Efficient Image Segmentation Incorporating Photometric and Geometric Information

Tien-Vu Nguyen, Trung Tran, Phong Vo, Bac Le

Abstract— This paper proposes a new segmentation approach based on graph cuts. However, we take advantage of it in a different way using photometric (color and texton) and geometric information to analyze and segment images. These additional information provide better knowledge of the image's internal structure. Hence, they lead to a significant improvement in segmentation quality, while require fewer instruction from user interactions. Some experimental results are presented in the variety contexts of image segmentation giving high accurate segmentation comparing with other methods.

***Index Terms*—geometric information, graph cut, interactive object segmentation, texton.**

I. INTRODUCTION

Image segmentation is a essential part of image processing applications, particularly in medical images analysis and photo editing. A wide range of computational vision algorithms can benefit from reliable and efficient image segmentation techniques. Image segmentation also has obvious applications in separating objects from its background (e.g., identification of specific structures in medical images, tracking objects in video, etc). The simple methods do not need prior knowledge but have limitations, whereas more effective methods are pretty complex and require training data. Though automated methods are being improved continuously, but they do not assure reliable results in all cases. Interactive segmentations are more preferred ways because it can segment image based on user directions actively. Image segmentation may have difficulties due to the semantic gap, where the level of segmentation depends on user's view of objects in the image. Consequently, interactive image segmentation is more customizable as it offers users a way to define what to separate out from the image.

Among interactive segmentation approaches, graph cut by Boykov and Jolly [4], which we summarize in section II, is one of the most powerful techniques.
in this paper is based on the model of graph cut. Nevertheless, we utilized photometric information (texton and color) and geometric information in the general framework of graph cuts. Fig.1. illustrates one example of our approach.

Tien-Vu Nguyen is now a student of the Department of Computer Science, University of Science Ho Chi Minh city, Viet Nam. Phone: 84-933419000; e-mail: ntienvu@gmail.com.

Trung Tran is now with the Department of Computer Science, University of Science Ho Chi Minh city, Viet Nam; e-mail: tntrung@fit.hcmus.edu.vn.

Phong Vo was with the Department of Computer Science, University of Science Ho Chi Minh city, Viet Nam. He is now a PhD student at the Télécom ParisTech, Paris, France; e-mail: vdphong@fit.hcmus.edu.vn.

Bac Le is the Head of the Department of Computer Science, University of Science Ho Chi Minh city, Viet Nam; e-mail: lhbac@fit.hcmus.edu.vn.

Fig.1. Example of our approach. Left: input image with little seeds. Middle: texton map of image. Right: result image with our approach proposing in this paper.

Section II is a brief review of some well-known segmentation methods. A detail description of our approach is presented in section III. Section IV shows our experiments. Conclusion and future work are discussed in section V.

II. RELATED WORK

In this section we shortly outline main features of prevailing state-of-the-art segmentation techniques. Current segmentation algorithms are far from comparing human performance for natural images. The challenge of this task and the limitations of the input data, which differ from their models due to noise, occlusion, clutter, often lead to poor results.

There are variety of image segmentation methods, including completely automated approaches [5], [6], model-driven approaches [7],[8], and semi-supervised (interactive) methods. With interactive methods, user defines the certain pixels as "object" and "background".

MagicWand is a common selection tool for almost image editor nowadays. User indicates points or regions to segment using color statistics of the specified region. Because the distribution in color space of object and background pixels have a considerable overlap, a satisfactory segmentation is not obtained.

Intelligent Scissors is a boundary-based method that computes minimum-cost path between user-specified boundary points. It treats each pixel as a graph node and uses shortest-path graph algorithms for boundary calculation. The main limitation of this tool is apparent: for highly texture (or untextured) regions many alternative "minimal" paths exist. Therefore many user interactions were necessary to obtain a satisfactory result [9].

