Robust Object Detection and Tracking in a Cluttered Scene

Xiuwei Zhang

G20904721

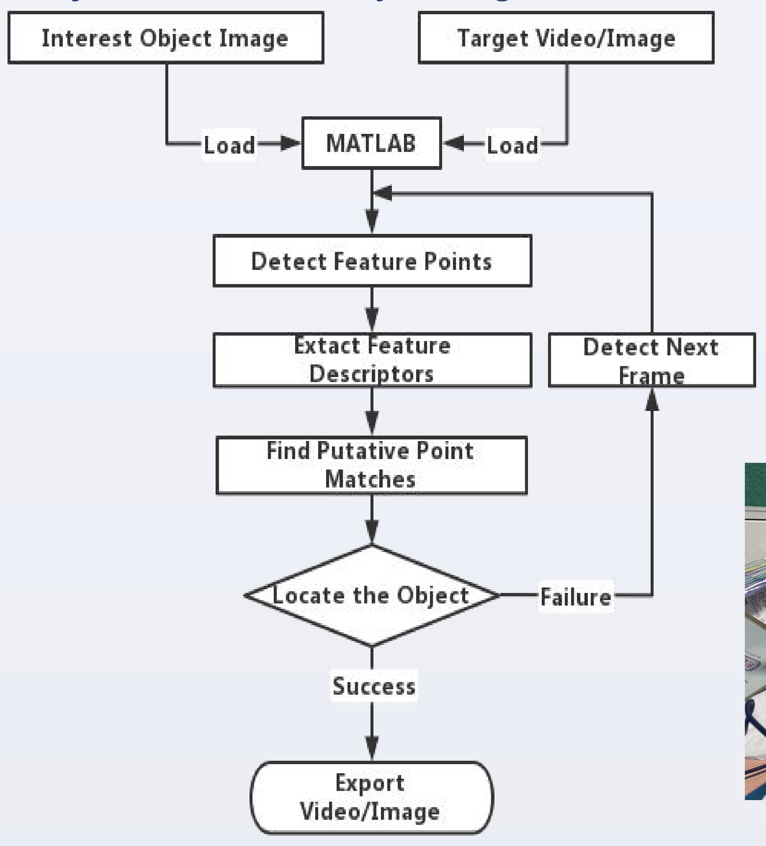
University of Central Lancashire

Abstract

This paper investigates how to perform robust object detection and tracking in a cluttered scene. This article describes the principles of SURF and SIFT algorithms and their workflow when performing object detection, and their applications to object detection in pictures, videos, and when using webcams are discussed. In addition, we analyze the impact of factors such as occlusion, brightness, and background clutter on the SURF and SIFT algorithms. Finally, the advantages and disadvantages of these algorithms are summarized. This research result is of practical application for object detection and tracking in security, and intelligent transportation fields.

Objective and Methods

The purpose of this article is to perform robust object detection and tracking in a cluttered scene by using SURF and SIFT algorithms. The following is the flowchart for object detection by using SURF and SIFT.



1. Read the image of interest and the video or image to be detected.
2. Follow the process shown in the flowchart to perform the detection.
3. Export the image.

Note: The principle of both video and webcam detection is to extract each frame of the video for loop detection.detection.

Figure 1:Result of webcam's object detection in cluttered environment

Result

Occlusion Area Result

As shown in the figure on the right, we divide the object to be detected into five horizontal sections and five vertical sections, each representing 20% of the occlusion area.We selected ten groups of image, which were detected at 20%, 40%, 60% and 80% of the masked area, respectively. And the experimental results were given in the following table.

Figure 2:Image divided into several equal parts

Image blocking area (%)	Success rate of SIFT(%)	Success rate of SURF(%)
20%	100%	70%
40%	100%	90%
60%	100%	90%
80%	40%	70%

