Segmenting the Uterus in Monocular Laparoscopic Images without Manual Input

Toby Collins, Adrien Bartoli, Nicolas Bourdel and Michel Canis

ALCoV-ISIT, UMR 6284 CNRS/Université d'Auvergne, Clermont-Ferrand, France

Abstract. Automatically segmenting organs in monocular laparoscopic images is an important and challenging research objective in computer-assisted intervention. For the uterus this is difficult because of high inter-patient variability in tissue appearance and low-contrast boundaries with the surrounding peritoneum. We present a new framework to accurately segment the uterus which is completely automatic, requires only a single monocular image, and does not require patient-specific prior knowledge such as a 3D model. Our main idea is to use a patient-independent uterus detector to roughly localize the organ, which is then used as a supervisor to train a patient-specific organ segmenter. The segmenter uses a physically-motivated organ boundary model designed specifically for illumination in laparoscopy, which is fast to compute and gives strong segmentation constraints. Our segmenter uses a lightweight CRF that is solved globally with a single graphcut. On a dataset of 228 patients our method obtains an average DICE score of 92.5%, and takes approximately one second per image on a standard desktop PC, without a GPU or much code optimisation.

1 Introduction and Background

The problem of segmenting organs in monocular laparoscopic images without any manual input is important yet unsolved for computer assisted laparoscopic surgery. This is challenging due to multiple factors including inter and intra-patient tissue appearance variability, low-contrast and/or ambiguous organ boundaries, texture inhomogeneity, bleeding, motion blur, partial views, surgical intervention and lens smears. In previous works a manual operator has been needed to identify the organ in one or more training images [3, 9]. From these images, models of patient-specific tissue appearance can be learned and used to segment the organ in other images. We present the first methodology to accurately segment an organ in laparosurgery *without any manual input*. Our solution is simple, fast and does not require separate training images, since training and segmentation is performed on the same image. We also do not require patient-specific prior knowledge such as a pre-operative 3D model. To use these models requires registration [9], where the model must be aligned to the image to give the segmentation (*i.e. segmentation-by-registration*). This shifts the problem burden to registration, which itself is hard to do automatically and reliably for soft organs and monocular laparoscopes [8]. Our approach uses recent work in patient-generic organ detection in laparoscopic images [11]. It was shown that the uterus can be reliably detected in an image *without* patient specific knowledge using a state-of-the-art 2D Deformable Part Model (DPM) detector [**?**,13] trained on a uterus image database. The problem of segmentation however was not considered, which is a fundamentally different problem.

For a given image our goal is to compute the binary label matrix $\mathcal{L}(\mathbf{x}) \in \{0, 1\}$ where $\mathcal{L}(\mathbf{x}) = 1$ means pixel $\mathbf{x}$ is on the organ and $\mathcal{L}(\mathbf{x}) = 0$ means it is not. We refer to these as the foreground and background labels respectively. We propose an energy minimisation-based approach to solve $\mathcal{L}$ that incorporates information from the DPM detector to define the energy function. The function is a submodular discrete Conditional Random Field (CRF) that is globally optimised with a *single* graphcut. Much inspiration has come from graphcut-based interactive image segmentation methods [2, 12, 10] where manual strokes or bounding boxes are used to guide the segmentation. Instead of user interaction, we do this using information from the DPM detector, which in contrast to user interaction information is inherently uncertain. A second major difference is that most graphcut-based methods use the contrast-sensitive Ising prior from [2], which encourages segmentation boundaries at strong intensity step-edges (*i.e.* points with strong first-order intensity derivatives). However step-edges do not accurately model the projection of an organ's boundary in laparoscopic images. We show that far better segmentations are obtained using a physically-motivated *trough-sensitive Ising prior*, which is computed from the response of a positive Laplacian of Gaussian (LoG$^+$) filter (*i.e.* a LoG filter with negative responses truncated to zero). This encourages segmentation boundaries at points with strongly positive *second-order* intensity derivatives.

