A Semantic Labeling Strategy to Reject Unknown Objects in Large Scale 3D Point Clouds

Huifang Ma[1], Lei Shi[2], Sarath Kodagoda[2], Rong Xiong[1]

1. State Key Laboratory of Industrial Control Technology and Institute of Cyber-Systems and Control, Zhejiang University, Zhejiang 310027, China
E-mail: hfma@iipc.zju.edu.cn, rxiong@iipc.zju.edu.cn

2. Centre for Autonomous Systems, The University of Technology, Sydney, Australia
E-mail: Lei.Shi-1@uts.edu.au, Sarath.Kodagoda@uts.edu.au

Abstract: In recent years, there has been a growing interest in the research of semantic labeling of indoor scenes represented by 3D point clouds. A fundamental problem that has largely been oversighted in the current research is the way of dealing with the unknown class which collectively includes all the objects that are of no interest to the application developer. In the training stage, these objects are either completely removed or labeled as unknown, resulting in a trained model which is not fully and fairly exposed to the actual sample space. In the test stage, the unknown objects are naturally present and provided to the classifier, causing a significant drop of the classification accuracy— usually 20%∼30%. Simply improving the features or the classifier will not address the root cause problem. In this paper, we propose a labeling framework combining both Conditional Random Field (CRF) and P_I-SVM to specifically solve the problem caused by the unknown class. First, we use a CRF to model the contextual relations in the 3D space, for which the parameters for both node potential and edge potential are learned from training data. Then, we make use of the rejection strategy of the P_I-SVM, which estimates an unnormalized probability for each class. Finally, we reinforce the result of CRF with the belief provided by the P_I-SVM, and the labeling result is based on the agreement of the two classifiers. The proposed method takes advantage of the global optimization of CRF and the advantage of unknown rejection of P_I-SVM. Experimental results on publicly available data set show that this method has improved the classification accuracy by 10.7% given the accuracy drop of 19.23% caused by the unknown.

Key Words: Semantic labeling, CRF, P_I-SVM, unknown rejection

1 Introduction

Understanding the content and the meaning of a perceived scene is a crucial capability to enable more intelligent autonomous behavior. Semantic scene labeling addresses the core of this problem, namely to decompose a scene into meaningful parts and assign semantic labels to them. For indoor scenes, this is a very challenging task, as they usually contain a large variety of different objects. Fully labeling all the objects is almost impossible since there are innumerable classes. The most commonly used methods manually select the objective classes and train models without unknown, as shown in Figure 1 an example from our experimental result. However, when testing on new scenes, this will cause a drop of accuracy, as the unknown objects will naturally present.

There is a finite set of known objects in myriad unknown objects, combinations and configurations – labeling something new, novel or unknown should always be a valid outcome. However, it is challenging to model the unknown objects as one class in the classifier. On one hand, the unknown class contains many sub-classes which request high generalization ability of the classifier. On the other hand, object classes has infinite potential negative classes which need a strong specialization of the classifier. Simply improving the classifier cannot solve the root of this problem, the unknown class need to be specifically tackled.

According to our knowledge, only a little effort has been paid on this critical problem. Since 2011, Walter *et al.* have been working on the unknown rejection in Support Vector Machine (SVM) [1]. They define the recognition problem with unknown as "open set" and the traditional recognition in which all testing classes are known during training are termed as "closed set". The result outperformed most of the