Table 1:The impact of the occlusion proportion on the detection

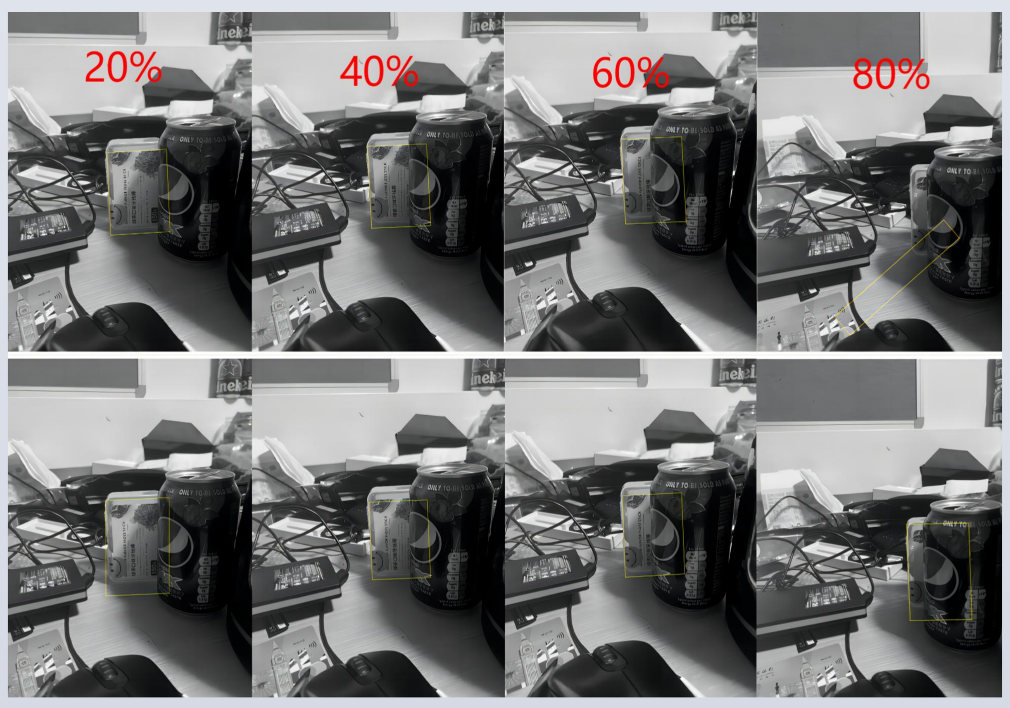


Figure 3：Detection results when using SIFT (top) and SURF (bottom) with different masking areas

When the image occlusion reaches 80%, the SURF algorithm outperforms SIFT because the SURF algorithm preserves as much as possible the number of key points identified while retaining the image structure information. However, when the occlusion rate is less than 60%, the SIFT algorithm has better stability than SURF because the SIFT algorithm can better retain the detailed information in the image, thus improving the matching accuracy.

Different Brightness Result

In this section, we calculate the brightness of the image through MATLAB, . HSV (Hue, Saturation, Value) is a color method tied to RGB, and here we use Value for evaluating brightness. The Value of "very dark", "dark", "normal", "bright" and "very bright" are set to 0.08~0.12, 0.2~0.35, 0.5~0.6 0.65~0.7 and greater than 0.75, respectively. In this experiment, a total of five photos groups of photos were taken, and the data were summarized in the table below.

Success Rate（%）	Very Dark	Dark	Normal	Bright	Very Bright
SURF	40%	80%	60%	60%	60%
SIFT	60%	100%	100%	80%	60%

Table 2: Detection in different lightness conditions

Combining the knowledge of the theory and the experimental results shown above, it can be concluded that SIFT is more robust than SIFT in different brightness environments.

