RESEARCH ARTICLE OPEN ACCESS

Object Detection and Localization Using SURF Supported By K-NN

Upendra Singh [1], Sidhant Shekhar Singh [2], Manish Kumar Srivastava [3]
UG Students [1] & [2], Assistant Professor [3]
Department of Computer Science and Engineering
Madan Mohan Malaviya University of Technology
Gorakhpur - 273010
UP - India

ABSTRACT
The ability to segregate objects from its background is really an important task for robots to interact with surroundings in real life scenario. The advent of geometrically rich feature based objects has given impetus to research in the field of visual image recognition. This paper aims at providing a method for object detection with the help of SURF (Speeded-Up Robust Feature) algorithm used with KNN (K-Nearest Neighbors), a supervised Machine Learning Algorithm . This paper purports a method for object detection in which a set of features are extracted from an image captured from different perspectives. To increase the segmentation accuracy, we preform fixated point type segmentation on image and next key points are estimated in each segment with the help of Speeded-Up Robust Features (SURF). These key points are used to carry out the matching task for every detected key point in segmented image of the scene. We analyze and evaluate the performance of our system with other recently proposed methods such as the scheme pro-posed by Mae et al.
Keywords:- Object Detection; K-NN; Segmentation; SURF; Feature Extraction; Voting.

I. INTRODUCTION

In the area of intelligent systems, object detection which is the task for searching and localizing objects in a scene is deliberated as prime feature for autonomy. This fact has impelled unprecedented research in this field and as a result several algorithms have been proposed in last two decades. The performance of all these algorithms has been promising up to some extent but the global acceptance of a single algorithm is still debatable.

It has been frequently regarded that human visual system works on the pattern in which it makes series of fixations at various conspicuous locations in an image while observing static or dynamic scene. It has spurred many researchers but unfortunately very little is known about operations carried out in human eye while fixation. In our proposed method, we postulate that during a fixation, before recognizing any object human eye first of all segments that object out of that image in lieu of segmenting the entire image at once. So, we perform region based segmentation with the help of fixated points.

For detection task, literature witnesses many algorithms have been suggested. All previous proposed algorithms share a common problem that they succumb in case of limited image information. Limited information can be understood as the diminutive size of object and occlusion. To tackle these problems, Mae et al. resorted on a local feature matching algorithm which uses local geometry consistency for detection task. This approach uses SIFT for feature extraction and matches with those of reference image. Despite the advantage of simplicity, it does bring some limitations i.e. non-planner surfaces pose an adverse effect on performance. So, the field of object detection still requires further research in order to achieve 100% accuracy. To meet the expectations and initiate further research in this field, this paper proposes an object detection algorithm based SURF algorithm assisted by fixation based segmentation. Dealing with real-life situations, due to perspective changes, various types of geometrical deformations are introduced in our image. The issue of required level of invariance needs to be tackled to enhance robustness of algorithm. In SURF emphasis is given on scale and image rotation invariant detector and descriptor. These descriptors are deemed robust enough to deal with second order effects like skewness, anisotropic scaling and perspective effects.

The contents of this paper are structured as follows- Section 2 Provides the brief overview of the proposed system. Various parts of section 2 i.e. 2.1 to 2.3 delve into the detailed description of various steps involved. Section 3 provides the description of detection. Crux of this paper is broached in section 4.

II. PROPOSED METHOD

The proposed system is based on fixation based segmentation and SURF algorithm along with K-NN voting mechanism. Fixation based segmentation strategy is used to segregate objects in an image. Here, segregation of objects means identification of boundaries of objects contained in the given image. Secluded objects from the given image are given as input to SURF which extracts all the feature points from each object. These extracted feature points are compared with the existing data in the database which has been constructed during training process. K-NN algorithm works on the voting principle which matches extracted features of objects with database and the trained object features stored in database matches at most is disclaimed as the detected object. Next subsections provide a detailed description of each step of proposed system.