Growcut [11] is an interactive segmentation algorithm. It uses Cellular Automaton as an image model. Automata evolution models segmentation process. Each cell of the automata has some label (in case of binary segmentation -

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

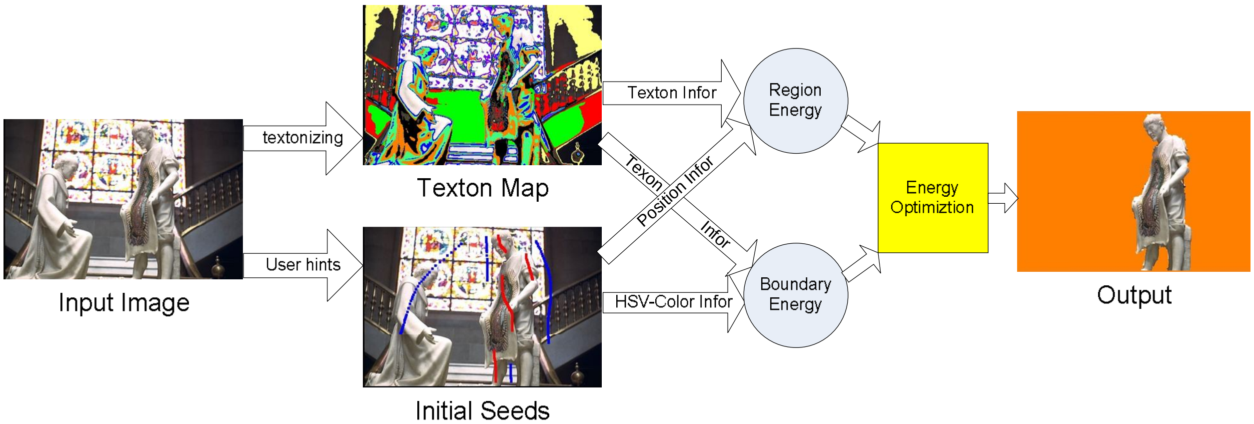

Fig. 2. Model of our approach.

object, background and empty). During automata evolution some cells invade their neighbors, replacing their labels. In GrowCut, a user draws some strokes inside the object of interest with an object brush, and outside the object with a background brush. In simple cases, only a few strokes are sufficient for segmentation.

Transduction [12] generates a segmentation of the entire image that is consistent with the seeds which is given as the representative of each region to be categorized in an image. It uses geometric information and the s-weighted Laplacian graph regularize, a powerful manifold learning tool that is based on the estimation of variants of the Laplace-Beltrami operator and is tightly related to broaden processes.

Graph Cut [4] by Boykov and Jolly is a powerful technique of optimization, which was applied to the task of image segmentation. Consider the image as a graph. Each pixel is a node. There are two types of edges: n-links and t-links. N-links connect pairs of neighboring pixels or voxels. Thus, they represent a neighborhood system in the image. Cost of n-links corresponds to a penalty for discontinuity between the pixels. T-links connect pixels with terminals (labels). The cost of a t-link connecting a pixel to a terminal matches to a penalty for assigning the corresponding label to the pixel. Then a graph can be computed efficiently by max-flow/min-cut algorithms. Given user-specified object and background seed pixels, the rest of the pixels are treated automatically. There are two other effective methods based on graph-cut : GrabCut [10], LazySnapping [16]. Grabcut extends graph-cut by introducing interactive segmentation scheme and uses graph-cut for intermediate steps. It is sufficient to mark the object with a rectangle to get the desired results. LazySnapping uses watershed algorithm for pre-segmented step to divide an image into small regions. The graph is established by considering each region as a graph node. Then graph cut is performed to segment this graph.

Graph Cut is a efficient technique rely on maximize energy to cut out objects from background. Nevertheless, it still has some minor problem about contours and edges in a low contrast image.

Our approach is based on Graph Cut. But we alter it in a different way to obtain better results. We will describe in detail in the following section.

Our contribution is that we modify graph cut components by using incorporated photometric information (color and texton) and geometric information. These other information supply necessary knowledge for achieving high accurate results with less user interaction in efficient time comparing with other technique on Microsoft GrabCut dataset [15] in Section IV.

III. Our Approach

Our approach is presented as follow. We calculate an image to get texton information. This pre-computed process, which introduced in the subsection below. The user marks some seeds as “background” or “foreground” respectively. Then the graph is established following Graph Cut by Boykov and Jolly [4]. However, we modify the energy functions rely on combining texton, color, and position information. Our approach model is illustrated in Fig. 2.

