ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 5, Issue 10, October 2016

Extraction of Foreground Objects Using Background Prior Propagation

M.Madhuri, B.Doss

***Abstract*— With an adaptive figure-ground segmentation algorithm that extract foreground objects in a generic environment. In this paper, initially a background mask is assigned and background prior is defined. Then by progressive patch merging several soft-label partitions are generated from different foreground priors. A foreground probability map is produced by joining (fusing) these partitions. By binarizing the probability map via threshold sweeping, multiple hard-label candidates will be created. The best candidate is determined by a score function that evaluates the segmentation quality. Generates multiple hypotheses segmentations from different evaluation scores. From the multiple hypotheses segmentations, the hypothesis with maximum local stability is propagated as the new background prior, and the segmentation process is repeated until convergence. The final segmentation is then automatically obtained by similarity voting from the multiple hypotheses. Experiments indicate that our method performs at or above the current state-of-the-art on several data sets, with particular success on challenging scenes that contain irregular or multiple-connected foregrounds. In addition, this improvement in accuracy is achieved with low computational cost.**

Index Terms— **Adaptive figure-ground segmentation, probability map, multiple hypotheses fusion, similarity voting.**

I. INTRODUCTION

Figure-Ground segmentation plays a key role in many vision applications. By producing a binary segmentation of the image, foreground regions are separated from their background. Popular approaches include interactive graph cut [2],[3], random walk, geodesic, information theory, variational solutions. Due to the broad diversity of visual cues in an image, automatic segmentation [1] is extremely difficult in a generic conditions. Interactive approaches are better to provide a priori knowledge to guide segmentation. Bounding box assignment and seed positioning are two representative methods. Multiple hypotheses and classifier fusion are the ideas applied to segmentation studies. Current state-of-the-art interactive segmentation methods suffer from restrictive assumptions about latent distributions, an inability to treat complicated scene topologies, or an inefficient similarity measure.

In this paper, an iterative adaptive figure-ground classification method gives promising solutions in a broadly applicable environment. The pipeline of the framework is illustrated in Fig: 1. Mask box can provide good statistical information about the background and image patches are formed using an adaptive mean-shift algorithm. By generating a large amount of hypotheses through an iterated background prior propagation routine, foreground extraction is achieved. The final result can be obtained by fusing most promising hypotheses. The algorithm obtains good result for challenging scenes in both segmentation accuracy and execution efficiency. This method is not sensitive to difficult scene topology or loose bounding box and reliably treats multi-connected, multi-hole foregrounds. The advantage of our method is that the spatial smoothness term is removed in popular conditional random fields approaches and hence no need of additional learning algorithm for tuning a smoothness parameter.

The rest of the paper is organized as follows. Section II presents related work. Section III presents our enhanced adaptive figure-ground classification framework. Section IV gives experimental results on several image datasets. Section V concludes the paper.

II. RELATED WORK

The four major aspects of figure-ground segmentation which are related to our work are the definition of prior knowledge, similarity measures, parameter tuning and evaluation of good hypotheses. Prior knowledge can provide information about either the foreground or background or both and it also includes bounding boxes or seed points to help define hard or soft constraints. Priors can propagate throughout the segmentation process.

Similarity measures are defined based on appearance cues like color, shape, texture and gradient. Besides traditional Euclidean distance, similarity measures are based on statistics or information theory.

Regarding parameter tuning, a common practice is to learn the parameters via an energy minimization framework using training data and supervised learning. The assumption exists a parameter setting to represent the images in the training set. This assumption is not necessarily true and bears the risk of training set bias and the generalization performance is bad. Adaptively set for each image prefers to find parameters.

Manuscript received Oct, 2016.