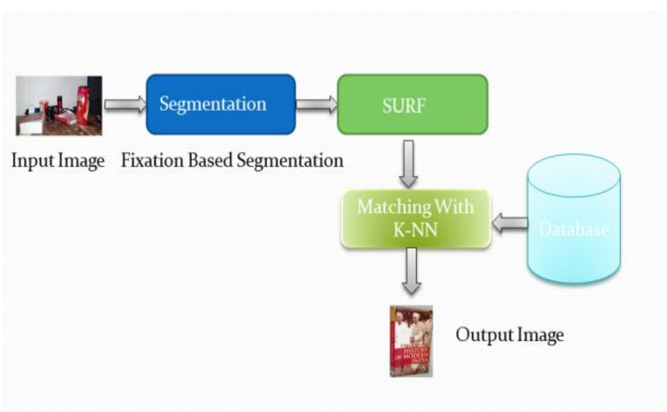


Fig. 2 Proposed Scheme

2.1 Segmentation:

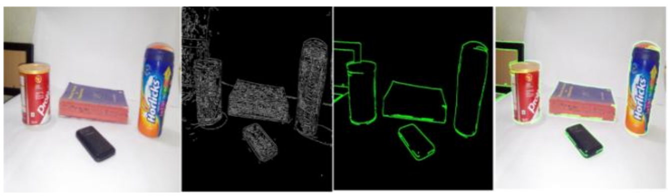
To interact with surroundings for robots, object perceptibility is equally important as the navigation capability. Here, object perception stands for identifying the boundary of objects. The segmentation strategy used here is a point based segmentation that aims at finding an optimal closed contour around a given point in an image. Here, boundaries can be object boundary or a non- object boundary. So, we have to employ techniques to filter out the non-object boundary. The concept of boundary ownership is used here, which means knowing their object sides. The information about boundary ownership is used to differentiate between object and non-object boundary.

Fig. 2.1 - Object and non-object

Overview of segmentation process used here is given as follows- First of all, a probabilistic boundary edge map is generated using the cue values (Color, Texture and Intensity). For the pixels which are likely to be the boundary pixels, the boundary ownership is also calculated. With the help of boundary ownership information, we can differentiate between objects and non-objects. Now, points within object boundaries are selected and closed contours are drawn around those points. All generated contours are refined to get the exact boundary of objects.

The probabilistic boundary edge map is used to determine whether a particular pixel is a boundary pixel or not. The higher is the probability, higher are the chances of a pixel being the boundary pixel. For the given image a 2-D optical flow map is generated with the help of cue values. Color and texture gradients of the pixels are calculated and all the edge pixels with non-zero color and texture gradients are overlaid on the flow map. The probability of calculated edge pixel to be edge or contact boundary is calculated and maximum of these two **($P_e = \max(P_d, P_c)$)** is assigned to edge pixels. Next step is to locate to fixating point for segmentation so that fixation based segmentation strategy [old fixation] can be used. Selection points for fixation are based upon the probability of edge pixel calculated in the previous step. Edge pixel having boundary probability greater than 0.5 is assumed to be boundary of some object in the image. Here, the subset of boundary edge pixel is given by O_b ,*$O_b(x,y)$={1 if P_e>0.5 ; 0 otherwise}.*

Now, on the basis of these fixated points many closed contours are generated. The emphasis is given on selecting the proper contour that corresponds to the objects in the image. Contours having higher coverage are selected because they are more likely to represent the boundary of objects in the image.



(a) (b) (c) (d)

Fig. 2.2 (a)input image (b)Canny edge detected (c) Boundary detection (d)Boundary superimposed on input image

2.2 Feature Extraction:

SURF (Speeded-Up Robust Features) is scale and rotation-invariant interest point detector and descriptor. Absolute details of SURF algorithm is beyond the scope of this paper. So we will review its most significant properties and describe its suitability for our proposed system. The merits of using SURF is that skew, scaling and perspective effects are covered to some extent due to the overall robustness of the descriptor. SURF mainly finds its application in finding similarity between two images of the same object. Before we delve into our proposed method, we first need to understand the key steps involved in SURF algorithm. SURF algorithm consists of these following steps: 1) detection of Interest key point; 2) estimation of a feature vector (descriptor); 3) matching between images of the given object. Interest point detection aims at finding the geometrically rich relevant feature points from the image of the object in order to extract valuable information from the image. The extracted information is fed to the descriptor which computes with the help of Hessian-Laplace matrix detectors.

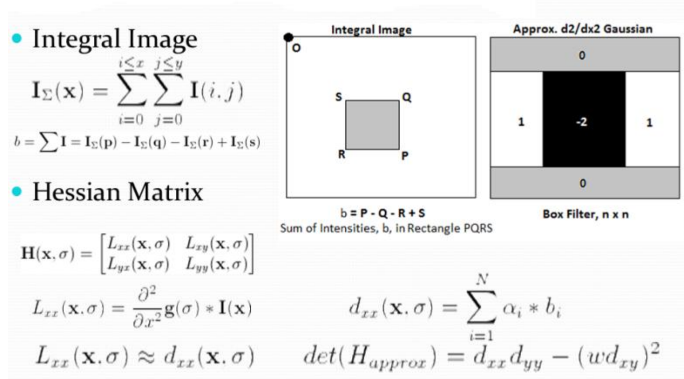


Fig 2.2.1 Hessian Matrix

Fig 2.2.2(a) depicts the image of an object and fig 2.2.2(b) represents the features extracted from the image of object.

