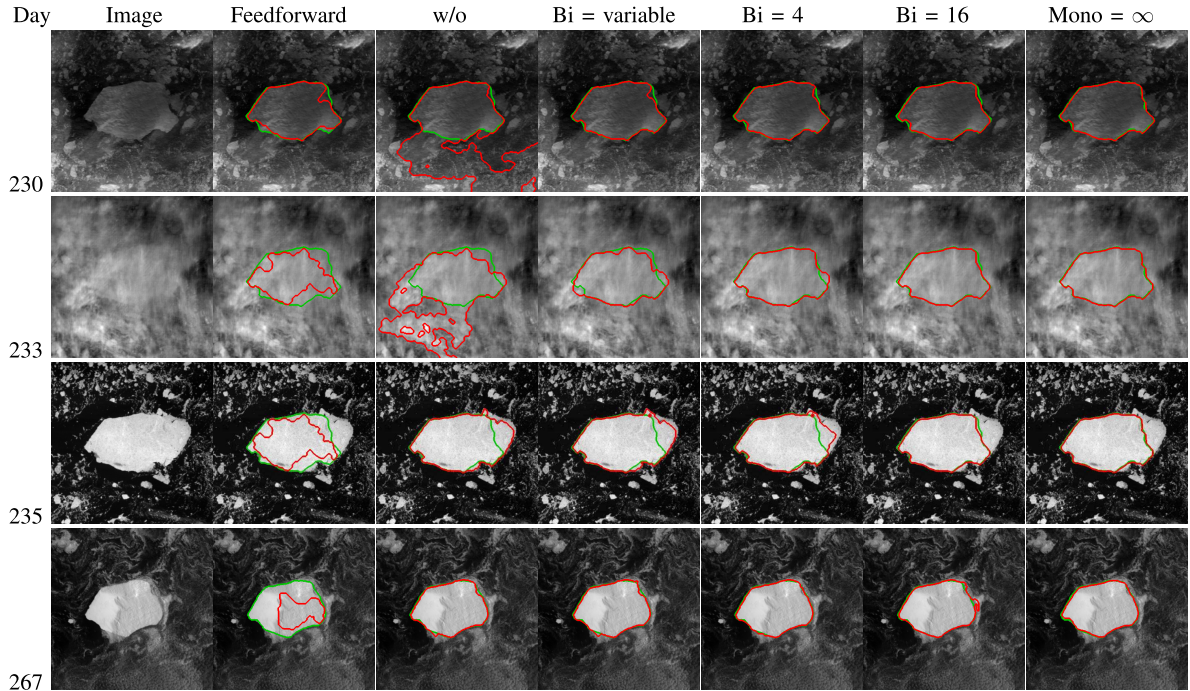


Fig. 3. (From top to bottom) Aligned images and segmentation contours (red) for four time moments (days 230, 233, 235 and 267, respectively) computed by the graph-cut methods: (From left to right) Aligned image, [Feedforward], [w/o], [Bi = variable], [Bi = 4], [Bi = 16], proposed method with monodirectional infinite links. Manual segmentation is shown in green. The rightmost part of the white area in the third row is not part of the object, but another ice floe who temporarily collided.

We explain in Fig. 3–4 the advantage of using monodirectional infinite links *versus* bidirectional links in the temporal dimension. Bidirectional edges with low values of  $w$  enforce only smoothness of variation of the contour in time, and yield segmentation errors in the case of low foreground/background contrast. For example, in the third image of Fig. 3 (day 235), the floe of interest collided temporarily with another ice floe. When using a weak smoothness constraint (see segmentation contour [Bi = 4]), the small encountered floe collided with the floe of interest during a certain number of consecutive frames would be considered as a part of the foreground. Enforcing more smoothness in space-time to avoid this has the undesirable effect of smoothing the foreground shape, so that the segmented foreground area is lower (underestimated) than the ground-truth for the first frames, and higher (overestimated) for the last frames (see results [Bi = 16] in Fig. 3 and-4(b)). With the increase of  $w$ , the estimated foreground tends to the constant shape for all time moments, and the Dice score decreases.

When the proposed shrinkage constraint is used instead, the segmentation accuracy increases with  $w$ , and  $w = \infty$  yields results with monotonous shrinkage of the shape area (see Fig. 4(a)). Moreover, the proposed method copes well

with rapid shrinkage events, without underestimating preceding images, or overestimating the event itself at onset. Another advantage of using monodirectional infinite links is that there are no additional parameters to quantify temporal coherency.

Fig. 5 depicts the computational time for the proposed graph-cut-based optimization as a function of the number of frames. The total computational time grows linearly with the number of frames, and is approximately twice the time that would be taken by the independent segmentation of each frame.

#### B. Application 2: Growing Burned Areas in Satellite Observations

Biomass burning has a significant impact on the Earth’s climate system. Satellite remote sensors acquire data for the continuous monitoring of burned areas at both regional and global scales. Thus, there is a need to develop methods for automated fire mapping. While most of the existing techniques for mapping burned areas analyze temporal evolution of each pixel in an image scene [27], recent studies have proved the advantage of considering spatial contextual classification for

