**Robust Object Detection and Tracking in a Cluttered Scene**

**CONTENTS**

[1 Introduction 3](#_Toc17534)

[2 Literature Review 4](#_Toc18742)

[2.1 Scale Invariant Feature Transform (SIFT) 4](#_Toc22778)

[2.2 Speeded Up Robust Features (SURF) 4](#_Toc21398)

[2.3 Histogram of Oriented Gradients (HOG) 5](#_Toc6904)

[3 Object Recognition in Cluttered Scenes Using Point Feature Techniques 6](#_Toc16480)

[3.1 Introduction of SIFT Technique and Comparison with SURF Technique 6](#_Toc14834)

[3.1.1Similarities Between SIFT and SURF 6](#_Toc14987)

[3.1.2 The Difference Between SIFT and SURF 6](#_Toc12483)

[3.2 Overview of the Principles of Image Detection Using SURF Technique 7](#_Toc26179)

[3.2.1 Fast-Hessian Detector 8](#_Toc15544)

[3.2.2 Construct the Scale Space 8](#_Toc22060)

[3.2.3 Feature Point Localization 8](#_Toc28234)

[3.2.4 Feature Point Orientation Assignment 9](#_Toc1969)

[3.2.5 Generate Feature Point Descriptors 9](#_Toc10407)

[3.2.6 Feature Point Matching 10](#_Toc1780)

[3.3 Detailed Explanation Object Detection in Picture 10](#_Toc20076)

[4 Tracking and Detection of Objects in Cluttered Scenes in Video 13](#_Toc7962)

[4.1 Detailed Explanation Object Detection in Video 13](#_Toc3389)

[4.3Impact of SIFT and SURF Algorithm on Detection Results 14](#_Toc25089)

[5 Using webcam to detect object in cluttered environment 15](#_Toc1826)

[6 Discussion of the technologies 17](#_Toc27301)

[6.1 Imperfections of SURF 17](#_Toc20438)

[6.2 Imperfections of SIFT 17](#_Toc6085)

[7 Conclusion 18](#_Toc9570)

[8 Reference List 18](#_Toc10367)

# 1 Introduction

Computer vision is a set of techniques for extracting information from images, videos, or point clouds (*MathWorks*, 2023). Object detection uses computer vision techniques to detect and track objects in images and videos and to separate the detected objects. Object detection technology is a major technology behind advanced driver assistance systems (ADAS), which enables vehicles system to detect driving lanes or perform pedestrian detection to improve road safety (*What is object detection?*, 2023). Besides, object detection technology is also widely used in various fields like video surveillance, image retrieval, etc.

In today's society, people have been processing all kinds of information, but the speed and accuracy of human processing information are pretty limited, especially in the case based on a large database, and there is obviously a gap between the efficiency of human processing data on a large scale and computers, especially in the field of processing images and videos. Therefore, it becomes a necessity to find a suitable algorithm for computers that can quickly process information in images and videos and improve the efficiency of information processing. And since object detection technology is the core and basis for processing information in images and videos, it is necessary to study and discuss it.

In this report, we focus on point-feature and similarity-based matching techniques for tracking and detecting objects in cluttered environments. This report is divided into eight parts, the first part is the Literature Review, which provides an introduction to some of the feature recognition techniques(SIFT and SURF, etc.). The second part introduces the principles of using point feature techniques and how to use them to detect specific objects in pictures in cluttered environments. The third part is about detecting and tracking specific objects in videos in cluttered environments. In this part, experiments are conducted using the SIFT technique and using the SURF technique, and the differences between the two experimental results produced are discussed. The fourth part is to connect a webcam and use it for real-time detection and tracking of specific objects in cluttered environments. The fifth part discusses the drawbacks of this technique. The sixth part is a discussion of other object recognition techniques. The seventh part discusses the application prospects of feature recognition techniques. The eighth part concludes this report.

# 2 Literature Review

## 2.1 Scale Invariant Feature Transform (SIFT)

Scale-invariant feature transform (SIFT), an intra-computer detection algorithm for identifying and defining local image features, has the main advantage of being highly robust to changes in scale, rotation, and being somewhat independent of viewpoint changes, affine transformations, and noise (Lowe, D.G., 1999). This algorithm was published by David Lowe (1999) and was refined and summarized five years later (Lowe, D.G., 2004).

