

Towards Dynamic Object Detection using Key-point matching and Super-pixel Segmentation

Ming Liu, Cédric Pradalier

[ming.liu, cedric.pradalier]@mavt.ethz.ch

April 25, 2010

Abstract

There are two-folded purposes of the work described in this paper. First, object-based environment reasoning is a hot-point of the concurrent researches for indoor robots. The recognition of dynamic objects will greatly help to understand the property and context of the specified environment. For example, it can be a reference for the background extraction. Second, because vision based navigation techniques could be easily affected by appearance changes in the environment, the detection of dynamic objects can be used to ignore the replacement of the objects (visual features) in the environment. Our work will search in the space of outliers defined by RANSAC result, and it manages to find vision features' displacement that can be grouped into the same groups, referring to the results of super-pixel segmentation.

1 Introduction

Nowadays, several key-point features have been greatly developed, such as SIFT [5] SURF [1] etc. These techniques enriched the description of objects and environment. Nevertheless, there are some inevitable inherent limitation of the single key-point features considering the application background of object-based methods. First, the self-similarity of features; second, the feature can only describe a relatively small area in the image without more general meanings in a larger scale. Usually, the first problem can be partly solved by different RANSAC methods, such as 8-point method [6], 5-point method [4], 1-point method [8] etc. The second problem can be analyzed at a methodology level, i.e. one single part of a union cannot represent the whole. In another word, the connection or description of the relationship between these single parts is missing. To work with this problem, the links between the features should be established. Considering the property of a solid object, usually the structure of a concrete object will not change significantly. It means that if we consider the color (ARG) attribute relational graph [2, 9, 7] of a certain object, the ARG is relatively fixed. Therefore, the relation graph of these attributes (colors) can represent the object. Furthermore, if we find a way to describe the fixed relations of colors, it will be a feasible way to link the single key-points, granting a more general meaning of the descriptor. One of the top cited works in the color based description is the color based super-pixel segmentation technique [3]. It used the possibility-based method to reason the neighbors of single pixels, looking for the similarities in color space.

Now that we have found a way to describe the objects, i.e. the super-pixel segmentation of the image combining with the single key-point descriptors. Another following problem in consequence is how can we select or filter the features that are potentially belonging to the same object. Regarding the application context, we can assume that only some objects can move in the environment, such as a chair, a box of milk on the table etc. In another word, most of the features which can be considered as background are static. When we do the scene images matching, RANSAC must be implemented to select the static related features from the two images. Usually, people pay much attention to the true positive matches, namely inliers, which can provide more ground truth. In this work, the outliers will be more interesting than inliers, because they are exactly what we are looking for, as they are the dynamic features against the static background. By grouping these outliers, the dynamic objects can be appeared.

The results of the detection of the dynamic objects can be used in two different topics. First it will help the feature selection for the vision based mobile robot navigation, e.g. removing the features from moving people; second it can define the mobility properties of objects and moreover the reasoning of the environment.

2 Approach

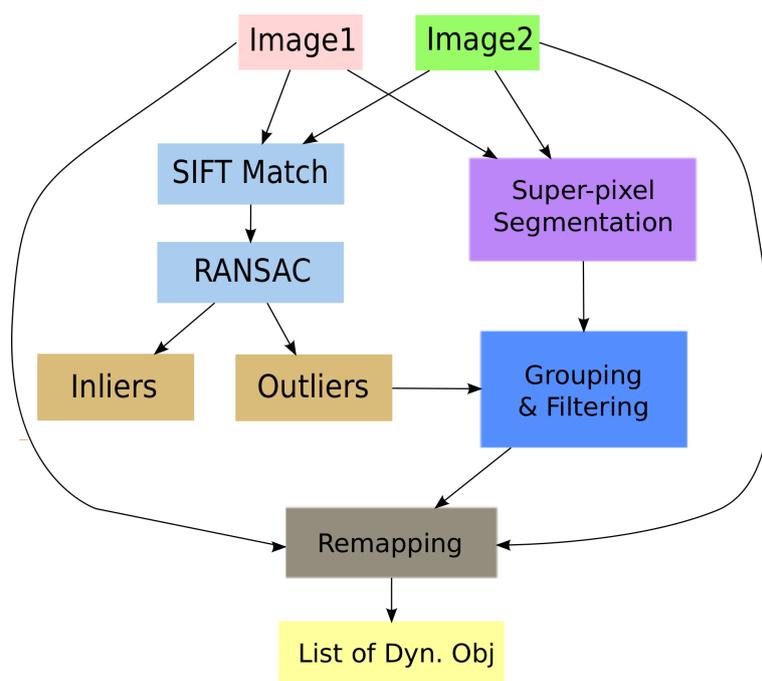


Figure 1: The process of dynamic object detection.