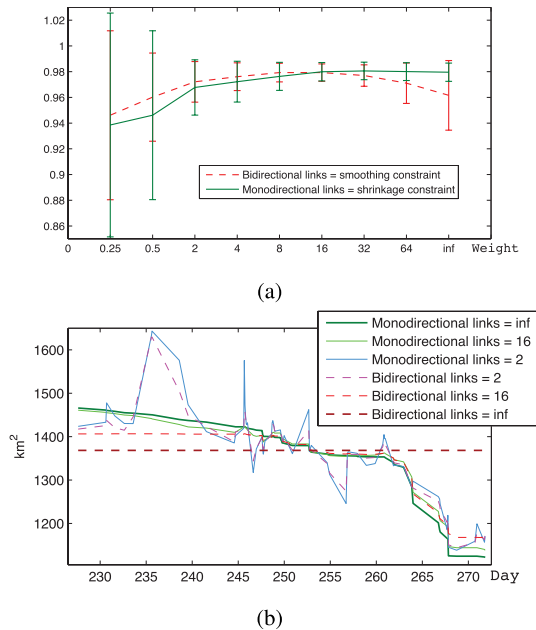


Fig. 4. Results for Application 1. (a) Mean and standard deviation for the dice score as a function of the temporal link's weight, when using mono- (green) and bidirectional (red) temporal links. (b) Area of a multiyear ice floe as a function of time, computed by using mono- and bidirectional links with different weights.

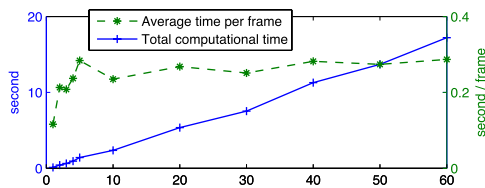


Fig. 5. Computational time for the proposed segmentation of the ice floe set as a function of the number of frames.

accurate fire classification [28], [29]. Both works [28], [29] map fires from MODIS data by detecting and classifying persistent changes in a daily vegetation-index time series. Giglio *et al.* [29] exploit the closest fixed pixel's neighborhood to refine fire classification. Lewis [28] segments and analyzes change detection maps between two consecutive time moments. Manual post-processing is needed to correct classification errors, which are a consequence of either a cloud cover, or low contrast between burned and unburned areas.

In our study we analyzed two time series of Terra MODIS atmospherically-corrected Level 2G daily surface reflectance measurements over the tropical savannas in the Northern Australia ("MOD09GA" product, tile h31v10), each of the data sets being acquired during forty days of the dry season in September-October (days 244–283) of 2011 and 2013, respectively. Wildfires in this region of Australia are frequent and extensive. We used MODIS band 5 (1.240  $\mu\text{m}$ ) 500-m land surface reflectance data as they provide the highest burned-unburned separability and are largely insensitive to smoke aerosols [27]. Each time series comprised a set of

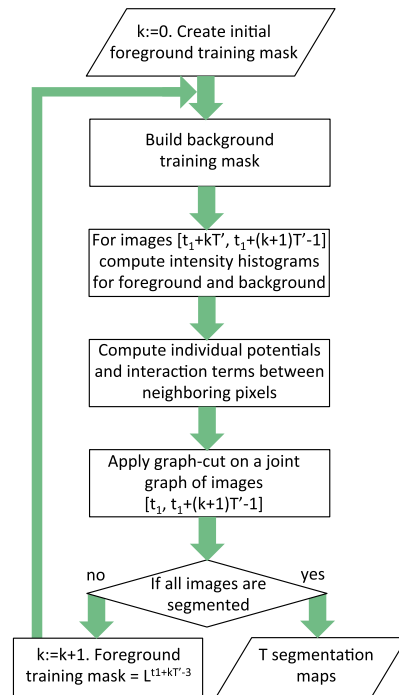


Fig. 6. Flowchart of the segmentation algorithm applied for Application 2.

$T = 40$  images with spatial dimensions  $W \times H = 400 \times 400$  pixels. Fig. 7(a) shows three images from each of the considered sets, where black pixels denote missing data (MODIS does not provide 100% daytime coverage of the terrestrial surface every day).

We used MODIS Collection 5.1 Direct Broadcast Monthly Burned Area Product (MCD64A1, see Fig. 7(b)) [29] for learning and testing the proposed method, *i.e.* for computing an initial histogram of burned areas and comparing fire maps, respectively. The MCD64A1 product contains fire classification maps, where each pixel is associated with either an estimated day of burn, or an unburned flag, or an unmapped flag due to insufficient data. These maps are computed by applying the approach from [29] on two 500-m MODIS channels coupled with 1-km MODIS active fire observations.

We segmented each of the considered image time series  $[t_1, \dots, t_T]$  by applying the following iterative procedure (the flowchart is shown in Fig. 6):

*Initialization:*  $k := 0$ . The initial training mask of burned areas  $R_k^B$  is built using MCD64A1 product, by selecting the pixels burned during the days  $[t_1 - D, t_1 - 1]$ . This mask can also be created based on ground observations of the considered area on the day  $(t_1 - 1)$ .

1. The training mask of unburned areas  $R_k^U$  is constructed by dilating  $R_k^B$  with a disk of radius  $r$  [30] and selecting the complementary of the resulting image.
2. For a subset of  $T'$  images  $t = [t_1 + kT', t_1 + (k+1)T' - 1]$ , intensity histograms of the MODIS band 5 for burned  $p^t(I|B)$  and unburned  $p^t(I|U)$  areas are computed, using the masks  $R_k^B$  and  $R_k^U$ , respectively. If the data for some pixels is missing, these pixels are not considered when computing histograms.



