

# A Survey: Object Identification Using Image Processing

Vinay V Kulkarni<sup>1</sup>, Dr.S. D. Lokhande<sup>2</sup>P.G. Student, Department of E&TC, Sinhgad College of Engineering, Vadgaon (BK), Maharashtra, India<sup>1</sup>Principle, Sinhgad College of Engineering, Vadgaon (BK), Maharashtra, India<sup>2</sup>

**ABSTRACT:** This survey paper presents the voluminous survey of different algorithms of detection of object from images and also classifies the detection approaches into different categories. Detection is generally bring off for higher level applications that feel necessity for the location or shape of the object in every image. As the identification of objects from video have improved significantly in the past few years; human motion and behaviour interpretation have naturally become the upcoming step. The research paper includes various approaches that have been used by different researchers for object detection.

**KEYWORDS:** Object Identification, SIFT, SURF, Feature Extraction, Matching, Detection, algorithm.

## I. INTRODUCTION

Now a days there are many technologies to detect and identify objects in image using local descriptors, such as Scale-Invariant Feature Transform and Speeded Up Robust Features have been proposed. These algorithms have enabled identification of objects of live world in images with robustness against differing camera viewpoints, occlusions, and lighting conditions. Local descriptors between images are correlated to determine keypoint matches, followed by a geometric verification to accurately match and localize common objects in the images. These techniques are becoming more advance so it is expected that various industrial applications using images will be enabled. However, most of the conventional methods using local descriptor matching focus on identifying only a few objects in the image. These methods cannot accurately detect and identify a large number of objects in the image. SURF (Speeded-Up Robust Features) [2] is the advance version of SIFT (Scale Invariant Feature Transform). SIFT [1] collects the gradients while SURF sums up wavelet responses. The output is high speed even in real time calculation and robust to rotation not in same plane. Another method for identification operation is HOG (Histogram of Oriented Gradients). HOG is utilized to human discrimination which involve Circular HOG and Rectangular HOG. The difference of both methods is that bins to collect HOG are not same. C-HOG has a center bin or multiple center bins, size of the bins having the different distance from the center are different also [9].

Mainly object detection method consists of two main steps. Preprocessing is initial step containing gray scaling, smoothing, and reducing image resolution and so on. Initial step is followed by filtering for removing the image noise present in the object. The filtering is performed by applying morphological operations like dilation and erosion. And finally on the filtered image connected component labeling is performed.

## II. RELATED WORK

A conventional method to detect and identify objects extracts local descriptors around keypoints detected in the query image, and computes local descriptor distances to find key point matches between query and database images. To identify objects from images various local descriptors, such as SIFT [1] and SURF [2] are implemented. These are robust to scaling and rotation of an object, have been proposed for identifying objects between query and database images. Then, it estimates a geometric model, such as Homography, of the key point matches using RANSAC (RANDOM SAMPLE CONSENSUS) [6] the object in the query image is considered as a match to that of the database image if the number of inliers for the estimated model is sufficient. The key point matches when the query image contains

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

multiple instances of the same objects. In this case, for key point matches of one object in the query image, the others are all regarded as outliers.

In other words, the number of inliers of each object is much smaller than that of outliers. It is well-known that robust fitting methods such as RANSAC perform poorly when the number of inliers falls much below 50% of that of the key point matches [7]. Therefore, it is difficult for the conventional method to detect and identify multiple identical objects in the image. In multiple object identification method [3] [4], objects in the query image are detected and identified by clustering the key point matches into individual objects. For each key point match, an object center position is estimated using scale and orientation associated with the keypoint. The estimated object centers are clustered using mean shift [5], and each cluster is regarded as a match to the object in the database image. Given  $n$  estimated object center positions  $x_i$  ( $i = 1, \dots, n$ ) and  $n$  initial positions of clusters  $y_i$ , mean shift [5] clusters  $n$  estimated object centers by updating  $y_i$  using a difference vector of  $y_i$  and a center of gravity of  $x$  near  $y_i$ .

