Research Journal of Applied Sciences, Engineering and Technology 5(2): 491-497, 2013
DOI:10.19026/rjaset.5.4979
ISSN: 2040-7459; E-ISSN: 2040-7467

Submitted: May 15, 2012 Accepted: June 08, 2012 Published: January 11, 2013

Research Article

Graph Cuts Based Image Segmentation with Part-Based Models

Wei Liu and Xuejun Xu
School of Hydropower and Information Engineering, Huazhong University of Science and Technology Wuhan, China

Abstract: This study proposed an improved pre-labeling method based on deformable part models and HOG features for interactive segmentation with graph cuts. Because of the complex appearance of foreground and background, the result of segmentation is unsatisfactory. Many priors have been introduced into graph cuts to improve the segmentation results and our work is inspired by the shape prior. In this paper we use the deformable part-based model and HOG features to pre-label the seeds before the graph cuts algorithm. The user involvement is reduced and the performance of the graph cuts algorithm is improved at the first iteration. Our assumption is based on the compact shape. We assume that the area between the center of the part filter and root filter belongs to foreground. If the area covered by more filters, it will more probably be the foreground. Our results show that our method can get more accurate result especially the appearance of the object and background is similar and the shape of the object close to rectangle and eclipse.

Keywords: Deformable part-based models, graph cuts, image segmentation

INTRODUCTION

Image segmentation is a fundamental research area of computer vision and many works have been done in this area. One of the famous frameworks is energy minimization using graph cuts. Because of the advantage of global optimization and compute efficiency and multidimensional, graph cuts becomes more and more popular in recent 10 years.

Boykov and Jolly (2001) proposed a very effective method for interactive segmentation based on graph cuts. Rother *et al*. (2004) made two enhancements to the graph cuts mechanism: "iterative estimation" and "incomplete labeling" which together allow a considerably reduced degree of user interaction for a given quality of result. Many priors are introduced into the graph cuts framework in order to improve the performance and the accuracy of the result recently. For example, label cost prior (Delong *et al*., 2010, 2011; Liu *et al*., 2008; Yuan and Boykov, 2010) is famous now. Delong *et al*. (2010) proposed an extension of α-expansion that also optimizes "label costs" with well characterized optimality bounds. Label costs penalize a solution based on the set of labels that appear in it.

Recently, many researchers have looked to shape priors to incorporate a prior shape information in order to further constrain the segmentation. Shape priors can be modeled by a known class of shapes or through statistical training. The approaches using shape prior have been divided into several classes. The first class uses the certain type of shape prior in particular situations including star shape prior (Veksler, 2008) and eclipse shape prior (Slabaugh and Unal, 2005) and tightness prior (Lempitsky *et al*., 2009). The Second class uses shape modeling. Many works (Leventon *et al*., 2000; Tsai *et al*., 2001; Rousson and Paragios, 2002) have been done on shape prior in the level set and curve evolution frameworks. These approaches can be quite robust when object shape is similar to model, but they are not applicable for classes of objects with high shape variability. The similar approaches are used in Lempitsky *et al*. (2008) and Cremers *et al*. (2008). In Lempitsky *et al*. (2008) the prior is defined by the set of exemplar binary segmentations, branch-and-bound is used to choose right prior from that set. Cremers *et al*. (2008) introduced a implicit representation of shape based on probability. The last class includes approaches which represent shape as a set of rigid parts that may have various positions w.r.t., to one another. In Felzenszwalb and Huttenlocher (2005) shape model is represented by a layered pictorial structure. Yangel and Vetrov (2011) proposed a shape prior that represents object shape via simplified skeleton graph, edges of the graph correspond to meaningful parts of an object. Our approach is inspired by this method.

The objective of this study is to improve the performance the interactive segmentation. We propose a pre-labeling method based on deformable part models and HOG features, our assumption is based on compact shape priori. The segmentation by graph

Corresponding Author: Wei Liu, School of Hydropower and Information Engineering, Huazhong University of Science and Technology Wuhan, China

cuts is an iterative process. After the pre-labeling, the user involvement is reduced and the performance of the graph cuts algorithm is improved at the first iteration.

