



Fig. 3. The windows employed in Exp_2 and Exp_3: (a) Case (ii), (b) case (iii).

Table 4. Average recognition rate with respect to three strategies of windows placement.

	P_R	P_FPD	P_FPD&R
Ave. rec. rate*	69.8	69.7	71.4

*Average recognition rate [%]

5. Discussion

In Exp_1, the positive performance of the VLAD was recognized as shown in Table 1. The idea of VLAD is to put emphasis on the VW which is characteristic to a particular object more than frequency, and it worked affirmatively in the recognition of the 10 objects employed in the experiment.

On the other hand, in Exp_2, the maximum recognition rate was 73.2% when strategy (iii) in section III was employed as windows location and 200 visual words were used with each of the 10 windows. When a single window, an image itself, is employed, which is the original way of using the BoF, the recognition rate is 63.0% as seen in Table 1. Although SPM is a multiple window method, the recognition rate is worse, 69.8%, than the proposed method which is more flexible in windows placement than SPM. This fact indicates the effectiveness of the proposed use of multiple windows in the BoF-based object recognition.

Finally, the proposed method, employing multiple windows and VLAD expression, achieved 74.8% of the recognition rate as given in Table 3. This is the best result at the moment.

As for the three strategies of windows placement, strategy (iii) seems to act better than the other two, which is seen in Table 4. It shows average recognition rates with respect to each windows placement in Table 2. It may, however, be necessary to perform more

experiments to make the superiority certain, since the difference is not very large.

In the employment of BoF, various weights could be considered including frequency of the SIFT feature points [1], VLAD [3], TF-IDF [5] and weighted BoF [6]. But they don't give very high recognition rates to general objects. One may need to improve this in some way.

6. Conclusion

In this paper, multiple window bag of features was proposed which considered positional relation of the feature points on an object. For BoF representation, the vector of locally aggregated descriptors, VLAD, was also employed for recognizing ten familiar objects in a traffic environment. By effective placement of the multiple windows on an image, 74.8% of recognition rate was achieved, which is satisfactory for general object recognition. However, the research should be continued to raise the recognition rate more in order to put the method into practical use.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 25350477.

References

1. G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in *Proc. ECCV: Learning in Computer Vision* (2004), pp. 1-22.
2. S. Lazebnik, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in *Proc. IEEE Computer Vision and Pattern Recognition* (2006), pp. 2169-2178.
3. H. Jégou, et al., Aggregating local descriptors into a compact image representation, in *Proc. CVPR* (2010).
4. D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60(2) (2004), pp. 91-110.
5. W. B. Frakes, R. Baeza-Yates, Information Retrieval: Data Structures & Algorithms. *Prentice-Hall, Engelwood Clifs*, (1992).
6. T. Manabe, J. K. Tan, H. Kim, S. Ishikawa, Recognizing indoor objects by weighted bag of features, in *Proc. IEEE Tensymp* (2014). (submitted)