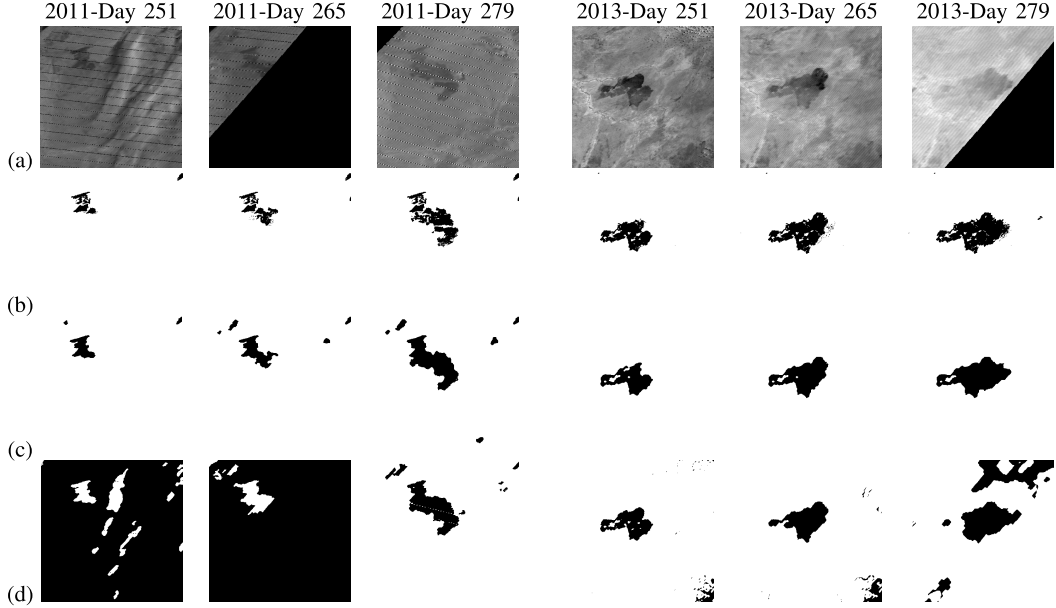


Fig. 7. Two time series of MODIS images acquired in 2011 and 2013 (left and right, respectively). (a) MODIS band 5 images for three time moments: days 251, 265 and 279. Black pixels denote missing data. (b) Maps from the MCD64A1 product of areas burned during the days 213–251, 213–265 and 213–279, respectively (white pixels = burned areas). (c-d) Segmentation maps for images (a) computed by: (c) the proposed method with growth constraint; (d) [w/o] method with no temporal constraints.

- For the images  $[t_1 + kT', t_1 + (k+1)T' - 1]$ , potentials are computed, assuming equal priors  $p^t(B) = p^t(U) = 1/2$ :

$$V_i^t(1) = -\ln[p^t(B|I_i^t)] = -\ln \left[ \frac{p^t(I_i^t|B)}{p^t(I_i^t|B) + p^t(I_i^t|U)} \right], \quad (7)$$

$$V_i^t(0) = -\ln[p^t(U|I_i^t)] = -\ln \left[ \frac{p^t(I_i^t|U)}{p^t(I_i^t|B) + p^t(I_i^t|U)} \right]. \quad (8)$$

If data  $I_i^t$  is missing for pixel  $i$  at time  $t$ , we set  $V_i^t(1) = V_i^t(0) = 0$  (no prior). Interaction terms are computed using Eq. 6.

- The graph-cut optimization is applied on a joint graph of the images  $[t_1, t_1 + (k+1)T' - 1]$ , yielding  $(k+1)T'$  segmentation maps.
- If the whole set of  $T$  images is segmented, exit the algorithm. Otherwise:  $k := k+1$ . The segmentation map  $L^{t_1+kT'-3}$  is used as the new training mask of burned areas  $R_k^B$ . We do not use the segmentation result of the last frame ( $t_1 + kT'$ ), because extreme frames benefit from less information from neighboring frames, and are therefore more subject to segmentation errors. Go to step 1.

We applied the described algorithm with the parameters empirically set as  $D = 31$ ,  $r = 20$ ,  $\beta = 2$  and  $T' = 20$ , and with different temporal regularizations, *i.e.*, [w/o], [Mono = 0.25...∞] and [Bi = 0.25...∞]. Neither [Feedforward] nor [Bi = variable] methods are suited for segmenting image sequences with missing data. Fig. 8(a, c) gives the resulting dice scores for both data sets. We find that

a weak temporal regularization (both bi-/monodirectional) outperforms a segmentation without temporal constraint. Increasing the bidirectional temporal regularization towards high values, however, decreases the performance. On the opposite, introducing monodirectional infinite links to impose shape growth yields the most accurate results. As can be seen from Fig. 7 and 8(b, d), the proposed approach achieves comparable results with the method [29] by using only one MODIS channel and no post-processing, while [29] used two MODIS bands coupled with 1-km MODIS active fire observations and post-processing. Furthermore, the new method copes better with the missing or noisy data thanks to the introduced spatio-temporal graph.

### C. Application 3: Growing Tumor in 3D Medical Scans

Glioma is the most frequent primary tumor of the brain. The tumor is known to grow steadily, and lesions are evaluated with respect to volume change in different magnetic resonance image (MRI) modalities. In our experiment we evaluated a set of 760 multimodal image volumes – each comprising T1 MRI, contrast-enhanced T1 MRI (T1c), T2 MRI, and T2 FLAIR MRI – acquired from ten patients initially diagnosed with low grade glioma. The time series have 3–14 time points, with 3–6 months in between any two acquisitions. All image volumes were rigidly registered and three 2D slices intersecting with the tumor center were manually annotated through an expert in every volume, representing an approximate truth. Full 3D segmentations for images of each individual time point were obtained using a generative model for multimodal brain tumor segmentation [31]. This algorithm models the lesion with a latent atlas class [15]

