Comparative Analysis of SURF and MSER based Detection Technique of Cluttered Objects in Multi-objects Scene

[1]Ms. Vimal S. Bodke, [2]Omkar S. Vaidya
[1]P.G. Student, [2]Assistant Professor,
Department of Electronics & Telecommunication Engineering,
Sandip Institute of Technology & Research Centre, Nashik, India.
Savitribai Phule Pune University, Pune.

Abstract: The completely unique methodologies for object detection in cluttered real pictures are proposed in this paper, where a single real time scenario is taken as model. The image edges are parted into contour segments & organized in an image representation that encodes their interconnections: the Contour segment Network. A computationally effective detection technique using Maximally Stable Extremal Regions (MSER) and Speeded up robust features (SURF) is developed to solve the issue of object detection, where paths are search through network & then resemble the model outlines with it. In this paper, real time cluttered images are considered and using proposed techniques such images to be detected from the complete scene. Simulated results show that both detection methods highlight accurate cluttered image. But, accuracy of matching the features among cluttered image and multiple objects in scene is highest in MSER while required computation time is lowest in SURF.

Index Terms **– Cluttered images, MSER, Object Detection, and SURF.**

I. Introduction

Efforts are being put to recognize the objects within fraction of time as researchers are gaining interest in object detection and recognition. Object detection and recognition is one of the toughest part of computer vision. Human vision discriminate the different patterns of objects speedily. This is a great question to present day Object Detection & Recognition in machines, because it is not corresponding to Human, hence the Human brains beats any algorithm addressable today. However there is need of enhancing object detection & recognition algorithms. They have legion applications in the present world. As world is moving towards automation object detection is gaining more attention [1].

The methods proposed in this paper use objects detection algorithms in cluttered scene. The method is related to recent & effective matching techniques & more particularly to, which deem only images & their gradients to detect objects. The proposed system uses point feature matching method which implies algorithms SURF & MSER for feature extraction & description & "Approximate nearest neighbor search" for feature matching.

In Point feature matching technique an "interesting" part of an image is called as Feature. There are three steps for detecting features:

i) Detection -Automatically discover interest points this must be done rugged like corner blob or edge.
ii) Description- Each interest point should have a unique description that does not depend on the features scale & rotation.
iii) Matching- Similar features are then retrieved based on predetermined interest points & object is detected.

The method of object detection works best for objects that display non-repeating texture patterns, which give rise to unique feature matches. This technique is not likely to work well for uniformly coloured objects, or for objects containing repeating patterns [2]. The proposed algorithm in [2] is designed for detecting a specific object, for example, the elephant in the reference image, rather than any elephant.

The outline of the paper is as follows: Section II discusses relevant research works carried out in object detection in cluttered environment. The proposed system is explained in Section II followed by mechanism of object detection algorithms in Section IV. The experimental results are discussed in Section V. Finally the paper is concluded in section VI.

II. Literature Survey

To define the balances between points that represent the object-of-interest & the input scene interpretation tress are used by authors in [3], however this methodology is computationally overpriced and neglects any other information, besides point locations. A shape descriptor, specifically the shape context, and specify the correspondences between the two purpose sets by minimizing the total of feature dissimilarities is proposed in [4]. In [5] a method proposed by Thayananthan et. al. shows that it cannot handle clutter, and impose a "figural continuity" constraint to enhance its performance. However, their methodology remains sensitive to serious litter and too large deformations of the target object. Moreover, their "figural continuity" constraint might need restricted impact once occlusion happens. An honest cause initialization of the target object near its actual location within the input image search through an outsized variety of model templates is needed generally in Chamfer matching ways. A way for detecting deformable shapes by representing them with triangulated polygons is explained in [6]; however the approach is computationally dearly-won. A flexible methodology is developed by authors in [7] which tolerate structural form changes, such as, the occlusion of some fingers once localizing instances within the input image. However, their approach is only driven by bottom-up mechanisms which could deform the target object to non-meaningful form instances.