The SIFT algorithm is executed in four main steps (Ansari, S., 2019):

1. Feature point localization: for each extreme point, determine whether it is a feature point by Gaussian curvature detection.
2. Orientation assignment: feature point orientation assignment based on local image gradients.
3. Feature description: generate a feature vector with 128 dimensions based on the image gradient orientation and magnitude around the feature point (Lowe, D.G., 1999).

According to the comparison given by Luo Juan et al (2009), SIFT is sluggish and not ideal for changes in illumination intensity (Juan, L. and Gwun, O., 2009), but insensitive to rotations, scale changes and affine transformations.

## 2.2 Speeded Up Robust Features (SURF)

The Speeded Up Robust Function (SURF), the robust image recognition and description algorithm was first proposed by Herbert Bay et al (2006) at the 9th European Conference on Computer Vision (ECCV). The algorithm is influenced by SIFT, and the standard version of SURF is several times faster than SIFT, and its authors claim that SURF is more robust than SIFT for different image transformations. The results of the study show that SURF outperforms SIFT in terms of speed and accuracy, and even outperforms other previous schemes. For example, under low-light conditions or weak contrast, SURF performs better than SIFT because it has higher robustness to noise and distortion in the image (Ansari, S. 2019). This is achieved using combined images for signal convolution and drawing on the advantages of current leading detectors and descriptors (K. Li and L. Cao, 2020).

There are three primary steps in the implementation of the SURF algorithm (Bay, H et al., 2006):

1. Feature point detection: SURF uses the Hessian matrix, which facilitates the use of integral images popularized by Viola and Jones (Viola P, Jones M., 2001), by which the computation time can be significantly reduced.
2. Direction assignment: for each feature point, the direction histogram of its surrounding image gradients is calculated and the main direction is assigned to it.
3. Feature description: a method similar to the SIFT algorithm (Lowe, D.G., 1999) is used to generate a feature vector with 128 dimensions.

At the same time, SURF is fast and has no less performance than SIFT, but SURF also has drawbacks like instability to changes in rotation and illumination.

## 2.3 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) technique was first conducted by Navneet Dalal and Bill Triggs (2005). Moreover, they proposed that the advantages of HOG features include better rotational invariance, scale invariance, etc., which makes the HOG technique perform well in the detection and recognition tasks of targets (Dalal, N. and Triggs, B., 2005) with high accuracy and robustness (Mizuno, K.et al., 2012).

The HOG technique has a wide range of applications in the field of computer vision, and this technique is broadly applied in the area of pedestrian detection, (Dalal, N. and Triggs, B., 2005). In addition, the HOG technique is also ~~used~~ utilized in face detection, which can effectively detect faces (Dalal, N.et al., 2006). This technique was initially used for human detection, but later it was employed in other research fields, such as Transportation (vehicles) and Animals Study (*Get Started with Cascade Object Detector*, 2023).

# 3 Object Recognition in Cluttered Scenes Using Point Feature Techniques

This section completes the recognition of images in cluttered environments using point feature techniques in images. The thought is to use feature points for matching, then find a hypothetical matching result, and finally, find the object in the background.

## 3.1 Introduction of SIFT Technique and Comparison with SURF Technique

### 3.1.1Similarities Between SIFT and SURF

The SIFT technique is not much different from SURF in terms of implementation steps, and in terms of scale space construction, both SIFT and SURF obtain images under different scale spaces by Gaussian difference pyramids, and then detect local maxima and minima in these images as feature point candidates (Lowe, D.G., 2004; Bay, H et al., 2006). In terms of feature point localization, both of them are used to get the location and scale of their key points by pinpointing the key point candidates (Bay, H et al., 2006; Mikolajczyk, K., and Schmid, C. 2005). Besides, both use gradient histograms to estimate their principal directions for rotational invariance. The couple obtain the feature descriptor of the key point by performing a Gaussian-weighted gradient histogram statistic on the image region around the key point (Bay, H et al., 2006; Mikolajczyk, K., and Schmid, C., 2005). The two of them are also used to match two images by calculating the Euclidean distance or Hamming distance between the descriptors of the key points in the two images (Lowe, D.G., 2004; Bay, H et al., 2006).

