© 2015 IJEDR | Volume 3, Issue 2 | ISSN: 2321-9939

An Efficient Approach for Detecting Thematic Object Using Surf

T.K.Puneethkumar[#1], R.Jeya[*2]
[#]Final year - M.Tech CSE, [*]Assistant Professor,
SRM University,
Chennai-603203, India

***Abstract* - The object that appears frequently in a set or sequence of images is defined as thematic object. In a collection of image sequences, thematic object as a key object. Detecting common objects that appear frequently in a set of images is an intriguing problem. It also lacks the prior knowledge in common pattern. System is proposed so that it overcomes the above problems. System consists of Pre-processing, which enhances the data images prior to computational processing. SURF algorithm is used for detecting, extracting and for matching the feature points with respect to geometric transformations. Obtained feature points from the different images are compared to determine the common feature from the data images. This solution helps to locate the thematic objects much more easily within less time.**

***Keywords* - thematic, feature, detecting, object, extraction**

I. INTRODUCTION

Given a number of collection of images, it is of great interests to clip the common objects that have many occurrences. Rather than mining and sectioning object categories in this system focus on mining similar common objects .There are many applications to discover and clip common objects from images, such as object retrieval, video summarization, and nearly repeating image detection. To automatically discover and locate common objects, there are two major challenges. First of all, there misses a anterior knowledge of the common pattern, thus not known in advance (i) the shapes and visual aspect of the common objects; (ii) the locations and scales of common objects in unique images .Moreover, each image may contain one or more common objects. The same object can look different when presented from different view-points, scales, or under different lighting conditions, not to mention partial closures. It is not trivial to handle its variations and accurately situate its occurrences in the image. Although constant local features greatly improve the image matching, accurate localization of common objects at the sub-image level remains a difficult problem.

To address the above problems, a novel bottom up approach for similar common object discovery. This concentrates on the accurate localization of the common objects, in order to clip them out from the background clutter. First of all, every image is characterized by a collection of local features, which is referred to as visual primitives. It matches the visual primitives and gradually expands them spatially to recover the whole common object. In the initialization phase, “uncommon” visual primitives that are of limited matches compared with the other images are discarded, because they will not belong to any common pattern. For each remained visual primitive, it consider its local spatial locality as a larger visual group and check the commonness score of this spatial pattern. Accompanying by a multi-layer commonness checking of different spatial scales, every local feature is gradually assigned a commonness score, which indicates its likelihood of belonging to a common object. The commonness score of any part of image is the summation of the scores of its local features. By searching the sub-image of highest commonness score in each image, it can locate and crop the common object. There are several advantages of our method. First, unlike top down productive models that based on a visual vocabulary for topic (*i.e.*, visual category) discovery, our method only relies on the matching of visual primitives. It can automatically detect and locate common objects without requiring and knowing their scales and shapes, as well as the total number of such objects. Moreover, it can handle object variations such as scale and slight point-of-view changes, color and lighting condition variations, and it is insensitive to partial occlusion. Finally, it does not require a large number of images for data mining and works well to detect common object from a very limited number of images. It can be tested on both image datasets and video sequences. The proposed system uses the pre-processing before applying SURF. This pre processing consists of median filter. This is used for removing the low contrast pixel, because the low contrast pixels have low brightness. The SURF is used to detect the descriptor points and used for feature extraction and feature matching .The detected box in the image defines the thematic object .The results are validated for the robustness and effectiveness of this method.

II. RELATED WORK

Literature survey is the most important step in the research oriented project. Before developing the system, it is necessary to determine time factor and cost which will denote the efficiency of the project. In this few reference papers are studied and analyzed, so that the problems in these approaches can be enhanced. The research orientation in this project is done to improve and to enhance the quality and efficiency of the present system. The time factor and cost factor can also be defined well and also the database with quality and trust will be made available so the extraction will provide quality results.

