4th Conference on Engineering Science and Technology-CEST-2021
Zliten-Libya 14th – 16th Dec 2021
Al Asmarya Islamic University – Faculty of Engineering
Conference homepage: www.asmarya.edu.ly/cest2021/

Object Detection Based on Template Matching and SURF Algorithm

B. Ghariba* *Department of Electrical and Computer Engineering, Faculty of Engineering, Elmergib University,Khoms, Libya. bmghariba@elmergib.edu.ly*
H. Alshrif *Department of Electrical and Computer Engineering, Faculty of Engineering, Elmergib University,Khoms, Libya. hmalshrif@elmergib.edu.ly*

ABSTRACT

Object Detection is one of the most important applications in the field of computer vision. Object detection algorithms generally use extracted features and learning algorithms for recognizing the objects. This paper presents a method based on template matching and the SURF algorithm to detect an object in the field of view. The purpose of using template matching is to find the similarity matching score between any unknown object and the labeled object. Also, the SURF algorithm provides other accuracy matching results. The combined method was successfully evaluated on synthetic and real images. The obtained results have been illustrated in the proposed method, that could detect the most of objects that are available in the scene with strong efficient.

Keywords:
Object decoction
Template matching
SURF

Corresponding Author Email: bmghariba@elmergib.edu.ly

1 INTRODUCTION

Object detection is the process of finding the location of the target object in an image. It is an important area of research in computer vision because many applications require the determination of the position of an object within an image [1-4]. Also, object detection is used in face detection, object tracking, image retrieval, and automated parking systems. Moreover, object detection is useful for robots to find the object, localization, and provide relate information. The main use of object detection is image classification or more precisely, image retrieval. Object detection is being used in several other fields such as defence and architecture, but it is used the most for medical purposes [5]. Modern work in this area can be divided into contour-based and appearance-based approaches. Contour-based methods work on the relationship between contour points and the shape that these points form. Appearance-based approaches detect interesting points and extract their patch descriptors using features such as colour, edges, texture, and gradient information [6] .

2 RELATED WORK

The visual object detection task also requires an algorithm to localize the objects from various classes present in an image. The traditional solutions for object detection are based on region proposals, such as Selective Search [7], CPMC [8] or MCG [9]. In addition, other methods based on sliding windows such as EdgeBoxes [10]. These methods are computationally expensive because rely on a large number of object proposals. Recently, most of the methods of object detection are based on the Convolutional Neural Networks, such as R-CNN [11] algorithm and explore two directions for further performance improvement. The first direction is to make the underlying CNN deeper. The second direction is to combine semantic segmentation [12]. Also, Other works focus on speeding up the computation time [13], [14], [15].

3 MATERIALS AND METHODS

3.1 TEMPLATE MATCHING

Template matching is a method in digital image processing to find small parts of an image which correspond to a template image. This method is widely used in object detection fields such as surveillance, object tracking, robotics, and medical imaging. There are two types of template matching methods. The first is based on the histogram method and the second is based on the feature extraction method. Template matching methods use an image patch (template), prepared to a certain feature of the search image, which we want to detect. This technique uses one of the matching score methods such as cross-correlation (CC) or Sum of absolute differences (SAD), which can be easily applied on grey images or edge images. The cross-correlation output will be the highest at places where the image frame matches the mask frame, where large image values get multiplied by large mask values. General classifications of template or image matching approaches are: Template-based approaches and Feature-based approaches

- Template- based approach

This method is sometimes called the area-based approach, correlation like methods, or template matching. This method is more suitable for the templates which have no strong features with images, since they operate directly on the bulk of values. Matches are evaluated based on the intensity values of both images and templates [16, 17].

- Featured-based approach

This method is appropriate when the template image and the reference has more correspondence with respect to features and control points. There are many kinds of features such as points, curves, surface model, and edges that should be matched. Here, the goal is to locate the pair wise connection between the template and the reference using their descriptor of features [18].

3.2 SIMILARITY MEASURE

The similarity measure is the measure of how similar two data objects are. The similarity measure in a data mining context is a distance with dimensions representing features of an object. If this distance is small, it wills the high degree of similarity where a large distance will be a low degree of similarity. Matching technique not only takes the similarity measures but also calculates the error between images depending on its difference using Mean Squared Error (MSE) metric:

$$MSE = \frac{\sum\sum[Temporary(x,y) - Target(x,y)]}{Number\ of\ pixels}, \quad (1)$$

Where, Temporary (x, y) and Target (x, y) are the intensity values of Template and input image, respectively.

