ROBUST DETECTION OF MOVING OBJECTS IN VIDEO

Lukáš Klicnar
Master Degree Programme (2), FIT BUT
E-mail: xklicn00@stud.fit.vutbr.cz

Supervised by: Vítězslav Beran
E-mail: beranv@fit.vutbr.cz

Abstract: This paper presents a method for an image segmentation to regions with a coherent motion. It is based on sparse features tracking, a short-range track repair improves the tracking robustness. The motion segmentation is performed by a RANSAC-based algorithm and the Voronoi tessellation provides a dense segmentation of the image. The work also compares the usage of different features.

Keywords: Motion segmentation, moving objects detection, feature tracking, track repair.

1 INTRODUCTION AND RELATED WORK

Motion segmentation is an important task, it allows an extraction of moving objects from a background. Commonly used algorithms are often based on background models and practically don't work with shots taken by a moving camera. This article contains overview of an approach that is capable to deal with a wide spectrum of scenes and provides a fast but approximate segmentation. It was designed for application in robotics with requirements to high speed and online processing.

This work is mainly based on two papers. In [2], authors benefit from a tracking of affine-covariant regions, which are far more discriminating than interest points, but computationally very intensive (about 130s to process a one frame). The main goal of this work is to adapt their approach to different features to achieve higher speed. A simplified method is used in [1], authors also introduce there a spatial proximity constraint and an object tracking capable of handling splitting and merging.

2 ONLINE MOTION SEGMENTATION ALGORITHM OVERVIEW

A block diagram is in Fig. 1. Features are tracked, so their trajectories are obtained, and a short-range track repair is used to increase the tracking robustness. Dominant motion groups of similar tracks are extracted and they are partitioned to meet a spatial proximity constraint of their tracks, that results to regions with a coherent-motion. Then their correspondence across frames is solved. Finally, the Voronoi tessellation is used to get the dense segmentation of the whole image (example in Fig. 2).

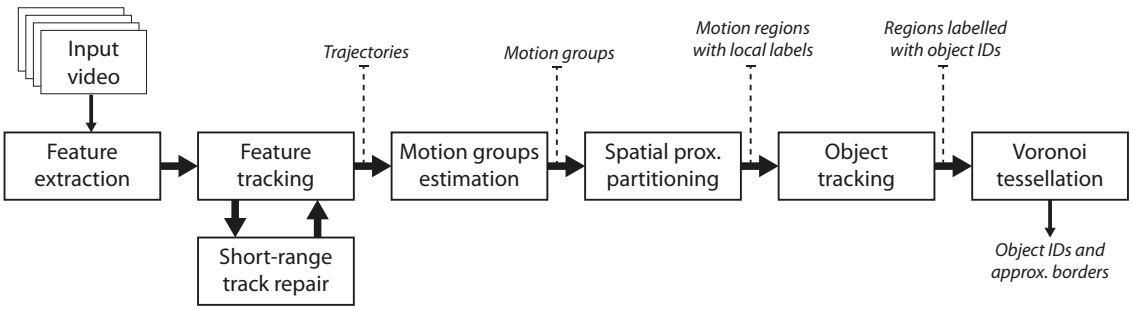

Figure 1: A block diagram of the proposed system for a moving objects detection.

2.1 FEATURE EXTRACTION AND TRACKING

The first step is the FAST feature tracking, their correspondence is estimated according to the BRIEF descriptors distance. Features are detected in every frame, with a grid-adapted detector for a better spatial distribution. This means that the frame is divided into a grid and a minimal number of features has to be present in each cell. This improves the overall frame coverage by tracked features.

The matching algorithm remains the same as in [2]. The feature correspondence is always estimated between the current and the previous frame. The corresponding feature is searched in a small distance from its position in the previous frame and the best match (the lowest distance of BRIEF) is found. Matches with a low normalized cross correlation are removed, that significantly reduces the number of outliers. Compared to [2], there is no RANSAC step, it discards only a very few outliers.

2.2 SHORT-RANGE TRACK REPAIR

To be robust, the motion segmentation needs trajectories ideally with no outliers and as long as possible. The first should be resolved by the matching algorithm, but the tracking can still fail (object occlusion, detection below a threshold, etc.). Some of them can be overcome by a short-range track repair from [2]. This means that terminated trajectories aren't considered dead immediately, but propagated into the next frames by estimating their position from spatially nearest trajectories. A small neighbourhood from the predicted position is searched for the feature occurrence, the matching is performed the same as in Sec. 2.2. If it's not successful in 5 frames, the trajectory is terminated.

