Class-Specified Segmentation with Multi-scale Superpixels

Han Liu[1], Yanyun Qu[1,⋆], Yang Wu[2], and Hanzi Wang[3]

[1] Computer Science Department, Xiamen University, China
[2] Academic Center for Computing and Media Studies, Kyoto University, Japan
[3] Center for Pattern Analysis and Machine Intelligence, Xiamen University, China

Abstract. This paper proposes a class-specified segmentation method, which can not only segment foreground objects from background at pixel level, but also parse images. Such class-specified segmentation is very helpful to many other computer vision tasks including computational photography. The novelty of our method is that we use multi-scale superpixels to effectively extract object-level regions instead of using only single scale superpixels. The contextual information across scales and the spatial coherency of neighboring superpixels in the same scale are represented and integrated via a Conditional Random Field model on multi-scale superpixels. Compared with the other methods that have ever used multi-scale superpixel extraction together with across-scale contextual information modeling, our method not only has fewer free parameters but also is simpler and effective. The superiority of our method, compared with related approaches, is demonstrated on the two widely used datasets of Graz02 and MSRC.

1 Introduction

This paper aims to segment an image into semantic objects. As a special case, we can extract foreground region from background region at pixel level. More generally, we can segment an image according to the class labels of its components as well; namely, label all objects in the image. Such a task is referred to as class-specified image segmentation in this paper.

Class-specified image segmentation is quite different from the unsupervised bottom-up image segmentation. A single region generated by the bottom-up image segmentation rarely represents a physical object, which is usually troublesome when used for higher level vision tasks. Moreover, bottom-up image segmentation is likely to be sensitive to the model parameters and the image data itself. Different choices of the parameters in a particular bottom-up image segmentation algorithm could generate segments with different quality on the same image. Therefore, the class-specified segmentation is proposed to overcome such problems. It segments an image according to the semantic information of the objects within it, which is expected to be consistent with humans perceptions.

[⋆] Corresponding author.

J.-I. Park and J. Kim (Eds.): ACCV 2012 Workshops, Part I, LNCS 7728, pp. 158–169, 2013.

The development of object localization has shed some light on class-specified segmentation. Dalal et al. [1] implemented a sliding window scheme combined with SVM classifiers to detect pedestrians. However, that method is time consuming. In order to solve this problem, Lampert et al. [2] proposed an efficient subwindow search method, which is based on the branch-and-bound scheme, to detect the generic object. Blaschko et al. [3] treated the problem of object localization as a regression problem, in which the objects location is an output of a learned objective function. However, the above-mentioned methods are all conditioned on the existence of an object template, which is hard to be made robust to the change of object appearance, such as rotation, illumination changes, occlusion, etc. Utilizing multiple templates might somewhat ease the problem, but for many objects in the unconstrained real images, such a strategy may lead to a significant increase of the total number of the required templates. Another flaw of those methods is that they only extract an object with a bounding box, thus being unable to provide accurate segmentation at pixel level.

Recent success in pixel-level categorization has shown a promise for object localization, in which one can label image pixels with the corresponding classes instead of roughly bounding an object with only a rectangle. Shotton et al. [4] constructed semantic texton forests (STF) to learn the local representation. They used a grid with small cells as the input to STF. However, their method is sensitive to the size of the cells and its accuracy decreases as it meets a higher speed demand. Fulkerson et al. [5] used superpixels instead of the regular patch grid for representation. They represented the local image information in an adaptive domain rather than in a fixed window and adopted Conditional Random Field (CRF) [6] to extract the object-level regions. Tighe et al. [7] proposed a similar method as [5]. The difference is that they used superpixel matching instead of classifying to compute the likelihood score for each class, while their commonness is to base themselves on the superpixels of a single scale. Therefore, both of their methods can only capture the context of neighboring superpixels, but not cover the across scale context of the informative superpixels of multiple levels in the scale space. Their performances are thus sensitive to the scale of superpixels and the range of superpixel neighbors, which results in a relative unstable object-level segmentation. Kohli et al. [8] proposed an image parsing method based on both pixels and unlabeled segments, encouraging pixels in the same segment to share the same label. Similar to [5][7], this method does not take into account the scale space context as well. The latest work that explored both the spatial coherency of neighboring superpixels in the same scale and the contextual information across scales was presented by Lubor et al. [9], in which a hierarchical CRF model was performed. However, this work has two shortcomings which limit its effectiveness and applicability. One is that its performance depends much on the goodness of the initial unsupervised segmentation, and the other is that it has many free parameters to be predefined, which is not a trivial task.