There are many methods of similarity measures such as: Euclidean Distance, Cross-Correlation (CC), Sum of Squared Differences (SSD), and Sum of Absolute Differences (SAD).

3.2.1 Euclidean Distance

This method is the most common use of distance. In most cases when we refer to distance, we will mention the Euclidean distance. This method is simple in calculation and is the best proximity measure. In addition, the Euclidean distance between two points is the length of the path connection to them. Fig. 1 shows how to calculate Euclidean distance.

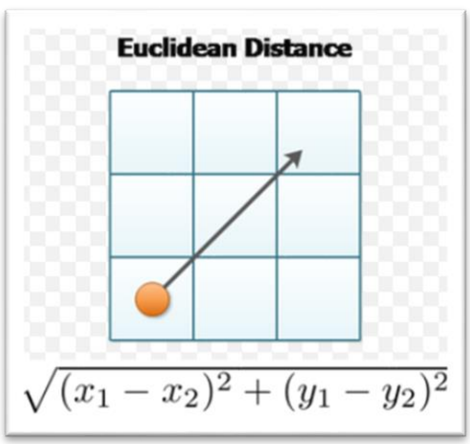

Figure. 1. The Euclidean distance

3.2.2 Cross-Correlation (CC)

Cross-correlation is the measure of similarity of two series as a function of the displacement of one relative to the other. This is also known as sliding dot product or sliding inner product. CC has applications in pattern recognition and signal processing. For continuous function f and g, the cross correlation is defined as:

$$(f * g)(\tau) = \int_{-\infty}^{\infty} f(t)^* g(t + \tau) dt, \quad (2)$$

Where $f(t)^*$ is denoted complicate conjugate of f, and the T is the displacement.

The cross-correlation is the same as the convolution of two functions. There is another name which called auto-correlation which is the cross- correlation of the signal with itself [7]. Fig. 2 explains the output of the cross-correlation.

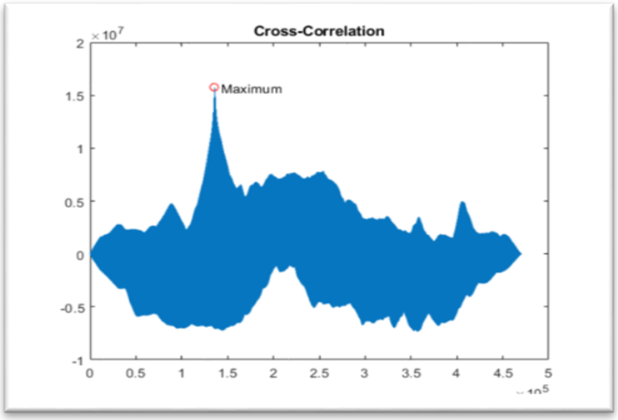

Figure. 2. The Cross-Correlation

3.2.3 Sum of Squared Differences (SSD)

This method is one of many measures of match that bases on pixel-by-pixel intensity differences between two images. SSD calculates the summation of squared for the product of subtracting pixels between two images. Generally, SSD directly uses the formulation the sum of square error. The differences are squared to remove the sign. The Sum of Squared Differences can be defined by next equation:

$$\sum_{j=-N/2}^{N/2} \sum_{i=-N/2}^{N/2} (I1(x+i, y+j) - I2(x+i, y+j)) \quad (3)$$

4 THEORY AND CALCULATION

4.1 SPEEDED-UP ROBUST FEATURES (SURF)

This algorithm is a patented local feature detector and descriptor. It was inspired from the Scale-Invariant Feature Transform (SIFT) algorithm. The standard version of SURF is three times faster than the SIFT algorithm. There are many applications that are using SURF such as: object detection, object recognition, image registration, and classification. SURF is good at handling images with blurred and rotate, but not good at handling viewpoint change and illumination change.

To detect key points, SURF uses an integer approximation of the determinant the Hessian blob detector, which can be computed with 3 integer operations using a pre-computed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. Figure 3 describes the SURF algorithm when it deals for interest point matching in two images [19].