Gangqiang Zhao and Junsong Yuan [1] propose a novel bottom-up approach to automatically cropping common objects from images. Rather than modeling each image as a visual document and discovering common patterns through conventional text mining, we evaluate each visual primitive and gradually expand them to recover the common object. To speed up the image data mining, they propose a multi-layer candidate pruning method to efficiently discard unqualified candidates of common patterns. Using the each visual primitive obtaining a commonness score, these local evidences of common patterns are finally fused through finding the bounding box of the highest commonness core. Experiments on both image datasets and video sequences show that their method can crop common objects automatically despite variations due to scale, view-point, and lighting

Sinisa Todorovic and Narendra Ahuja [2] have formulated a new problem, which is that of completely unsupervised extraction and learning of a visual category frequently occurring in a given arbitrary image set, and presented its solution. The visual category is defined as a set of sub images characterized by similar geometric, photometric, and topological properties. Unsupervised means that the objective category is not defined by the user and whether and where any examples of the category appear in a specific image is not known. To discover category occurrences in the unlabeled image set, they have proposed using many-to-many matching algorithm that finds matching sub images within every pair of images. They have defined a new similarity measure between matching sub images that is recursively computed in terms of differences in geometric, photometric, and topological properties of sub regions embedded within the sub images. This similarity measure fuses the information of similarities of the embedded sub images, where the similarities are weighted with respect to their relative significance to recognition. They have presented an algorithm for estimating these weights without using any supervision. They also proposed computing a union of all matching sub images in the image set, interpreted as category instances, and thus obtaining the category model. The category model registers all (partial) views of category occurrences in the image set, yielding a representation of the complete (unconcluded) object. Empirical validation on seven benchmark data sets which present challenges such as object articulation, occlusion, and significant background clutter demonstrates high recall and precision of category detection and recognition, as well as high accuracy of segmentation of category occurrences, in completely unsupervised settings. In weakly supervised settings, using the same experimental procedures as those presented in prior work, our approach outperforms existing baseline methods in object detection and segmentation on almost all categories tested, with one\ exception where our performance is slightly inferior within standard deviation. Our qualitative empirical evaluation demonstrates that the learned category model correctly captures the recursive containment and spatial layout of regions comprising the category instances in the image set.

Junsong Yuan, Zhu Li, Yun Fu, Ying Wu, and Thomas S. Huang [3] have presented an effective common spatial pattern mining algorithm for image database. With the visual primitive database size of N, its complexity is around O (N2). Compared with global image features like color histograms, this method is more robust to the pattern variations and is color- insensible by using the SIFT descriptor points as visual primitives. Moreover, the proposed method does not have the local optimal limitation as the EM-algorithm, and can discover multiple common patterns simultaneously. Discovering such common spatial pattern among images is useful for measuring the matching between images and for indexing image database through an unsupervised manner. This gives a rich and efficient method for detecting common spatial patterns from images. Each image **I**i is described by a set of visual primitives, **I**i = {p1... pm} where p = {x, y, **d**} represents a visual primitive; (x, y) denotes the spatial location; and **d** is the descriptor vector of the visual primitive. A common spatial pattern P ⊆ **I**i is a set of spatially co-located visual primitives that has good matches in other images. Rather than searching all possible pattern candidates P in the image dataset, we detect P by gradually pruning those visual primitives that do not belong to any P. Such cropping process is probably correct since it does not discard qualified solutions. And this method is robust to different pattern variations by using local invariant visual features, and is only of a quadratic complexity of the total number of visual primitives in the database.