The computational complexity of mean shift increases with the square of  $n$ . Therefore, it takes an enormous amount of time for identifying a large amount of identical objects. It has been shown that SIFT descriptors can be used to achieve fairly robust object detection from images which are stationary [2]. SIFT has been used in several other related applications like [3] metric robot localization and bio-medical imaging [4]; and it was shown to be one of the best comparative study for currently available descriptors [5].

### III. METHODS FOR FEATURE EXTRACTION

**1. HOUGH TRANSFORM METHOD:** The process includes image fuzzification, membership modification, and image defuzzification. To adjust the brightness of image int operator is used. If the brightness value is greater than 0.5 then it will increase brightness of image if it is less than 0.5 then it decreases the brightness. For this fuzzy image enhancement procedure is followed. Fuzzy regions are divided in two parts: fuzzy edge and smooth region [8].

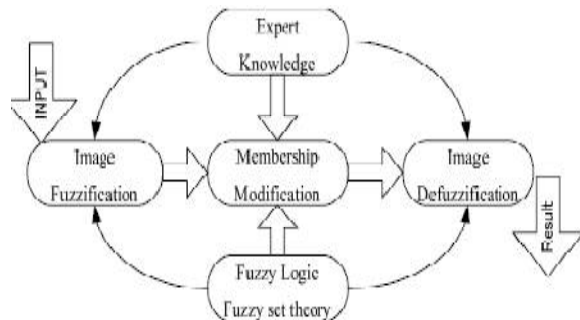


Fig 1. Fuzzy image processing [2]

If the pixels have difference values of gray scale level more than the neighboring pixels causing the result is a fuzzy edge region. The image histogram difference calculates the threshold value. The difference histogram was built from the images by calculating difference of highest value of neighbors. The difference histogram is needed to find by using the formula,

$$D(x, y) = \max (I(x, y) - I(x + i, y + j)) \dots\dots(1)$$

When  $I(m, n)$  is value of gray level of the input image at  $(m, n)$ ,  $d$  is the most different values at  $(m, n)$  close to 8 pixels surrounding, and  $I(m + i, n + j)$  surrounding with  $I(m, n)$  when  $i = -1, 0, 1$  and  $j = -1, 0, 1$ . When  $S(m, n)$  is the value derived from the Sobel operator  $R(m, n)$  is the result of the pixel at  $(m, n)$ , and  $\mu_{SF}, \mu_{EF}$  are the membership functions of SF and EF. If the result is not in either case, check the condition and then repeat to get results.

**2. Scale Invariant Feature Transform:** Lowe proposed SIFT algorithm in 2004 [6] to solve the image affine deformation, scaling, illumination changes, also has strong robustness. The SIFT algorithm has four main steps:

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

- (A) Detection of Scale Space Extrema
- (B) Key point Localization
- (C) Orientation Assignment
- (D) Description Generation

The first stage is to identify location and scales of key points using scale space extrema in the DoG functions with different values of  $\sigma$ , the DoG function is convolved of image in scale space which is separated by constant  $k$  as in the following equation.

$$D(m, n, \sigma) = (G(m, n, k\sigma) - G(m, n, \sigma)) \times I(m, n) \dots\dots (2)$$

Where,  $I$  is the image and  $G$  is nothing but Gaussian function.

Now the Gaussian images are subtracted to produce a DoG, the Gaussian image is subsampled with factor 2 and produce DoG for sampled image. A pixel compared of  $3 \times 3$  neighborhood to detect the local maxima and minima of  $D(m, n, \sigma)$ . In the key point localization step, key point candidates are localized and refined by removing the key points where they rejected the low contrast points. In the step of orientation assignment, the orientation of key point is calculated which is based on local image gradient. In description generation stage is to calculate for every keypoint the local image descriptor which rely on image gradient magnitude and orientation [2]; these samples build a 3D histogram of gradient orientation and rotation; with  $4 \times 4$  array location grid and 8 orientation bins in each sample.