### 3.1.2 The Difference Between SIFT and SURF

In terms of scale space construction, SIFT uses a Gaussian difference pyramid to find scale-invariant feature points (Ansari, S., 2019), while SURF uses a Hessian matrix to find scale-invariant feature points (Bay, H et al., 2006). the descriptors of SIFT are directional histogram statistics based on local gradients (Mikolajczyk, K., and Schmid, C., 2005), while the descriptors of SURF are directional histogram statistics based on Haar wavelet responses (Bay, H et al., 2006). The direction estimation method used by SIFT determines the principal direction by computing the direction of the gradient, while the direction estimation method used by SURF determines the principal direction by computing the Haar wavelet response of the pixels around the feature point. both feature point detection and descriptor generation of SIFT are slower than SURF (Juan, L., & Gwun, O., 2009), and this will be discussed specifically in a later paper.

## 3.2 Overview of the Principles of Image Detection Using SURF Technique

SURF is an improved version of the SIFT algorithm proposed by David Lowe in 1999, which, like SIFT, uses similarity feature matching and improves the execution efficiency of the algorithm, offering the possibility of applying the algorithm to real-time computer vision systems. Like SIFT, target detection using the SURF algorithm can be divided into three steps: extraction of local feature points, description of feature points, and matching of feature points (Bay, H et al., 2006).

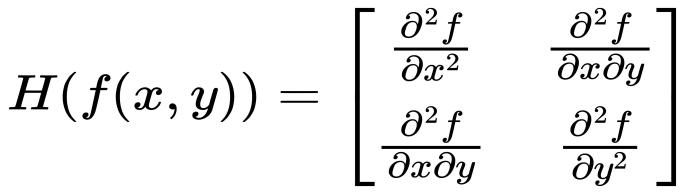
In order to conduct a better job of feature extraction, extracting some good features is needed, which should have the following characteristics：

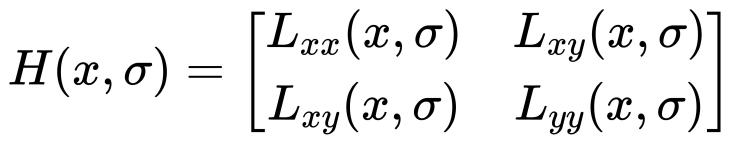
1. Repeatability: The same area in different images can be detected repeatedly under the influence of other factors such as rotation, blur, lighting, etc.
2. Distinguishability: Different detection points should be distinguishable.
3. Fine-tuning: Select the appropriate number of feature points. Too few points will result in incomplete feature detection, too many points will lead to too many feature points, making it difficult to handle the corresponding points matching.
4. Accuracy: The algorithm’s matching result should be close to the actual answer.
5. Real-time: The higher the efficiency of detection, the better.

The main techniques of SURF and the process it performs for feature recognition ~~are~~ will be discussed below.

### 3.2.1 Fast-Hessian Detector

The Hessian matrix was developed in the 19th century by the German mathematician Ludwig Otto Hesse and later named after him. Hesse originally used the term "functional determinants" (*Hessian matrix Wikipedia*, 2023). Hessian Matrix is a square matrix composed of second-order partial derivatives of multivariate functions, constructed because of its good performance in terms of computational time and accuracy. For an image f(x,y), its Hessian matrix can be expressed in the following form,



After Gaussian filtering, the Hessian matrix becomes of the following form:

As the main technology of the SURF technique, Hessian matrix is often used to express image processing operators in computer vision and image processing.

### 3.2.2 Construct the Scale Space

Suppose we have an original image and blur it with different degrees of Gaussian filters to obtain a series of images of different scales, and then construct a Gaussian difference pyramid, each layer of the Gaussian difference pyramid is obtained by the Gaussian difference of two adjacent layers of images. Specifically, for the nth and n+1th layers, we perform Gaussian filtering using different convolution kernels to obtain two images and then differ the two images to obtain the Gaussian difference image of the nth layer (Bay, H et al., 2006). After which, the corresponding images of each layer of the Gaussian difference pyramid are stitched together to obtain a scale space.