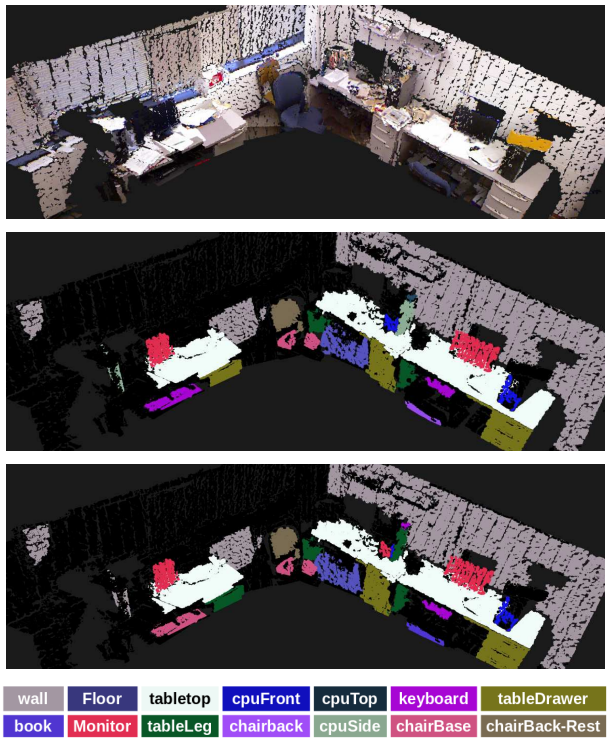


Fig. 1: The topmost is the original point cloud which has been over segmented in the object level, the middle is the ground truth labeling where black represents all unwanted segments and the downmost is the point cloud with predicted labels.

This work was supported by Zhejiang University Lu Graduate Student Education Foundation for International Exchange and the National Nature Science Foundation of China (Grant No.61473258)

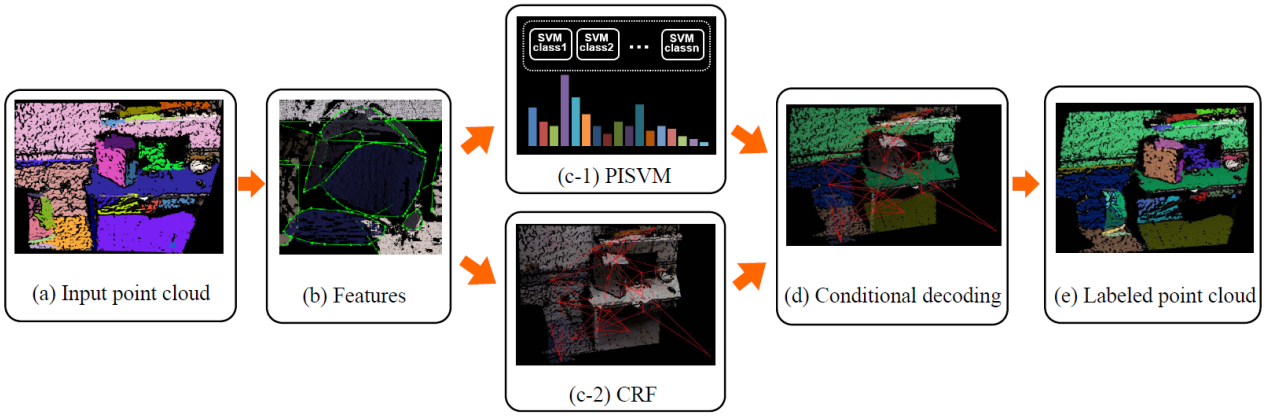


Fig. 2: Outline of the proposed framework. (a) The input data to our algorithm. The point cloud is over-segmented in object level as shown in different colors. (b) First, we extract various features for each segment and segment pair to capture local properties and contextual relations. The picture here shows the convex hull for each segment. (c) Then we generate two labeling results from both P_I-SVM and CRF, using the features extracted in the last step. (d) The consistent labels of the two classifiers (the colored segments) are considered as prior knowledge to conditional decoding in the remaining segments in CRF. (e) The final result is based on the agreement of conditional decoding with the former P_I-SVM. Details about the separate steps are given in section 3.

SVM based classifiers in both binary and multi-class object recognition. However, their exploration aims at improving the stand alone SVM classifier which is designed for independent and identically distributed data. The state-of-the-art methods in semantic labeling also incorporates the neighboring information using graphical models to capture mode information and produce more accurate results. SVM does not take the neighboring information into account and is improper to be used in the semantic labeling directly. Motivated by the partial success of these two methods in this application, we developed a labeling strategy that does not only consider the neighboring informations but also explicitly reject the unknown or unwanted objects.

Our basic idea is modeling the scene using a Conditional Random Field (CRF). CRF captures both the local features and neighboring informations of objects and finds the optimal global labeling result. In dealing with the unknown, CRF has very limited success due to modeling the unknown as one class, which is hard for the reasons mentioned before. In this paper, we combine CRF with the P_I-SVM, a variation of SVM proposed in [2] to deal with the multi-class open set recognition. We test our method on the Cornell office data set [3], which contains 24 segmented point clouds of different office environments. We choose the 17 most commonly seen objects as target classes and regard the remaining as unknown. Informative apparent features have been extracted for each segments (e.g. height, area) and for neighboring segments (e.g. minimal distance) which are later used in the classification stage. Experimental results show that the proposed framework has largely improved the drop of labeling accuracy caused by the unknown.