### Construction of SIFT Descriptor

Figure 2 illustrates the calculation of the key point descriptor. Initially the image gradient magnitudes and around the key point location orientations are sampled, by the use of scale of the key point for selecting the level of Gaussian blur for the image. For achieving invariance of orientation, the descriptor's coordinates, then the gradient orientations are rotated relative to the key point orientation. Fig no. 1 calculated at each sample location on the left side with small arrows. By finding interesting local keypoints SIFT features are generated in an image.

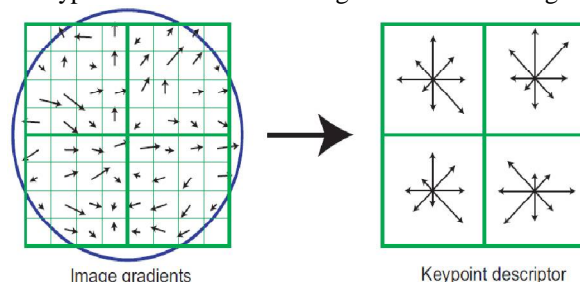


Fig 2. A keypoint descriptor using gradient magnitude [3]

A very efficient way of making these keypoints is locate the minima and maxima of Difference-of-Gaussians in the scale-space pyramid of image. This is done by calculating Gaussian blur's different levels on the input image and then calculating the difference of neighboring octaves. A canonical orientation for each keypoint vector can then be estimated, thus giving a complete keypoint coordinate of  $x$  and  $y$ , orientation and scale. A SIFT feature is calculated by combining the orientation histograms of locations which are closely surrounded by the keypoint in scale-space which is 128-dimensional vector. They are invariant to scale and rotation is advantage of SIFT keypoints, and relatively robust to perspective changes. For object recognition tasks this is an ideal descriptor. One problem of SIFT is its relatively high dimensionality, which makes it less suitable for closest neighbor against a training dataset. PCA-SIFT [1] is an extension to SIFT which aims to reduce high dimensionality of SIFT after applying PCA.

The goal of the recognition stage is to detect and to find all instances of the guided object in an incoming video stream. We again start out by generating all PCA-SIFT keypoints for every observation received from the input image. We then

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

perform a nearest neighbor lookup from the reduced training set T for each of these features against every feature. A simple linear nearest-neighbor search does current implementation, mainly for simplicity reasons. Any other more sophisticated data-structure used for more important dimensional data (such as KD-trees) could be used equally well if additional performance is required. The Euclidean distance in PCA-SIFT space between a feature and its nearest neighbor detects threshold  $\theta$  is defined as from the training dataset. Choosing a good value for  $\theta$  is critical. A smaller value of  $\theta$  will give less, but more accurate matches while a larger value of  $\theta$  will give large matches while increasing the likelihood of false positives. The object's center is calculated by counting it as a vote in the two dimensional voting space. This is called a heuristic approximation. It will not be confirmed that the same feature can come off at a different relative location towards the center of object than in the training data. However, here it can be supposed that most votes do in fact approximately match the center of object and that our voting algorithm should care of potential outliers. An interesting side-effect of this kind of algorithm is that it solves the problem of occlusion.

**3. Speeded Up Robust Features:** SURF (Speed Up Robust Features) algorithm, is based on multi-scale space theory and the feature detector is based on Hessian matrix. Since Hessian matrix gives better and accuracy and performance. In image I,  $x = (x, y)$  is the given point, the Hessian matrix  $H(x, \sigma)$  in  $x$  at scale  $\sigma$ , it can be define as

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{yx}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \dots (3)$$

Where  $L_{xx}(x, \sigma)$  is the convolution result of the second order derivative of Gaussian filter of  $(\sigma)$  with the image I in point  $x$ , and similarly for  $L_{xy}(x, \sigma)$  and  $L_{yy}(x, \sigma)$ . SURF produces a stack without down sampling for higher levels in the pyramid resulting in images whose resolution is same. By using integral images, SURF filters the stack using a box filter approximation of second-order Gaussian partial derivatives. Since integral images permits the finding of rectangular box filters in near constant time. In Figure 3 shows partial derivatives in y-direction and xy-direction of the Gaussian second orders.