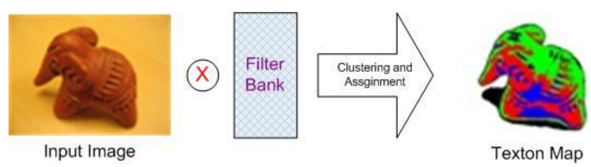

Fig. 3. Textonizing image process.

A. Textons

Textons present human texture perception [1], and are very useful for categorizing materials [3]. In order to apply texton information into our method, we have to textonize an image for texton map. The textonizing process is illustrated in Fig. 3. In general, we use the same process as [2]. In our method, we textonize each image, respectively. Unlike [2], it calculates for the whole training images. The image is convoled with a 17-dimensional filter-bank at scale k. Then the 17-d responses for all pixels are whitened (to give zero mean and unit covariance), and the unsupervised clustering method is performed afterward.
J. Shotton et al [2] operate the Euclidean-distance K-means clustering algorithm. Finally, each pixel in image is assigned to the nearest cluster center, producing the texton map. They denote the texton map as T where pixel i has value $T_i \in \{1, . . ., K\}$.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Fig. 4. Example of texton map.
Left: original image. Right: texton map.

The textonizing process is pre-calculated before our segmentation process. Segmentation algorithms need properties which represent for the range of different appearances of an object that is texton. In our approach, texton property plays an important role for expanding the texton-similar region. One example of texton map is showed in Fig 4.

B. *Incorporation of photometric and geometric properties*

Considering that P is the whole pixels in an image and N be a set of all unordered pairs if neighborhood pixels. To segment a given image we create a graph $G = \langle V, E \rangle$, with nodes corresponding to pixels $p \in P$ of the image. There are two additional nodes: an "object" terminal (a source S) and a "background" terminal (a sink T) as Boykov and Jolly [4]. Consequently, V = $P \cup \{S, T\}$. Let A = $(A_1,\ldots,A_p,\ldots A_{|P|})$ be a binary vector where component A_p denotes foreground or background assignment to pixel p in P. Hence, vector A defines segmentation. The cost function E(A) is

$$E(A) = \lambda.R(A) + (1-\lambda).B(A) \quad (1)$$

where

$$R(A) = \sum_{p \in P} R_p(A_p) \quad (2)$$

$$B(A) = \sum_{\{p,q\} \in N} B_{p,q}(A_p, A_q) \quad (3)$$

We use the same model of the state-of-the-art graph cut by Boykov and Jolly [4], but we calculate R(A) and B(A) differently using geometric (position) and photometric (texton).

$$B_p \sim \exp\left(-\frac{\| C_q - C_p \|^2}{\sigma^2}\right).\frac{1}{dist(p,q)}.\gamma \quad (4)$$

where

$$\gamma = \begin{Bmatrix} u \ \ if \ \ T_p = T_q \\ \frac{1}{u} \ \ if \ \ T_p \neq T_q \end{Bmatrix}$$

with u is a parameter indicating the relationship between texton information and boundary. Parameter γ helps to increase B_p if p and q have the same texton index. Conversely, γ will reduce the cost B_p. Other parameters are considered in a similar way as Boykov and Jolly [4].

$$R_p('obj') \sim -\ln \Pr(T_p / 'obj').l_p \quad (5)$$

$$R_p('bkg') \sim -\ln \Pr(T_p / 'bkg').l_p \quad (6)$$

where l_p is location of pixel p comparing with 'object' or 'background' in an image. Finally, the edge weighting table, which is assigned to build a graph, is used similar to [4].

In case of user want to adjust their segmentation results, they can add extra seeds. A maximum flow on a new graph is accomplished without re-computing from scratch.