2 Related Work

There are many studies in the area of semantic labeling using Conditional Random Filed [4]. The robot learning lab of Cornell University has done a significant amount of work ([3][5][6][7][8]) in this area. They segment the point cloud in object level to make full use of contextual relations in the scene, including both the relations between objects and the relations between objects and human. Their CRF model has reached impressive labeling precision, but when testing new scenes [6], they faced about 30% drop in the recall rate.

The work reported in [9] and [10] uses Random Forest Tree (RFT) to initialize the unary potentials of a densely connected CRF. RFT can significantly improve the performance of CRF by increasing the confidence of the more structural classes. However, in their implementation point clouds are over-segmented into supervoxels, so that the node scale is too small to capture informative contextual features. In addition, when including the unknown class, [9] and [10] received a drop of about 30% and 20% in global accuracy respectively.

The advantage of graphical models is that they integrate both the local features and the global information, but alone cannot deal with the unknown objects properly. Although, the Hidden Conditional Random Field [11] can cope with the unknown as hidden nodes in testing, they assume the possible states of the hidden nodes are finite and go through all the potential states to maximize a conditional probability. Thus the total number of classes are actually fixed and known which does not apply in the real world.

Another prevalent technique in scene classification and object recognition is the Deep Learning framework ([12][13][14][15][16]). The main idea is using a Convolutional Neural Network (CNN) to learn a high dimensional features for the image (RGB or depth). One advantage of Deep Learning is that it does not need to manually define any feature. Feature extraction is learned in a unified optimization framework. However, the downside is that Deep Learning cannot give a clear explanation to the meaning of the features, and it is data-consuming, usually need tens of thousands of training samples. The CNN does not explicitly handle the unknown class and according to our knowledge may not be feasible to the large scale unorganized point clouds used in our experiment.

In the domain of machine learning, efforts have been made to solve the unknown class problem, for example, the proposed 1-vs-set machine in [17] for independent, identically

distributed data. Based on the work of Support Vector Machine [18], it models the risk of a sample being far away from the training data in the feature space and adds another hyper plane to reject the unknown. It demonstrates better generalization and specialization ability than the binary one-class SVM [19] which can also reject the unknown class by only modeling the positive samples. The P_I-SVM proposed by Jain *et al* in [2] is a multi-class classifier with unknown rejection considered, and it is built up on single binary classifiers. To avoid the constraint of the Bayes' rule, they calculate a probability in each single classifier separately and leave the probabilities unnormalized. By comparing these unnormalized posterior probabilities, they find the most probable class or reject as unknown if all the values are below a threshold. However, all the above-mentioned SVM-based classifiers treat samples as independent cases without incorporating contextual information. In the scene labeling applications, each test sample is a part of the whole scene and therefore, if they are treated independently, the classifier does not necessarily exploit all the information.

3 Approach

A general overview on our proposed labeling strategy is given in Figure 2. Our approach operates on segmented 3D point clouds. First of all, we extract a feature vector for each segment and segment pair which captures color, shape and pose properties. Then the extracted features are fed to a P_I-SVM as well as a CRF to generate a first stage labeling results. We find the most confident labels according to the agreement of the two classifiers. The relevant nodes of these labels are then treated as prior conditions in the same CRF to relabel the remaining segments. If the relabeled results still disagree with the P_I-SVM, we reject them as unknown. The following sections describe the separate steps of our framework in more detail.

3.1 Features

There are two kinds of features in our model: node features for single segment and edge features between segments. They both capture color informations and geometric properties. Table 1 summarizes the features used in our experiments.

Let's consider λ_{i0}, λ_{i1} and λ_{i2} denote the first three eigenvalues of the scatter matrix, v_{i0}, v_{i1} and v_{i2} denote the corresponding principle components. As for the histogram of base colors, we divide the color space into 10 subspaces−*red*, *orange*, *yellow*, *green*, *cyan*, *blue*, *purple*, *white*, *grey* and *black*, and accumulate colors of every point in the segment to form the bin. The well-known HOG feature which efficiently captures borders in images is not adopted in this work, because in our case the input point clouds are already segmented based on the smoothness and continuity of 3D surfaces.