In this paper, we propose a new approach for class-specified segmentation which inherits the virtues of the existing methods while at the same time avoids

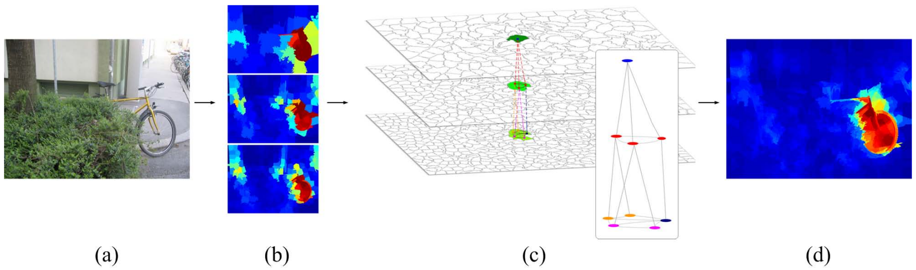


Fig. 1. The framework of the class-specified segmentation method. a) the original image; b) the classification of the segments at three scales where the red color means a high probability of the corresponding superpixels belonging to the bike and the blue color means a high probability of the corresponding superpixels belonging to the background; c) graph construction on the multi-scale superpixels; d) the obtained confidence map. (The figure is best viewed in color.)

their shortages. The proposed approach follows the idea of using CRF to integrate both the spatial coherency and across-scale consistency of multi-scale superpixels, but in a simpler and more effective way than the one presented in [9]. More precisely, instead of using appearance for representing the across-scale contextual information, we use the overlapping ratio which is proved to be more efficient and more effective. Our model has only one single free parameter: the number of scales, which is not sensitive to the input data, as to be witnessed in our experiments. All the other parameters of our model can be learned in the training stage. Besides its applicability, its superiority in terms of segmentation performance will be demonstrated in this paper, especially when it is compared with the most related method [5].

The rest of the paper is organized as follows. In section 2, we give the details about our method. In section 3, the experimental results are given to show the performance of our method. Conclusions are given in the section 4.

2 Class-Specified Segmentation

The framework of our method is shown in Figure 1. We firstly obtain the superpixels at multiple scales by changing the number of segments at each scale. Then an adaboost classifier for the foreground object is learned on the labeled training data. After that, the confidence values of superpixels are computed using the classifier. We employ the CRF model [6] to enforce the spatial consistency between the superpixels and their neighbors both in the same scale and in the consecutive levels in the scale space. Finally, we obtain the class-specified segmentation of an image, as shown in Figure 1(d).

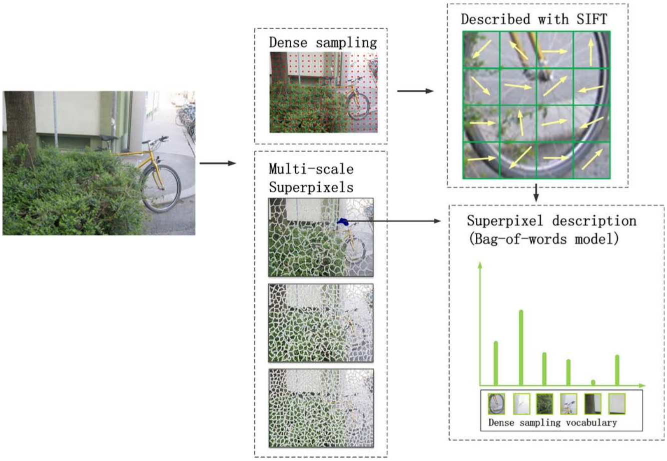

Dense sampling
Described with SIFT
Multi-scale Superpixels
Superpixel description (Bag-of-words model)
Dense sampling vocabulary

Fig. 2. The flowchart of superpixel's description with bag of words model. We get the SIFT descriptors for all the pixels. With the vocabulary of visual words, we describe each superpixel by the word frequency of the pixels within it.