Hough-style voting scheme is implemented [8] to initialize the cause of the target object within the input image and then perform non-rigid form matching employing a variation of [9]. Due to the initialization stage in their method it may run into difficulties once the target object consists of small subparts whose configuration will vary considerably between instances of the object-of-interest. Authors in [10] describe Part based object detection. HOG (Histogram of Gradient) detectors were used by them for individual object half. Support Vector Machine was to classify individual half to a category (object). 3D depth cue is also one of half-based method proposed in [11] which uses 2D object detection for depth cue. The color of objects is employed by [12] to notice objects. They convert RGB image to high intensity pictures (HSI) for detection functions, they conjointly use varied basic shapes for sleuthing & recognizing the objects. Another necessary paper is [13], during this analysis paper Spatio-temporal graph-based approach is employed for detecting moving object. Their experimental results show that a small-sized object is detected. The detection of moving object is completed using ROI (Rectangular Region of Interest). In [14] authors extract pro HOG (Proportionate HOG) from pictures and use pro HOG descriptor to notice the objects. Object Detection & Recognition may assist blind folks for detecting & recognizing objects in their surroundings. Backnotes recognition was planned by [15]. They use adaptive thresholding morphological operations for separating the foreground & the backgrounds from the notes. The technique for Object Detection in Aerial images & videos was proposed by [16]. Researchers take away the foremost frequent colors and use Harris Corner detector for feature extraction for edge detection.

Authors in [17] projected algorithm for detection a particular object based on finding point correspondences between the reference the target image. This notice objects despite a scale amendment or in plane rotation. This technique of object detection works best for objects that exhibit non-repeating texture patterns which create pleasantness matches. For objects containing continuation patterns and uniformly coloured objects this method does not work properly. To find the correspondences between the reference the target image the proposed algorithm is used.

Authors in [18] discussed and compared various feature extraction techniques like SIFT, SURF, PCA-SIFT, HOG, ORB and FAST based on qualitative and quantitative performance parameters. After feature description process, features are matched between image pair using RANSAC algorithm. It is mathematical model which fit only inliers and remove the outliers of matched keypoint features. In [19] researcher simulated feature descriptor techniques like KLT, Harris and SIFT. The experimental results shows that percentage matching rate of SIFT is highest while Harris corner detector take lowest computation time to match the features.

Authors in [20] describes a machine learning approach for visual object detection that's capable of processing pictures very quickly & achieving high detection rates. This work is distinguished by three key contributions. The primary is that the introduction of a brief new image illustration referred to as the "Integral Iinage" that permits the choices utilized by our detector to be computed very quickly.

In [21] authors describe bar graph based mostly ways of manufacturing object likelihood distributions. The mode of associate object's likelihood distribution at intervals a video scene are trace by the author. Since the likelihood distribution of the thing will change & move dynamically in time, the mean shift rule is changed to influence dynamically dynamical likelihood distributions. The changed rule is named the continuously adaptive Mean Shift (CAMSHIFT) rule. It is then used as associate interface for games & graphics. It is easy, computationally efficient likelihood distribution based object tracker quick enough to be used as a part of a sensory activity program. CAMSHIFT scales its search window to object size & so scales its potential trailing speed with object distance from camera. CAMSHIFT is used to ignore the outliers.

III. PROPOSED SYSTEM

The proposed block diagram is shown in Fig. 1. Initially, it detects feature points in both images and then visualizes the strongest feature points found in the reference image and the target image. After that, feature descriptors extract feature at the interest points in both images and Putative Point Matches are obtained. Then, it locates the Object in the Scene Using Putative Matches and calculates the transformation relating the matched points, while outliers are eliminated. This transformation allows localizing the object in the scene. Finally, the matching point pairs by removing outliers are displayed and obtained the bounding polygon of the reference image. The polygon is transformed into the coordinate system of the target image that suggests the location of the object in the scene.