As for the local shape features, $\lambda_{i0} - \lambda_{i1}$ (linearness), $\lambda_{i1} - \lambda_{i2}$ (planarness), and λ_{i0} (scatterness) are commonly used in the spectral analysis of point clouds. These features do not work effectively in our application. The first two indicators are influenced by the segment size so we change them to $\lambda_{i0}/\lambda_{i1}$ (linearness) and $\lambda_{i1}/\lambda_{i2}$ (planarness). As we have seen the importance of segment area in point cloud registration [20], we replace the scatter with the segment area. To estimate area, the 3D points are projected onto the plane perpendicular to its normal. Then we extract the convex hull of the 2D points and the area is estimated as a polygon.

Table 1: FEATURES

Node features for segment i	
Description	***Dimension***
Average HSV color values	3
Histogram of base color values	10
Convex hull area	1
Linearness $\lambda_{i0}/\lambda_{i1}$, planarness $\lambda_{i1}/\lambda_{i2}$	2
Vertical position of centroid	1
Vertical component of the normal v_{i2_z}	1
Vertical component of the principal direction v_{i0_z}	1
Distance from the scene boundary	1

Edge features between segment i and segment j	
Description	***Dimension***
Difference of average HSV color values	3
Minimal distance between two segment d_{ij}	1
Angle between normals $v_{i2} \cdot v_{j2}$	1
Difference in angle with vertical $acos(v_{i2}) - acos(v_{j2})$	1
Coplanarity	1

The pose features of the segment are also extracted. The data is aligned in a way that $z-$axis is vertical and the ground is at zero height, so the vertical position of centroid is exactly its height which is very informative in our labeling task. Other features include the vertical component of normal to capture horizontal pitch/roll angle, the vertical component of principle direction to capture vertical deviation angle and the minimal distance to the scene boundary to capture horizontal location.

3.2 P_I-SVM

P_I-SVM [2] is a multi-class classifier with unknown rejection option. It is actually a combination of binary SVM classifiers which can estimate the posterior probability of inclusion. Let $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ be the training samples, where x_i is the segment feature vector extracted in the last step and y_i is the corresponding label. First, a kernelized SVM h is trained for each single class l, to generate an uncalibrated hypothesis score s:

$$s = h(x) = \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b \tag{1}$$

where α_i are support vectors, $K(x_i, x)$ is a radial basis function kernel, and b a bias term. All the positive training scores are then fitted with a Weibull distribution [21] to get the single class probability distribution parameters $\theta_l = [\tau_l, \kappa_l, \lambda_l]$. The estimated posterior probability of class l for the input x and class label y conditioned on the parameters θ_l can be calculated as:

$$\begin{aligned} P_I(y|x, \theta_l) &= \xi\rho(l)P_I(x|l, \theta_l) \\ &= \xi\rho(l)(1 - e^{-(\frac{x-\tau_l}{\lambda_l})^{\kappa_l}}) \end{aligned} \tag{2}$$

where $\rho(l)$ is the prior probability of class l, and ξ should be the normalization constant. As the model does not assume that all the classes are known, so the ξ is set to 1 and the posterior estimation is left unnormalized.

For multi-class open set recognition, P$_I$-SVM relies on a minimum threshold δ on class probability to select

$$\begin{aligned} y^* &= \underset{y \in C}{\operatorname{argmax}} P_I(y|x, \theta_l) \\ s.t. \quad & P_I(y^*|x, \theta_{y^*}) \geq \delta \end{aligned} \tag{3}$$

if all the probabilities of inclusion for target classes are below δ, then the corresponding target will be rejected and labeled as unknown.

3.3 Conditional Random Field

In this work the 3D structure of a scene is modeled using a pairwise Conditional Random Field. Given a segmented point cloud $\mathbf{x} = (x_1, ..., x_n)$ consisting of segments x_i, every segment is a node in the CRF model. Any two segments having an intersection in the 3D space will be virtually linked by an edge in CRF. The label y_i for variable x_i is determined by the global observation $\mathbf{x}$. The overall prediction $\hat{\mathbf{y}}$ is optimized as the argmax of a discriminant function $f_{\mathbf{w}}(\mathbf{x}, \mathbf{y})$