2.1 Superpixel and Description

The motivation of using multi-scale superpixels is to capture the context of the superpixels of multiple levels in the scale space which may be critical for stable object-level segmentation. In our method, we firstly use SLIC superpixels [10] to oversegment an image with different numbers of segments and obtain the multi-scale superpixels. We have evaluated the following image segmentations in our framework: graph-based-segmentation [11], quickshift [12] and SLIC [10], and we found that the superpixels obtained by SLIC achieved the best performance with our model. It probably dues to the fact that the parameters in quickshift segmentation and graph-based-segmentation are relatively less sensitive to the change of color, which results in many repeated superpixels.

We employ the bag-of-words model (BOW) to describe these superpixels. Since sparse sampling may end up with a representation of superpixels which is not informative and stable enough, we use the dense description instead and describe each pixel by a SIFT descriptor [13] as shown in Figure 2, which is similar to [5,7]. These descriptors are then mapped to a vocabulary of visual words which are computed using vector quantization based on the K-means scheme. Before representing a superpixel, we dilate each superpixel region by four pixels in order to enforce the boundary information to the superpixel descriptor following [5,7]. To represent superpixels, we build a histogram of word frequency for each superpixel with the vocabulary. Moreover, we use the color cue as well. We compute the average color for each superpixel in the Lab color space for its high discriminative ability on colors. Finally we simply concatenate the histogram of visual words and the average color to form a high dimension feature vector for each superpixel.

2.2 Classification

In order to compute the confidence value of superpixels, we learn adaboost classifiers [14] based on the superpixels contained in the labeled object regions in the training datasets. The label of each training superpixel is decided by the labels of the pixels in the superpixel. If the pixels in the region of a superpixel belong to several different classes, the label of the superpixel is determined by the label shared by the largest number of pixels in the superpixel region. In the case of binary segmentation, we learn a single binary adaboost classifier using the labeled training data. For the case of the multi-class segmentation, we learn the adaboost classifiers with multiple weak learners trained in a one-vs-rest way, and the confidence of predicted label for each superpixel is decided by calculating the votes from all these weak learners.

2.3 Graph Construction

Considering the spatial consistency between superpixels, we construct an three-dimensional adjacency graph $G(S, E)$ to encode the spatial constraints, in which S is the set of nodes, indicating all the superpixels from all scales, while E is the set of edges connecting pairs of superpixels (s_i, s_j) being adjacent either spatially in the same scale or across consecutive scales. As shown in Figure 1(c), we define these two types of edges as horizontal edges and vertical edges. We connect the pairwise superpixels in the same scale of an image with a horizontal edge if they share a boundary, which represents the spatial context. And we connect the pairwise superpixels in the multiple levels in the scale space with a vertical edge if they share pixels, which stands for the scale context. Compared with [5], we add the vertical edges which enable our method to capture the context of multiple levels in the scale space and extract a stable object-level region. In contrast, the performance of [5] is sensitive to the size of superpixels.

2.4 Inferring with CRF

We introduce CRF to carry out inference on the graph we built. Let $P(c|G, \omega, \nu)$ be the conditional probability of predicting label $\{c_1, \cdots, c_n\} \in C$ given the adjacent graph $G(S, E)$ and the weights ω and ν:

$$-log(P(c|G,\omega,\nu)) = \sum_{s_i \in S} \psi(c_i|s_i) + \omega \sum_{(s_i,s_j)\in E_h} \phi(c_i, c_j|s_i, s_j) + \nu \sum_{(s_i,s_j)\in E_v} \varphi(c_i, c_j|s_i, s_j) \tag{1}$$