### 3.2.3 Feature Point Localization

Each pixel point processed by Hessian matrix is compared with 26 points in the two-dimensional image space and scale space neighborhood to initially locate the feature points (Lowe, D.G., 2004), and then the final stable feature points are obtained after filtering out the feature points with weaker energy and the mislocalized feature points.

### 3.2.4 Feature Point Orientation Assignment

Directional matching of the SURF algorithm is employed to improve the rotational invariance of the feature point descriptors. During feature point detection, the SURF algorithm quickly locates feature points by computing the integral image of the Haar wavelet response. Then, when computing the feature point descriptor, gradient information around the feature point is applied to compute the principal direction of the feature point and convert the descriptor to a rotationally invariant descriptor concerning that principal direction. Regarding computing the principal direction of the feature point, the SURF algorithm uses a gradient histogram that computes the principal direction by weighting the gradient directions (V. Hedau et al., 2008). This approach makes it possible to improve the rotation invariance of feature point descriptors while maintaining high computational speed, thus not only guarantee the accuracy but also maintain a desired time-consumption.

### 3.2.5 Generate Feature Point Descriptors

In the key point localization process, the SURF algorithm utilizes a Hessian matrix to identify feature points with large-scale spatial variations. The direction assignment process determines the principal direction of the feature points by computing the Haar wavelet response while ensuring the rotational invariance of the descriptors. In the feature point description process, a small 4x4 square segmented image is used and the directional gradient histogram of the Haar wavelet response within each subregion is calculated, which in turn generates the special point descriptors (Bay, H et al., 2006). Because the SURF algorithm uses some efficient computer vision techniques like integral images and Gaussian difference pyramids, it can therefore generate descriptors efficiently with good computational performance and robustness.

### 3.2.6 Feature Point Matching

In SURF technique, feature point matching is done by calculating the similarity of feature point descriptors in two images. Specifically, the matching process is to match the two parts that are closest in distance in Gaussian space by selecting a feature point in one image and then finding the feature point that is most similar to it in the other image.

In the matching process, the nearest neighbor algorithm is usually used to compare the descriptors of the feature point to be matched with the descriptors of all the candidate feature points and select the most similar one (Bay, H et al., 2006). The similarity here is normally calculated via metrics such as Euclidean distance (De Maesschalck, R. et al., 2000) or Hamming distance (Hamming, R.W., 1950). Also, some screening, such as selecting sub-similarity points and ratio testing, is required to improve the accuracy and robustness of matching.

In practical scenarios, feature point matching is a crucial application of SURF technology, which is often used in image stitching, object tracking, 3D reconstruction and other fields. Especially in computer vision, SURF technique holds a solid position since its accuracy and robustness.

## 3.3 Detailed Explanation Object Detection in Picture

After loading the images into MATLAB, the RGB image is converted to grayscale using the "rgb2gray" function to remove the hue and saturation information of the image while maintaining the brightness of the image. This is because that the point feature technique requires grayscale images to handle the recognition, and the original image is an RGB image that cannot be recognized directly with feature points. This function extracts features from grayscale images using the "Speeded-Up Robust Features (SURF)" algorithm."selectStrongest" function is applied to pick the most representative points which extracts the 600 strongest features (below).



Figure 1: The original image and the grayscale image after extracting 600 feature points

The next step is to employ the "extractFeatures" function to extract the feature points of the template (the box here) and the feature points of the target (box in the cluttered scene), use the "matchFeatures" function is utilized to match the feature points of them. And finally, use the "showMatchedFeatured" function to match the feature points of the ~~box~~ template in the grayscale image with that of the target. Due to the principle of nearest neighbour, the closest and most similar feature points (the closest Gaussian space, i.e., the point that is actually most similar) in the two grayscale images are matched and then connected with yellow lines to get the following figure.