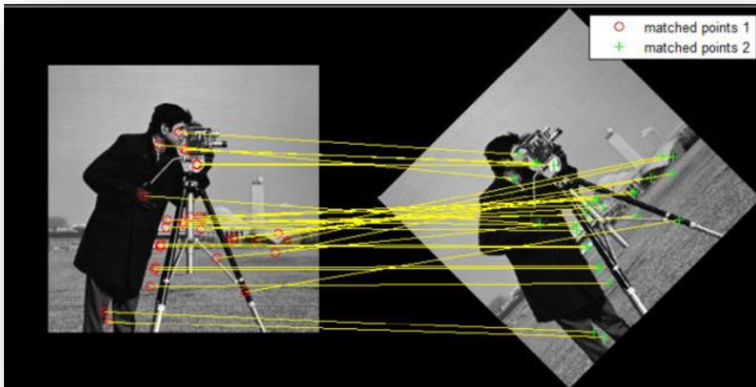

Figure .3. SURF interest point matching

4.2 PROPOSED WORK

In our work, template matching, and the SURF algorithm are used for detecting an object in the field of view. The general block diagram of the proposed system is shown in the figure 4.

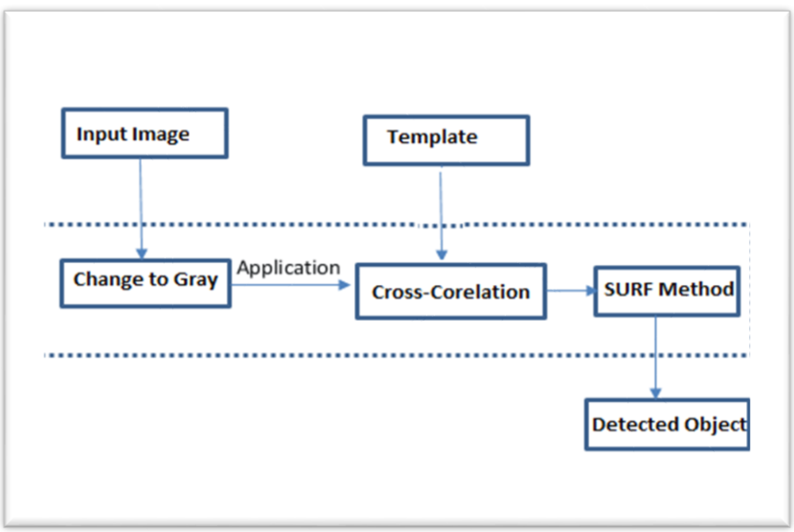

Figure. 4. Block Diagram of the Proposed System

The proposed system consists of two stages. The first stage is based on applied cross-correlation between a whole image and a template (object of interest). The second stage uses the SURF algorithm between the entire image and the detected object in the first stage to find the best match.

5 RESULTS AND DISCUSSION

There are many steps to implement the proposed method, which can be explained by the following points:

- **Gray Image**

This step converts a colour image to a gray image. Figure 5 shows an example of an original colour image, and the figure 6 explains the gray scale of the selected original image

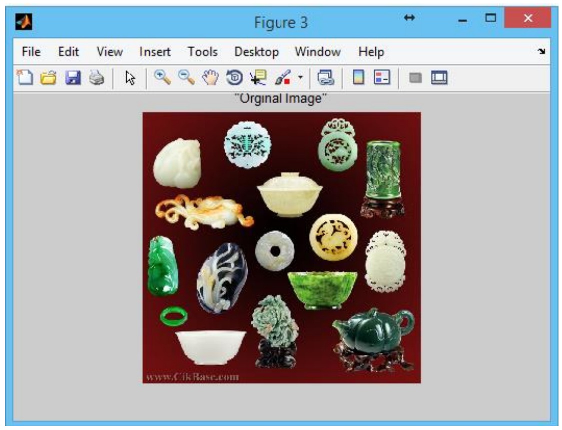

Figure. 5. The Original Image

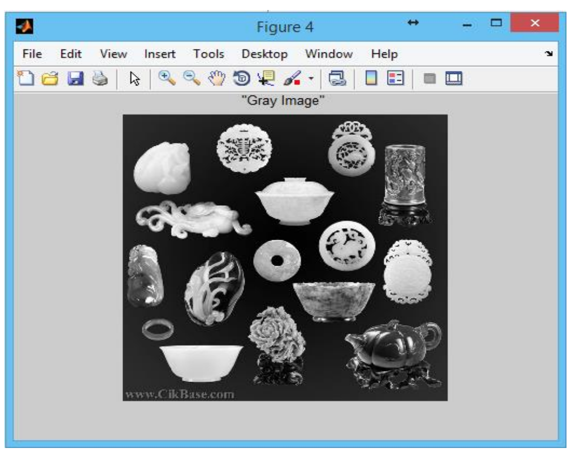

Figure. 6. The Gray Image

The reason for differentiating such images from any other sort of colour image is that less information needs to be provided for each pixel. In other words, Grayscale images are most used in image processing because smaller data enables developers to do more complex operations in a shorter time

- **Cross-Correlation**

This part of the experiment gives a similarity measure between the template (Interested object) and any object in the image. The high similarity means the selected template is matching an object in the image with a high degree, as shown in figure 7. In this figure we can conclude the location of the highest match as a spotlight. This spot line indicates to the center of the object that we want to detect [5] [6].