Kyle Heath, Natasha Gelfand, Maks Ovsjanikov, Mridul Aanjaneya and Leonidas J. Guibas [4] introduce the notion of capturing and exploiting global connectivity in large and dense image collections. This link images through image regions discovered by cosegmentation and use a measure of connectivity from spectral graph theory as a tool for improving the web construction process. They demonstrated a number of initial applications that exploit the global connectivity of such an Image Web. Just as the World Wide Web has established the value of interlinked text documents and the power of link analysis techniques, we believe there is value in having interlinked sets of images and in studying how to exploit their connectivity structure

T. Quack, V. Ferrari, B. Leibe, and L. Van Gool [5] propose a novel method to filter the large mass of features. It classifies features which have high chance of lying on instances of the object class of interest. This technique is intended as an intermediate layer between feature extraction and object detection. The filtered set of features our method delivers can then be fed into a higher-level object detector. They presented an efficient data mining approach to detect frequent and distinctive feature configurations, representative for an object class. Moreover, it shows how to exploit the mined configurations to measure how likely it is for features of novel test images to lie on an instance of the object class.

III. ALGORITHM

In feature detection, SURF is faster than SIFT which is the main requirement of the today’s real time application. It is the robust image detector and descriptor. SURF detector is mainly based on the approximated Hessian Matrix. On the other hand, the descriptor gives a distribution of Haar-wavelat responses within the interest point’s neighborhood. Both the detector and descriptor are used to reduce the computation time because descriptor has low dimensionality. So that SURF is better than previously used schemes with respect to repeatability, distinctiveness, robustness and speed.

SURF creates a “stack” without 2:1 down sampling for higher levels in the pyramid resulting in images of the same resolution. Due to the use of integral images, SURF filters the stack using a box filter approximation of second-order Gaussian partial derivatives, since integral images allow the computation of rectangular box filters in near constant time. In key point matching step, the nearest neighbour is defined as the key point with minimum Euclidean distance for the invariant descriptor vector. The Gaussian second order partial derivative box filters of Dyy and Dxy are show in Fig.1

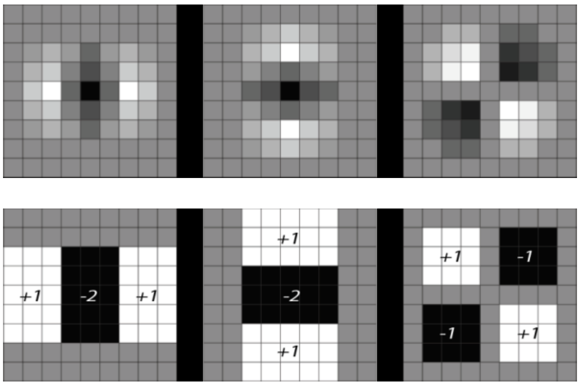


Fig. 1 The Gaussian Second order partial derivative box filters in y-(Dyy) and xy-direction (Dxy)

Fast Hessian Detector

Surf detector is based on the Hessian metrics which causes good performance and also good accuracy. Suppose in the image I, X = (x, y) is the given point, then the Hessian metrics H(x, σ) [2] for the X having the Scale σ, is defined as given in equation (1).

$$H(x,\sigma) = \begin{bmatrix} \mathrm{Lxx}(x,\sigma) & \mathrm{Lxy}(x,\sigma) \\ \mathrm{Lyx}(x,\sigma) & \mathrm{Lyy}(x,\sigma) \end{bmatrix} \ldots\ldots.. (1)$$

Where Lxx(x, σ) is the convolution of the Gaussian second order derivative ($\frac{\partial^2}{\partial x^2}$ g (σ)) with the image I in x and same for Lxy(x, σ), Lyx(x, σ) and Lyy(x, σ).

In approximated Hessian detector, an approximated Hessian matrix using box filter is used instead of only Hessian matrix as shown in Fig. 1. Here, box filter is used having σ = 1.2. Normally, the filter response is normalized with respect to the mask size.