METHODOLOGY

Segmentation with graph cuts: Image segmentation problems can be seen as labeling problems and labeling problems can be seen as probability problems. In the MRF-MAP framework, we formulated this as an energy minimization such that for a set of pixels P and a set of labels L, the goal is to find a labeling f:P→L that minimize the energy. According to the MRF-GRF equivalence, the energy is given by:

$$E(f) = \sum_{p\in P} D_p(f_p) + \lambda \sum_{pq\in N} V_{p,q}(f_p, f_q) \tag{1}$$

where,

N : The neighborhood system

$D_P(f_P)$: The penalty of assigning label $f_P \in$ L to p, $V_{P,q}(f_P, f_q)$ is the penalty of labeling the pair p and q with labels $f_P, f_q \in$ L

According to Kolmogorov and Zabih (2004), a globally optimal binary labeling for Eq. (1) can be found via graph cuts if and only if the pairwise interaction potential $V_{P,q}$ satisfies:

$$V_{p,q}(0,0) + V_{p,q}(1,1) \le V_{p,q}(0,1) + V_{p,q}(1,0) \tag{2}$$

and the minimum E (f) can be computed efficiently with graph cuts.

Graph cuts: Let G = (V, E) be a graph with vertices V and edges E. Each edge e ∈ E in G is assigned a non-negative cost W_e. There are two special vertices called terminals identified as the source, s and the sink, t. A cut C ⊂ E is a subset of edges, such that if C is removed from G, then V is partitioned into two disjoint sets S and T = V - S such that s ∈ S and t ∈ T. The cost of the cut C is the sum of its edge weights: $|C| = \sum_{e\in C} W_e$. The minimum cut is the cut with the smallest cost. Many standard polynomial time algorithms for min-cut/max-flow have been developed.

These algorithms can be divided into two main groups: "push-relabel" style methods (Goldberg and Tarjan, 1988) and algorithms based on augmenting paths by Ford and Fulkerson (1962).

We use the max-flow algorithm of Boykov and Kolmogorov (2004) which is designed specifically for computer vision applications and has linear time performance in practice.

Segmentation algorithm: Boykov and Jolly (2001) proposed the famous algorithm of image segmentation with graph cuts. In that study, the problem of segmenting an object from the background is interpreted as a binary labeling problem. Each pixel in the image is assigned a label l where, l ∈ L, L = {0, 1}, 0 stands for the background and 1 stands for the object.

In Eq. (1), the first term is called the regional or data term. The regional term assumes that the the individual penalties for assigning pixel p to "object" and "background", correspondingly $D_P(''obj'')$ and $R_P(''bkg'')$, are given. The more likely f_P is for p, the smaller is $D_P(f_P)$. The second sum in Eq. (1) is called the boundary or smooth term. The boundary term comprises the "boundary" properties of segmentation. $V_{P,q}$ should be interpreted as a penalty for a discontinuity between p and q. Typically, $V_{Pq}(f_P, f_q) = W_{Pq}$ I $(f_P \neq f_q)$, where I(·) is 1 if $f_P = f_q$ and 0 otherwise. Normally, $W_{P,q}$ is large when pixels p and q are similar (e.g., in their intensity) and $W_{P,q}$ is close to zero when the two are very different. In Boykov and Jolly (2001) $W_{P,q}$ is defined as:

$$exp(-\frac{(I_p - I_q)^2}{2\sigma^2})\square\frac{1}{dist(p,q)}$$

Parameter λ≥0 in Eq. (1) weights the relative importance between the regional and boundary terms, (Peng and Veksler, 2008) introduced how to select proper parameter λ.

Deformable part models: Deformable part models such as pictorial structures (Felzenszwalb and Huttenlocher, 2005) provide an elegant framework for object detection. Felzenszwalb *et al.* (2010) described an object detection system based on mixtures of multiscale deformable part models. The model represents objects by a collection of parts arranged in a deformable configuration. Each part captures local appearance properties of an object while the deformable configuration is characterized by spring-like connections between certain pairs of parts.