2 Methodology

Segmentation pipeline. The main components of our method are illustrated in Fig. 1, which processes an image in five stages. In stage 1 we detect the presence of the organ with the DPM uterus detector from [11]. We take the detector's highest-confidence detection and if it exceeds the detector's threshold we assume the organ is visible and proceed with segmentation. The highest-confidence detection has an associated bounding box $\mathcal{B}$, which gives a rough localisation of the organ. In stage 2 we use $\mathcal{B}$ to train rough appearance models for the organ and background, which are used in the CRF as colour-based segmentation cues. Similarly to grabcut we use Gaussian Mixture Models (GMMs) with parameters denoted by θ_{fg} and θ_{bg} respectively. However unlike grabcut, we do not iteratively recompute the GMM parameters and the segmentation. This is because with our organ boundary model, the first segmentation is usually very accurate even if the appearance parameters are not. This has the clear advantage of reduced computation speed since we only perform one graphcut.

In stage 3 we use the detection's bounding box to extract a Region Of Interest (ROI) $\mathcal{R}$ around the organ, and all pixels outside $\mathcal{R}$ are labelled background. This reduces computation time because pixels outside $\mathcal{R}$ are not included in the CRF. One cannot naively set $\mathcal{R}$ as the detection's bounding box because there is no guarantee that it will encompass the whole organ, as seen in Fig. 2, bottom row. We then normalise $\mathcal{R}$ to have a default width of 200 pixels, which gives sufficiently high resoluiton to accurately segment the uterus. The normalisation step is important because it means the CRF energy is independent of the organ's scale. Therefore we do not need to adapt any parameters depending on the organ's physical size, distance to the camera or camera focal length. In stage 4 we construct the CRF which includes information from three important sources. The first is colour information from the foreground and background colour models. The second is edge information from the response of a LoG^+ filter applied to $\mathcal{R}$. The third are spatial priors that give energy to pixels depending on where they are in $\mathcal{R}$. All of the CRF energy terms are submodular which means it can be solved globally and quickly using the maxflow algorithm. In practice this takes between 20-50ms with a standard desktop CPU implementation.

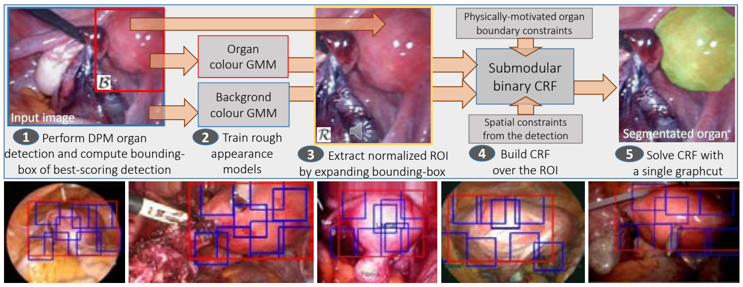

Input image
$\mathcal{B}$
Organ colour GMM
Backgrond colour GMM
$\mathcal{R}$
Physically-motivated organ boundary constraints
Submodular binary CRF
Spatial constraints from the detection
Segmentated organ
1 Perform DPM organ detection and compute bounding-box of best-scoring detection
2 Train rough appearance models
3 Extract normalized ROI by expanding bounding-box
4 Build CRF over the ROI
5 Solve CRF with a single graphcut

Fig. 1. Proposed framework for segmenting the uterus in a monocular laparoscopic image without manual input. The top row shows the five processing stages and the bottom row shows example uterus detections using the DPM detector [11, 6].

The CRF energy function. The CRF is defined over the ROI $\mathcal{R}$, which is computed by enlarging the bounding box to encompass all likely foreground pixels. This is done by scaling the bounding box about its centre $\mathbf{x}_b$ by a factor of $x\%$. We set this very conservatively to $x = 60\%$, which means all foreground pixels will be within $\mathcal{R}$ when the bounding box of the detection overlaps the ground truth bounding box by at least $\approx 40\%$. In practice we do not normally obtain detections with less than approximately 50% overlap with the ground truth bounding box, because the corresponding detection score would normally be too low to trigger a detection. The CRF energy function E is conditioned on the image content in $\mathcal{R}$ and the detection's bounding box $\mathcal{B}$. This has the following form:

$$\begin{aligned} &E(\mathcal{L};\mathcal{R},\mathcal{B}) \overset{\text{def}}{=} E_{app}(\mathcal{L};\mathcal{R}) + \lambda_{edge}E_{edge}(\mathcal{L};\mathcal{R}) + \lambda_{spatial}E_{spatial}(\mathcal{L};\mathcal{R},\mathcal{B}) \\ &E_{app}(\mathcal{L};\mathcal{R}) \overset{\text{def}}{=} \textstyle\sum_{\mathbf{x}\in\mathcal{R}} \mathcal{L}(\mathbf{x})E'_{app}(\mathbf{x};\theta_{fg}) + (1-\mathcal{L}(\mathbf{x}))E'_{app}(\mathbf{x};\theta_{bg}) \end{aligned} \quad (1)$$

