Moving Object Detection from a Moving Camera

Masahiro Ogawa, Atsushi Yamashita

Yamashita An Lab, Department of Precision Engineering, University of Tokyo, Japan
Mini thesis
July 12, 2023

ogawa@robot.t.u-tokyo.ac.jp
yamashita@robot.t.u-tokyo.ac.jp

Abstract

We develop the state of the art accuracy moving object detection from a moving camera. The basic idea of our algorithm is if the camera is moving straight forward, we extract moving object as strange flow considering the computed focus of expansion, otherwise we assume the camera is rotating and extract moving objects as undominant flow area. At the same time, to reduce false positives, we incorporate optical flow and semantic segmentation information. Our approach accomplish state of the art accuracy on our qualitative evaluation.

Figure 1. Example output of our moving object detection from a moving camera.

1   Introduction

Separating moving and static objects is very important for real environment applications. Such applications include obstacle avoidance for autonomous vehicles, and scene understanding for assistant robots.

2   Rerated Work

There have been several papers on moving object detection from a moving camera. One of the famous review paper about moving object detection from a moving camera is Chapela et al's one [1]. They categorized moving object detection methods, but did not do quantitative evaluation nor create a ranking. Another famous review paper is X.Zhao et al's one [8]. They categorized methods and provided rankings. The highest-ranking method is Motion U-Net2 [3]. They used grayscale monocular video as an input and combined three channels; grayscale image, background subtraction image, and flux mask. However, this method only works for slightly moving cameras and cannot handle large camera motions.

On the other hand, the best method on paperswithcode web site of the category; "Semi-Supervised Video Object Segmentation on Davis (no YouTube-VOS training)" is "HMMN" [4]. However this method requires human inputs of target initial positions, which means this method is not moving object detection in our definition but is just a tracking algorithm.

In addition to search reviews, we also search state of the art approaches.

W.Zhang et al [7] developed a optical flow orientation based approach. They used the orientation of the optical flow between adjacent frames and calculates the background orientation field. The motion saliency is obtained by the difference between the reconstructed background orientation and the region orientation. Their results looks good, but they didn't provide the quantitative comparison between current state of the art methods.

Hu an Uchimura [2] developed Focus of Expansion (FoE) based approach to extract moving objects from a moving camera. This method estimates the camera rotation and translation using precomputed FoE. Then, if the rotation compensated optical flow is not the one which expected for static objects, it will consider the pixel comes from moving objects. However, this method assumes a fixed FoE because it assumes car mounted camera. Therefore if the camera is handheld, as our assuming case, it will not work anymore.

One of the current state-of-the-art approach is Yang et al's deep adversarial network based approach [6]. They used an adversarial network for optical flow information. The generator tries to create a mask that is hard to estimate the inside mask flow, while the in-painter tries to estimate the flow inside the mask. The loss is computed by the information reduction rate. Although this method performed well on their evaluation dataset, but had some issues when we tested on our dataset. There are some false positives in low-textured areas and large object motion regions, where is hard to estimate optical flow from surrounding flow.

3 Proposed Method

The concept of our method for detecting moving objects from a moving camera is 1. extract self motion using FoE, 2. combine optical flow and texture cues. To detect moving objects, we found that estimating self camera motion at the same time is necessary. The overall abstract procedure is in
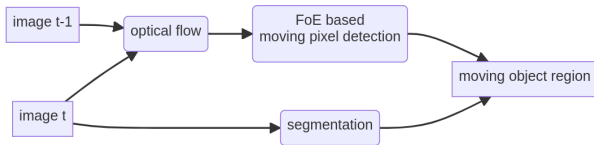
Figure 2. Flow chart of proposed method. Rectangle box means an object and rounded box means a process.

Figure 2.

And the detail of "FoE based moving object pixel" in Figure 2 is in Figure 3.
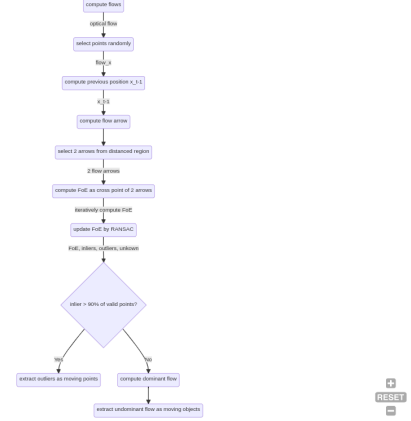
Figure 3. Detail flow chart of FoE based moving pixel detection

To compute optical flow, we adopt top ranked method in paperswithcode, which is named UniMatch[5]. This approach extracts features by CNN and match features by transformer.

—from here, just voice input—

We assume a small area for computing the focal length as the camera is assumed to be rotating. We use the dominant flow in the background as the background flow and extract the different orientation flows for moving objects. By computing the focal length and having a sufficiently large area, we use the outliers as moving objects and combine them with segmentation information to extract the final moving object region. Our approach does not involve refining the rotation as we cannot estimate both the focal length and rotation simultaneously. We use the RunSAC algorithm to compute the focal length and rotation for each candidate.

4 Evaluation

4.1 Qualitative Evaluation

We used our own dataset to compare the results of our method with state-of-the-art methods, including Rosario Network. Our approach successfully extracted moving object areas compared to the failure of Rosario Network in large floor areas.

4.2 Quantitative Evaluation

For quantitative evaluation, we used the Davis dataset, which is commonly used for moving object detection. However, we have not yet finished the implementation, so there are no quantitative evaluation results available at this time.

5 Summary

In this paper, we have presented a state-of-the-art accuracy method for extracting moving objects from a moving camera. Our method combines flow and segmentation information to achieve better results compared to existing state-of-the-art methods like Rosario Network.

5.1 Future Work

In future work, we plan to:

- Estimate counter motion using deep neural networks instead of using traditional methods.

- Utilize nerve object scene representation to extract security representation and objective information.

- Apply this method to 3D reconstruction.

References

[1] M.-N. Chapel and T. Bouwmans. Moving objects detection with a moving camera: A comprehensive review, 2020.

[2] Z. Hu and K. Uchimura. Multiple moving objects detection and simultaneous tracking from the time-varied background. *IEEJ Transactions on Industry Applications*, 120:1134–1142, 09 2000.

[3] G. Rahmon, F. Bunyak, G. Seetharaman, and K. Palaniappan. Motion u-net: Multi-cue encoder-decoder network for motion segmentation. In *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*, pages 8125–8132. IEEE, 2020.

[4] H. Seong, S. W. Oh, J. Lee, S. Lee, S. Lee, and E. Kim. Hierarchical memory matching network for video object segmentation. *CoRR*, abs/2109.11404, 2021.

[5] H. Xu, J. Zhang, J. Cai, H. Rezatofighi, F. Yu, D. Tao, and A. Geiger. Unifying flow, stereo and depth estimation. *arXiv preprint arXiv:2211.05783*, 2022.

[6] Y. Yang, A. Loquercio, D. Scaramuzza, and S. Soatto. Unsupervised moving object detection via contextual information separation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[7] W. Zhang, X. Sun, and Q. Yu. Moving object detection under a moving camera via background orientation reconstruction. *Sensors*, May 2020.

[8] X. Zhao, G. Wang, Z. He, and H. Jiang. A survey of moving object detection methods: A practical perspective. *Neurocomputing*, 503:28–48, 2022.