Models and matching: The models are defined by the response of filters and feature maps. A filter is a rectangular template defined by an array of d-dimensional weight vectors. A feature map is an array whose entries are d-dimension feature vectors

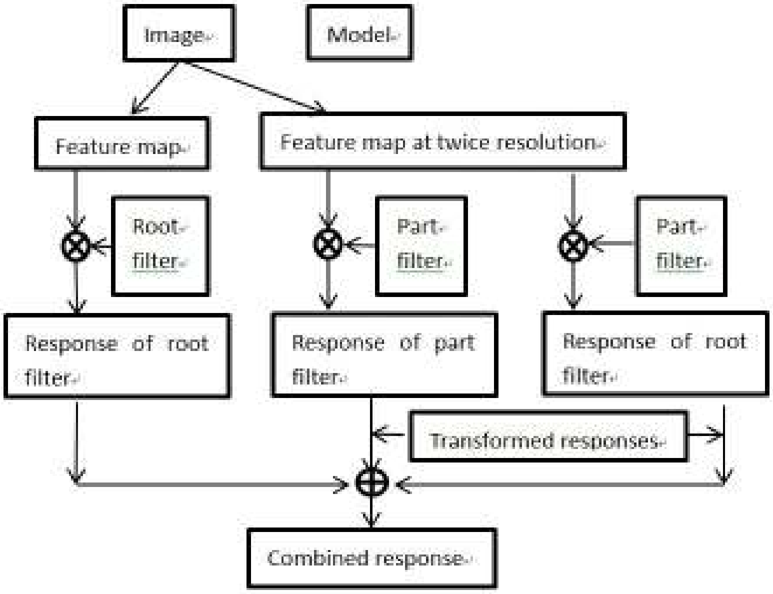

Fig. 1: Part-based model

computed from a dense grid of locations in an image. Then the filters are applied to dense feature maps. The response of a filter F at a position (x, y) in a feature map G is the "dot product" of the filter and a subwindow of the feature map with top-left corner at (x, y):

$$\sum_{x',y} F[x',y'] \bullet G[x+x',y+y'] \qquad (3)$$

Figure 1 shows the process of the part-based model. The filters contain a coarse root filter that approximately covers an entire object and higher resolution part filters that cover smaller parts of the object. Consider building a model for a car. The root filter could capture coarse resolution edges such as the car boundary while the part filters could capture details such as wheels, windows, head and rear. We denote the root filter by F_0 and the n part filters by $F_1 \dots F_n$. Each part model is defined by a 3-tuple (F_i, v_i, d_i) where F_i is a filter for the i-th part, v_i is a two-dimensional vector specifying an "anchor" position for part i relative to the root position, and d_i is a four dimensional vector specifying coefficients of a quadratic function defining a deformation cost for each possible placement of the part relative to the anchor position. We use $Pi = (x_i, y_i, 1_i)$ to specify the level and position of the i-th filter. The response of a object hypothesis is given by:

$$R = \sum_{i=0}^{n} F_i \bullet \phi(H,p_i) - C_d + b \qquad (4)$$

where,

C_d : A deformation cost
b : A bias term and we can get the details refer to Felzenszwalb *et al*. (2010)

To detect objects in an image we compute an overall score for each root location according to the best possible placement of the parts. High-scoring root locations define detections while the locations of the parts that yield a high-scoring root location define a full object hypothesis.

Hog feature: HOG (Histograms of Oriented Gradients) feature is a kind of pixel-level features which was first proposed in Dalal and Triggs (2005). This technique is used to calculate the statistical value of local image gradient orientation. HOG is similar to edge orientation histograms, SIFT descriptors and shape contexts, but HOG is calculated based on a dense grid of uniformly spaced cells and uses the technical of overlapping local contrast normalization.