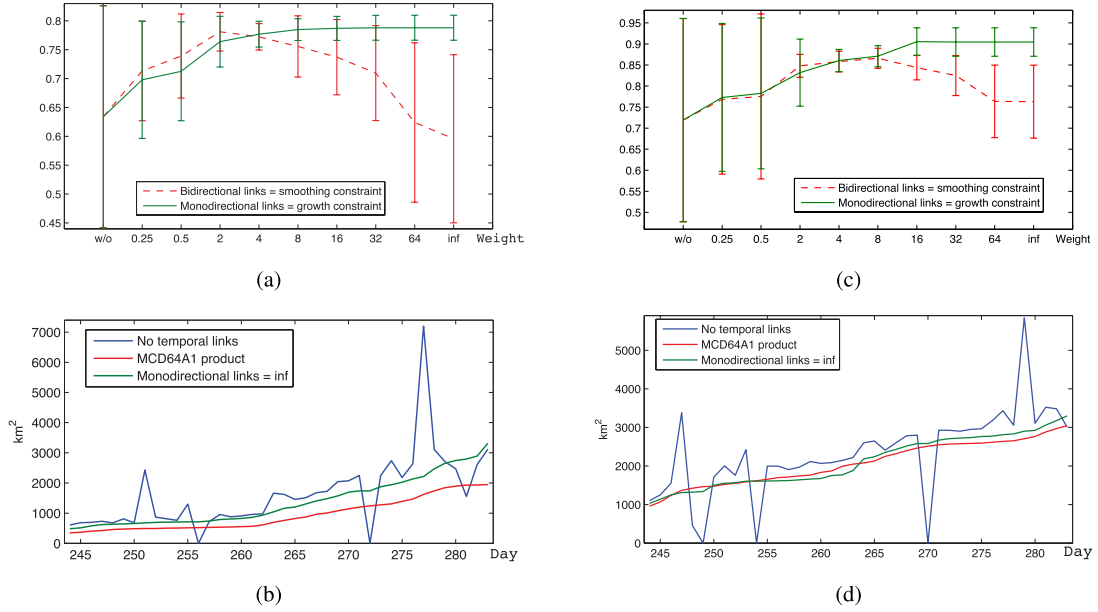


Fig. 8. Results for Application 2, for two MODIS time series acquired in 2011 (a–b) and 2013 (c–d), respectively. (a, c) Mean and standard deviation for the dice score as a function of the temporal link’s weight, when using mono- (green) and bidirectional (red) temporal links. (b, d) Burned area as a function of time, when using no temporal links (blue), monodirectional infinite links (green) and MCD64A1 product (red).

amending the tissue atlas of the standard EM segmenter [32], [33]. We applied it to each multimodal data set of each time point in an independent fashion. The segmentation model delineates the lesion individually in each modality. It assumes that changes of the core (visible in T1c) will occur within the larger edema regions (visible in T2 or FLAIR) and, hence, to only have class transitions from *healthy* to *edema* and from *edema* to *core*. As the tumor grows steadily, we can assume that negative volume changes stem from imaging artifacts, such as local intensity changes, a common problem in MRI. To this end we model the tumor to be either stable, in this case regularizing the segmentation along time and suppressing noise, or to expand in volume between any two time points.

We identified the foreground label  $F$  with tumor (*edema* and *core*) and background  $B$  with healthy tissue. Then the potential  $V_i^{s,t}(L_i^{s,t})$  of label  $L_i^{s,t}$  at voxel  $i$ , time point  $t$ , and imaging sequence  $s$  is equal to  $V_i^{s,t}(0) = p_{s,t}(F|I^{s,t})$  and  $V_i^{s,t}(1) = p_{s,t}(B|I^{s,t}) = 1 - p_{s,t}(F|I^{s,t})$ . The tumor probabilities were calculated from image volumes  $I$  using the generative model [31], and we identified tumor subclasses with  $p(F|I^{s=T1,t})$  for *core*, and  $p(F|I^{s=T2,t})$  with *edema*. We modeled the 3D spatial constraints through a 26 neighborhood ( $\mathcal{N}$ ) linking the central voxel with all its immediate neighbors. Interaction terms  $W_{i,j}^{s,t}(L_i^{s,t}, L_j^{s,t})$  between neighboring voxels in each sequence  $s \in [T1, T2]$  are computed from the channel-specific intensity differences as

$$\delta_{L_i \neq L_j} \beta \frac{\alpha(i, j)}{\alpha_{\text{tot}}} \exp\left(-\left(\frac{I^{s,t}(i) - I^{s,t}(j)}{A}\right)^2\right)$$

with  $\beta = 0.5$ ,  $\alpha(p, q) = 1/\text{distance}(p, q)$ ,  $\alpha_{\text{tot}} = \sum_{q \in \mathcal{N}(\text{pixel } p)} \alpha(p, q)$  and  $A = 1/3 (\max I^{s,t} - \min I^{s,t})$ . We impose growth constraint in 3D+t as explained in

Sec. II-B, and inclusion constraints as in Sec. II-C: the foregrounds in T1 and T1c modalities are required to be included in the one of T2, which is included in the one of FLAIR.