where E_h is the set of the horizontal edges, and E_v is the set of the vertical edges. Moreover, ψ is the unary potential and ϕ is the horizontal pairwise potential, while φ is the vertical pairwise potential. There are two weights ω and ν used in our model: ω is the tradeoff parameter between the unary potentials and the horizontal edge potentials, and ν is the tradeoff parameter between the unary potentials and the vertical edge potentials. Since each graph may contain more than one thousand nodes and thousands of edges, it could take several days to train the parameters if we use gradient decent scheme. Alternatively, we use an approximate scheme called stochastic gradient descent [15] to train the parameters ω and ν. For each iteration t, this scheme randomly selects a sample which contains about 5 to 20 batches of points, and computing its gradient by optimizing the maximum-likelihood estimation of C with $P(c|G, \omega, \nu)$. Then update the current parameters with the gradient by a small step. Repeat this process until it converges or iterates sufficient times. It is very fast and efficient.

We define the unary potential $\psi(c_i|s_i)$ by the confidence value obtained from Adaboost which is operated on the superpixels obtained in subsection 2.2. The horizontal pairwise edge potential ϕ is defined as:

$$\phi(c_i, c_j|s_i, s_j) = \frac{1}{1 + \|s_i - s_j\|} \cdot [c_i \neq c_j] \qquad (s_i, s_j) \in E_h \tag{2}$$

and the vertical pairwise edge potential φ is defined as:

$$\varphi(c_i, c_j|s_i, s_j) = \frac{|s_i \cap s_j|}{|s_i \cup s_j|} \cdot [c_i \neq c_j] \qquad (s_i, s_j) \in E_v \tag{3}$$

where $[c_i \neq c_j]$ is the zero-one indicator function. $\|s_i - s_j\|$ is the norm of the color distance between superpixels in Lab color space. The vertical pairwise edge potential is the ratio of the intersection area $|s_i \cap s_j|$ and the union area $|s_i \cup s_j|$ of the pairwise superpixels.

In our experiments we find that the vertical edges have contributed more than horizontal ones, because ν/ω is greater than 1 in most cases. It indicates that the context across scales is more important for object segmentation. As we mentioned before, the total number of the nodes in the graph is usually over a thousand, thus an exact inference is intractable. Therefore, we carry out approximate inference by employing the loopy belief propagation (LBP) [16], which is simple and efficient.

2.5 Across-Scale Label Confidence Integration

For each test image, we get the superpixels and the corresponding descriptions as mentioned in section 2.1. And then all the superpixels are tested through the Adaboost classifier obtained in section 2.2. After that, the CRF inference is carried out with the graph constructed in section 2.3, and the confidence value of each superpixel is obtained. Based on the CRF inference result, we can construct a pixel-wise confidence map for each category by averaging the class-specified confidence values from all the corresponding superpixels, that is, the confidence map is an image whose dimension is equal to the number of classes. Finally, a pixel is labeled according to its the maximum value of the labels, as shown in Figure 3(d) and 5(d).

3 Experimental Results

We evaluate our method on two publically available databases: Graz-02 and MSRC, and all of our results have been released on our website [1].

[1] https://sites.google.com/site/hanliupers/research/image-parsing

3.1 Graz-02

There are three categories in the Graz-02 dataset: car(300), bike(300) and person(300). It is a challenging dataset because the objects significantly vary with rotation, occlusion, scales, etc. We mainly compare our work with Fulkerson et al.'s work [5] because it is most related to our method, and we use the same training and testing data (i.e. the oddly indexed images are used as the training set, while the evenly indexed ones are used as the testing set). Some representative results are shown in Figure 3(d). As a result, our method has achieved better performance than Fulkerson et al.'s work (see Figure 3(c) and Table 1). In their work, different dilate sizes have been applied to superpixels, which results in different segmentation accuracies. On the contrast, in our work we dilate each superpixel with four pixels. In Table 1, we compare our segmentation accuracy with the best result of [5]. We present the results of our method with various numbers of scales (NS) to test the influence of NS on the performance. The results show that our approach work best when the number of scales is equal to 5. Compared with [5], our method achieves 11% higher accuracy on car, while the accuracy improvements on bike and person are 7% and 9%, respectively.