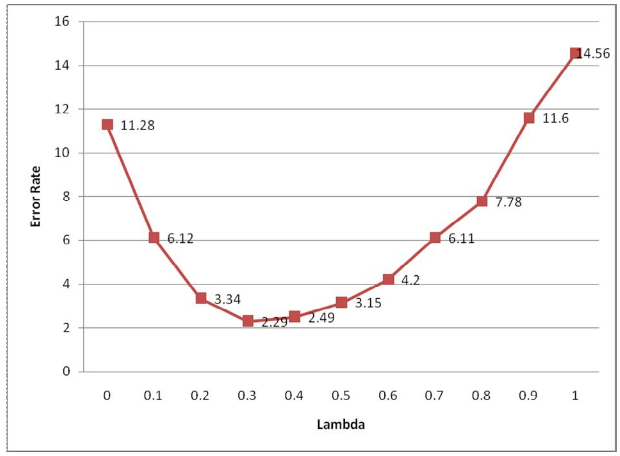

Fig. 5. Lambda (λ) versus Error Rate. A good value for λ is 0.3.

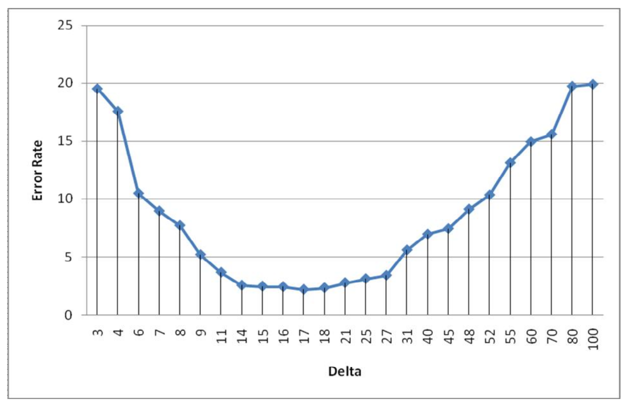

Fig. 6. Delta (δ) versus Error Rate. A good value for δ is 17 with λ = 0.3.

IV. Experiments

This section presents our experimental results. Testing of our approach is conducted in Microsoft GrabCut dataset and arbitrary images as well. Fig. 7 shows one example of segmentation in arbitrary image.

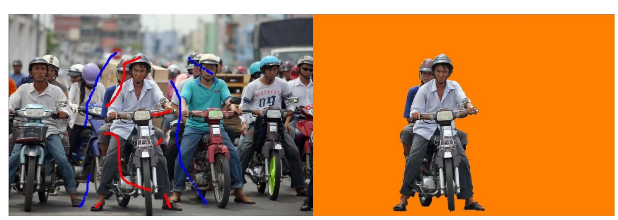

Fig. 7. Result in arbitrary image.

Lambda (λ) in equation (1) shows the relationship between boundary properties and region properties in general. More specifically, it presents the connection of

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

color information versus texton and position information. The coefficient λ is the important parameter in graph-cut segmentation. Different images can be in need of alternative lambda values. When λ=0, the cost function in equation (1) is merely boundary. In contrast, only region properties is concerned if λ=1. Fig. 5 demonstrates the variation of segmentation results when lambda is changed from 0 to 1, representing through error rate. The good value for λ is 0.3. Fig. 6 presents error rate when delta is changed. The appropriate number for delta is approximately 17. In terms of textonizing image, we use K=50 for texton index. This parameter was derived from our experimental results. HSV color channel is used to represent color information for giving high results.

Fig. 8. Example of graphcut base on superpixel.
Left: superpixel image.
Right: GraphCut-superpixel based results.

Graph Cut based on superpixel

We also developed other way which based on the idea of lazy snapping [16] for comparing with our approach. Lazy snapping used watershed algorithm to pre-segmented image into small regions. But watershed has a drawback when giving the bad contour regions. Unlike watershed, superpixel [17] has many advantages such as:

It is computationally efficient: it reduces the complexity of images from hundreds of thousands of pixels to only a few hundred superpixels.

It is also representatively: pairwise constraints between units, while only for adjacent pixels on the pixel-grid, can now model much longer-range interactions between superpixels.

The superpixels are perceptually meaningful: each superpixel is a sensitively consistent unit, i.e. all pixels in a superpixel are most likely uniform in color and texture.