In our test we first segment the images from each time point independently, and calculate Dice scores averaged over all images of a time series as an estimate of the baseline performance. Then we test different regularizations in time, *i.e.*,  $[w/o]$ ,  $[Mono]$  and  $[Bi]$ , calculate Dice scores, and compare them against the baseline results from the unregularized segmentations. Fig. 9 shows results for two exemplary time series, and Fig. 10(a) reports differences between Dice scores of regularized segmentations and baseline segmentations. In this experiment “weak” regularizations refer to small regularization parameters ( $w \ll 1$ ), while “strong” regularizations refer to  $w \gg 1$ , representing the infinite mono- and bi-directional links. Fig. 10(a) shows results for all ten time series. As for the previous two applications, both bi-/monodirectional temporal regularization with low values of  $w$  yields better accuracies when compared to the results without temporal links. Enforcing more smoothness with bidirectional links decreases segmentation accuracies, while introducing monodirectional “growth” regularization through infinite links improves performance (Fig. 10(a)). The volume-time graphs (Fig. 10(b)) of the segmented tumor structures show as very regular pattern (red), even being smoother than the manual segmentation (green). Moreover, the  $\log(\text{volume})$ -time graph (Fig. 10(b), right) shows the exponential growth of the tumor core (dotted lines) that is associated with this disease [34], for this patient indicating a rapid tumor progression starting at about day 500. Longitudinal image segmentation, as obtained for the ten time series in our test, can be further analyzed, for diagnosis and treatment monitoring, *e.g.*, through algorithms estimating the speed of the tumor outlines under

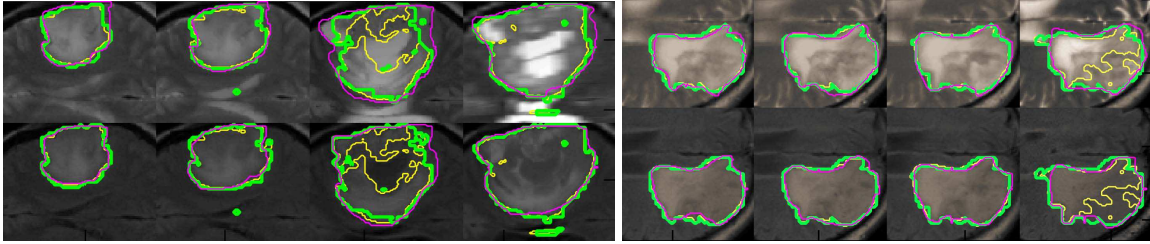


Fig. 9. Two time series of T2 and FLAIR MR image volumes, each of which acquired about 3-6 months apart. The left case is rapidly growing between the second and fourth scene, the right case displays intensity modifications in the last scene, leading to a suboptimal performance of the initial multimodal segmentation (yellow) [31]. The proposed multi-temporal segmentation with growth constraints (green) delineates areas similar to the manual evaluation (magenta), being more robust against intensity variations of the MR images. It does not smooth out outlines of rapidly growing tumors as conventional bi-directional temporal constraints would do.

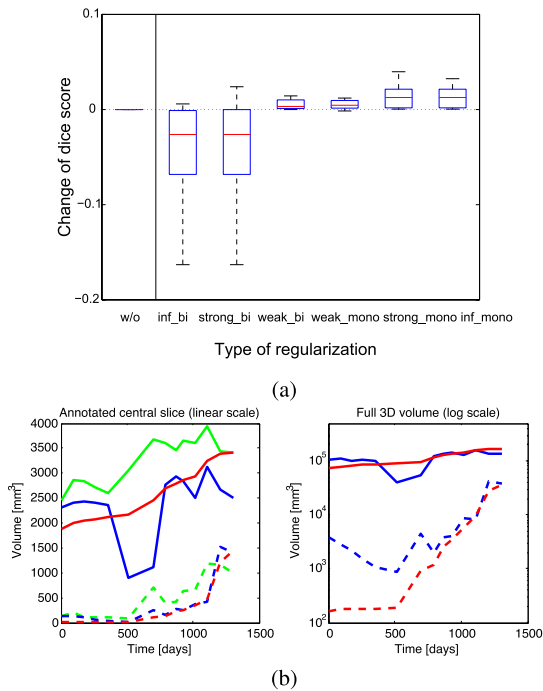


Fig. 10. Results for Application 3. (a) Boxplots report changes in the average segmentation performance of the ten image sequences when testing different regularization approaches. (The box indicated quartiles, the whiskers outliers.). Using a strong monodirectional regularization acting as a growth prior yields the best results. (b) Volume-time plot for a patient with 14 observations (cmp. video in supplement). Solid lines indicate *edema*, dashed indicate tumor *core* that starts growing with constant rate at around day 500. The segmentation with growth constraint (red) returns results similar to the manual segmentation (green). Segmentations obtained by evaluating image volumes of each time point individually (blue) show significant variation, even obscuring the overall trend.

anatomical constraints [35]. Extension of the current 5D segmentation could integrate this speed estimation, or extend the multimodal tumor segmentation [31] for longitudinal data sets.

#### IV. CONCLUSIONS AND FUTURE WORK

In this work, we addressed the problem of shape segmentation in 2D and 3D sequences of very noisy/low-contrast images, where shapes monotonously grow or shrink in time. In order to enforce shape growth or shrinkage, we proposed a new graph-cut-based method for computing the

globally-optimal spatio-temporal segmentation satisfying that constraint. The main idea was to introduce monodirectional infinite links between pixels at the same spatial locations in successive image frames, which prohibit a shape to shrink or grow over time, and then to perform a graph cut optimization on the constructed graph. The limitation of the proposed method is that it can be applied only to a time series of images on the same scale and perfectly aligned with respect to the foreground object, so that in the case of shape growth the foreground at the moment  $(t + 1)$  contains all the foreground pixels at the previous moment  $t$ , sharing the same spatial locations. Thus, if the foreground object moves over time, images must be aligned before applying the new graph-cut-based technique. We also demonstrated the possibility to impose inter-sequences object inclusion constraints by adding directed infinite links to the joint graph associated to all sequences.