SURF Descriptor

In the first step of the SURF descriptor, to extract the feature points, fix a reproducible orientation based on information from a circular region around the interest point. After, it built a square region aligned to the selected orientation. To become the invariant to rotation, it calculates the Haar- Wavelet which responses in x and y direction as shown in Fig. 2

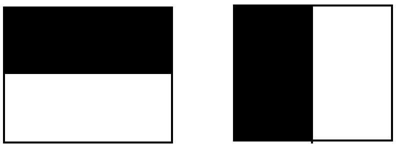


Fig. 2 Haar-wavelet response in x and y direction

Haar-wavelet response in x and y direction It can be processed in a circular neighborhood of radius 6s around the interest points, where s is the scale at which the interest points are detected. By calculating the sum of all responses within a sliding orientation window which covering a 60 degree, the dominant orientation is estimated. Then the horizontal and vertical responses within the window are summed and the resulting is called a new vector. The longest such vector lends its orientation to the interest point. The size of the window is 20s. The whole region is split up regularly into 4*4 square sub regions. A few simple features are computed at 5*5 spaced sample points. Here, dx is the Haar wavelet response in horizontal direction and Dy is the Haar wavelet response in vertical direction. So, this descriptor having the low dimensionality that reduces the computation time.

Affine Transformation

The most commonly used registration transformation is the affine transformation which is sufficient to match two images of a scene taken from the same viewing angle but from different position. It is composed of scaling, translation, and rotation. It is global transformation which is rigid. Affine transformations are more general than rigid. The equation for P'can be given as,

$$P' = Ap + t \ldots\ldots\ldots\ldots\ldots.. (2)$$

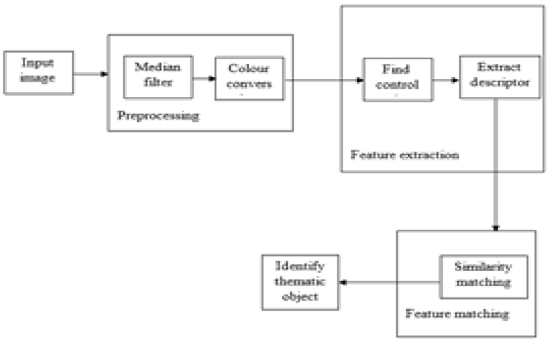
The general 2D affine transformation can be given as,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} tx \\ ty \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \ldots\ldots (3)$$

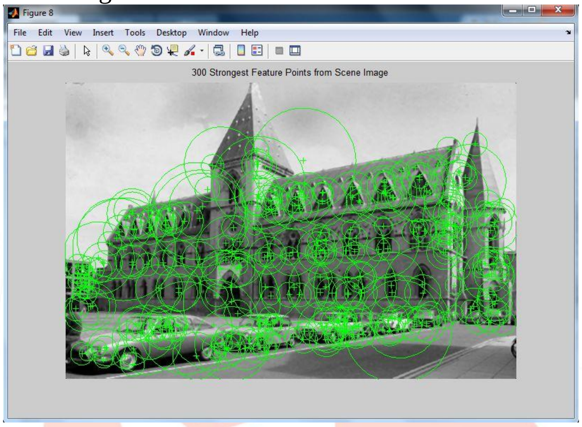
Angles and lengths are not preserved. Parallel lines remain parallel