. (a) (b)

Fig. 2.2.2 (a) Training image (b) Feature points extracted from training image

2.3 Training:

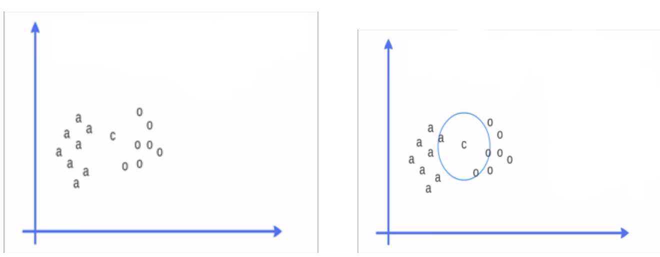
Training phase starts with taking different pictures of objects with several perspectives. Each clicked image is fed to SURF for feature extraction. Extracted points of interest contain some unwanted features that do not belong to the object. So, emphasis is given on to remove such phony features with the help of annotation mask. Furthermore, feature descriptor is generated with these reduced points of interests. Now, simple K-nearest neighbor machine learning algorithm is fed these feature descriptors which labels these descriptors with the name of training object.

2.4. Matching Mechanism:

The key points estimated that are inside the localized object boundary are considered in matching process. K-nearest neighbor algorithm is used as matching mechanism. K-NN is a type of instance-based learning, or lazy learning, where first of all local functions are calculated and computation is done after classification.The *k*-NN algorithm is a simple machine learning algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).K-NN works on a principle that the points (extracted feature vectors) which are close in the space belong to the same class. The algorithm assimilates all training samples and predicts the response for a new sample by analyzing a certain number (**K**) of the nearest neighbors of the sample by using some similarity measure such as Euclidean distance measure etc., the distance between two neighbors using Euclidean distance can be found using the given formula.

$$Dist(X,Y) = \sqrt{\sum_{i=1}^{D}(Xi - Yi)^2}$$

Working of K-NN is demonstrated in fig 2.4.1 (a) and (b) where N training vectors are given and K-NN algorithm identifies the k-nearest neighbor of ‘c’. Feature ‘c’ needs to be classified either in class ‘a’ or class ‘o’. For instance, the value of k is taken 3. So, K-NN algorithm looks for 3 nearest neighbor for feature ‘c’. And it turns out two features of class ‘o’ and one feature of class ‘a’ resides in the very vicinity of ‘c’. Since, class ‘o’ has outvoted class ‘a’, feature ‘c’ is declared to have belonging with class ‘o’. Selection of K’s value is critical. The value of ‘K’ is chosen in such a way that it is always odd and is not the multiple of the number of classes to avoid the ties.



(a) **(b)**

Fig 2.4.1 Working of K-NN

2.5 Proposed Algorithm:

In this section, we give a succinct description of our proposed algorithm:

1) Load training image
2) SURF extracts interest points from training image
3) We generate training interest point descriptors
4) Feed descriptor to K-NN for training
5) Load Test image for detection
6) Apply fixation based segmentation scheme to break test image and extract object contour
7) While (Contour)
   Detect interest points
   Extract descriptors
   Match interest points with training points (via K-NN)
   If (Matching Points > Threshold)
   Compute Homographic Transform Box
      Draw Box on Object and Display
   Else Continue
End While

III. EMPIRICAL RESULTS

This section endues the results of empirical evolution of proposed system. The experiment is conducted in a room using 12 different objects as shown in figure 3.1(a). In order to build the training dataset for K-NN algorithm, user captures the pictures of each object from different angle and furthermore feature vectors are generated for learning phase as shown in figure 3.2(b).

3.1 Detection Performance:

In order to test the capability of our proposed system, a few experimental setups are constructed. Figure 3.1.2(a) represents the setup1 where the scene contains 4 objects; only three of them belong to database which means that our system will only be able to detect three of them. Fig 3.1.2(b) depicts the output image of fixation based segmentation that contains the localized objects. Extracted feature points are shown in fig 3.1.2(c). Figure 3.1.2 (d), (e) and (f) contain the objects that have been detected by K-NN matching mechanism.