***M.Madhuri**, Department of Electronics and Communication Engineering, JNTU College of Engineering, Anantapur, India, Mobile No.7093016553.*

***B.Doss**, Department of Electronics and Communication Engineering, JNTU College of Engineering, Anantapur , India, Mobile No.9985221365.*

Finally, it is difficult to segment all types of natural images in a single criterion that contain a broad variety of visual patterns. So, in recent works, good evaluation strategies have developed to judge and combine multiple candidates [9],[10]. In this paper, by merging good hypotheses, the foreground will extracts. Our method is distinct from previous multi-hypothesis approaches in the following two aspects: First, in the previous work, multiple hypotheses will generates by varying segmentation parameters or employing different over-segmentation algorithms and results in multiple K-way segmentations. In our task foreground extraction is a binary segmentation task. By following multiple evaluation, a novel method is proposed to generate binary hypotheses using a tree-structured likelihood propagation. Since it is unknown which part of the bounding box contains the foreground, by initializing the tree with various regions as the foreground to generate candidate segmentations. Second, due to lack of effective mechanisms to choose a better one from multiple hypotheses in general environment from the most previous approaches and can only resort to some extra learning process. In our method, the idea of similarity voting is used that does not require learning process to fuse soft-segmentation into a probability map [5] and hard-segmentation into a final segmentation.

In this paper, the presence of new method improves over the original in the following aspects: 1) Soft label schemes are used based on foreground likelihoods that leads to significant improvement in the segmentation quality of fine details. 2) An iterative scheme is introduced to propagate the background prior that increases the accuracy of the segmentation. 3) A novel score function is defined by using maximally stable external regions (MSER) for goodness evaluation and background prior propagation that effectively prevent over-propagation of the background and better handles loose bounding boxes. 4) Probability map [5] is generated by extending the similarity voting and hypotheses set selection that contains a robust classifier fusion from multiple hypotheses.

III. ENHANCED ADAPTIVE FIGURE-GROUND CLASSIFICATION FRAMEWORK

In this section, we propose an enhanced adaptive figure-ground classification framework with background prior propagation. Our segmentation algorithms is based on fusing multiple candidate segmentations and is guided by two underlying principles: 1) Candidates are generated by voting or fusion of multiple candidates often has a better chance than optimization of a single score function in classification tasks. 2) The hypotheses set with higher intra similarity, to yield a result that satisfies more participants than by regarding the fusion strategy. The fusion strategy of multiple candidate segmentations denotes as similarity voting. Similarity voting plays an important role throughout the algorithm.

A. Algorithm Overview

The pipeline of our figure-ground segmentation algorithm is shown in fig 1. Our algorithm consists of two main stages: 1) hypothesis segmentation generation, and 2) similarity voting & fusion. In the first stage, the user box indicates the initial background , and a large number of candidate segmentations are created, best hypothesis segmentations are selected from the set of candidate segmentations. The new background prior is defined by using one of the hypotheses and several hypotheses sets are formed by repeating the segmentation process. In the second stage, the best hypothesis set is automatically selected by comparing the intra similarities of the set, and the final segmentation is formed by fusing the corresponding hypotheses.

B. Bounding Box Assignment

Our algorithm is based on a user-specified mask box that defines the initial background prior, as in previous approaches. Either side of the box can be defined as the background mask that contains only background pixels. The complement of the background mask makes the foreground mask, that contains both foreground and background elements. The mask box can flexibly handle different cases of multiply connected foregrounds.

C. Image Patches by Adaptive Mean-Shift

The segmentation [4] is defined by grouping the non-overlapping regions have become popular due to its advantages in information transfer and computational efficiency. We choose the mean-shift algorithm to do over-segmentation since mean-shift patches are better described statistically in comparison to other super-pixel generators. To remain consistent with the underlying probabilistic framework of the mean-shift algorithm, we model each mean-shift patch as a multivariate normal distribution. A feature vector in the 5D joint color-spatial feature space is given by

$$f_{x,y} = (L(x, y), a(x, y), b(x, y), x, y) \quad (1)$$

where (x, y) are the 2D pixel coordinates and (L, a, b) are the pixel values in the Lab color space. The Lab color space is used because it is better modeled by a normal distribution in comparison to RGB. Finally, we treat the 3D color features and the 2D spatial features identically, and hence distances defined over the feature space, compare both visual similarity and spatial adjacency of the patches.