The most important idea of HOG is: the shape and appearance of the local object in an image can be well described by dense distribution of the gradient orientation. The implementation detail is: first the image is divided into connected regions which called cell units, then we collect the histogram of gradient orientation of pixels in each cell unit, and these

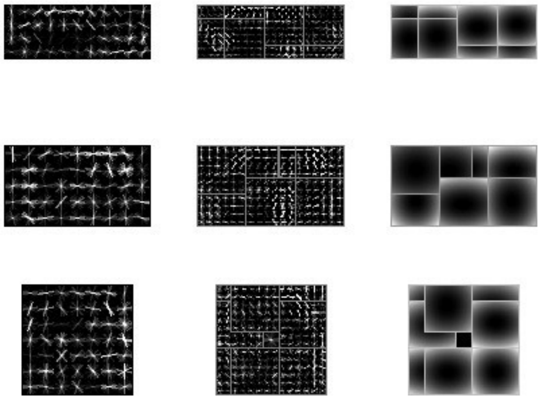

Fig. 2: HOG feature of car model

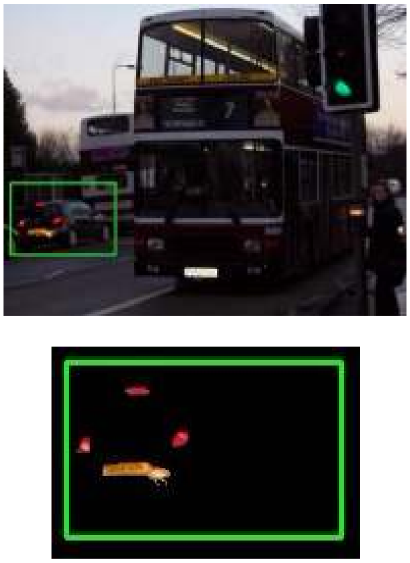

Fig. 3: A failed example of grab cut algorithm

histograms compose the feature descriptors at last. Figure 2 shows the model of the car.

HOG descriptor has several advantages compared with others. It captures edge or gradient structure that is very characteristic of local shapeand it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations: translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size. For human detection, under the conditions of coarse spatial sampling, fine orientation sampling and strong local photometric normalization, the small deformations of the limbs and body don't influence the detection result and can be ignored provided that they maintain a roughly upright orientation.

RESULTS AND DISCUSSION

Our work: The main idea of this study is that we can label the foreground/background and most probably foreground/background pixels before the graph cuts algorithm. The part-based model is introduced to locate these seeds so the user involvement will be reduced and the efficiency of graph cuts will be improved.

Seed points in interactive segmentation: In Boykov and Jolly (2001) the user has to initially select a few foreground and background seeds. After running the algorithm the user has to inspect the quality of the segmentation.

In the grab cut algorithm (Rother *et al*., 2004) the user needs only specify the background region T_B, leaving $T_F = 0$. No hard foreground labeling is done at all. The initial T_B is determined by the user as a strip of pixels around the outside of the marked rectangle. The initial incomplete user-labeling is not always sufficient to allow the entire segmentation to be completed, so further user editing is needed too.

The grab cut algorithm is based on the color feature, so when foreground is not clear with the background, the algorithm can't get the desirable result, as shown in Fig. 3. Our work is based on the grab cut algorithm. The purpose of our work is to select the initial seeds automatically. We use HOG feature and deformable part model to pre-label the object. The process of pre-labeling only needs the user to provide the object bounding box. Our method will label the foreground, background, most probably foreground and most probably background automatically. As a result, the user involvement will be reduce effectively and the algorithm will be more efficiency. The user can also modify the pre-labeling result and the accuracy will be improved.

Bounding box prior and compact regions: User-provided object bounding box is a simple and popular interaction paradigm considered by many existing interactive image segmentation frameworks. Study (Lempitsky *et al*., 2009) discussed how the bounding box can be further used to impose a powerful topological prior, which prevents the solution from excessive shrinking and ensures that the user-provided box bounds the segmentation in a sufficiently tight way. The main idea is the desired segmentation should have parts that are sufficiently close to each of the sides of the bounding box.