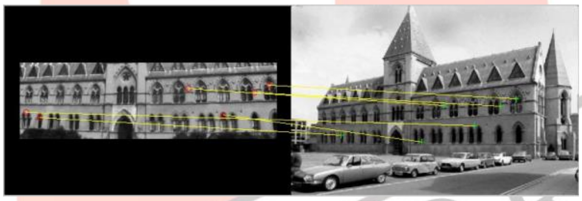
IV. ARCHITECTURE

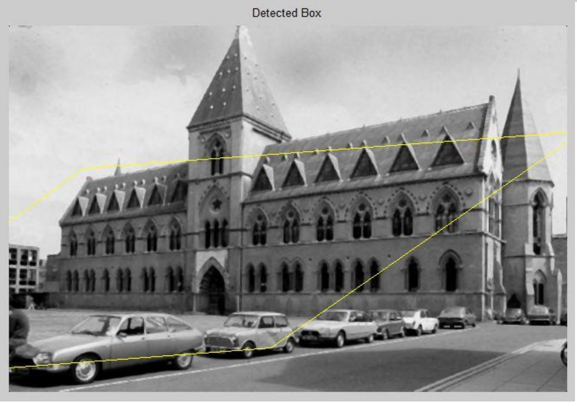
The Architecture has four major modules where the image processing takes place. Pre-processing is the technique of enhancing the data images prior to computational processing. Median filtering is used for reducing the "salt and pepper" noise while preserving edges. It is a non linear digital filtering methodology which is carried out with the aid of a window comprising of an odd number of samples. The values present within the window are arranged into numerical order. The median value, the sample in the center of the window is chosen as output .the oldest sample is abandoned, a new sample is obtained, and the calculations are re done .in general, the principle behind the medial filtering is that it modifies the pixels based on median function of a local neighborhood of the pixels i.e., running medians are computed. Color conversion is used for converting image from one color space to another color space.



In the second module the surf features are detected from the given image and also the same surf features for the database images also detected by using feature detection. Further the strongest feature values are extracted from the detected feature values from the input image and also from the database images next if any geometrical transformation is needed it was done. After the above surf feature detection and extraction from the input image and current image of the database further feature matching was done and the thematic object of the database images was found.







VI. CONCLUSION

In this paper give a new system for both feature matching and feature detection, the proposed design may compute features in lesser time as compared with existing SIFT. We reduce lot of time required for extracting feature. For feature matching we remove the drawback of SIFT matching also we reduce the time require for matching the features as compared with traditional nearest neighbor matching method by using SURF features matches features more accurate and consumes less time .our future work includes more fast and more accurate techniques for feature matching and feature extraction..

REFERENCES

[1] G. Zhao and J. Yuan, “Mining and cropping common objects from images, “in Proc. ACM Multimedia, 2010.
[2] N.Ahuja and S.Todorovic, “Unsupervised category modeling, recognition, and Segmentation in images,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 30, no. 12, pp. 2158–2174, Dec. 2008.

[3] J.Yuan, Z, Z, Z.Li, Y. Fu, Y.Wu, and T. S. Huang, "Common spatial pattern discovery "By efficient candidate pruning," in Proc. IEEE Conf. Image Process., 2007.
[4] K. Heath, L. J. Guibas, M. Ovsjanikov, M. Aanjaneya and N. Gelfand "Image webs: Computing and exploiting connectivity in image collections," in Proc. IEEE Conf. Compute. Vis. Pattern Recog. 2010.
[5] B. Leibe, L. Van Gool, T. Quack and V. Ferrari, "Efficient mining of frequent and distinctive feature configurations," in Proc. IEEE Int.Conf. Computer. Vis., 2007.
[6] O. Chum and J. Matas, "Large scale discovery of spatially related images," EIEE Trans. Pattern Anal. Mach. Intel., vol. 32, no. 2, pp. 371–377, Feb. 2010.
[7] A. Zisserman, P. Perona, and R. Fergus, "Object class recognition by unsupervised scale-invariant learning," in Proc. IEEE Conf. Computer. Vis. Pattern Recog., 2003.
[8] Mikolajczyk, K., Schmid, C.: An affine invariant Interest point detector. In: ECCV. (2002) 128 –142
[9] S. U. R. F. (SURF), "Herbet bay, andreas ess, tinne tuytelaars, luc van gool," Elsevier preprint, 2008.
[10] M. Zhou and V. K. Asari, "Speeded-up robust features based moving object detection on shaky video," in Computer Networks and Intelligent Computing. Springer, 2011, pp. 677–682.
[11] A. C. Murillo, J. Guerrero, and C. Sagues, "Surf features for efficient robot localization with unidirectional images," I Robotics and Automation, 2007 IEEE International Conference on. IEEE, 2007, pp. 3901–3907.