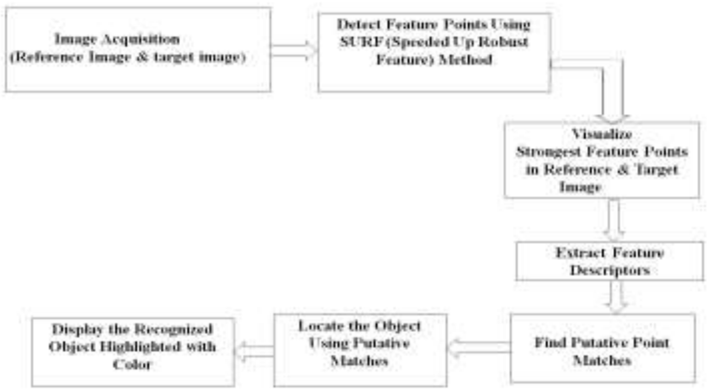

Fig. 1 Block Diagram of Proposed System

IV. System Development

4.1 Speeded up robust feature (SURF)

SURF is a robust detector of local characteristics of an image presented by Herbert Bay in 2006. SURF can be used in the field of object recognition and 3D reconstruction in Computer vision. It is partially inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and its authors say that it is more robust between different image transformations than SIFT. To detect points of interest, SURF uses an integer approximation of the determinant of the Hessian blob detector. It is used for extracting, description and matching the images. It is a patented local feature detector and descriptor. This is achieved by relaying on integral image for image convolution and approximations for achieving higher speed than SIFT. Integral image allows fast computation of box type convolution filters. The entry of an integral image I(x) at location x = $(x;y)^T$ represent the sum of all pixels in the input image I within rectangular region formed by origin and x. It is shown in Eq. (1)

$$I\Sigma(X) = \sum_{i=0}^{i\le x} \sum_{j=0}^{j\le y} I(i,j) \qquad (1)$$

Once Integral image has been formed it takes three additions to calculate sum of intensities over rectangular area of any size. SURF is used because of it should provide better results, uses a hessian based blob detector to find interest points.

4.2 Maximally Stable Extremal Regions (MSER)

MSER is method for Blob detection algorithm extracts from an image a number of co-variant regions. MSER is based on the idea of taking regions which stay nearly the same through a wide range of thresholds. All the pixels below a given threshold are white and all those above or equal are black. If we are shown a sequence of threshold images It with frame corresponding to threshold, it would see first a black image, then white spots corresponding to local intensity minima will appear then grow larger. These white spots will eventually merge, until the whole image is white. The set of all connected components in the sequence is the set of all extremal regions. The word extremal refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary are attached to the MSERs by fitting ellipses to the region this region descriptor are kept as features.

V. Performance Analysis

The experimentation is carried out on i3 Processor with MATLAB 2013a as computing platform and Computer Vision Toolbox to extract features, feature descriptiors and feature matching. Figure 2 & 3 depicts the reference image and Fig. 4 depicts the cluttered scene image from which we have to detect the reference image using the feature point matching method.

Fig. 2 Reference image 1

Fig. 3 Reference image 2

Fig. 4 Target image indicating cluttered part of reference images

Figure 5 and Fig. 6 shows the inliers matched point calculated for the reference image from the clutter image using SURF Method. After finding the match points the image is detected from the clutter image as show in Fig. 7 & Fig. 8.

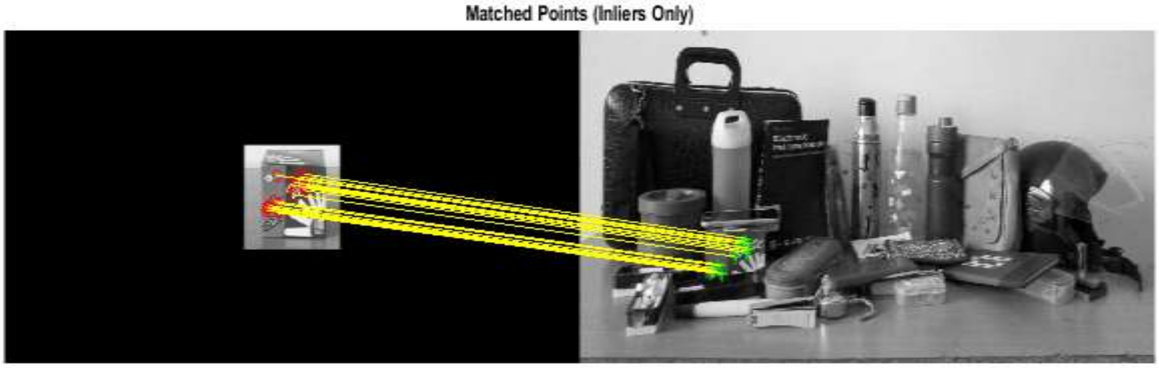

Fig. 5 Matching of Chalk Box with Cluttered image by removing outliers (SURF Method)