In this study we also use the bounding box. Method in Felzenszwalb *et al.* (2010) shows the

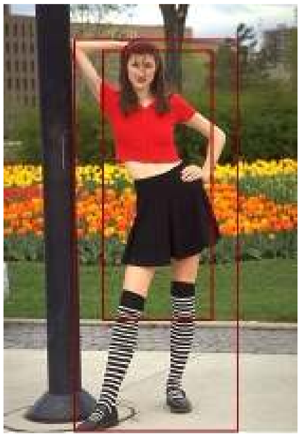

Fig. 4: The wrong detection of deformable part-based model

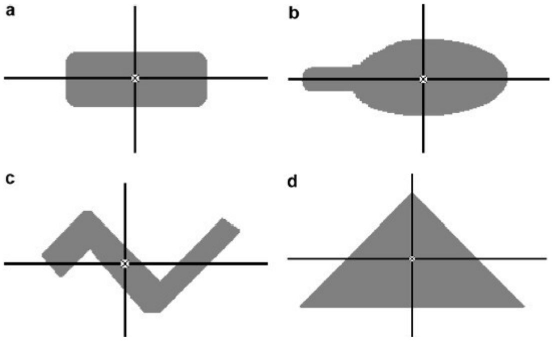

Fig. 5: A and b are examples of objects which are compact in shape, c and d are examples of objects which are not compact in shape. b and c are not convex shapes

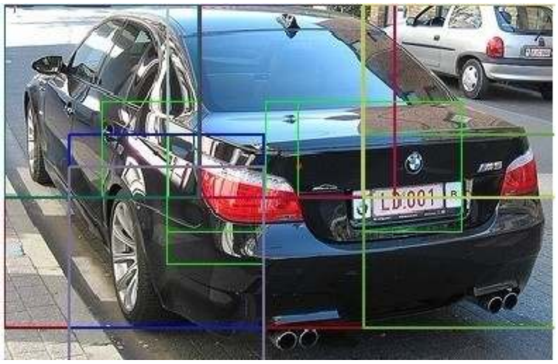

Fig. 6: The detection result of part-based model

detections by a set of bounding boxes. In the segmentation process we have known where the object is, the algorithm don't need to search the hypothesis. The time cost of matching of the deformable part-based model will be reduce effectively. We know that the bounding box provided by the user contains the result bounding box by the deformable part-based model and they are most close to each other. And it will reduce the wrong detection of the deformable part-based model. As shown in Fig. 4.

In Veksler (2002), they chose the word compact to reflect the fact that for compact shapes, the perimeter to area ratio tends to be small. Das *et al.* (2009) introduced a compact shape as a hard constraint in segmentation. The compact shape is similar to the convex shape but they are actually quite different. Figure 5 shows the difference. Based on the idea of compact shape, we make our assumption: for the object with compact shape, the areas between the center of root filter and the centers of part filters are defined to be the foreground. This is shown in Fig. 6. The rectangles in green are the areas between the center of root filter and the centers of part filters.

Pre-labeling: We propose the pre-labeling method based on our assumption. We denote the bounding box of the root filter by B_r, the bounding box provided by the user by B_u, the bounding box of the root filter by B_p. B_r is always smaller than B_u. We make the areas between the center of B_r and the centers of B_p the foreground. If an area is covered by B_u but not by B_r we think this is the background. If an area is covered by many B_p, we think it is most probably foreground, otherwise it's most probably background.

Experiments: We present our experimental results in this section. First we summarize the experimental setup. The user provides a bounding box contains the object need to be segment. Then the part-based model gives the pre-labeling result based on the matching of HOG feature. At last the graph cuts algorithm uses the bounding box and pre-labeling information to do the segmentation.

The compact shape encourages objects with boundaries that are relatively simple. Our method is appropriate for these shapes, especially rectangles and ellipses. The deformable part-based models are trained using a discriminative procedure and achieved state-of-the-art results on the PASCAL VOC benchmarks (Everingham *et al.*, 2009) and the INRIA Person dataset (Dalal and Triggs, 2005).

Some of the images are from the Berkeley database and PASCAL VOC benchmarks (Everingham *et al.*, 2009) and others are from the Web. Because of the assumption of compact shape and the training results of Felzenszwalb *et al.* (2010), we choose images about cars, buses and human in special gesture.