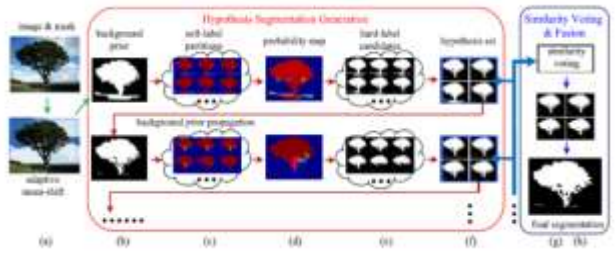


Fig.1. The Pipeline of enhanced figure-ground classification with background prior propagation (EFG-BPP): (a) Original image with box mask(Section III-B),and image patches from adaptive mean-shift(Section III-C); The inputs of the "hypothesis segmentation generation" are the image patches. (b) the box mask helps define the initial background prior; (c) Different foreground priors generate multiple soft-label partitions (Section III-E); (d) all soft-labels are combined into one foreground probability map (Section III-F); (e) Thresholding the probability map forms a set of hard-label candidates (Section III-G); (f) A set of hypotheses is selected using evaluation by various score functions

(Section III-G). The lower-right hypothesis of (f) is the result using the MSER score function, and is propagated as the background prior for the next segmentation round (section III-H). The inputs of the "similarity voting & fusion" block are multiple hypothesis set. (g) The winning hypothesis set is selected using similarity voting; (h) The final segmentation is obtained by fusing the hypothesis set (Section III-I).

D.Similarity Measure Between Patches

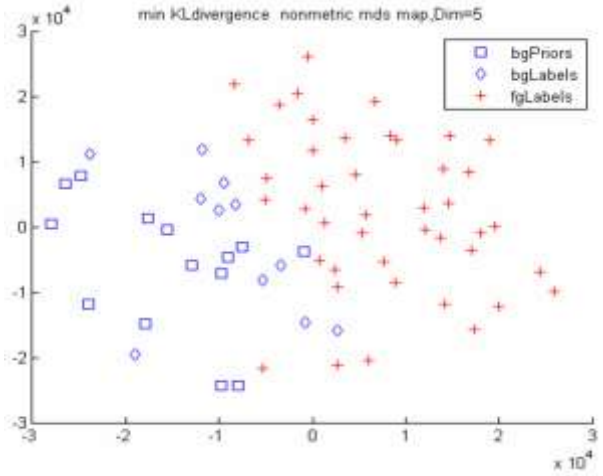
In the next stage of the segmentation pipeline, patches are gradually assigned to the current background or foreground regions, based on their distances to each region. A region will represent as the set of its patches. Hence, we must first define a suitable distance measure between two patches, and between a patch and a region, i.e., a set of patches. We model each mean-shift patch pi as a multivariate normal distribution $N(\mu_1,\sum_1)$ in the 5D feature space defined in the above equation. The mean vector μ_1 and the covariance matrix $\sum_1$ are estimated by patch statistics. All patches are eroded with a radius-1 disk structuring element to avoid border effects. In this paper, we consider two distance measures. The first is the minimum KL-divergence between two patches,

$$D_k(N_1,N_2) = \min\left(KL(N_1,N_2), KL(N_2,N_1)\right) \quad (2)$$

where N_1 and N_2 are two Gaussians, $N(\mu_1,\sum_1)$ and $N(\mu_2,\sum_2)$ and the KL-divergence between two d-dimensional Gaussians.

$$KL(N_1,N_2) = \frac{1}{2}[(\mu_1-\mu_2)^T\sum_2^{-1}(\mu_1-\mu_2) + Tr(\sum_2^{-1}\sum_1) + \log\left(\frac{|\sum_2|}{|\sum_1|}\right) - d] \quad (3)$$

is a symmetrized version of the KL divergence. The graphical representation of KL-divergence in 5D is shown below