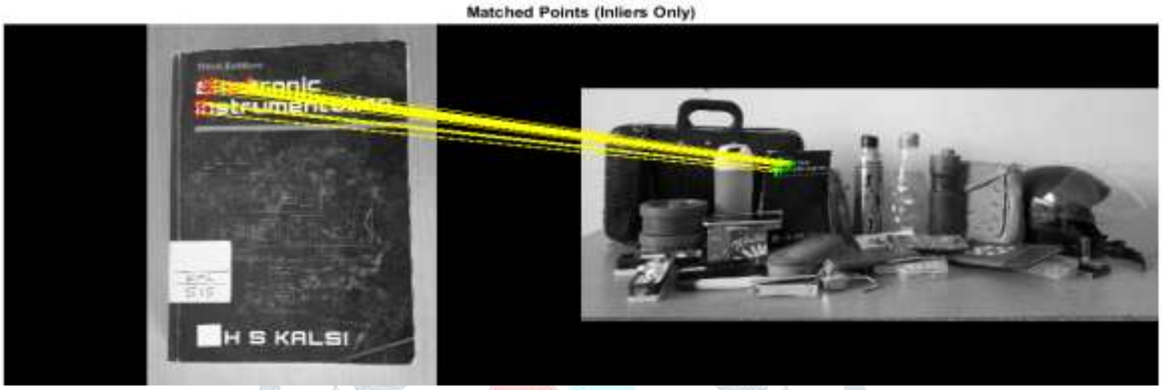

Fig. 6 Matching of Kalsi Book with Cluttered image by removing outliers (SURF Method)

Fig. 7 Detected Chalk Box image in cluttered image (SURF Method)

Fig. 8 Detected book image in cluttered image (SURF Method)

Table 4.1 and table 4.2 show the results obtained by implementing MSER and SURF on our database. The results show that MSER performance better than SURF in terms of higher percentage matching rate although the detection time of reference object from cluttered scene is lowest in SURF.

Table 4.1: Comparison of SURF & MSER based detection technique for Chalk Box image

Methodology	Number of feature point of Reference Image	Number of feature point of Target Image	Number of feature matched between Reference and Target	Number of Inliers Match	Number of Outliers Match	Processing time (sec)	Percentage Matching Rate
SURF	389	3779	52	21	31	17.01	48.38%
MSER	171	1129	34	17	17	46.56	50.00%

Table 4.2: Comparison of SURF & MSER based detection technique for Kalsi Book image

Methodology	Number of feature point of Reference Image	Number of feature point of Target Image	Number of feature matched between Reference and Target	Number of Inliers Match	Number of Outliers Match	Processing time (sec)	Percentage Matching Rate
SURF	173	3779	39	15	24	5.08	38.46%
MSER	99	1129	21	13	8	7.41	61.90%

VI. CONCLUSION

The proposed techniques work well to detect cluttered images from multiple objects scene accurately without false detection. The strongest features helped to get maximum feature match count that leads to identify the exact position of cluttered objects. In this paper, two cluttered objects are considered with 30 % part are hidden in multi-objects scene. The system is designed for detecting a specific object which is given in the reference image, rather than any other similar image like that & the objects are detected based on finding point correspondences between the reference & the target image. It can detect objects despite a scale change or image rotation. It is also racy to small amount of out-of-plane image rotation & occlusion. The experimental results show that SURF is better than MSER in case of processing time and MSER is better than SURF in case of percentage matching rate.
In future this technique will likely to work well for uniformly coloured objects, or for objects containing repeating patterns, the algorithm can be improvised for detecting all kinds of objects and also it will detects the objects that exhibit repeating texture patterns.

REFERENCES

[1] Pomerleau, D. A., "Knowledge-based Training of Artificial Neural. Networks for Autonomous Robot Driving", Book Section/Chapter, Robot Learning, Kluwer Academic Publishers, pg: 14-43, January 1993.

[2] Qiang Tong, Song-Hai Zhang, Shi-Min Hu, Ralph R. Martin, "Hidden images", NPAR '11: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering, August 2011.

[3] S. Umeyama, "Parameterized Point Pattern Matching & Its Application to Recognition of Object Families", IEEE Transactions on Pattern Analysis & Machine Intelligence, 15(2):136–144, 1993.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching & Object Recognition Using Shape Contexts", IEEE Transactions on Pattern Analysis & Machine Intelligence, 24(4):509–522, 2002.