We validated the performance of the proposed approach for the segmentation of growing (burned areas) and shrinking (ice floe) shapes from 2D time series of satellite images, and for the segmentation of growing 3D tumor volumes from MR sequences. The new method proved to be robust to important noise and low contrast, and to cope well with missing data. Moreover, it showed linear complexity in practice, so that globally optimal shape-consistent segmentations of image time series are obtained in a matter of seconds.

We plan to apply the proposed framework to other applications, such as the segmentation of wrinkles during the cosmetological treatment. We are also currently extending our work on tumor segmentation to the case of several medically-motivated classes  $C_i$ , expressing tumor evolution stages, that satisfy an inclusion constraint  $C_i \subset C_{i+1}, \forall i$ , considering additional relations between tumor substructures.

#### APPENDIX

##### SUBMODULARITY PROOF

*This proof is better viewed in colors.*

$W_{i,j}$  can be written as a sum of terms over time  $t$ , which is, if  $\tau_i \leq \tau_j$ :

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{cases} W_{i,j}^t(0, 0) & \text{if } t < \min(\tau_i, \tau_j) & (A) \\ W_{i,j}^t(1, 0) & \text{if } \tau_i \leq t < \tau_j & (B) \\ W_{i,j}^t(1, 1) & \text{if } t \geq \max(\tau_i, \tau_j) & (D) \end{cases}$$

and

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{cases} W_{i,j}^t(0, 0) & \text{if } t < \min(\tau_i, \tau_j) \\ W_{i,j}^t(0, 1) & \text{if } \tau_j \leq t < \tau_i \\ W_{i,j}^t(1, 1) & \text{if } t \geq \max(\tau_i, \tau_j) \end{cases} \quad (A)$$

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{cases} W_{i,j}^t(0, 1) & \text{if } \tau_j \leq t < \tau_i \\ W_{i,j}^t(1, 1) & \text{if } t \geq \max(\tau_i, \tau_j) \end{cases} \quad (C)$$

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{cases} W_{i,j}^t(0, 0) & \text{if } t < \min(\tau_i, \tau_j) \\ W_{i,j}^t(1, 1) & \text{if } t \geq \max(\tau_i, \tau_j) \end{cases} \quad (D)$$

otherwise.

We can thus represent  $W_{i,j}(\tau_i, \tau_j)$  by the sequence of cases A, B, C or D chosen for each  $t$ :

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{array}{c} \tau_i \qquad \tau_j \\ \hline \text{AAAAA BBBBDDDDDDDDDD} \end{array} t$$

when  $\tau_i \leq \tau_j$ , and otherwise :

$$W_{i,j}(\tau_i, \tau_j) = \sum_t \begin{array}{c} \tau_j \qquad \tau_i \\ \hline \text{AAAAA CCCCCC DDDDDDD} \end{array} t$$

Now, to represent the submodularity condition, let  $\tau_i, \tau_j, \tau'_i, \tau'_j$  be any times, satisfying  $\tau_i \leq \tau'_i$  and  $\tau_j \leq \tau'_j$ . We will suppose moreover that  $\tau_i \leq \tau_j$ : otherwise consider  $W_{j,i}(\tau_j, \tau_i)$  instead and reverse names  $i$  and  $j$ . Three cases are possible:

- 1)  $\tau_i \leq \tau'_i \leq \tau_j \leq \tau'_j$
- 2)  $\tau_i \leq \tau_j \leq \tau'_i \leq \tau'_j$
- 3)  $\tau_i \leq \tau_j \leq \tau'_j \leq \tau'_i$

In case (1), the terms of the submodularity condition write:

$$\begin{array}{l} W_{i,j}(\tau_i, \tau_j) : \text{AAA BBBBDDDDDDDDDD} \\ W_{i,j}(\tau'_i, \tau'_j) : \text{AAAAAABBBBBDDDD} \\ \hline W_{i,j}(\tau'_i, \tau_j) : \text{AAAAAABBBBBDDDDDD} \\ W_{i,j}(\tau_i, \tau'_j) : \text{AAA BBBBDDDDDDDDDD} \end{array} t$$

and the submodularity condition is satisfied if the sum of the two first rows is less than or equal to the sum of the two last rows. It turns out that in the particular case above, the sums are equal for each instant  $t$ , and consequently the submodularity condition is satisfied as being an equality.

In case (2), the terms of the submodularity condition write:

$$\begin{array}{l} \tau_i \qquad \tau_j \qquad \tau'_i \qquad \tau'_j \\ \hline W_{i,j}(\tau_i, \tau_j) : \text{AAA BBBBDDDDDDDDDD} \\ W_{i,j}(\tau'_i, \tau'_j) : \text{AAAAAABBBBBDDDD} \\ \hline W_{i,j}(\tau'_i, \tau_j) : \text{AAAAA CCCCCDDDDDD} \\ W_{i,j}(\tau_i, \tau'_j) : \text{AAA BBBBDDDDDDDDDD} \end{array} t$$

This time, the two sums, at any instant  $t$ , are equal if  $t < \tau_j$  or  $t \geq \tau'_i$ . For other values of  $t$ , that is to say when  $\tau_j \leq t < \tau'_i$ , the sum  $A+D$  is to be compared with  $B+C$ , *i.e.* the question is to compare  $W_{i,j}^t(0, 0) + W_{i,j}^t(1, 1)$  with  $W_{i,j}^t(0, 1) + W_{i,j}^t(1, 0)$ . As  $E^t$  is binary submodular by hypothesis, and by definition of binary submodularity in equation (2), we have  $A + D \leq B + C$ , and thus the sum of the two first rows is less than or equal to the sum of the two last ones, which means precisely  $W_{i,j}(\tau_i, \tau_j) + W_{i,j}(\tau'_i, \tau'_j) \leq W_{i,j}(\tau_i, \tau'_j) + W_{i,j}(\tau'_i, \tau_j)$ , and thus the submodularity condition for  $E$  is checked.