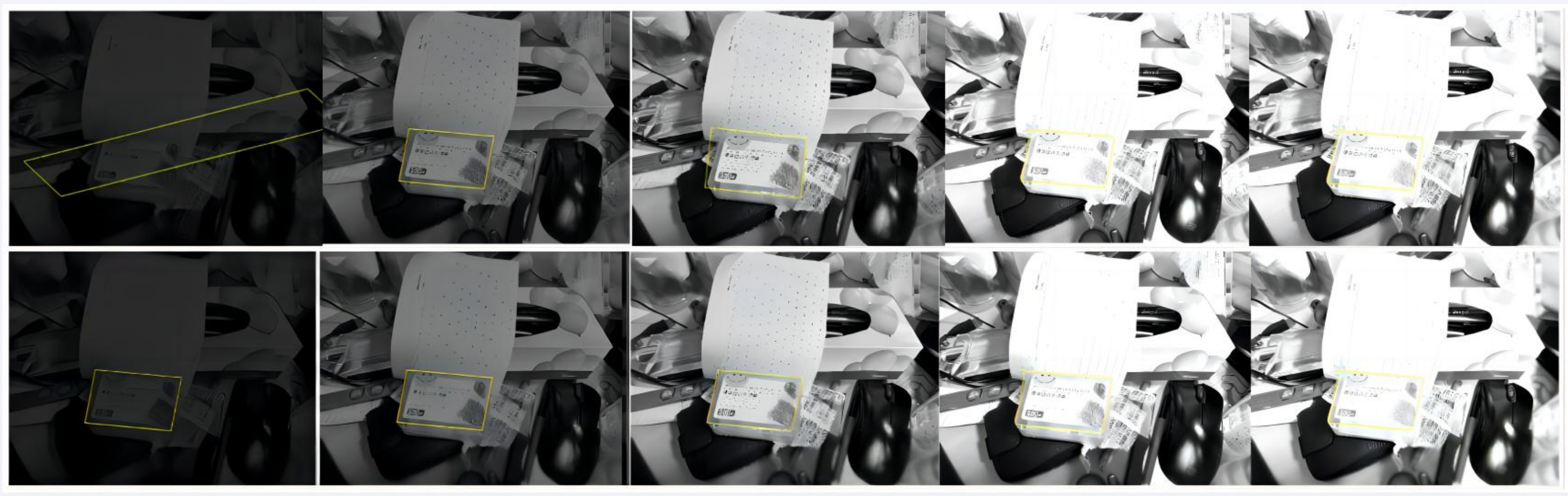


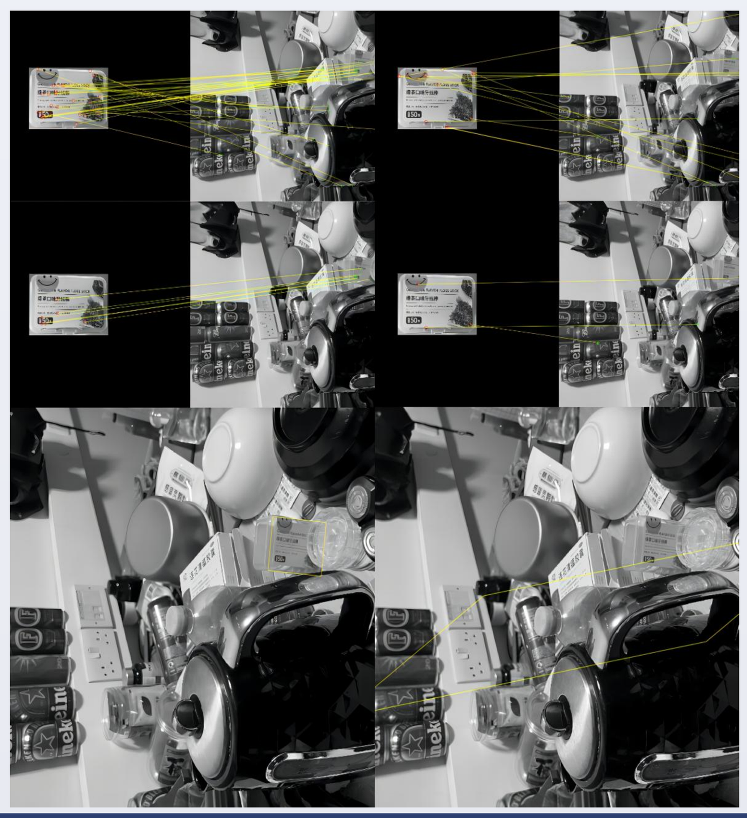
Figure 4: Experimental results of SURF(Top) and SIFT(Bottom) at different brightness

Different Clutter Levels Result

In this section, we will detect the impact of different cluttered backgrounds on object detection, which can be roughly divided into three parts:

1. Simple backgrounds: backgrounds with less than three irrelevant objects.
2. More clutter backgrounds: backgrounds with five to six irrelevant objects.
3. Very clutter background: a background with more than ten irrelevant objects.

After detecting all the images we can get the statistics shown in the table below.



Success Rate(%)	Simple backgrounds	More clutter backgrounds	Very clutter background
SURF	100%	100%	100%
SIFT	100%	80%	80%

Table 3： Object detection in different backgrounds

Meanwhile, the above figure shows that SURF detects more feature points than SIFT in the cluttered environment, and SURF smoothly excludes similar feature points of other objects when clearing outlier matching, which also shows that SURF has stronger robustness to rotational changes of images.

Figure 5: Feature point extraction and detection results of SURF (left) and SIFT (right) in cluttered environment

Conclusion

This poster briefly introduces the detection steps of SURF and SIFT, and the results of using SURF and SIFT to detect objects with different occlusion area, brightness and background clutter, in summary, we can conclude that

1. The detection of SIFT is better than SURF when the occlusion area is less than 60%; the detection of SURF is better than SIFT when the occlusion area is at 80%.
2. SIFT is more robust than SURF in the case of more obvious changes in brightness.
3. SURF has better detection than SIFT in different degrees of clutter and different object placement angles.

Reference List

1.Ansari, S. (2019, February). A review on SIFT and SURF for underwater image feature detection and matching. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-4). IEEE.

2.Bay, H., Tuytelaars, T. and Van Gool, L., 2006. Surf: Speeded up robust features. Lecture notes in computer science, 3951, pp.404-417.

3.Dalal, N., Triggs, B. and Schmid, C., 2006. Human detection using oriented histograms of flow and appearance. In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9 (pp. 428-441). Springer Berlin Heidelberg.

4.K. Li and L. Cao, "A Review of Object Detection Techniques," 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT), Nanchang, China, 2020, pp. 385-390, doi: 10.1109/ICECTT50890.2020.00091.

5.Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60, pp.91-110.

Acknowledgement

I would like to express my gratitude to all the people who helped me. I want to thank my supervisor Bogdan, for his assistance and guidance, which helped me navigate the challenges I encountered in the experiments. Without your inspiration, I can't imagine how I would have completed this experiment. Thank you for your patience and academic rigor. Thank you for explaining all the issues I need to pay attention to, pointing out my shortcomings. I want to thank my parents and all the professors who taught me， I appreciate your help.