Fig 3.2(b) Learning Phase, objects used for training.

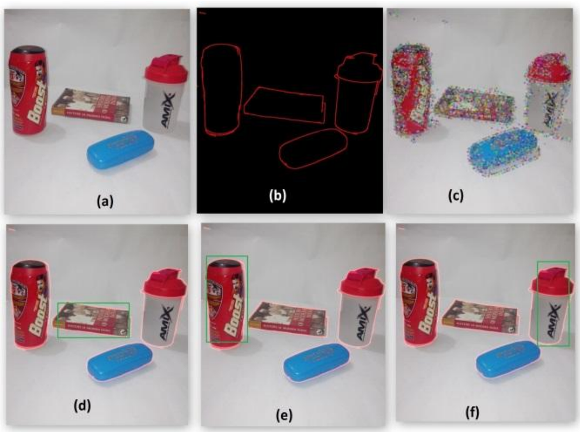


Fig 3.1.2

Fig 3.1.2 – (a)input image (b))Segmented image (c)Feature extracted from input image (d)Prediction of book detection by feature matching (e) Prediction of boost box detection (f) prediction of mixer box detection.

Figure 3.1.3(a) represents the setup1 where the scene contains 4 objects; only three of them belong to database which means that our system will only be able to detect three of them. Fig 3.1.3(b) depicts the output image of fixation based segmentation that contains the localized objects. Figure 3.1.2 (d), (e) and (f) contain the objects that have been detected by K-NN matching mechanism.

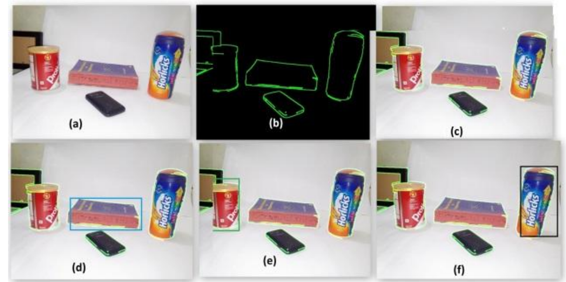


Fig 3.1.3– (a)input image (b))Segmented image (c)Feature extracted from input image (d)Prediction of book detection by feature matching (e) Prediction of 'protinex' box detection (f) prediction of horlix box detection.

Above two different experimental setups show that if background and objects are contrasting and object prsented in the scene reside in database then objects are detected and localized in each scenario.

IV. CONCLUSIONS

This paper proposes a new way to effectively recognize and locate multiple object in image which utilizes SURF with K-NN. This proposed approach can be a helping hand to those systems which seek object recognition as prime feature. Experimental results illustrate that it can be used in an online robot system for search task in real time environment. Experimental results show that this proposed algorithm has shown potential results in majority of scenarios, although its performance degrades when it required detecting transparent object or objecting having same color as the background. SURF algorithm does have some limitations. It requires object to be geometrically rich and the performance degrades in case of objects having low geometrical features. In that scenario, contour based techniques outperform this technique.

REFERENCES

[1]. Ajay Mishra , Yiannis Aloimonos , Cheong Loong Fah, Active Segmentation with Fixation,Computer Vision, 2009 IEEE 12th International Conference.

[2]. Mishra, Ajay K. and Aloimonos, Yiannis. "Visual Segmentation of Simple Objects for Robots" Paper

presented at the meeting of the Robotics: Science and Systems, 2011.

[3]. Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool. *Computer Vision and Image Understanding (CVIU)*(*2008*) D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, vol. 2, No. 60, pp. 91-110, 2004.

[4]. Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K.. "kNN Model-Based Approach in Classification." Paper presented at the meeting of the International Conference on Ontologies and Databases and Applications of Semantics and ODBASE 2003, 2003

[5]. J. Schlecht and B. Ommer Contour Based Object Detection, In Proceedings of the British Machine Vision Conference (BMVC), Sept 2011.

[6]. Stefan Zickler and Manuela Veloso. Detection and Localization of Multiple Objects. In Proceedings of Humanoids 2006, Genoa, Italy, December 2006.

[7]. Miriam Lopez-de-la-Calleja, Takayuki Nagai, Muhammad Attamimi, Mariko Nakano-Miyatake, Hector Perez-Meana , Object Detection Using SURF and Superpixels*, Journal of Software Engineering and Applications, 2013, 6, 511-518

[8]. S. Helmer, D. G. Lowe, Object recognition with many local features, Workshop on Generative Model Based Vision, Washington, D.C., 2004.