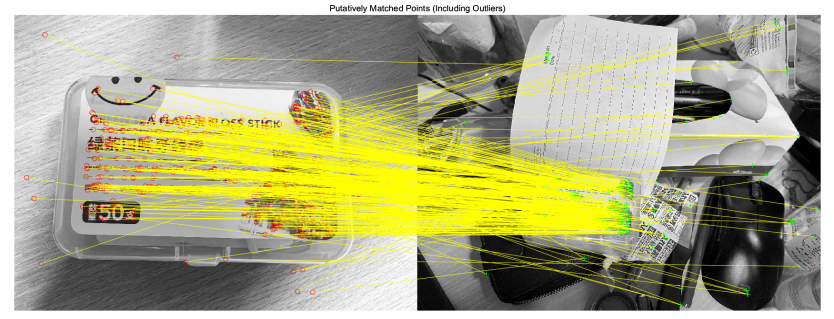


Figure 2:Initial matching feature point concatenation (Including Outliers

Because there are many feature points connected in the initial matched line that are not related to the object to be detected, therefore, it is necessary to use the "estgeotform2d" function to detect and exclude these outliers. The "showMatchedFeatured" function again is used here to match the feature points of the template in the grayscale image with the feature points of the target in the cluttered scene, and then connect them with yellow lines. The lines connecting the filtered boxes with the feature points of the boxes in the cluttered scene are shown below.



Figure 3: Feature point matching linkage after filtering outliers

After getting the boundary polygons of the reference image, setting their dimensions and converting the polygon to the coordinate system of the target image, the yellow border is generated, representing the position of the object in the scene, which in turn completes the case of a cluttered environment detection in the image.

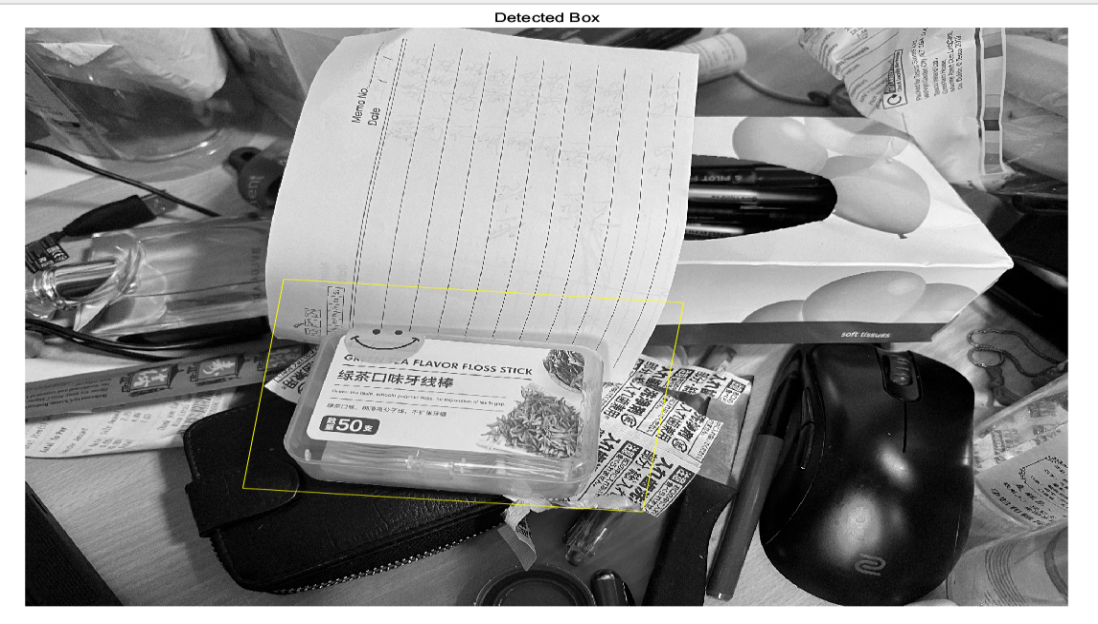


Figure 4:Detected box

# 4 Tracking and Detection of Objects in Cluttered Scenes in Video

In this section, we complete the tracking and detection of objects in cluttered scenes in videos by using SIFT and SURF. A comparison and brief discussion of the results produced by them will be given.

## 4.1 Detailed Explanation Object Detection in Video

The principle of object recognition in video is the same as the basic principle of object recognition in the picture, object recognition in the video needs to use a while loop to recognize each frame of the picture in the video, and the specific operation steps are as follows.