Finally, in case (3), the terms of the submodularity condition write:

$$\begin{array}{l} \tau_i \qquad \tau_j \qquad \tau'_j \qquad \tau'_i \\ \hline W_{i,j}(\tau_i, \tau_j) : \text{AAA BBBBDDDDDDDDDD} \\ W_{i,j}(\tau'_i, \tau'_j) : \text{AAAAA CCCCCDDDD} \\ \hline W_{i,j}(\tau'_i, \tau_j) : \text{AAAAA CCCCCDDDD} \\ W_{i,j}(\tau_i, \tau'_j) : \text{AAA BBBBDDDDDDDD} \end{array} t$$

The two sums are equal for all instants  $t < \tau_j$  or  $t \geq \tau'_j$ . When  $\tau_j \leq t < \tau'_j$ , the sum  $A + D$  is to be compared with  $B + C$ , as previously, and consequently the submodularity condition for  $E$  is checked again.

In all cases (1), (2) and (3),  $E$  is proven to be submodular, and this concludes the proof.

#### ACKNOWLEDGMENT

The authors would like to thank L. Giglio and D. Morton from the NASA GSFC, USA, for providing the MCD64A1 data and for useful discussions about fire classification, and A. Ivanoff from the NASA GSFC, USA, for discussions about multiyear ice floe analysis.

#### REFERENCES

- [1] O. Lézoray and L. Grady, *Image Processing and Analysis With Graphs: Theory and Practice* (Digital Imaging and Computer Vision), O. Lézoray and L. Grady, Eds. Boca Raton, FL, USA: CRC Press, 2012.
- [2] H. Winnemöller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1221–1226, Jul. 2006.
- [3] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," in *Proc. ACM SIGGRAPH*, Oct. 2007, p. 103.
- [4] S. Paris, "Edge-preserving smoothing and mean-shift segmentation of video streams," in *Proc. 10th Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2008, pp. 460–473.
- [5] Y. Tarabalka, L. Brucker, A. Ivanoff, and J. C. Tilton, "Shape-constrained segmentation approach for arctic multiyear sea ice floe analysis," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2012, pp. 4958–4961.
- [6] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Proc. 6th Int. Conf. Comput. Vis.*, Washington, DC, USA, Jan. 1998, pp. 1154–1160.
- [7] D. Dementhon, "Spatio-temporal segmentation of video by hierarchical mean shift analysis," in *Proc. Statist. Methods Video Process. Workshop (SMVP)*, 2002.
- [8] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph based video segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2141–2148.
- [9] R. Wolz *et al.*, "Measurement of hippocampal atrophy using 4D graphcut segmentation: Application to ADNI," *NeuroImage*, vol. 52, no. 1, pp. 109–118, Aug. 2010.
- [10] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr, "Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 295–313, Dec. 2002.
- [11] M. Rousson and N. Paragios, "Shape priors for level set representations," in *Proc. 7th Eur. Conf. Comput. Vis. (ECCV)*, May 2002, pp. 78–92.
- [12] P. Felzenszwalb, "Representation and detection of shapes in images," Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2003.
- [13] T. Schoenemann and D. Cremers, "Globally optimal image segmentation with an elastic shape prior," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–6.
- [14] T. Schoenemann and D. Cremers, "Globally optimal shape-based tracking in real-time," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–6.
- [15] T. Riklin-Raviv, K. V. Leemput, B. H. Menze, W. M. Wells, and P. Golland, "Segmentation of image ensembles via latent atlases," *Med. Image Anal.*, vol. 14, no. 5, pp. 654–665, Oct. 2010.
- [16] Y. Tarabalka, G. Charpiat, L. Brucker, and B. H. Menze, "Enforcing monotonous shape growth or shrinkage in video segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2013.



- [17] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [18] G. B. Dantzig and D. R. Fulkerson, "On the max-flow min-cut theorem of networks," *Ann. Math. Studies*, vol. 38, pp. 215–221, 1956.
- [19] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *J. ACM*, vol. 35, no. 4, pp. 921–940, Oct. 1988.
- [20] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1333–1336, Oct. 2003.
- [21] A. Delong and Y. Boykov, "Globally optimal segmentation of multi-region objects," in *Proc. 12th Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 285–292.
- [22] K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images—A graph-theoretic approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 119–134, Jan. 2006.
- [23] A. Delong and Y. Boykov, "A scalable graph-cut algorithm for N-D grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
- [24] J. Darbon, "Global optimization for first order Markov random fields with submodular priors," in *Proc. 12th Int. Workshop Combinat. Image Anal.*, Apr. 2008, pp. 229–237.
- [25] L. Dice, "Measure of the amount of ecological association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
- [26] D. Haverkamp and C. Tsatsoulis, "Using temporal information in an automated classification of summer, marginal ice zone imagery," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, May 1996, pp. 109–111.
- [27] D. P. Roy, P. E. Lewis, and C. O. Justice, "Burned area mapping using multi-temporal moderate spatial resolution data—A bi-directional reflectance model-based expectation approach," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 263–286, Nov. 2002.
- [28] J. B. Lewis, "Fire mapping for managers in North Australian savanna; An object orientated classification model for MODIS imagery," in *Proc. Austral. Remote Sens. Photogram. Conf. (ARSPC)*, Fremantle, Australia, 2004, pp. 1–12.
- [29] L. Giglio, T. Loboda, D. P. Roy, B. Quayle, and C. O. Justice, "An active-fire based burned area mapping algorithm for the modis sensor," *Remote Sens. Environ.*, vol. 113, no. 2, pp. 408–420, Feb. 2009.
- [30] P. Soille, *Morphological Image Analysis*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.
- [31] B. H. Menze, K. van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, "A generative model for brain tumor segmentation in multi-modal images," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Sep. 2010, pp. 151–159.
- [32] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, no. 4, pp. 429–442, Aug. 1996.
- [33] T. Kapur, W. E. L. Grimson, W. M. Wells III, and R. Kikinis, "Segmentation of brain tissue from magnetic resonance images," *Med. Image Anal.*, vol. 1, no. 2, pp. 109–127, Jun. 1996.
- [34] F. G. Blankenberg *et al.*, "The influence of volumetric tumor doubling time, DNA ploidy, and histologic grade on the survival of patients with intracranial astrocytomas," *Amer. J. Neuroradiol.*, vol. 16, no. 5, pp. 1001–1012, 1995.
- [35] E. Konukoglu *et al.*, "Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic Eikonal equations," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 77–95, Jan. 2010.