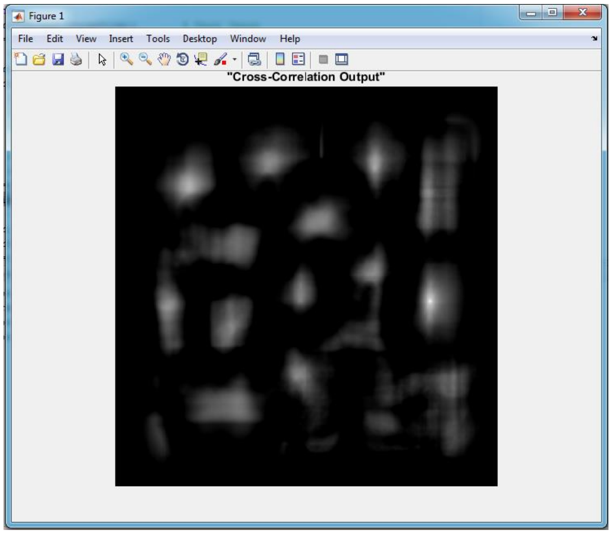

Figure.7. Cross-Correlation output

- **Bounding–Box**

The bounding box is a rectangular shape that describes the location of the object in the original image, as shown in the next figure (figure 8). There are a variety of use cases for image processing and bounding boxes, such as Self-driving cars, Insurance claims, Agriculture, and Healthcare. Bounding boxes are used in all these areas to train algorithms to identify patterns [20].

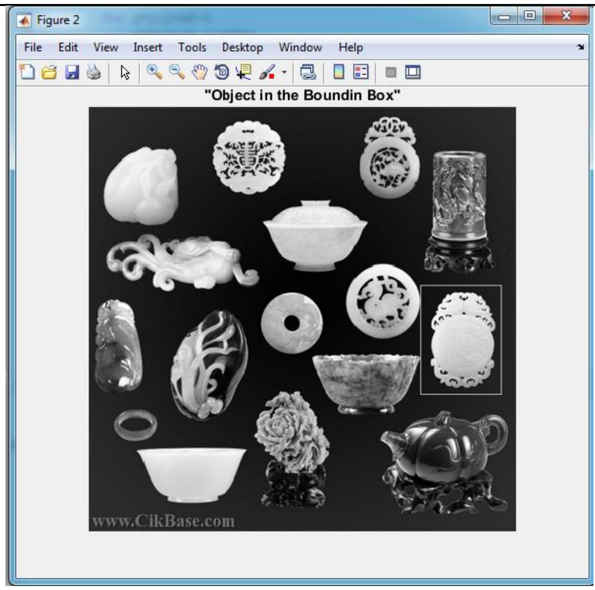

Figure. 8. The Bounding–Box

- **SURF method**

The SURF algorithm provides the method for matching two objects in different images. The object that has many corresponding points with an image means the object that we want to detect. Figure. 9 shows the SURF technique.

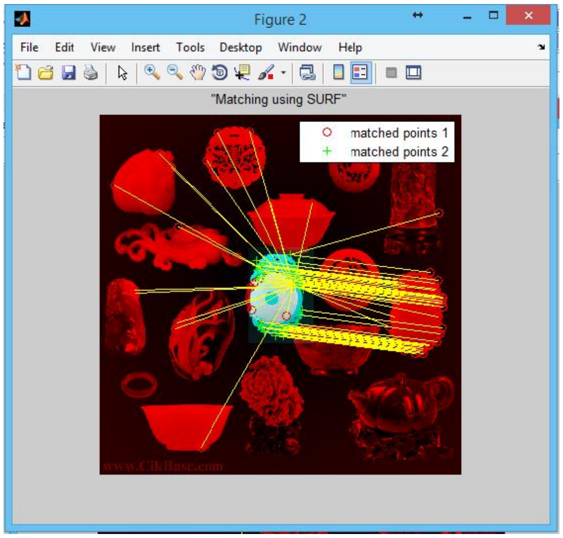

Figure. 9. The SURF algorithm

Based on the experimental results, we can say the Template matching method provides efficient solution to find the interest object in the field of view. Moreover, the SURF algorithm provides good accuracy matching results. Next figure (figure 10) shows the flowchart for SURF Feature detection.