Import the image of the target object into MATLAB, and use the same method as the object recognition in the picture to convert the image to a grayscale image and extract its features. Employ the “VideoReader” function to read the video to be detected. In the while loop, after utilizing the “hasFrame” function to pull out each frame of the video, each frame is put into the function using the “readFrame” function.

To facilitate the writing of the function, define a sub-function: put the code used for the scene detection part of the image recognition into it. A “try” exception capture structure is used, if the specified object is detected it is circled in yellow, and the “catch” function means that the condition defined by the "try" function is not fulfilled, and if it is not detected, this frame detection will be skipped.

Finally, after running the function an ideal detection result can be obtained as shown in the figure below.

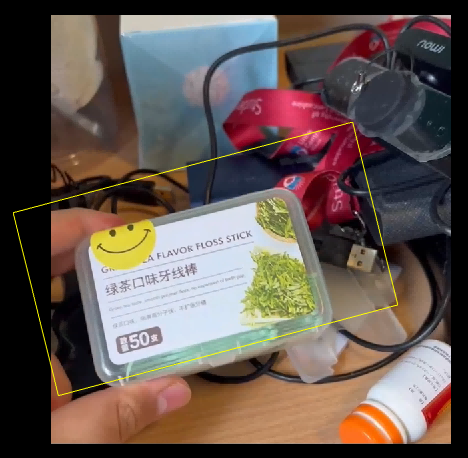


Figure 5:Object detected in the video (SURF)

## 4.3Impact of SIFT and SURF Algorithm on Detection Results

Simply changing the function "detectSURFFeatures" to "detectSIFTFeatures", we can get the results of object detection in the cluttered environment using SIFT technique. According to the results, the videos detected by SURF are smoother than those detected by SIFT for the following two reasons:

First, in feature point detection, SURF uses the Gaussian difference pyramid method to detect feature points at different scales, which can detect enough feature points in a faster time (Bay, H et al., 2006). SIFT, on the other hand, uses a scale-space extreme value detection method, where a large number of Gaussian fuzzy and differential operations are required, resulting in a large computational effort (Lowe, D.G., 2004).

The second reason lies on the descriptor generation step, SURF uses a fast Haar wavelet transform to generate descriptors (Bay, H et al., 2006), while SIFT uses a more spatially complex Gaussian Weighted Gradient Histogram (GWH) method (Lowe, D.G., 2004). But the one also leads to a larger computational effort. So, SURF is more fluid than SIFT in detection.

However, the detection of videos with SIFT for detection is more stable than SURF, which can be reflected in the video, i.e., the recognition of SIFT is better than that of SURF in different luminance environments. This is because more local feature information is used, which is more stable in detecting objects. In contrast, SURF uses a scale-space extreme value detection algorithm, which may lead to false and missed detections, thus affecting the stability of detection (Turcsany, D. et al., 2013).

/private/var/folders/y0/sfjs7g3s2tvbjbkqcrx5wy5h0000gn/T/com.kingsoft.wpsoffice.mac/picturecompress_20230318182934/output_1.pngoutput_1

Figure 6: Memory of the original video (Left), memory of the SIFT processed video (Mid), memory of the SURF processed video (Right)

In terms of the memory footprint of the saved “. fig" files, the SIFT-processed video takes up more memory than the SURF-processed video when processing the same video. This also proves that the above theoretical description is correct in disguise.

# 5 Using webcam to detect object in cluttered environment

The principle of object detection in the cluttered environment using webcam is the same as in the previous two chapters, and they both use point feature techniques for object recognition. Using the "cam=webcamlist" function, you can find all of the webcams connected to the device, and then select the webcam you want to use, here the "snapshot" function is employed, which is expressed in MATLAB as

img = snapshot(cam)

Acquires a single image from the webcam object cam and assigns it to the variable ‘img’. The snapshot function returns the current frame. Calling snapshot in a loop returns a new frame each time. The returned image is always an RGB image. ‘Snapshot’ uses the camera’s default resolution or another resolution that specify using the Resolution property (*Cam*,2023).