Figure. 1 shows the process of our approach. It starts with the loading of two scene images taken at the same place. Our test sensor is an omni-directional camera, therefore it can capture the 360° surroundings. The wide field of view in horizontal direction indicates that it will record the changes in a environment without taking the alignment into account. SIFT extraction will be implemented on both images. A RANSAC method is performed after the matching. As we mentioned previously, the outliers are our target features. They will be grouped by super-pixel segmentation. One important fact is that usually the outliers are caused by the self-similarity of the descriptors of the features, but these features can be considered as pure noise and usually

show up as single isolated point in the image. Therefore, they can be ignored directly. After we defined which features can be grouped within one super-pixel, the rest of the work is to remap the coordinates of the selected points to the original image and localize the dynamic objects.

Our method will try to detect maximum 10 dynamic objects in the image. Due to the fact that sometimes one object may be divided into more than one super-pixels, there are duplicated detections in the result.

3 Evaluation



Figure 2: Result of Dynamic object detection. The red rectangles mark the detected dynamic objects from the image pair. The green circles indicate the inliers of the key-point matching, while the red ones indicate the outliers of this scene matching.

Considering the pair of images shown in Figure 2, we can intuitively see that the yellow bag has been moved from left to the right. The intermediate result of the super-pixel segmentation

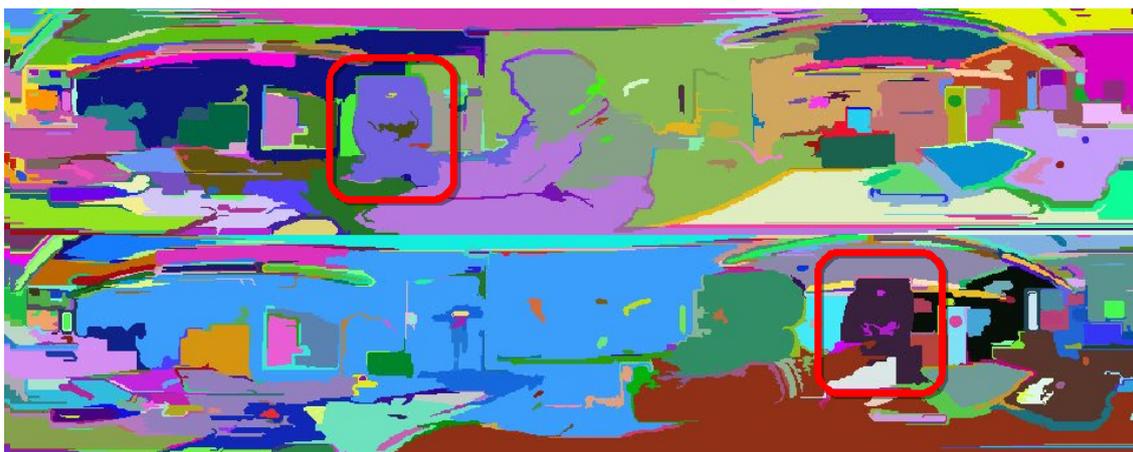


Figure 3: Result of the super-pixel segmentation. The red rectangles mark the detected dynamic objects.

is shown in Figure 3. By selecting proper parameters for the segmentation according to the nature of the scene image, most of the object-parts with the same color can be divided into the same super-pixel. The distribution of the outliers in Figure 2 also shows that the outliers can

be diverse into two types. Some of them can be called “true outliers”, which are caused by the self-similarity of the feature; the rest are caused by the dynamic objects, which implies the potential dynamic objects. The former can be refined by counting the number of features in one super-pixel. After the refinement, the result is marked by red rectangles. The segmentation time is around 865ms and the time for the dynamic object detection is 14.9ms.

4 Outlook

The following tests are still needed to be carried out:

1. A real-time test on mobile robot.
2. A test of the images taken at not exactly same places, and with different orientation of the objects.
3. Fusing the information of dynamic object detection with other tasks of the robot.

References

- [1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Lecture notes in computer science*, 3951:404, 2006.
- [2] M. Eshera and KS Fu. A similarity measure between attributed relational graphs for image analysis. In *Proc. 7th Int. Conf. Pattern Recognition*, pages 75–77, 1984.
- [3] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [4] M. Lhuillier. Automatic structure and motion using a catadioptric camera. *sign*, 8:d0T.
- [5] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [6] D. Nistér. Preemptive RANSAC for live structure and motion estimation. *Machine Vision and Applications*, 16(5):321–329, 2005.
- [7] B.G. Park, K.M. Lee, S.U. Lee, and J.H. Lee. Recognition of partially occluded objects using probabilistic ARG (attributed relational graph)-based matching* 1. *Computer Vision and Image Understanding*, 90(3):217–241, 2003.
- [8] D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In *EEE International Conference on Robotics and Automation*. Citeseer.
- [9] D.Q. Zhang and S.F. Chang. Detecting image near-duplicate by stochastic attributed relational graph matching with learning. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 877–884. ACM, 2004.