E. Soft-Label Partitions

With the patch distances defined in Section III-D we build our foreground extraction algorithm. In light of the assumption that the pre-assigned mask provides sufficient background statistics, we first initialize the background and foreground priors, and then gradually refine the segmentation by merging patches into the foreground.The first step is to obtain the background prior B by selecting patches overlapped with the background mask. Next, patches are selected for the initial foreground prior [7],[8] F if the distance from patch p_i to the background prior B is greater than a threshold D_t,

$$L(p_i) = 1, if\ D(p_i,B) > D_t \quad (4)$$

To start the initial foreground set F, first, all unlabeled patches are sorted in descending order by their distances from the background prior B. Patches are then labeled in turn by comparing them with the current background and foreground sets,

$$L(p_i) = \begin{cases} 1\ if\ D(p_i,F) \le D(p_i,B) \\ 0\ if\ D(p_i,F) > D(p_i,B) \end{cases} \quad (5)$$

F. Foreground Probability Map

By fusing all soft-label partitions to build the foreground probability map. The fusion occurs based on the idea of similarity voting. F_i denotes the i_{th} soft-label partition from the previous stage, andF_i^m denotes the likelihood value of the m_{th} pixel in F_i , where take the likelihood pixels of their corresponding patches. The similarity between two soft-label partitions F_i and F_j can be denoted as

$$d(F_i,F_j) = (\sum_{m=1}^{M}|F_i^m - F_j^m|)/(\sum_{m=1}^{M} sign(F_i^m + F_j^m)) \quad (6)$$

Where M is the total number of pixels, and sign(x) is 0 when x=0 and 1 when F_i and F_j are hard-label partitions.

G.Hypotheses Segmentation set

The segmentation procedure in the previous section generates a large set of candidate segmentations by varying the threshold D_t. Taking into account the fact that perceptually meaningful segmentations may correspond to different cost functions, we generate multiple segmentation hypotheses by selecting the best candidate segmentations according to several evaluation scores. In particular, we consider three score functions that maximize the global distance between background and foreground patches, using the distance measure D: sum-cut [2],[3], average-cut [2],[3] and maxmin-cut [2],[3] (abbreviated as s-cut, a-cut, and m-cut). Other score functions could also be easily incorporated to enclose any available prior knowledge.The sum-cut [2],[3]score function is defined as the sum of the distances D(f, B) from each foreground patch f to the background set, i.e. the selected threshold is given by

$$D_t^{(s)} = argmax_{D_t\in[D_l,D_u]}\sum_{f\in F(D_t)} D(f,B(D_t)) \quad (7)$$

Where $F(D_t)$ and $B(D_t)$ are respectively the foreground and background sets in the final segmentation map computed from the threshold D_t .Taking the average instead of the sum yields the average-cut score function

$$D_t^{(a)} = argmax_{D_t\in[D_l,D_u]}\frac{1}{|F(D_t)|}\sum_{f\in F(D_t)} D(f,B(D_t)) \quad (8)$$

Finally, the maxmin-cut [2],[3] score maximizes the minimum distance between foreground and background patches,

$$D_t^{(m)} = argmax_{D_t \in [D_l, D_u]} D\left(F(D_t), B(D_t)\right) \quad (9)$$

H.Iterated Background Prior Propagation

A set of hypothesis segmentations are resulted in the section III-G. Background prior propagation (BPP) [6] can be defined as the one of the background region of these segmentations that can be used as the background prior for a new segmentation. For BPP, the use of background from the MSER segmentation because of its favourable properties mentioned in Section III-G. The process iteratively continues until the background prior reaches a stable point. The convergence of BPP is guaranteed because image patches can only be added to the background prior in each segmentation. To prevent the over-propagation of the background, MSER contributes the bigger foreground regions.