We have tried several different vocabulary sizes K=[100, 200, 400, 600, 800, 1000] in our experiment, and the results show that larger K tends to result in a better performance. However, when K gets greater than 200, it has little effect on the performance improvement. Therefore, we select $K = 600$ in our method.

Table 1. The comparison results in terms of the recall=precision points between [5] and the proposed method with different numbers of scales on the Graz-02 dataset

		Car	Bike	Person
The method in [5]		72.2%	72.2%	66.3%
The proposed method with different numbers of scales.	NS=3	79.5%	76.4%	72.2%
	NS=4	**83.5**%	77.1%	72.9%
	NS=5	81.4%	**79.4**%	**75.1**%
	NS=6	81.2%	78.2%	74.8%

3.2 MSRC

The MSRC dataset contains twenty-three categories. Similar to [4], we discard two categories: horse and mountain because they have too few samples. In this dataset, misclassifications usually happen between similar categories. For example, some parts of a cow can be misclassified as those of a sheep. Shotton [4] suggested to use an image-level prior (ILP) to solve this problem, considering that one may have some prior knowledge about what an image possibly contains before image parsing. To evaluate the ILP in our experiment, we simply describe each trained image with the bag of words model of Spatial Pyramid scheme [17], and learn a classifier from the training images. For each test image, we compute the prior probability(ILP) $P(c)$ on twenty-one categories with the

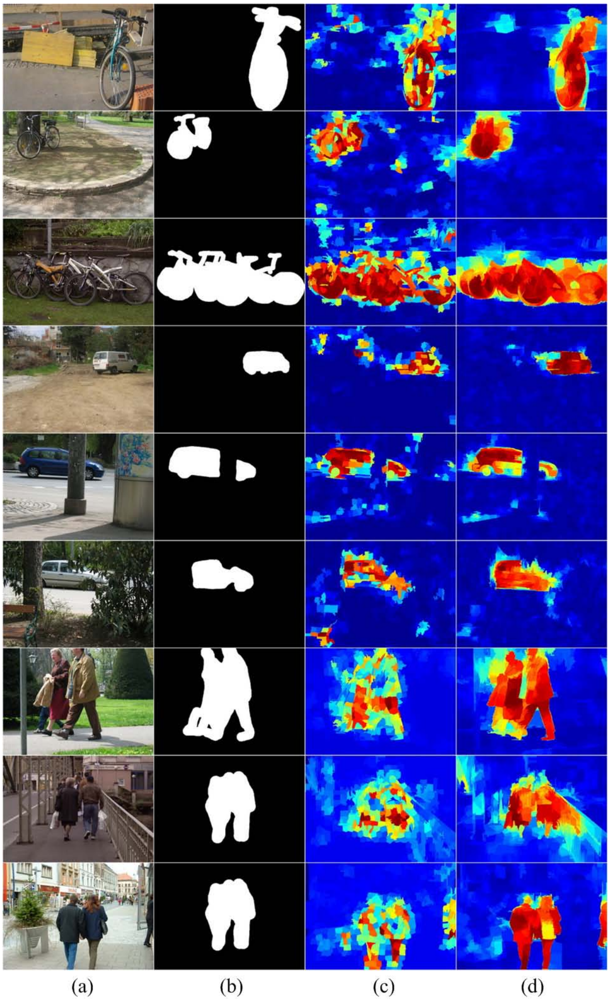


Fig. 3. Representative results on the Graz-02 dataset in which the red pixels represent the predicted foreground region; a) original images; b) ground-truth labels; c) the results of [5]; d) our results. (They are best viewed in color.)