The first term E_{app} denotes the *appearance energy*, which is a standard unary term that encourages pixel labels to agree with the foreground and background

GMM models [12]. The term $E'_{app}(\mathbf{x};\theta)$ denotes the negative density of a GMM parameterised by θ. The terms E_{edge} and $E_{spatial}$ denote the edge and spatial energies, which are unary and pairwise clique energies respectively. The terms λ_{edge} and $\lambda_{spatial}$ are weights that govern the relative influence of the energies.

A physically-motivated edge energy model based on the LoG$^+$ filter. The purpose of the edge energy is to encourage a smooth segmentation whose boundary is attracted to probable organ boundaries. In nearly all graphcut-based optical image segmentation methods, this is based on the step-edge model, which says that a transition between labels should occur at regions with high first-order intensity derivatives [2]. However this model does not match well with the physical image formation process in laparoscopic images. This is a combination of the fact that the scene is illuminated by a proximal light source centred close to the camera's optical center, and that because organs are smooth, discontinuities in surface orientation are rare. To see this, consider a point $\mathbf{p}$ on the organ's boundary with a normal vector $\mathbf{n}$ in camera coordinates. By definition $\mathbf{n}$ must be orthogonal to the viewing ray, which implies $\mathbf{n}$ is approximately orthogonal to the light source vector, so $\mathbf{p}$ necessarily reflects a very small fraction of direct illumination. Consider now the image intensity profile as we transition from the organ to a background structure (Fig. 2(h)). We observe a smooth intensity fall-off as the boundary is reached, and then a discontinuous jump as we transition to the background. Due to imperfect optics we measure a smooth version of this profile, which is characterised by a smooth intensity trough at a boundary point. *Likely organ boundaries are therefore those image points with strongly positive second-order intensity derivatives*, and this can be computed stably with the LoG$^+$ filter. One issue is that edge filters such as LoG$^+$ are also sensitive to superficial texture variation of the organ. An effective way to deal with this is to apply the filter on the red channel only, because red light diffuses deeper into tissue than blue and green light [4]. Fig. 2 illustrates the effectiveness of the LoG$^+$ filter for revealing the uterus boundaries, which we compare to the Sobel step-edge filter (manual segmentations are overlaid in red).

We define E_{edge} in a similar manner to [2] but replace the intensity difference term by the LoG$^+$ response at the midpoint of two neighbouring pixels $\mathbf{x}$ and $\mathbf{y}$:

$$\begin{aligned} E_{edge}(\mathcal{L}) &\overset{\text{def}}{=} \textstyle\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{N}} w_{\mathbf{x},\mathbf{y}}(\mathcal{L})\exp\left(-\text{LoG}^{+}((\mathbf{x}+\mathbf{y})/2)/2\sigma\right) \\ w_{\mathbf{x},\mathbf{y}}(\mathcal{L}) &= \begin{cases} 1/d(\mathbf{x},\mathbf{y}) & \text{if } \mathcal{L}(\mathbf{x}) \neq \mathcal{L}(\mathbf{y}) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

where $\mathcal{N}$ denotes the set of pixel neighbour pairs. The term $w_{\mathbf{x},\mathbf{y}} \in \mathbb{R}$ assigns energy when the labels of $\mathbf{x}$ and $\mathbf{y}$ are different that decreases as the LoG$^+$ response at their midpoint increases. The function d gives the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$, which reduces the influence of neighbours that are further away. Inspired by [2] we set σ automatically as the standard deviation of LoG$^+$ across all pixels in $\mathcal{R}$. The LoG$^+$ has a free parameter σ_N that pre-smoothes the image to mitigate noise. We have found that results are not highly sensitive to σ_N, and in all experiments we use $\sigma_N = 3$ pixels with a filter window of 7 pixels.