The difference with video detection is that the code for detection using webcam has to change "VideoReader" to "cam ('webcam's name') ". Change the function in the loop that extracts each frame of the matching video to a "snapshot" function. After making above changes, the object detection using webcam in a cluttered environment can be obtained (Figure 7).

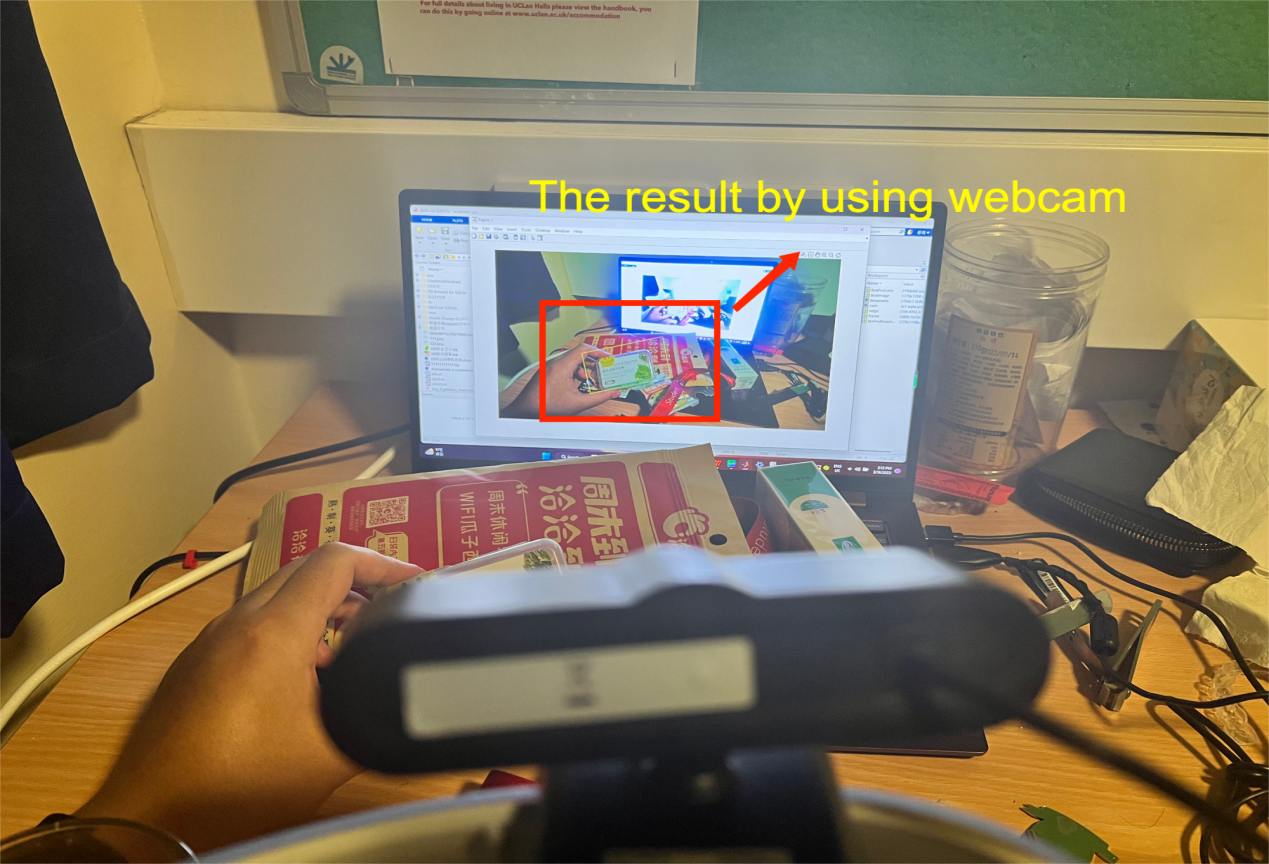


Figure 7: Using webcam to detect object in cluttered environment

If we switch to SIFT for detection, the detection results are the same as the object detection in the video, and SIFT still has more stable detection results. However, the lagging problem of SIFT still exists, so it is necessary to choose whether to use SIFT or SURF according to different situations.

