Semantic Perception, Mapping and Exploration

Acquisition and modeling of semantic information is a key requisite for mobile robots to be deployed in human environments. In this field, fundamental aspects faced by research are: the recognition of places and objects, the construction of semantic models and the exploration strategies to enrich contextual knowledge. In the remainder of this section, we will present relevant work that focused on the mentioned problems, namely: semantic perception, semantic mapping and semantic exploration.

Semantic Perception

Extracting semantic information from visual data is one of the fundamental problems of computer vision. Scene understanding can be decomposed into sub-tasks, depending on the information one is interested to extract from the input data. These sub-tasks can be organized on a progression that goes from coarse to fine grained inference.

Image classification is the task of assigning a semantic label to an input image from a fixed set of categories. Ulrich and Nourbakhsh [24] propose an appearance-based place recognition system for topological localization. They use colour histogram features [21] and a simple voting scheme for nearest-neighbor matching. In a similar fashion, Torralba *et al.* [23] derive an hidden Markov model (HMM) for place recognition and new place categorisation based on the global statistic feature retrieved from texture [15]. In contrast, Lisin *et al.* [11] propose to model classes of images as a probability distribution over local features, in order to be combined with global features. This method has proven to perform well in applications where a rough segmentation of objects is available.

Object detection consists of making a prediction not only of object categories but also of their spatial locations. A seminal work can be considered that of Viola and Jones [25], who proposed a fast and robust face detection. Their method makes use of Haar-like features [17] to search for likely face candidates, which can then be refined using a cascade of more expensive but selective detection algorithms [8]. Likewise, a well-known example of pedestrian detection has been proposed by Dalal and Triggs [5], who use a set of overlapping Histogram of Oriented Gradients (HOG) descriptors fed into a Support Vector Machine (SVM) [4].

Image segmentation is the task of finding groups of pixels that possess some "similarity" and is one of the oldest and most widely studied problems in computer vision. Early techniques focus on local region merging and splitting [14, 1], while, more recent algorithms often optimize some global criterion, such as intra-region consistency and inter-region boundary lengths or dissimilarity [3, 19, 7, 2, 16].

Despite the popularity of the presented methods, a recent breakthrough in scene understanding has been the adoption of Convolutional Neural Networks (CNNs) [9]. Krizhevsky *et al.*in [10] present the pioneering deep CNN that,

despite its simplicity, won the Imagenet 2012 classification challenge with wide margin on the closest competitor. Similarly, different object detection methods based on deep neural netowrks have shown to outperform the state-of-the-art [18, 6, 12]. Consequently, the capabilities of such networks have been also investigated in pixel-level labeling problems like semantic segmentation. In this context, a milestone is the work of Long *et al.* [13] who transformed existing classification models ([20, 22]) into fully convolutional ones to output spatial maps instead of classification scores. One of the main reason behind its popularity is that, with this approach, CNNs can be trained end-to-end and efficiently learn to make dense predictions with inputs of arbitrary size.

Semantic Mapping

Semantic Exploration

References

[1] Claude R Brice and Claude L Fennema. Scene analysis using regions. *Artificial intelligence*, 1(3-4):205–226, 1970.

[2] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277, 2001.

[3] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(5):603–619, 2002.

[4] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893. IEEE, 2005.

[6] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. Scalable object detection using deep neural networks. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2147–2154, 2014.

[7] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *Intl. Journal of Computer Vision (IJCV)*, 59(2):167–181, 2004.

[8] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[9] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[11] Dimitri A Lisin, Marwan A Mattar, Matthew B Blaschko, Erik G Learned-Miller, and Mark C Benfield. Combining local and global image features for object class recognition. In *Computer vision and pattern recognition-workshops, 2005. CVPR workshops. IEEE Computer society conference on*, pages 47–47. IEEE, 2005.

[12] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, pages 21–37. Springer, 2016.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[14] Ron Ohlander, Keith Price, and D Raj Reddy. Picture segmentation using a recursive region splitting method. *Computer Graphics and Image Processing*, 8(3):313–333, 1978.

[15] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Intl. Journal of Computer Vision (IJCV)*, 42(3):145–175, 2001.

[16] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[17] Constantine P Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, pages 555–562. IEEE, 1998.

[18] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

[19] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(8):888–905, 2000.

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[21] Michael J Swain and Dana H Ballard. Color indexing. *Intl. Journal of Computer Vision (IJCV)*, 7(1):11–32, 1991.

[22] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, et al. Going deeper with convolutions. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[23] Antonio Torralba, Kevin P Murphy, William T Freeman, and Mark A Rubin. Context-based vision system for place and object recognition. In *null*, page 273. IEEE, 2003.

[24] Iwan Ulrich and Illah Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, volume 2, pages 1023–1029. Ieee, 2000.

[25] Paul Viola and Michael J Jones. Robust real-time face detection. *Intl. Journal of Computer Vision (IJCV)*, 57(2):137–154, 2004.