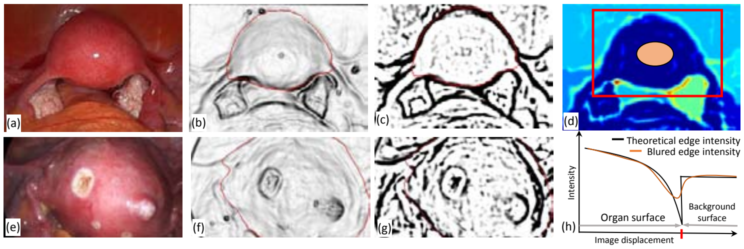


Fig. 2. Laparoscopic images of two uteri with different edge filter response maps. Sobel step edge responses are shown in (b,f) and LoG$^+$ responses are shown in (c,g). The LoG$^+$ geodesic distance transform $\mathcal{D}$ for (a) is shown in (d), with the detection's bounding box and central ellipse $\mathcal{S}$ overlaid. An illustration of the edge intensity profile across an organ boundary edge is shown in (h).

Hard labels and spatial energy. We assign hard labels to pixels in the image that we are virtually certain of either being on the organ or on the background. The job of this is to prevent complete over or under-segmentation in instances when the organ's appearance is very similar to the background. We assign pixels within a small region around the bounding box center $\mathbf{x}_b$ the foreground label, which is valid because the main body of the uterus is always highly convex. Specifically we define a small elliptical region $\mathcal{S}$ by $\mathbf{x} \in \mathcal{S} \Leftrightarrow s^2(\mathbf{x}-\mathbf{x}_b)^\top \text{diag}(1/w, 1/h)(\mathbf{x}-\mathbf{x}_b) \leq 1$, and assign all pixels in $\mathcal{S}$ the foreground label. This is an ellipse with the same aspect ratio as the bounding box, where w and h are the width and height of the bounding box. The scale of $\mathcal{S}$ is given by s, which is not a sensitive parameter and in all experiments we use $s = 0.2$. To prevent complete over-segmentation we assign pixels very far from the bonding box the background label. We do this by padding $\mathcal{R}$ by a small amount by replication (we use 20 pixels), and assign the perimeter of the padded image the background label.

The spatial energy encodes the fact that pixels near the detection's center are more likely to be on the organ. We measure distances to the detection's center in terms of geodesics $\mathcal{D}(\mathbf{x}) : \mathcal{R} \rightarrow \mathbb{R}^+$ using the LoG$^+$ filter response as a local metric. This is fast to compute and more informative than the Euclidean distance because it takes into account probable organ boundaries in the image. We compute $\mathcal{D}(\mathbf{x})$ by measuring the distance of $\mathbf{x}$ to $\mathcal{S}$ using the fast marching method. We give a visualisation of $\mathcal{D}$ for the image in Fig. 2 (a) in Fig. 2 (d), with the central ellipse overlaid in red. Dark blue indicates lower distances, and the darkest shade corresponds to a distance of zero. One can see that for most pixels either on the uterus body, or connected to the uterus body by ligaments or the Fallopian tubes, the distance is zero, because for these points there exists a path in the image to $\mathcal{S}$ that does cross an organ boundary. We therefore propose a very simple spatial energy function, which works by decreasing the energy of a pixel $\mathbf{x}$ if it is labelled foreground and has $\mathcal{D}(\mathbf{x}) = 0$. We do this for all pixels within the detection's bounding box, and define the spatial energy as:

$$E_{spatial}(\mathcal{L};\mathcal{D},\mathcal{B}) \overset{\text{def}}{=} \sum_{\mathbf{x}\in R} \begin{cases} 1 & \text{if } \mathcal{L}(\mathbf{x})=0 \text{ and } \mathcal{D}(\mathbf{x})=0 \text{ and } \mathbf{x}\in\mathcal{B} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The effect of $E_{spatial}$ is to encourage pixels within the bounding box to be labelled foreground if they can reach the detection's center by a path that does not cross poins that are likely to be organ boundaries. To improve the computation speed for $E_{spatial}$ we compute $\mathcal{D}$ on a down-sampled version of $\mathcal{R}$ (by a factor of two). On a standard desktop PC this means $E_{spatial}$ can be computed in approximately 100 to 200ms, and there is no real impact on segmentation accuracy.