# 6 Discussion of the technologies

Through this study of object detection in cluttered environment and theoretical knowledge, it can be seen that although SURF and SIFT have many advantages, at the same time they also have some shortcomings.

## 6.1 Imperfections of SURF

Feature point matching accuracy may be inferior to SIFT: Although SURF algorithm performs better in speed, it may be inferior to SIFT algorithm in feature point matching accuracy, especially in the presence of large scale and rotation variations.

Not robust enough to geometric deformation and viewpoint changes: SURF algorithm is more sensitive to geometric deformation and viewpoint changes, which may lead to less stable detected feature points and affect the performance of the algorithm.

## 6.2 Imperfections of SIFT

High computational complexity: The SIFT algorithm needs to construct Gaussian pyramids and difference pyramids to detect feature points in the image and perform operations such as gradient calculation, direction assignment and feature descriptor generation for pixels around each feature point, which is computationally intensive and leads to a slow running speed of the algorithm, especially in large-scale image processing.

The number of feature points is unstable: the number of feature points detected by the SIFT algorithm in an image depends on the setting of the threshold and the complexity of the image itself; if there are too many feature points, it will increase the complexity of subsequent processing; if there are too few feature points, it may lead to insufficient feature extraction and affect the accuracy of matching and recognition.

# 7 Conclusion

The report focuses on the use of SURF to detect objects in cluttered environments and examines and discusses the principles of SURF and SIFT and other techniques. The results produced using SIFT and SURF for object detection in video and webcams are compared and the shortcomings of the two techniques compared to each other are discussed.

# 8 Reference List

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.

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

Computer vision (2023) MATLAB &amp; Simulink. Available at: https://uk.mathworks.com/discovery/computer-vision.html?s\_tid=srchtitle\_computer\_5#why-it-matters(Accessed: March 15, 2023) .

Cam (2023) Acquire single image frame from a webcam - MATLAB snapshot - MathWorks United Kingdom. Available at: https://uk.mathworks.com/help/supportpkg/usbwebcams/ug/webcam.snapshot.html?s\_tid=srchtitle\_snapshot\_1 (Accessed: March 18, 2023).

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Vol. 1, pp. 886-893).

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.

De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D.L., 2000. The mahalanobis distance. Chemometrics and intelligent laboratory systems, 50(1), pp.1-18.

Get Started with Cascade Object Detector - MATLAB &amp; Simulink - MathWorks United Kingdom. Available at: https://uk.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html (Accessed: March 16, 2023).

Hessian matrix Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Hessian\_matrix (Accessed: March 17, 2023).

Hamming, R.W., 1950. Error detecting and error correcting codes. The Bell system technical journal, 29(2), pp.147-160.

Juan, L. and Gwun, O., 2009. A comparison of sift, pca-sift and surf. International Journal of Image Processing (IJIP), 3(4), pp.143-152.

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.

Lowe, D.G., 1999, September. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157). Ieee.

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

Mizuno, K., Terachi, Y., Takagi, K., Izumi, S., Kawaguchi, H. and Yoshimoto, M., 2012, October. Architectural study of HOG feature extraction processor for real-time object detection. In 2012 IEEE Workshop on Signal Processing 4Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence, 27(10), 1615-1630.

Mahiddine, A., Seinturier, J., Boï, J.M., Drap, P. and Merad, D., 2012, June. Performances analysis of underwater image preprocessing techniques on the repeatability of SIFT and SURF descriptors. In WSCG 2012: 20th International Conference on Computer Graphics, Visualization and Computer Vision.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence, 27(10), 1615-1630.

Turcsany, D., Mouton, A. and Breckon, T.P., 2013, February. Improving feature-based object recognition for X-ray baggage security screening using primed visualwords. In 2013 IEEE International conference on industrial technology (ICIT) (pp. 1140-1145). IEEE.

Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Ieee, 2001, 1: I-I.

V. Hedau, H. Arora and N. Ahuja, "Matching images under unstable segmentations," 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587759.

What is object detection? (2023) What Is Object Detection? - MATLAB &amp; Simulink. Available at: https://uk.mathworks.com/discovery/object-detection.html(Accessed: March 15, 2023) .