It is near-complete: because superpixels are results of an over segmentation, most structures in image is conserved. There is very little loss in moving from the pixel-grid to the superpixel map.

Thus, we use superpixel instead of watershed. The obtained results are acceptable with little seeds and without any adjustment after segmenting. Fig. 8 demonstrates our superpixel-based segmentation results.

Table 1: Error rate comparison between our approach with other GraphCut-based methods in GrabCut dataset

Segmentation Methods	**Error Rate**
GraphCut-pixel based	10.19%
GraphCut-superpixel based	7.34%
Our Approach	**2.29%**

We compare our performance with other methods in Microsoft GrabCut dataset 50 images. The error formula as follow [15]:

$$\varepsilon = \frac{no.\, missclassified\ pixels}{no.\ pixel\ in\, unclassifiedregion}$$

The operating time is acceptable. It takes totally two minutes and ten seconds to perform 50 images either 640x480 pixels or 450x600 pixels in Microsoft GrabCut data set on Dual Intel 1.73 GHz. Each image takes roughly from 2 seconds to 3 seconds. The textonizing process is pre-computed so it does not affect the real computational time.

Like other segmentation methods, the drawback is imperfect contours in some cases of relating to ragged, low contrast, and sophisticated image. The real time is not achieved.

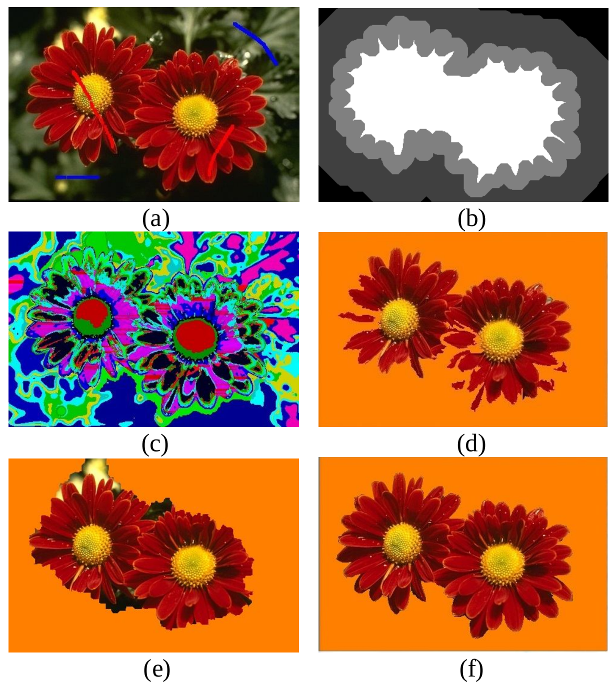

Fig. 9. Display the comparison of our approach versus other state of the art methods. (a): original image with initial seeds. (b): ground truth image. (c): texton map of original image. (d): GraphCut Boykov and Jolly [4] (using color mixture model instead of grayscale information). (e): superpixel segmentation. (f): our approach.

V. CONCLUSION AND FUTURE WORK

Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share the certain visual characteristics. Its duty is computed in the real time with the perfectly segmented result. We have presented a new approach segmentation base on graph cut. Utilizing photometric information and geometric information for high accuracy results with less user-interaction in economical time. However, it also has some problem about the contour in some cases.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