I. Hypotheses Selection and Fusion

Given the multiple hypotheses, we borrow the idea of classifier fusion to automatically obtain the final segmentation. Two methods were proposed for determining the final segmentation by pooling over all the hypotheses. The first scheme, which we denote as similarity voting, is based on the assumption that a good solution is likely to be selected by different score functions. In the case of multiple winners, we select the one most similar to the other hypotheses. The similarity between two binary segmentations H_1 and H_2 is defined as the Jaccard index,

$$s(H_1, H_2) = |H_1 \cap H_2| / |H_1 \cup H_2| \quad (10)$$

The second scheme, which we denote as probability map that computes the final foreground map F by a simple pixel-wise majority vote,

$$F = ((H_{MSER} + H_{acut} + H_{mcut1} + H_{mcut2}) \geq 2) \quad (11)$$

IV. Experimental Results

In this section, we evaluate our algorithm. Experiments were run on a notebook computer with an Intel core-i7 CPU 2.7Ghz processor and 4GB RAM. Our algorithm is implemented in MATLAB and is available online3.

A. Evaluation of Segmentation Results

In this experiment, 4 datasets were considered. The first two datasets are from the Weizmann evaluation dataset containing 100 1-obj images and 100 2-obj images. We also use the IVRG dataset(1000 images) and the grab cut dataset (50 images).

Enhanced figure-ground classification uses soft-label partitions as background prior propagation as EFG-BPP and the performance of hard-label partitions tests and denotes as EFG-BPP (hard-label).The user assigns and fix the initial mask box by comparing the methods.

EFG-BPP successfully labels background holes and multiply connected components, and identifies many details missed in the manual-made truths. EFG-BPP exhibits better performance, by segmenting complicated foreground and background shapes. The result of the first iteration is sufficiently good, even in the absence of background prior propagation. By employing background prior propagation and automatic hypotheses selection the result becomes better. Similarity comparison plays an important role in hypothesis selection. Finally, all hard-label candidates forms an upper bounds for the figure-ground classification method and the foreground map is closest to the ground truth in all hard-label.

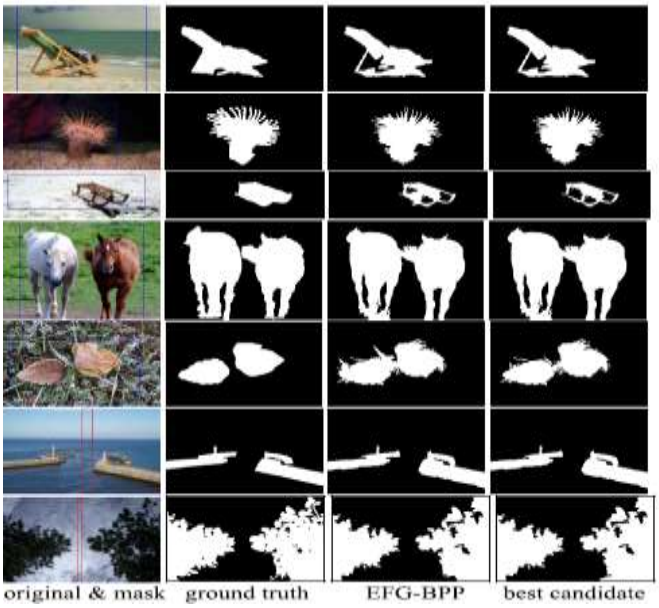


Fig.2 Weizmann examples. Rows 1-3 are 1-obj examples. Rows 4-8 are 2-obj examples. The EFG-BPP result is equally good as the user selection for rows 1,2,4,6 & 7; and slightly worse on the remaining rows. The outside of the blue boxes or the inside of the red boxes define the background masks.

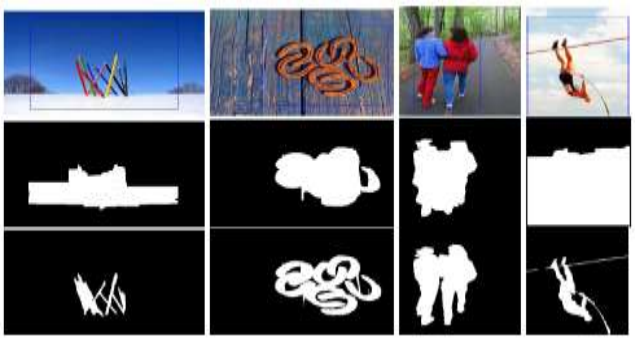


Fig.3 IVRG images. Top: image & mask; middle: grabcut results; bottom: EFG-BPP results.