3 Experimental Results

We have evaluated on a new dataset consisting of 228 uterus images of 82 different patients, which extends the 40-patient database from [11] (Fig. 3). The images were gathered from patients at our hospital (12 patients) and demonstration and tuition images from the web (70 patients). 27.7% of the patients had uteri with pathological shape, caused mostly by uterine fibroids. For each image we computed the best-scoring detection from the uterus detector, using the accelerated code of [5]. A detection was considered a true positive if the overlap between the detection's bounding box and the manually-computed bounding box exceeded 55% (which is a typical threshold in object detection literature). In total 201 images had true positive detections. In the other 27 images false positives were caused nearly always by very high amounts of tool occlusion. We then segmented all images with true positive detections. Because our method is the first to achieve completely automatic organ segmentation in laparoscopic images, there is not a direct baseline method to compare to. We therefore adapted a number of competitive interactive and seed-based segmentation methods, by replacing manual inputs with the output of the uterus detector. These were as follows. (*i*) *Grabcut-I* [12]: we replaced the user-provided bounding box required in grabcut with the bounding box from the detection, and replaced hard labels from the user with the same hard labels as described above. (*ii*) *Non-iterative Grabcut* (GrabCut-NI): This was the same as GrabCut-I but terminating after one iteration (*i.e.* the appearance models and segmentation were not iteratively refined). (*ii*) *Growcut* [13]: we used GrowCut with $\mathcal{S}$ as the foreground seed region and the perimeter of $\mathcal{R}$ as the background seed region. (*ii*) *Edge-based Levelset Region growing* (ELR) [7]: we used a well-known region growing method based on levelsets, passing it $\mathcal{S}$ as the initial seed region. For Grabcut-I, GrabCut-NI, Growcut and our method, we test with RGB and chromaticity (*i.e.* illumination-invariant) colourspaces. We found negligible differences between different chromaticity spaces, so report results with just one (CrCb). The free parameters of the baseline methods were tuned by hand to maximise their performance on the dataset. The free parameters of our method (λ_{edge} and $\lambda_{spatial}$) were tuned using a set of 20 images held out from the dataset, which gave $\lambda_{edge} = 60$ and $\lambda_{spatial} = 4$. The hold-out set did not include patients from the main dataset. We did not use a hold-out set for tuning the baseline method parameters, which meant we could could measure their best possible performance on the dataset.

We present the DICE score boxplots and summary statistics in Fig. 4. The methods labelled Proposed-NI (RGB) and Proposed-NI (CrCb) represent our method using RGB and CrCb colourspaces respectively for the appearance models. We also investigated whether our method could be improved by iteratively retraining the colour models and resegmenting, like GrabCut. We label this Proposed-I (RGB) and Proposed-I (CrCb). !!THINGS TO NOTE!! (a) colourspace makes no real difference for the methods. (b) Doing colour model/segmentation iterations does not lead to improved accuracy. (c) Our mean and median is much higher than the rest. (d) our max is higher than grabcut. This is because our edge model makes the segmentation well to the contours. Fig. 3 show 12 representative images from the dataset, with our automatic segmentations overlaid in green. The images show typical difficulties including pathalogical shape, surgical changes, partial occlusions, strong light fall-off, low-contrast bondaries and oversaturation.

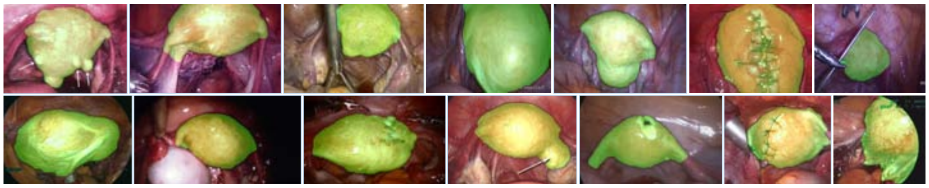
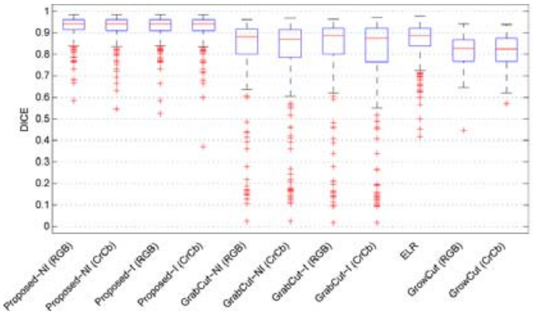