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmax}} f_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) \tag{4}$$

where $f_{\mathbf{w}}(\mathbf{x}, \mathbf{y})$ is consisted of a set of parameterized node features and edge features and $\mathbf{w}$ is the parameter set. The optimization problem can be efficiently solved using the Loopy Belief Propagation (LBP) approximation. Let node features and edge features extracted in section 3.1 be denoted as $\phi_n(i)$ and $\phi_e(i,j)$ for segment i and edge(i,j) respectively, the discriminant function is:

$$\begin{aligned} f_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) &= \sum_{i \in V} \sum_{l=1}^{L} y_i^l \cdot e^{w^l \cdot \phi_n(i)} \\ &+ \sum_{(i,j) \in E} \sum_{l=1}^{L} \sum_{k=1}^{L} y_i^l y_j^k \cdot e^{w^{lk} \cdot \phi_e(i,j)} \end{aligned} \tag{5}$$

CRF will assign each segment a label as one of the preset target classes. When testing on new scenes, it will make wrong prediction on all unknown segments.

3.4 Combining CRF and P$_I$-SVM

In this subsection, the proposed framework which combines both the CRF and the P$_I$-SVM will be described. The similarity between the two classifiers is that they can both give a probability estimation for each trained class, but CRF lacks a proper mechanism to reject unknown while P$_I$-SVM does not have the ability to consider contextual informations.

The proposed framework is illustrated in Figure 3. In the first step, CRF and P$_I$-SVM separately produce two sets of labeling results. CRF use both node features and edge features. On one hand, it will benefit a lot since segments of known classes provide more neighboring informations. On the other hand, the unknown will mislead a larger scale segments through the edge connections. Result from P$_I$-SVM is only based on the node features, but it has considered the unknown in the probability estimation. Although P$_I$-SVM

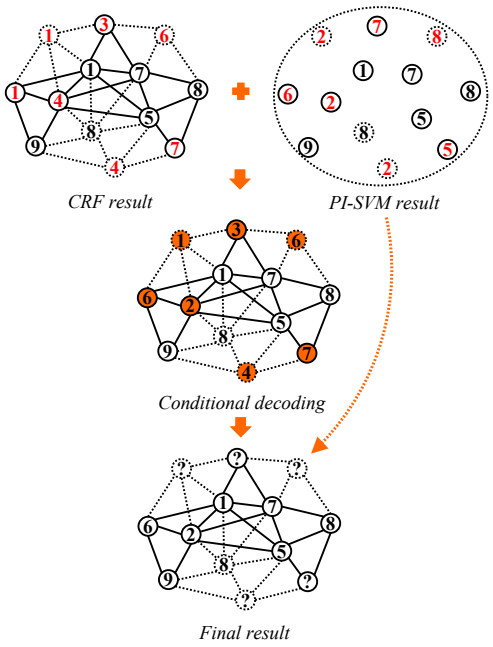


Fig. 3: CRF+P$_I$-SVM labeling framework

alone can efficiently reject unknown, relying on a predetermined threshold is risky when features get less distinct. So firstly we generate a labeling result of P$_I$-SVM according to the maximum probabilities, no matter how small they are.

In a second step, we find the consistent labels of the two classifiers and fix these labels down as conditions in the CRF to conditionally decode the remaining nodes. The consistent labels have a greater degree of confidence and they can provide more certain informations for the graph. If the labels of the remaining nodes have changed to be consistent with P$_I$-SVM, they will be treated as one of the target classes. Otherwise they will be rejected and labeled as unknown.

4 Experiments, Results and Discussion

Our approach is designed to improve performance for semantic labeling with unknown. We want to show two things through our experiments – First, when classifiers do not consider the unknown during training but test data contain them, the existence of unknown will cause a big drop of labeling accuracy; Second, our proposed method can significantly improve the labeling performance with the presence of unknown objects.

4.1 Data

We label object segments in stitched 3D point clouds of the Cornell office data as shown in Figure 1. The data contains 24 scenes with a total of 1108 segments. 17 most commonly seen objects are chosen as the target classes which include {*wall, floor, tableTop, chairBackRest, cpuFront, monitor, paper, tableLeg, keyboard, chairBase, tableDrawer, chairBack, printerFront, book, cpuTop, cpuSide, printerSide*}. These classes have made up 60% of the total data set, and the remaining 40% segments are unknown to our model.