Fig:2 displays some example segmentations from the Weizmann dataset. Our enhanced f-g classification successfully labels background holes and multiply connected components. It even identifies many details missed in the manual-made truths (rows 1, 2, 4, 6, 7 of Fig: 2). Fig: 3 shows some example segmentations from the IVRG dataset. Our method better segments foregrounds with strongly irregular contours, compared to grab cut. The performance on each image is evaluated using F-measure, F=2PR/(P+R), where P and R are the precision and recall values. Table 1 reports the 95% confidence intervals of the average F-measure on each of the 4 datasets. First, looking at the single hypothesis segmentations produced by each of the four score functions and 2 distances, we note that there is no clear best score

function or distance measure for all scenarios. For example, s-cut is best for Weizmann 1-obj, but does poorly on Weizmann 2-obj.The last row of Table I shows the results of several reference algorithms. EFG-BPP performs slightly better than the state-of-the-art techniques for single connected foregrounds (Weizmann 1-obj), and outperforms the state-of-the-art on multiple connected foregrounds (Weizmann 2-obj).

Type	Weizmann 1-obj	Weizmann 2-obj	IVRG images	Grabcut images
EFG BPP(hardlabel)	0.93±0.010(4.77s)	0.89±0.019(2.12s)	0.94±0.005(4.02s)	0.93±0.018(12.09s)
First iteration	0.93±0.011(3.21s)	0.88±0.021(1.91s)	0.94±0.004(3.74s)	0.91±0.027(10.38s)
Last iteration	0.92±0.014	0.88±0.021	0.93±0.006	0.92±0.021
EFG-BPP	**0.94±0.010(4.82s)**	**0.90±0.017(3.09s)**	**0.95±0.003(5.80s)**	**0.93±0.017(25.99s)**
First iteration	0.93±0.011(3.28s)	0.89±0.017(2.34s)	0.94±0.003(4.90s)	0.92±0.023(15.90s)
Second iteration	0.92±0.014	0.88±0.022	0.93±0.006	0.92±0.021
User-select(upper bound)	0.95±0.009	0.91±0.015	0.96±0.002	0.95±0.013
Nearest competitors	0.85±0.035(5.67s)	0.81±0.044(3.95s)	0.93±0.006(4.96s)	0.89±0.035(12.95s)
	0.93±0.009	0.68±0.053		
	0.87±0.010	0.66±0.066		

TABLE I
F-MEASURES ON FOUR IMAGE DATASETS.BOLD INDICATES BEST ACCURACY AMONG ALL METHODS, EXCLUDING USER-SELECT

B. Evaluation of Soft and Hard Label Schemes

The soft-label scheme performs better than the hard-label scheme at the cost of slightly slower running time (see Table I). For simple scenes the outputs of the two schemes are often the same or have only minor differences. For cluttered scenes the two schemes are more likely to output different results. The soft-label scheme has better chance of keeping fine details, due to the soft likelihoods that are transferred to the probability map.

C. Evaluation of Background Prior Propagation

Overall, about 80% of the time, EFG-BPP converges within three iterations of BPP, and most of them select the first round as the winner. This suggests that our method still can perform well without BPP. Nevertheless, we have seen in Table I that the output of the first loop is not necessarily the best one.

D. Evaluation of Initial Mask Box

The background prior propagation mechanism makes EFG-BPP tolerant to loose initial mask boxes around the foreground subject, since for each iteration the background region [6] can move further into the initial box. IVRG images[500] allow to test the looser bounding boxes while keeping parts of the background prior. For each image, we manually assign the maximally allowed range of the 4 edges of the mask box such that some common parts of the background remain, and test various box sizes within the range.