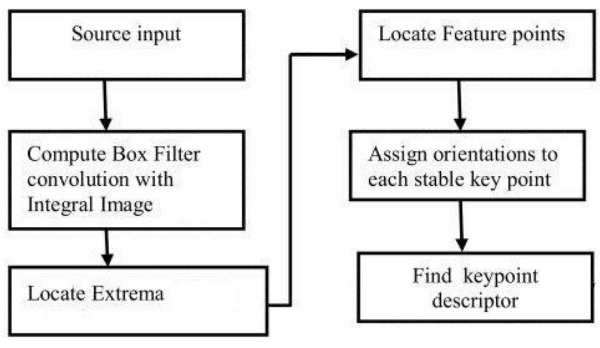

Figure.10. Flow Chart for SURF Feature Detection

6 CONCLUSIONS

In this paper, we have presented an object detection system based on a template matching and SURF algorithm. The template matching is simple technique, low complexity, and fast implementation when compare with other technique that using for object detection. Also, the SURF algorithm gives the ability to matching two objects in different images. The SURF algorithm is used to extract the feature points because SURF features are invariant to image translation, scaling, and rotation. Moreover, the results have indicated that the performance of the proposed method is strong efficient. In the future work, the authors intend to extend this work (object detection) in the video image. In addition, using deep learning technique instead the template matching, SURF algorithm in order to get a better result for the object detection.

7 REFERENCES

[1] Bouti, A., et al. *Robust system for road sign detection and recognition using template matching*. in *2017 Intelligent Systems and Computer Vision (ISCV)*. 2017. IEEE.

[2] Erhan, D., et al. *Scalable object detection using deep neural networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

[3] Li, Z., Y. Cao, and X. Xing. *Object detection based on template matching by using enhanced global-best ABC*. in *The 27th Chinese Control and Decision Conference (2015 CCDC)*. 2015. IEEE.

[4] Singh, N., A. Daniel, and P. Chaturvedi. *Template matching for detection & recognition of frontal view of human face through Matlab*. in *2017 International Conference on Information Communication and Embedded Systems (ICICES)*. 2017. IEEE.

[5] He, L., H. Wang, and H. Zhang. *Object detection by parts using appearance, structural and shape features*. in *2011 IEEE International Conference on Mechatronics and Automation*. 2011. IEEE.

[6] Luo, P., L. Lin, and X. Liu, *Learning compositional shape models of multiple distance metrics by information projection.* IEEE transactions on neural networks and learning systems, 2015. **27**(7): p. 1417-1428.

[7] Uijlings, J.R., et al., *Selective search for object recognition.* International journal of computer vision, 2013. **104**(2): p. 154-171.

[8] Carreira, J. and C. Sminchisescu, *CPMC: Automatic object segmentation using constrained parametric min-cuts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011. **34**(7): p. 1312-1328.

[9] Pont-Tuset, J., et al., *Multiscale combinatorial grouping for image segmentation and object proposal generation.* IEEE transactions on pattern analysis and machine intelligence, 2016. **39**(1): p. 128-140.

[10] Zitnick, C.L. and P. Dollár. *Edge boxes: Locating object proposals from edges*. in *European conference on computer vision*. 2014. Springer.

[11] Girshick, R. *Fast r-cnn*. in *Proceedings of the IEEE international conference on computer vision*. 2015.

[12] Dai, J., K. He, and J. Sun. *Instance-aware semantic segmentation via multi-task network cascades*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[13] Ren, S., et al., *Faster r-cnn: Towards real-time object detection with region proposal networks.* Advances in neural information processing systems, 2015. **28**: p. 91-99.

[14] Redmon, J., et al. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[15] Najibi, M., M. Rastegari, and L.S. Davis. *G-cnn: an iterative grid based object detector*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[16] Astua, C., et al., *Object detection techniques applied on mobile robot semantic navigation.* Sensors, 2014. **14**(4): p. 6734-6757.

[17] Perveen, N., D. Kumar, and I. Bhardwaj, *An overview on template matching methodologies and its applications.* International Journal of Research in Computer and Communication Technology, 2013. **2**(10): p. 988-995.

[18] Hashemi, N.S., et al., *Template matching advances and applications in image analysis.* arXiv preprint arXiv:1610.07231, 2016.

[19] Hai-Yan, Y. and H. Yu-Xin. *An image retrieval algorithm based on SURF for embedded system*. in *2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*. 2017. IEEE.

[20] Dai, J., K. He, and J. Sun. *Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation*. in *Proceedings of the IEEE international conference on computer vision*. 2015.