Fig. 3. Example images from the test dataset and segmentations from our method.



	max	min	mdian	mean	sd
Proposed-NI (RGB)	98,58	58,43	94,38	92,49	6,03
Proposed-NI (CrCb)	98,61	54,72	94,36	92,22	6,58
Proposed-I (RGB)	98,63	52,33	94,30	92,40	6,49
Proposed-I (CrCb)	98,67	37,07	94,30	91,99	7,45
GrabCut-NI (RGB)	97,91	2,54	88,44	81,78	19,01
GrabCut-NI (CrCb)	97,97	2,54	87,21	77,41	24,31
GrabCut-I (RGB)	97,66	1,68	88,58	81,22	20,13
GrabCut-I (CrCb)	98,19	1,68	87,64	77,18	25,34
ELR	98,06	41,92	88,58	85,76	9,76
GrowCut (RGB)	94,26	44,63	82,69	81,21	7,31
GrowCut (CrCb)	93,95	56,12	82,56	81,06	7,94

Fig. 4. DICE performance statistics of our proposed method in four diferent configurations against baseline methods.

4 Conclusion

We have presented a new method for segmenting the uterus in monocular laparoscopic images that requires no manual input and no patient-specific prior knowledge. The method is based on using a patient-independent uterus detector to supervise the training of an CRF-based patient-specific segmenter. There are several possible directions for future work. Firsly we aim to properly integrate tool segmentation such as [1] with our method. Secondly because our method produces patient-specific appearance models, we can combine this information with the patient-generic detector to make a patient-specific detector. We expect this will reduce false positive detections in later frames. Finally, our segmenter can also be used for performing 3D shape-from-silhouette, and if we combine this with Sructure-from-Motion we are likely to obtain better invivo 3D reconstructions than SfM alone.

Acknowledgements. This research was funded by the EU FP7 ERC research grant 307483 FLEXABLE and Almerys Corporation.

References

1. M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, , and D. Stoyanov. 2d-3d pose tracking of rigid instruments in mis. In *Int Conf Information Processing in Computer Assisted Interventions*, 2014.
2. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary amp; region segmentation of objects in N-D images. In *ICCV*, 2001.
3. A. Chhatkuli, A. Malti, A. Bartoli, and T. Collins. Monocular live image parsing in uterine laparoscopy. In *ISBI*, 2014.
4. T. Collins and A. Bartoli. Towards live monocular 3d laparoscopy using shading and specularity information. In *IPCAI*, 2012.
5. C. Dubout and F. Fleuret. Exact acceleration of linear object detectors. In *ECCV*, 2012.
6. P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE PAMI*, 2010.
7. C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.*, 2010.
8. A. Malti, A. Bartoli, and T. Collins. Template-based conformal shape-from-motion-and-shading for laparoscopy. In *IPCAI*, 2012.
9. M. Nosrati, J.-M. Peyrat, J. Abi-Nahed, O. Al-Alao, A. Al-Ansari, R. Abugharbieh, and G. Hamarneh. Efficient multi-organ segmentation in multi-view endoscopic videos using pre-op priors. In *MICCAI*, 2014.
10. B. L. Price, B. S. Morse, and S. Cohen. Geodesic graph cut for interactive image segmentation. In *CVPR*, 2010.
11. K. Prokopetc, T. Collins, and A. Bartoli. Automatic detection of the uterus and fallopian tube junctions in laparoscopic images. In *IPMI*, 2015.
12. C. Rother, V. Kolmogorov, and A. Blake. Grabcut - Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 2004.
13. V. Vezhnevets and V. Konushin. Growcut - Interactive multi-label n-d image segmentation by cellular automata. In *GraphiCon*, 2005.