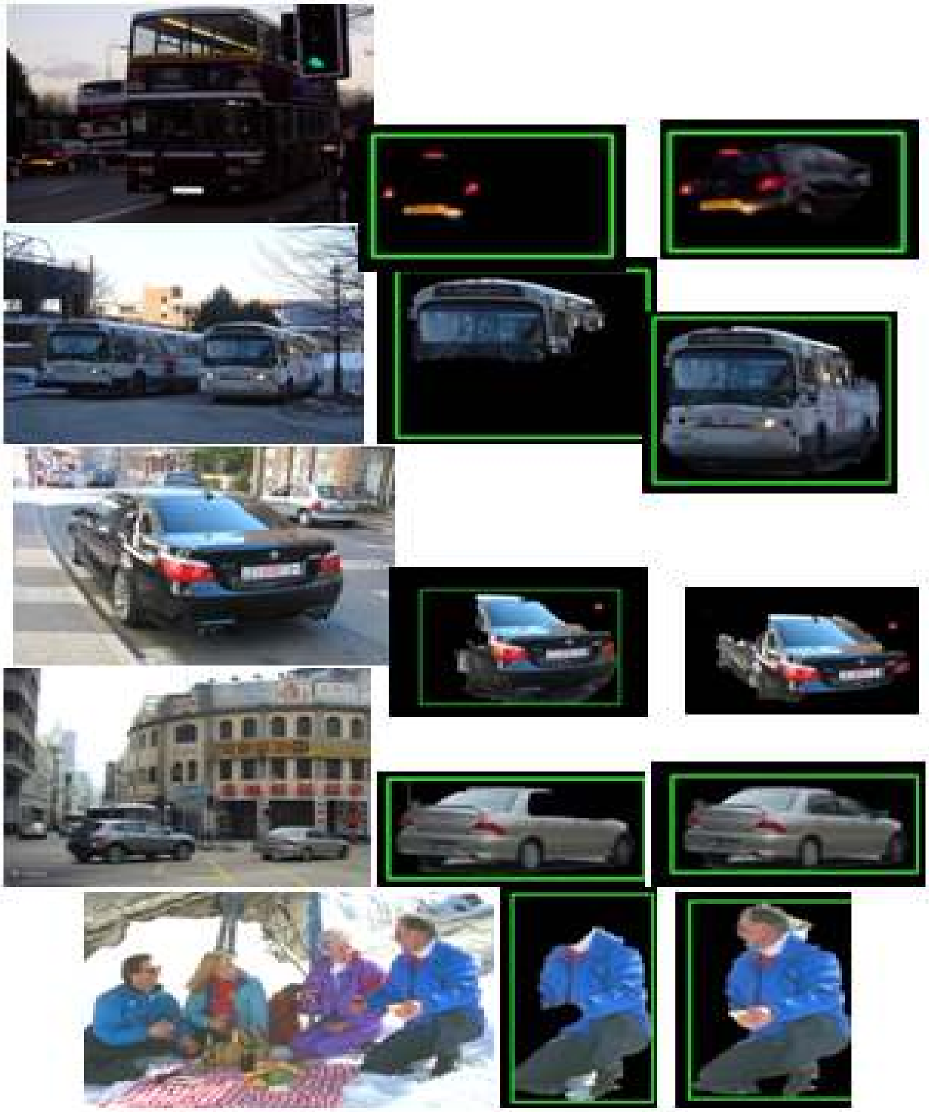

Fig. 7: Some results

We compare our method with the grab cut method. Both need a bounding box provided by user. Figure 7 shows results. The first column is original pictures, the second column is results by grab cut algorithm, the third column is results of our method.

The results are very promising, especially the appearance of foreground and background are similar to each other, as shown in the first and second row in Fig. 7. In the fourth and fifth row the grab cut algorithm can't give the window of the carand the face of the human. But in our method these parts are defined to be foreground by the result of the part based model. Our method gives more information than the grab cut algorithm, so at the first iteration the result is much better without the user involvement.

ACKNOWLEDGMENT

This study is supported in part by National Natural Science Foundation of China No. 41001236.

REFERENCES

Boykov, Y. and M.P. Jolly, 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. Eighth International Conference on Computer Vision (ICCV), 1: 105-112.

Boykov, Y. and V. Kolmogorov, 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE T. Pattern Anal., 26(9): 112-1137.

Cremers, D., F.R. Schmidt and F. Barthel, 2008. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), Dept. of Comput. Sci., Univ. of Bonn, Bonn, pp: 1-3.

Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1: 886-893.