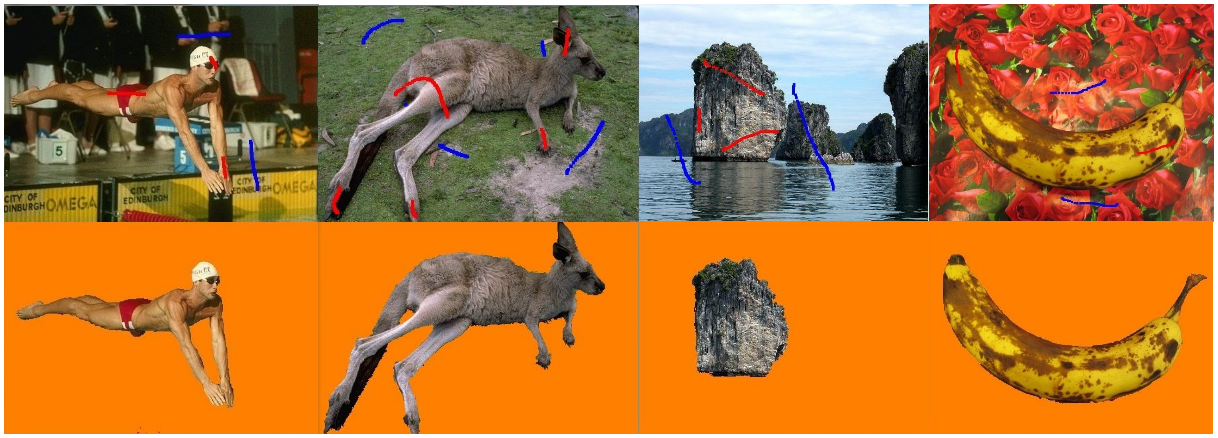

Fig. 10. More experiments. First Row: Images with initial seeds. Second Row: Output Images

Future work

We integrate other information into our model to exploit thoroughly image knowledge. In term of getting efficient contour, we will apply some new edge detection techniques; utilize entropy of pixels on boundary. In addition, computational time can be reduced by using multicore methods for sparse system.

REFERENCES

[1] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. Int. J. Computer Vision, 43(1):7–27, June 2001.

[2] J. Shotton, J. Winn, C. Rother, A. Criminisi. TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context.

[3] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. Int. J. Computer Vision, 62 (1-2):61–81,April 2005.

[4] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In ICCV, volume 1, pages 105–112, July 2001.

[5] J. Shi and J. Malik. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):888–905, 2000.

[6] F. J. Estrada and A. D. Jepson. Quantitative evaluation of a novel image segmentation algorithm. In CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) -Volume 2, pages 1132–1139, Washington, DC, USA, 2005. IEEE Computer Society.

[7] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based space. In International Conference on Computer Vision and Pattern Recognition, volume 2, pages 699–704, Madison, Wisconsin, USA, June 2003.

[8] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Transactions on Medical Imaging, 22(2):137–154, 2003.

[9] C. Rother, V. Kolmogorov, and B. A. Grabcut – interactive fore-ground extraction using iterated graph cuts. ACM Transactions on Graphics, SIGGRAPH 2004, 2004.

[10] ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. Grabcut - interactive foreground extraction using iterated graph cuts. Proc. ACM Siggraph.

[11] Vezhnevets, V. and Konouchine, V. GrowCut: Interactive multi- label N-D image segmentation by cellular automata, Proc. of Graphicon, pp. 150–156, 2005.

[12] Duchenne, O., Audibert, J.Y., Keriven, R., Ponce, J., Ségonne, F. Segmentation by transduction. In: CVPR. (2008)

[13] L. Grady and G. Funka-Lea. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. In M. Sonka, I. A. Kakadiaris, and J. Kybic, editors, ECCV Workshops CVAMIA and MMBIA, volume 3117 of Lecture Notes in Computer Science, pages 230–245. Springer, 2004.

[14] Yuri Boykov and Vladimir Kolmogorov. In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), September 2004.

[15] A. Blake, C. Rother,M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In ECCV06, pages I: 428–441, 2006.

[16] Yin Li, Jian Sun, Chi-Keung Tang and Heung-Yeung Shum. Lazy Snapping. SIGGRAPH 2004, Vol. 23, pp. 303-308.

[17] Superpixels and Supervoxels in an Energy Optimization Framework, O. Veksler, Y. Boykov, P. Mehrani, in European Conference on Computer Vision (ECCV), 2010.

[18] Efficient Approximate Energy Minimization via Graph Cuts. Y. Boykov, O. Veksler, R.Zabih. IEEE TPAMI, 20(12):1222-1239, Nov 2001.

[19] What Energy Functions can be Minimized via Graph Cuts? V. Kolmogorov, R.Zabih. IEEE TPAMI, 26(2):147-159, Feb 2004.

[20] An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. Y. Boykov, V. Kolmogorov. IEEE TPAMI, 26(9):1124-1137, Sep 2004.