4.2 Result and Discussion

In this subsection we report the result of our experiments. We perform the leave-one-out experiment on the 24 scenes

Table 2: UNKNOWN INFLUENCE IN CRF

Data component	Accuracy				
	Micro P	***Micro R***	***Macro P***	***Macro R***	***Macro F-score***
without unknown class	**66.27**	**66.27**	**55.93**	**52.48**	**54.15**
with unknown class	36.46	59.76	28.93	44.05	34.92

Table 3: UNKNOWN INFLUENCE IN P_ISVM

Data component	Accuracy				
	Micro P	***Micro R***	***Macro P***	***Macro R***	***Macro F-score***
without unknown class	**58.73**	**58.73**	**51.10**	**46.49**	**48.69**
with unknown class	36.20	57.84	33.18	44.99	38.19

Table 4: LABELING RESULTS WITH UNKNOWN

Algorithm	Accuracy				
	Micro P	***Micro R***	***Macro P***	***Macro R***	***Macro F-score***
CRF	36.46	**59.76**	28.93	44.05	34.92
P_ISVM	36.20	57.84	33.18	**44.99**	38.19
CRF+P_ISVM	**58.66**	48.08	**63.00**	35.75	**45.62**

in our implementation, using {*micro precision, micro recall, macro precision, macro recall*} described in [6] to analyze different approaches, and we use the *Macro F-score* as an evaluation criteria for different methods.

$$F\text{-}score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{6}$$

Table 2 and Table 3 show the influence of unknown class in CRF and P_I-SVM respectively. We train a CRF and a P_I-SVM for the 17 target classes and then test on two different data components, one with unknown segments manually removed and another one with unknown segments included. Both models have not consider the unknown during training. When testing on the data with unknown segments, performance of classifiers are supposed to drop significantly compared to the test on the data without unknown, since all the results of unknown will be false positive. The results have confirmed it by having dropped both of the precision and recall. The *Macro F-score* drop about 20% in CRF and 10% in P_I-SVM.

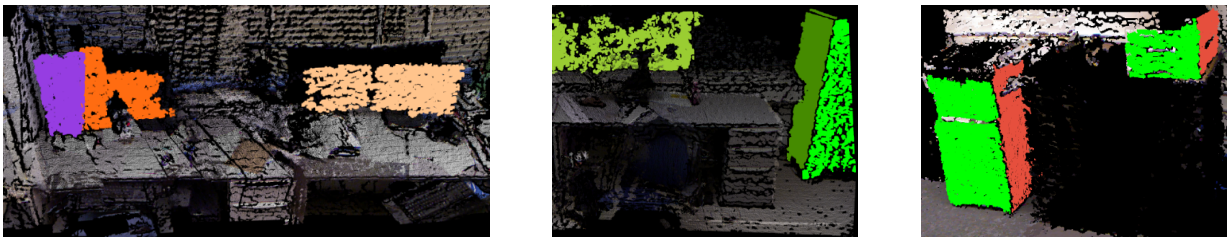
Table 4 shows the labeling result of our proposed method comparing with the results of CRF and P_I-SVM. The CRF+P_I-SVM has much better precision but lower recall, which is reasonable. When the model does not consider the unknown but the data contains them, the unknown objects will be wrongly labeled as a targeted class. This influences a lot on the precision but little on recall, as the results suggest in Table 2 and Table 3. Our proposed method intends to reinforce the strong predictions of the known classes by combining CRF and P_I-SVM, and then reject the less confident segments as unknown. The consistency from two classifiers is a little bit strict for the known which reduces the misclassification among known classes but rejects more segments as unknown, thus increased the precision but decreased the recall. Nevertheless, the *Macro F-score* has improved about 10%.

Discussion a) The overall labeling accuracy is not high mainly because of two reasons. First is the influence of data imbalance. The unknown alone consists of 39% of the data, but almost all the targeted known classes are below 10% except for the *wall* which is 15.1%, and the least known class *cpuTop* only takes a share of 0.7%. Hence test samples are more likely to be labeled as unknown. The other reason is that intra-class variation is bigger than inter-class variation among the known classes. In Figure 4(a), the ground truth of all the three colored segments are monitors, they are of different shapes and sizes. For example the left-most segment, it is more like a cpu-front than a monitor. In Figure 4(b), all the colored segments in the left one are actually wall, while in Figure 4(c), the green segments are printer-front and the pink segments are the printer-side, but both of them look like walls.