trained classifier, and then multiply $P(c|G, \omega, \nu)$ by the posterior probability $P(c)$ as:

$$P'(c|G) = P(c|G, \omega, \nu) \cdot P(c)^{\alpha} \tag{4}$$

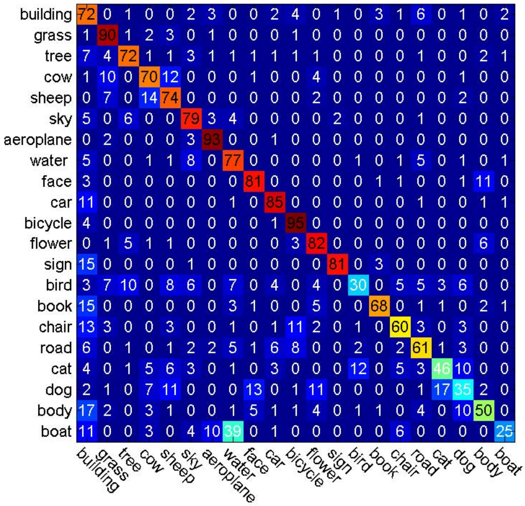
where α is used to soften the prior probability. $P'(c|G)$ is the final confidence value on each category. In section 2.4 we mentioned that the vertical edge is more important than the horizontal edge. To prove it, we show the value of

Table 2. The comparison of the tradeoff parameter ratios between vertical pairwise potential and horizontal pairwise potential for the 21 categories on the MSRC dataset

	building	grass	tree	cow	sheep	sky	airplane	water	face	car	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat
ν/ω	2.0	3.4	2.2	0.3	4.8	1.4	1.3	4.3	2.1	1.2	1.3	3.4	2.1	5.7	1.3	3.3	1.8	6.7	1.5	3.0	1.8

the tradeoff ν/ω for each category trained by stochastic gradient descent [15] in Table 2. We have $\nu/\omega > 1$ for all the 21 categories except cow.

We present the pixel-level confusion matrix of our method on MSRC dataset in Figure 4 and show some representative segmentation results in Figure 5. Since Fulkerson et al. [5] did not experiment on the MSRC dataset, to compare with it, we test their method with our own implementation and show its results in Figure 5(c) and Table 3. Besides of that, we also compare our method with [4] and [18] in terms of segmentation accuracy. As shown in Table 3, our method achieves both the highest global accuracy (total proportion of correctly predicted pixels) of 75% and the highest averaged accuracy of 68%, and performs better than the other methods on 11 categories (more than half of 21). In terms of efficiency, our method takes an average of 8 seconds to process an image, while [18] required 3 minutes per image. Unlike [4] relies on learning from a large pool of features, our method only uses quite simple ones.



	building	grass	tree	cow	sheep	sky	aeroplane	water	face	car	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat
building	72	0	1	0	0	2	3	0	0	2	4	0	1	0	3	1	6	0	1	0	2
grass	1	90	1	2	3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
tree	7	4	72	1	1	3	1	1	1	1	1	1	0	0	0	0	1	0	0	2	1
cow	1	10	0	70	12	0	0	0	1	0	0	4	0	0	0	0	0	0	1	0	0
sheep	0	7	0	14	74	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0
sky	5	0	6	0	0	79	3	4	0	0	0	0	2	0	0	0	1	0	0	0	0
aeroplane	0	2	0	0	0	3	93	0	0	1	0	0	0	0	0	0	0	0	0	0	0
water	5	0	0	1	1	8	0	77	0	0	0	0	0	1	0	1	5	0	1	0	1
face	3	0	0	0	0	0	0	0	81	0	0	0	0	0	1	1	0	0	1	11	0
car	11	0	0	0	0	0	0	1	0	85	0	0	0	0	0	0	1	0	0	1	1
bicycle	4	0	0	0	0	0	0	0	0	1	95	0	0	0	0	0	0	0	0	0	0
flower	0	1	5	1	1	0	0	0	0	0	3	82	0	0	0	0	0	0	0	6	0
sign	15	0	0	0	0	1	0	0	0	0	0	0	81	0	3	0	0	0	0	0	0
bird	3	7	10	0	8	6	0	7	0	4	0	4	0	30	0	5	5	3	6	0	0
book	15	0	0	0	0	0	0	3	1	0	0	5	0	0	68	0	1	1	0	2	1
chair	13	3	0	0	3	0	0	1	0	1	11	2	0	1	0	60	3	0	3	0	0
road	6	0	1	0	1	2	2	5	1	6	8	0	0	2	0	2	61	1	3	0	0
cat	4	0	1	5	6	3	0	1	0	3	0	0	0	12	0	5	3	46	10	0	0
dog	2	1	0	7	11	0	0	0	13	0	1	11	0	0	0	0	1	17	35	2	0
body	17	2	0	3	1	0	0	1	5	1	1	4	0	1	1	0	4	0	10	50	0
boat	11	0	0	3	0	4	10	39	0	1	0	0	0	0	0	6	0	0	0	0	25