E. Failure Cases

Failure cases of EFG-BPP were shown in fig.4. In general, if the background prior does not match the true background well then the method will fails. The similar appearances of foreground and background were shown in Fig. 4a, or too cluttered background prevents successful background prior propagation [6] were shown in Fig. 4b. These can be improved by employing a more flexible initial mask.

Fig.4 Example failure cases of EFG-BPP. (a) Failure caused by similar foreground and background appearances; (b) failure caused by cluttered background.

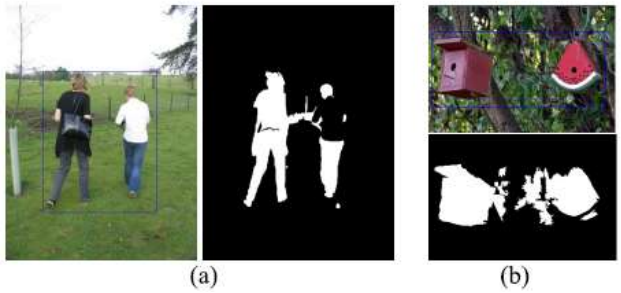


Fig.4 Example failure cases of EFG-BPP. (a) Failure caused by similar foreground and background appearances; (b) failure caused by cluttered background.

V. CONCLUSION

In this paper, an enhanced figure-ground classification algorithm which is based on the principles of generating multiple candidate segmentations, selecting the most promising using several scoring functions, and then fusing them with similarity voting. Specifically, image patches are generated by using an adaptive mean shift, and by using tree-structured likelihood propagation, soft-segmentations are produced. By the idea of similarity voting, multiple foreground map hypotheses are generated, and select the most promising ones. To improve robustness, the background prior propagates iteratively and generate multiple hypothesis sets. The most promising hypothesis set is automatically determined by similarity voting, and fused them to yield the final foreground map. Our method produces state-of-the-art results on challenging datasets, and is able to segment the fine details in the segmentation, as well as background holes and multiply-connected foreground components.

REFERENCES

[1] Z. Tu, X. Chen, A. L. Yuille, and S.-C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," Int. J. Comput. Vis., vol. 63, no. 2, pp. 113–140, 2005.
[2] Y.Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in Proc. 8th IEEE ICCV, vol. 1. Jul. 2001, pp. 105–112.
[3] I. B. Ayed, H.-M. Chen, K. Punithakumar, I. Ross, and S. Li, "Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the Bhattacharyya measure," in Proc. IEEE CVPR, Jun. 2010, pp. 3288–3295.
[4] S. Bagon, O. Boiman, and M. Irani, "What is a good image segment? A unified approach to segment extraction," in Proc. 10th ECCV, 2008, pp. 30–44.
[5] Y. Liu and Y. Yu, "Interactive image segmentation based on level sets of probabilities," IEEE Trans. Vis. Comput. Graphics, vol. 18, no. 2, pp. 202–213, Feb. 2012.
[6] R. Szeliski et al., "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
[7] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004.
[8] B. Micusik and A. Hanbury, "Automatic image segmentation by positioning a seed," in Proc. ECCV, vol. 2. 2006, pp. 468–480.
[9] L. Grady, M.-P. Jolly, and A. Seitz, "Segmentation from a box," in Proc. IEEE ICCV, Nov. 2011, pp. 367–374.
[10] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, "Using multiple segmentations to discover objects and their extent in image collections," in Proc. IEEE CVPR, Jun. 2006, pp. 1605–1614.

Ms.Madhuri is currently studying M.Tech in Digital Electronics and Communication Systems at JNTU College of Engineering, Anantapur, A.P. Her areas of interest are wireless communication and digital electronics

Mr.B.Doss is working as Ad-Hoc Lecturer , Department of ECE, JNTU College of Engineering ,Anantapur. He obtained B.Tech degree in Electronics and communication engineering from Rajiv Gandhi Memorial College of Engineering. He received his M.Tech degree from Visvesvaraya Technological University, Belgam. He did his Ph.D program in JNTU Anantapur. He has taught a wide variety of courses for UG & PG students and guided several projects. He has published several papers in National & International Journals & Conferences.