2.3 MOTION GROUPS ESTIMATION

The goal is to divide trajectories into groups with a coherent motion. I assume that tracks with the similar motion belong to a single object, which is basically true for rigid objects only. The frame-to-frame motion of every point is described by a homography. The initial motion segmentation for every frame is obtained by the following RANSAC-based algorithm (a modified approach from [1]):

Four-tuples of tracks are randomly taken and a homography matrix (representing a motion between the current and the previous frame) is estimated. This is done iteratively until the desired minimal reprojection error or the maximal number of iterations is reached. In opposite to [1], the most dominant motion is chosen as a homography with the minimal reprojection error, not with the highest number of inliers. This allows to stop iterating if a sufficient transformation is found, which leads to a higher speed. Inlying trajectories are then removed as a motion group and the rest of them goes through the same process, so a one dominant motion group is extracted in each iteration.

Motion groups are based only on a motion similarity. For example, if two objects are moving in a similar way, they are grouped together even though they occur at different places. This can be resolved by a partitioning groups according to their spatial proximity (from [1]). Every group G is divided, that for each point from G, the distance to its nearest neighbour is lower than a threshold T.

2.4 MOVING OBJECTS TRACKING AND IMAGE SEGMENTATION

To track moving objects, the correspondence of motion regions across frames must be solved. This is considered to be 1:N, so multiple regions can match a single (oversegmented) object. For now, only a simple approach based on [1] is used. Every motion group is matched with an object from the previous frame with the highest number of common trajectories. If there is none, new label is generated. This approach cannot handle a region splitting, e.g. a formerly still object remains assigned to a background even when it started moving, its label is propagated from the former occurrence. A dense image segmentation is obtained by the Voronoi tessellation (Fig. 2). The input are the current positions of trajectories in this frame, all pixels in each cell are labelled the same as the cell generator.

(a) (b) (c)

Figure 2: An example of output from particular steps of the motion segmentation. Trajectories of tracked features (a) are divided into motion groups (b) – the color of trajectory corresponds to a motion group membership. A dense segmentation is obtained by the Voronoi tessellation (c).

3 CURRENT RESULTS

Currently, I am focusing on a performance of FAST+BRIEF features compared to others. Experiments were made to show a feature stability (track length), results are summarized in Table 1. FAST+BRIEF are much worse suitable for tracking than SIFT/SURF, but sufficient after the short-range repair. A huge advantage is their speed – 6.6x higher than using SIFT. The motion segmentation fails mostly on poor-quality data (camera shaking, object motion blur). In this case, no feature is detected, which leads to the tracking failure. Other detectors suffer less to this problem.

Feature	**Track length (without repair)**	**Speed (fps)**
SIFT	67.3 (13.4) frames	1.0x (0.25 fps)
SURF	64.8 (17.2) frames	1.8x (0.45 fps)
SURF+BRIEF	63.0 (17.0) frames	2.1x (0.53 fps)
SIFT+BRIEF	52.8 (12.5) frames	2.2x (0.54 fps)
FAST+BRIEF	38.4 (7.4) frames	6.6x (1.64 fps)

Table 1: Average track lengths for tracking different features. Only trajectories longer than 2 frames are accounted. All detectors were set to provide approx. 1000 features in every frame. The speed includes the whole motion segmentation. A set of videos (720x406px) containing various moving vehicles was used, total approx. 20000 frames, moving objects last for average approx. 140 frames.

4 CONCLUSION

FAST+BRIEF proved to be usable for the motion segmentation. The slowest part, RANSAC segmentation, takes about 50 % of running time. I will try to reduce it by introducing a track locality to the random sampling and constraining homographies for fast discarding of invalid transformations.

ACKNOWLEDGEMENT

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the European Community's 7th Framework Artemis JU grant agreement n. 100233 (R3-COP).

REFERENCES

[1] Basharat, A., Zhai, Y., Shah, M.: *Content Based Video Matching Using Spatiotemporal Volumes*. Computer Vision and Image Understanding, June 2009.

[2] Sivic, J., Schaffalitzky, F., Zisserman, A.: *Object Level Grouping for Video Shots*. In Proceedings of the 8th European Conference on Computer Vision, April 2004.