Fig. 4. The pixel-level confusion matrix of our method on the MSRC dataset

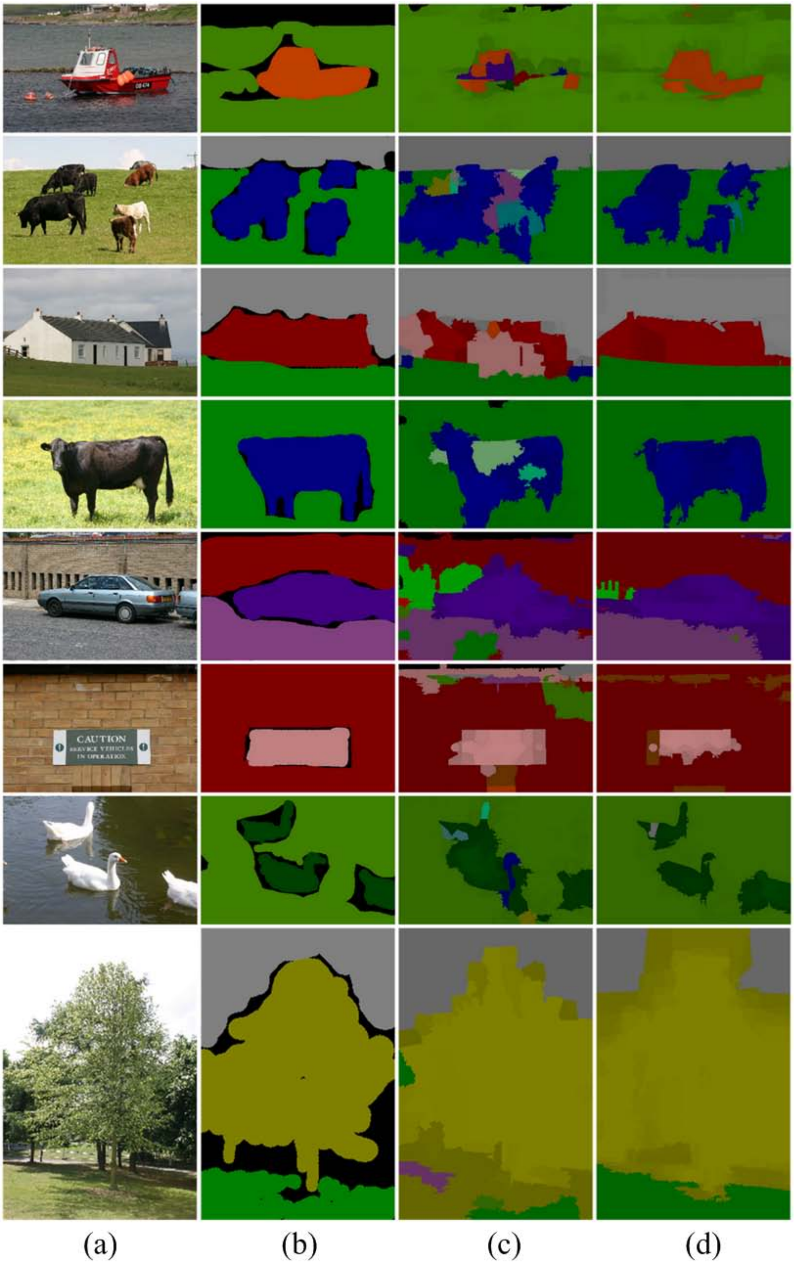


Fig. 5. The results of segmentation and classification for the MSRC dataset; a) original images; b) ground-truth labels; c) the results of [5]; d) our results, each map shows the most confident class at pixel level. (The figure is best viewed in color.)