**Yuliya Tarabalka** (S'08–M'10) received the B.S. degree in computer science from Ternopil Ivan Pul'uj State Technical University, Ternopil, Ukraine, in 2005, the M.Sc. degree in signal and image processing from the Grenoble Institute of Technology (INPG), Grenoble, France, in 2007, and the Ph.D. degree in signal and image processing, and electrical engineering from INPG and the University of Iceland, Reykjavik, Iceland, in 2010.

She was a Researcher with the Norwegian Defence Research Establishment, Norway, from 2007 to 2008. From 2010 to 2011, she was a Post-Doctoral Research Fellow with the Computational and Information Sciences and Technology Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA. In 2012, she was a Post-Doctoral Research Fellow with the French Space Agency and Inria Sophia Antipolis-Méditerranée, Nice, France. She is currently a Researcher with the AYIN team, Inria Sophia Antipolis-Méditerranée. Her research interests are in the areas of image processing, pattern recognition, hyperspectral imaging, and the development of efficient algorithms. She is member of the IEEE Society and the IEEE Geoscience and Remote Sensing Society.



**Guillaume Charpiat** is a Researcher with Inria Sophia-Antipolis, Nice, France, in the Stars team-project. He studied mathematics and physics at the École Normale Supérieure (ENS), Paris, France, computer vision and statistical learning at ENS Cachan, Cachan, France, (M.Sc.), and theoretical physics in Paris. He received the Ph.D. degree in computer science from ENS, Inria Sophia-Antipolis, and École Nationale des Ponts et Chaussées, France, in 2006, affiliated with the doctoral school of the École Polytechnique, on the topic of distance-based

shape statistics for image segmentation with priors. He then held one year post-doctoral position with the Max-Planck Institute for Biological Cybernetics, Tübingen, Germany, on the topics of medical imaging (MR-based PET prediction) and automatic image colorization. Since then, he has been a Researcher with INRIA, Rennes, France. His main interests are computer vision, machine learning, and optimization.



**Ludovic Brucker** received the M.S. degree in physics from the University of Clermont-Ferrand, Clermont-Ferrand, France, in 2006, and the Ph.D. degree from the Laboratoire de Glaciologie et Géophysique de l'Environnement, Grenoble University/Centre National de la Recherche Scientifique, Grenoble, France, in 2009. He joined the NASA Goddard Space Flight Center, Greenbelt, MD, USA, in 2010. His research focuses on the investigation of climate evolution in polar and subpolar regions by

interpreting space-borne microwave measurements of snow-covered surfaces (i.e., sea ice, ice sheet, and terrestrial snowpacks). The goal of his work is to contribute to the comprehension of the relationships between both passive and active microwave air- and space-borne observations and snow/ice physical properties using modeling approaches to provide climate-related variables to the community for the satellite era. He has also participated in the International Polar Year in 2008 with a deployment to North Quebec, and has been deployed on the Antarctic and Greenland ice sheet.



**Bjoern H. Menze** is an Assistant Professor and the Head of the Image-based Biomedical Modeling and Computational Physiology Group at the Department of Computer Science, Technische Universität München, Munich, Germany, and a Visiting Research Scientist of the Asclepius team at the Inria Sophia-Antipolis, Nice, France. He is also a Research Affiliate of the Medical Vision group at CSAIL, Massachusetts Institute of Technology (MIT), Cambridge, MA. He received the M.Sc. degree in physics from Uppsala University, Uppsala,

Sweden, in 2002, the M.Sc. degree from Heidelberg University, Heidelberg, Germany, in 2004, and the Ph.D. degree in computer science from the Interdisciplinary Center for Scientific Computing, Heidelberg, in 2007. He subsequently moved to Boston, MA, USA, where he was a Post-Doctoral Researcher at Harvard University, Cambridge, Harvard Medical School, Boston, and MIT. He was then as a Researcher at Inria Sophia-Antipolis, Nice, France, and a Senior Researcher and Lecturer at ETH Zurich, Zurich, Switzerland. He is developing methods in medical image computing at the interface of computational pathophysiology, medical computer vision, and machine learning. In this, he focuses on applications in multimodal clinical neuroimaging and the patient-adaptive modeling of tumor growth. He organized workshops at MICCAI, NIPS, and CVPR in the fields of medical computer vision and neuroimage processing, served as a Guest Editor of *Medical Image Analysis*, and as a Program Committee Member of the International Conference on Medical Image Computing and Computer Assisted Intervention.