Discussion b) CRF is the most traditional way in semantic labeling while P_I-SVM is a relatively new idea in dealing with the unknown. When the unknown are manually removed, CRF shows better result since it contains contextual relations. When considering the unknown, P_I-SVM has a less decrease in *F-score* since it models the probability of unknown. The two classifiers have complementary advantages for this particular problem. The proposed approach makes use of the global knowledge of CRF and the exclusiveness of P_I-SVM and outperforms both of the classifiers in *F-score*. It is a promising attempt in semantic labeling to remove the unknown, although its recall rate needs further improving due to the reasons mentioned before.

5 Conclusion

When a robot is required to perform tasks in new scenes, it will surely face plenty of new objects. In addition, in some task-related work, robot may care more about certain object



(a) (b) (c)

Fig. 4: Some misclassified objects. (a) all the colored segments are monitors; (b) all the green segments are walls; (c) the green segments are printerFront and the pink segments are printerSide.

sets in different situations. Being able to reject the unknown can let robot behave more intelligently and with more flexibility.

In this paper, we addressed the semantic labeling problem using a combination of CRF and P_I-SVM. It has better performance than CRF or P_I-SVM alone when the test data contains large percentage of unknown objects. To our knowledge, this is the first semantic labeling strategy to explicitly reject the unknown segments. We make full use of the probability estimation ability of P_I-SVM and the global optimization ability of CRF. We have shown in our experiment that the unknown class does have a big influence on the labeling result. By using the proposed approach, we have largely solved this problem and increased the overall labeling accuracy.

The main task for our future work is to go further into modeling the unknown in CRF, to improve the CRF by the thought of P_I-SVM. Establishing a rejection mechanism in CRF itself seems to be a more fundamental and promising way in semantic labeling.

References

[1] F. de O. Costa, M. Eckmann, W. J. Scheirer, and A. Rocha, "Open set source camera attribution," in *XXV SIBGRAPI - Conference on Graphics, Patterns and Images*, August 2012.

[2] L. P. Jain, W. J. Scheirer, and T. E. Boult, "Multi-class open set recognition using probability of inclusion," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 393–409.

[3] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, "Semantic labeling of 3d point clouds for indoor scenes," in *Advances in Neural Information Processing Systems*, 2011, pp. 244–252.

[4] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[5] Y. Jiang, M. Lim, C. Zheng, and A. Saxena, "Learning to place new objects in a scene," *The International Journal of Robotics Research*, p. 0278364912438781, 2012.

[6] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena, "Contextually guided semantic labeling and search for three-dimensional point clouds," *The International Journal of Robotics Research*, p. 0278364912461538, 2012.

[7] Y. Jiang and A. Saxena, "Infinite latent conditional random fields for modeling environments through humans." in *Robotics: Science and Systems*, 2013.

[8] C. Wu, I. Lenz, and A. Saxena, "Hierarchical semantic labeling for task-relevant rgb-d perception," in *Robotics: Science and systems (RSS)*, 2014.

[9] A. Hermans, G. Floros, and B. Leibe, "Dense 3d semantic mapping of indoor scenes from rgb-d images," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2631–2638.

[10] D. Wolf, J. Prankl, and M. Vincze, "Fast semantic segmentation of 3d point clouds using a dense crf with learned parameters," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4867–4873.

[11] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 1848–1852, 2007.

[12] C. Couprie, C. Farabet, L. Najman, and Y. LeCun, "Indoor semantic segmentation using depth information," *arXiv preprint arXiv:1301.3572*, 2013.

[13] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in Neural Information Processing Systems*, 2014, pp. 487–495.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] W. Ouyang, X. Wang, X. Zeng, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, C.-C. Loy *et al.*, "Deepid-net: Deformable deep convolutional neural networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2403–2412.

[16] Y. Liao, S. Kodagoda, Y. Wang, L. Shi, and Y. Liu, "Understand scene categories by objects: A semantic regularized scene classifier using convolutional neural networks," *arXiv preprint arXiv:1509.06470*, 2015.

[17] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult, "Towards open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 36, July 2013.

[18] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[19] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[20] J. Xiao, B. Adler, J. Zhang, and H. Zhang, "Planar segment based three-dimensional point cloud registration in outdoor environments," *Journal of Field Robotics*, vol. 30, no. 4, pp. 552–582, 2013.

[21] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001, vol. 208.