Table 3. Segmentation accuracy (in percentage) for each class on the MSRC dataset

	Ours	[5]	[18]	[4]
building	**72**	54	62	49
grass	90	73	**98**	88
tree	72	66	**86**	79
cow	70	65	58	**97**
sheep	74	68	50	**97**
sky	79	**89**	83	78
airplane	**93**	90	60	82
water	**77**	65	53	54
face	81	75	74	**87**
car	**85**	76	63	74
bicycle	**95**	89	75	72
flower	**82**	69	63	74
sign	**81**	78	35	36
bird	**30**	24	19	24
book	68	50	92	**93**
chair	**60**	59	15	51
road	61	46	**86**	78
cat	46	53	54	**75**
dog	**35**	31	19	**35**
body	50	55	62	**66**
boat	**25**	23	7	18
Global	**75**	65	71	72
Average	**68**	62	58	67

4 Conclusions

We propose a class-specified segmentation method, which utilizes CRF to integrate the information of multi-scale superpixels under spatial constraints. The proposed method can be used to segment foreground objects from background, and it can also be used for image parsing. The experimental results on the widely used Graz02 and MSRC datasets show that the proposed method is superior to the related methods [4,5,18] and is more simpler and more effecient than the work [9].

Acknowledgement. This research work was supported by Fundamental Research Funds for the Central Universities (2010121067), National Defence Basic Scientific Research program of China(B1420****55), National Natural Science Foundation of China (61170179), Xiamen Science & Technology Planning Project Fund (3502Z20116005) of China, and "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Special Coordination Fund for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
2. Lampert, C., Blaschko, M., Hofmann, T.: Beyond sliding windows: Object localization by efficient subwindow search. In: Computer Vision and Pattern Recognition, pp. 1–8 (2008)
3. Blaschko, M.B., Lampert, C.H.: Learning to Localize Objects with Structured Output Regression. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 2–15. Springer, Heidelberg (2008)
4. Johnson, M., Shotton, J.: Semantic Texton Forests. In: Cipolla, R., Battiato, S., Farinella, G.M. (eds.) Computer Vision. SCI, pp. 173–203. Springer, Heidelberg (2010)
5. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: Computer Vision and Pattern Recognition, pp. 670–677 (2009)
6. Sutton, C., Mccallum, A.: An Introduction to Conditional Random Fields for Relational Learning. In: Getoor, L., Taskar, B. (eds.) Introduction to Statistical Relational Learning, MIT Press (2006)
7. Tighe, J., Lazebnik, S.: Superparsing: Scalable nonparametric image parsing with superpixels. International Journal of Computer Vision (2012)
8. Kohli, P., Ladický, L., Torr, P.H.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vision 82, 302–324 (2009)
9. Ladicky, L., Russell, C., Kohli, P., Torr, P.H.S.: Associative hierarchical crfs for object class image segmentation. In: ICCV 2009, pp. 739–746 (2009)
10. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods. Pattern Analysis and Machine Intelligence (2012)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient Graph-Based Image Segmentation. International Journal of Computer Vision 59, 167–181 (2004)
12. Vedaldi, A., Soatto, S.: Quick Shift and Kernel Methods for Mode Seeking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 705–718. Springer, Heidelberg (2008)
13. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004)
14. Friedman, J., Hastie, T., Tibshirani, R.: Additive Logistic Regression: a Statistical View of Boosting. The Annals of Statistics 38 (2000)
15. Vishwanathan, S.V.N., Schraudolph, N.N., Schmidt, M.W., Murphy, K.P.: Accelerated training of conditional random fields with stochastic gradient methods. In: International Conference on Machine learning, ICML 2006, pp. 969–976. ACM, New York (2006)
16. Murphy, K.P., Weiss, Y., Jordan, M.I.: Loopy belief propagation for approximate inference: an empirical study. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, UAI 1999, pp. 467–475. Morgan Kaufmann Publishers Inc., San Francisco (1999)
17. Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178 (2006)
18. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. International Journal of Computer Vision 81, 2–23 (2009)