[5] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. "Shape Context & Chamfer Matching in Cluttered Scenes", Proceeding of the IEEE Conference on Computer Vision & Pattern Recognition, 2003.

[6] P. Felzenszwalb, "Representation & Detection of Deformable Shapes", IEEE Transactions on Pattern Analysis & Machine Intelligence, 27(2):208–220, 2005.

[7] V. Athitsos, J. Wang, S. Sclaroff, & M. Betke, "Detecting Instances of Shape Classes that Exhibit Variable Structure", Proceeding of the European Conference on Computer Vision, 2006.

[8] V. Ferrari, F. Jurie, & C. Schmid, "Accurate Object Detection with Deformable Shape Models Learnt from Images", Proceeding of the IEEE Conference on Computer Vision & Pattern Recognition, 2007.

[9] H. Chui & A. Rangarajan, "A New Point Matching Algorithm for Non-Rigid Registration", Computer Vision & Image Understanding, 89(2-3):114–141, 2003.

[10] Hui Hang, Hong Zhang, "Object Detection by Parts Using Appearance, Structural & Shape Features", IEEE International Conference Mechatronics & Automation (ICMA), pp. 489-494, 2011.

[11] Hong Pan, Yaping Zhu, Siyu Xia, Kai Qin, "Improved Generic Categorical Object Detection Fusing Depth Cue with 2D Appearance & Shape Features", IEEE International Conference on Pattern Recognition (ICPR), pp. 1467 – 1470, 2012.

[12] Tran Thi Trang, Cheolkeun Ha, "Irregular Moving Object Detecting & Tracking Based on Color & Shape in Real-time System", IEEE International Conference Computing, Management & Telecommunication, pp. 415 – 419, 2013.

[13] Houari Sabirin & Munchurl Kim, "Moving Object Detection and Tracking Using a Spatio-Temporal Graph in H.264/AVC Bitstreams for Video Surveillance", IEEE Transactions on Multimedia, Vol. 14, No. 3, pp. 657- 668, 2012.

[14] Richard Palmer, Geoff West, Tele Tan, "Gradients for Object Detection in Co-registered Visual & Range Data", IEEE International Conference of Digital Image Computing Techniques & Applications (DICTA), pp. 1-8, 2012.

[15] Zóra Solymár, Attila Stubendek, Mihály Radványi, Kristóf Karacs, "Banknote Recognition for Visually Impaired", IEEE 20th European Conference on Circuit Theory & Design (ECCTD), pp. 841 – 844, 2011.

[16] Hsu-Yung Cheng, Member, IEEE, Chih-Chia Weng, & Yi-Ying Chen, "Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks", IEEE Transactions On Image Processing, Vol. 21, No. 4, APRIL 2012.

[17] K. Rasool Reddy, K. V. S. Krishna, V. Ravi Kumar "Detection of Objects in Cluttered Scenes Using Matching Technique" International Journal of Electronics & Communication Technology (IJECT) Vol. 5, special issue - 3, pp. 42-44, Jan - March 2014.

[18] Omkar S. Vaidya, Dr. Sanjay T. Gandhe, "The Study of Preprocessing and Postprocessing Techniques of Image Stitching", International Conference On Advances in Communication and Computing Technology (ICACCT), Amrutvahini College of Engineering, Sangamner, Ahmednagar, India, pp. 431-435, Feb 8-9, 2018.

[19] Omkar S. Vaidya, Dr. S. T. Gandhe, Pooja B. Sonawane, PERFORMANCE ANALYSIS OF KLT, HARRIS AND SIFT FEATURE DETECTOR FOR IMAGE STITCHING, International Journals of Electrical and Electronics Engineers (IJEEE), vol. 9, issue 1, pp. 833-840, Jan-June 2017.

[20] P. Viola & M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision & Pattern Recognition. CVPR 2001, 2001, pp. I-511-I-518 vol.1.

[21] G. R. Bradski, "Real time face & object tracking as a component of a perceptual user interface", Applications of Computer Vision, 1998. WACV '98. Proceedings of Fourth IEEE Workshop on, Princeton, NJ, 1998, pp. 214-219.