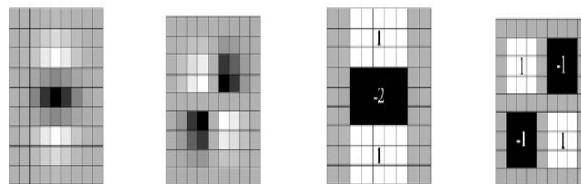


Fig 3. The Gaussian second orders partial derivatives in y-direction and xy-direction [4]

In descriptors, As compared to other descriptors SIFT gives good performance. The proposed SURF descriptor is based on similar properties. The first step includes fixing of reproducible orientation which is based on information from a circular region around the interest point. And second constructs a square region which is aligned to the selected orientation, and extract the SURF descriptor from it. For avoiding dependency upon rotation, it estimates the Haar-wavelet responses in x and y direction shown in figure 3. To detect features across scale we have to examine several octaves and levels, where SIFT scales the image down for each octave and use progressively larger Gaussian kernels, the integral images allows the SURF algorithm to calculate the responses with arbitrary large kernels. Increasing the size of kernels, and keeping the lobes correctly scaled this does pose two challenges, how to scale the approximated kernels and how this influences the possible values for  $\sigma$ . The kernels has to have an uneven size to have a central pixel and the rectangular areas, the lobes, has to have the same size. The SURF paper [2] goes into detail with these

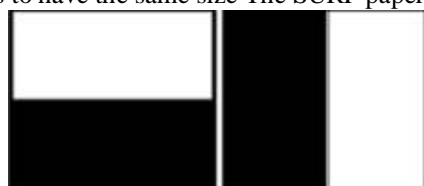


Fig.4 Haar wavelet types used for SURF [5]

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Considerations. The result is that division of scale space into levels and octaves becomes fixed. The filter size will be large and with a regular Gaussian kernel if the convolution were to be done this would be prohibitively expensive. The use of integral images make this feasible as well as it fast, and without the need to downscale the image. It should be noted that this approach with large box filters can preserve and be sensitive to high frequency noise [4].

## IV.CONCLUSION

In this paper, we have discussed various object detection techniques. The template matching technique requires large database of image templates for correct object recognition. Hence it must be used only when limited objects are to be detected. SIFT, Global features and shape based method can give better result and are efficient as compared to local features. These techniques help in easy access of the images. They also find their application in fields such as biometric recognition, medical analysis, surveillance, etc. A method for multiple object detection is also presented.

## REFERENCES

- [1] T. Gao, B. Packer, D. Koller, "A Segmentation-aware Object Detection Model with Occlusion Handling", In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2011.
- [2] W. Hu, A.M.Gharuib, A.Hafez, "Template Match Object Detection for Inertial Navigation Systems," Scientific research (SCIRP), pp.78-83, May 2011.
- [3] E.Shechtman, M.Irani, "Matching Local Self-Similarities across Images and Videos," In IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.
- [4] F. Khan, R. Muhammad , et.al., "Color Attributes for Object Detection," In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3306 – 3313, 2012.
- [5] T. Gevers, A. Smeulders, "Color-based object recognition," Pattern Recognition, 1999.
- [6] P. Viola and M. Jones, "Robust real-time object detection," International. Journal of Computer Vision, 57(2), pp.137–154, 2004.
- [7] R. Fergus, P. Perona, A. Zisserman, "Weakly supervised scale-invariant learning of models for visual recognition," International Journal of Computer Vision, 2006.
- [8] G. de Croon, "Active Object Detection," In 2nd International conference on computer vision theory and applications (VISAPP 2007), Barcelona, Institute for Systems and Technologies of Information, Control and Communication (INSTICC), pp. 97–103, 2007.
- [9] A. Berg, T.Berg , J. Malik, "Shape Matching and Object Recognition using Low Distortion Correspondences," In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 26 – 33, 2005.