Das, P., O. Veksler, V. Zavadsky and Y. Boykov, 2009. Semiautomatic segmentation with compact shape prior. Image Vision Comput., 27(1-2): 206-219.

Delong, A., A. Osokin, H.N. Isack and Y. Boykov, 2010. Fast approximate energy minimization with label costs. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Dept. of Comput. Sci., Univ. of Western Ontario, London, ON, Canada, pp: 2173-2180.

Delong, A., L. Gorelick, F.R. Schmidt, O. Veksler and Y. Boykov, 2011. Interactive segmentation with super-labels. EMMCVPR'11 Proceedings of the 8th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer-Verlag Berlin, Heidelberg, pp: 147-162.

Everingham, M., L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman, 2009. The PASCAL Visual Object Classes Challenge (VOCC). Retrieved from: http://www.pascalnetwork.org/.

Felzenszwalb, P.F. and D.P. Huttenlocher, 2005. Pictorial structures for object recognition. Int. J. Comput. Vis., 61(1): 55-79.

Felzenszwalb, P.F., R.B. Girshick, D.A. McAllester and D. Ramanan, 2010. Object detection with discriminatively trained part-based models. IEEE T. Pattern Anal., 32(9): 1627-1645.

Ford, L. and D. Fulkerson, 1962. Flows in Networks. Princeton University Press, Princeton U.P., pp: 194.

Goldberg, A.V. and R.E. Tarjan, 1988. A new approach to the maximum-flow problem. J. ACM, 35(4): 921-940.

Kolmogorov, V. and R. Zabih, 2004. What energy functions can be minimized via graph cuts? IEEE T. Pattern Anal., 26(2): 147-159.

Lempitsky, V., A. Blake and C. Rother, 2008. Image segmentation by branch-and-mincut. In Tenth European Conference on Computer Vision (ECCV), 4: 15-29.

Lempitsky, V., P. Kohli, C. Rother and T. Sharp, 2009. Image segmentation with a bounding box prior. IEEE 12th International Conference on Computer Vision (ICCV), pp: 277-284.

Leventon, M.E., W.E.L. Grimson and O.D. Faugeras, 2000. Statistical shape influence in geodesic active contours. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Artificial Intelligence Lab., MIT, Cambridge, MA, 1: 1316-1323.

Liu, X., O. Veksler and J. Samarabandu, 2008. Graph cut with ordering constraints on labels and its applications. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Univ. of Western Ontario, London, pp: 1-8.

Peng, B. and O. Veksler, 2008. Parameter selection for graph cut based image segmentation. British Machine Vision Conference (BMVC).

Rother, C., V. Kolmogorov and A. Blake, 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. ACM Trans. Graph., 23(3): 309-314.

Rousson, M. and N. Paragios, 2002. Shape priors for level set representations. Proceedings of the 7th European Conference on Computer Vision-Part II, ECCV' 2, Springer-Verlag London, UK, pp: 78-92.

Slabaugh, G.G. and G.B. Unal, 2005. Graph cuts segmentation using an elliptical shape prior. IEEE International Conference on Image Processing (ICIP), Dept. of Intelligent Vision and Reasoning, Siemens Corp. Res. Inc., Princeton, NJ, USA, 2: 1222-1225.

Tsai, A., A. Yezzi, W. Wells, C. Tempany, D. Tucker, A.C. Fan, W.E.L. Grimson and A.S. Willsky, 2001. Model-based curve evolution technique for image segmentation. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Dept. of Electr. Eng. and Comput. Sci., MIT, Cambridge, MA, USA, 1: 463-468.

Veksler, O., 2002. Stereo correspondence with compact windows via minimum ratio cycle. IEEE T. Pattern Anal., 24(12): 1654-1660.

Veksler, O., 2008. Star shape prior for graph-cut image segmentation. Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV' 08, Springer-Verlag Berlin, Heidelberg, 3: 454-467.

Yangel, B. and D. Vetrov, 2011. Image segmentation with a shape prior based on simplified skeleton. 11 Proceedings of the 8th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, pp: 247-260.

Yuan, J. and Y. Boykov, 2010. Tv-Based Multi-Label Image Segmentation with Label Cost Prior. British Machine Vision Association (BMVC), pp: 1-12.