RESEARCH ARTICLE OPEN ACCESS

Binary Descriptor Approach- Local Difference Binary Approach For Object Recognition

Anjana Mittal[1] , Dr. Sidhhartha Choubey[2]
[1]M.E. Scholar [2]Associate Professor
[1,2]Department of Computer Science & Engineering
[1,2]SSGI Bhilai,(C.G.)

ABSTRACT
The problem in object recognition is to determine a given set of objects appear in a given image or image sequence. Thus object recognition is a problem of matching models from a database with representations of those models extracted from the image luminance data. The machine recognition provides poor results if pixel differences based approach can be applied for calculation face differences. Large scale memory is important for storing all the patterns which increases space complexity of the algorithm.

In this project, a binary descriptor, named local difference binary (LDB) is applied for obtaining the region of interest achieving computational speed and robustness. This approach is done in following three schemes. First, LDB utilizes both average intensity and first-order gradients of grid cells within an input image. Second, LDB employs a multiple gridding strategy to encode the structure at different spatial granularities. Coarse-level grids remove high frequency noise, while fine-level grids can capture detailed local patterns, thus enhancing distinctiveness. Third, LDB selects a set of salient features for minimizing distances between matches while maximizing the mismatch values, optimizing the performance of LDB.

Keywords: Object recognition, binary descriptor, spatial granularity, local difference binary and multiple –gridding.

I. INTRODUCTION

The object recognition is a challenging task in the real world [1]. The problem is that each object makes a different perspective to the human mind with the position of the object, pose and background. A human brain can effortlessly by its own hypothesis and biological inspirations. A method to solve the problem is to create a large database of the similar images with the variations is done. The different methods have developed over the years for the recognition of the object from the given set of images.

Some of the basic components for the object recognition are:

a) Base Model: The model describes the functional descriptions to the feature information. The functional information models for the size, shape and color of the features. Certain schemes are applied for facilitating the candidate objects for consideration.

b) Feature detector: A feature detector detects the objects locations forming the hypotheses. The recognition depends on the organization of the database the category of objet to be recognized.

c) Hypothesis formation and verification: A verifier or classifier can be considered for finding the object with the maximum relevance refining the likelihood of the correct object.

The object recognition depends on the many factors as the multiple entities exist in a scene [2].

Scene Constancy: The various conditions such as illumination, background and viewpoint that affect the images should remain. The performance of the detectors varies with the change in the scene conditions.

Model Space: The three-dimensional image when modeled into two-dimensional changes in characteristics. The selection of the feature is done in the two-dimensional would become different representation in three-dimensional space.

Number of objects: The number of objects decides the formation of the hypothesis. If the number of objects is less the hypothesis formation id not requires rather only the matching does the recognition. Otherwise, if the number of objects is more, an effort is required for the selection of features for recognition.

Number of possible occlusions: Occlusions present in an image hides the features and generates unexpected features. The presence of occlusions needs to be handled in the hypothesis verification stage. In the object

Recognition, if the number of object present in the scene is one the object is completely visible. If the multiple objects exist, the occlusions may appear and can lead problem in recognition.

II. LITERATURE SURVEY

In 2012 Reza Oji [2], proposed a method for object recognition with boundary detection with the ASHIFT (Affine Scale Invariant Feature Transform) and region merging. Initially the segmentation is done o partition the image into smaller portions. The images are simulated with two camera orientations. The key point descriptor relative to the orientation can get invariance in rotation. The histogram is created of local gradients to select the scale. The highest peak shows the direction of the local gradient. The regions in the image that corresponds to the object are identified with the Euclidean measure. The lower the distance hence more is the similarity. The merging process is done after the ASHIFT is performed. The basic aim of merging is to merge the non-object regions so that the object gets highlighted.

In 2012, V.Subbaroyan, Dr.Selvakumar Raja [3], proposed an approach for the recognition of the multiple objects by considering the features from spatial as well as frequency domain. The input image is converted into gray scale image. The Sobel operator is used to find the gradient of the image intensity. The color features of the image are obtained with the help of the global color histogram. The two level wavelet transform is done to obtain the feature with higher frequency. The spatial features are obtained from three color moments for the basic colors. Both features are fused to obtain the recognized object.

In 2012, Shulin Yang et al [4], found that the different species and in different poses makes the focus to decompose a ne-grained object recognition problem into two sub-problems, first aligning image regions that contain the same object part and second extracting image features within the aligned image regions. In the proposed work, a template represents a shape pattern, and the relationship between two shape patterns is captured by the relationship between templates, which reflects the probability of their co-occurrence in the same image. This model is learned using an alternative algorithm, which iterates between detecting aligned image regions, and updating the template model.

Michael Calonder et al [5], proposed a method where the image patches are classified with respect to the pair wise intensity values. The Nave Bayesian Classifier recognizes the patches from the different viewpoints. A clear advantage of binary descriptors is that the Hamming distance. The individual bits are compared with the intensities of pair of points. The positions of the pixels are pre-selected randomly according to a Gaussian distribution around the patch center. The kernels are used to smooth the patches before intensity difference and spatial pair of images.

Stefan Leutenegger et al [6], proposed an algorithm that addresses the problem such as low memory, computation complexity, scaling efficient, rotation and noise. The method builds the descriptor bit-stream; a limited number of points in a specie sampling pattern is used. Each point contributes to many pairs. The pairs are divided in short-distance and long-distance subsets. The long-distance subset is used to estimate the direction of the key point while the short-distance subset is used to build binary descriptor after rotating the sampling pattern.

Chiverton et al**.** [7] proposed a new fully automatic object tracking and segmentation framework. The framework consists of a motion-based bootstrapping algorithm concurrent to a shape-based active contour. They propose two approaches of increasing computational intensity and accuracy that statistically estimate the foreground from the potential mixture of foreground and background enveloped by an alpha hull.

Zhenjun et al. [8] used combined feature set which is built using color histogram (HC) bins and gradient orientation histogram (HOG) bins considering the color and contour representation of an object for object detection. The combined feature set is the evolvement of color, edge orientation histograms and SIFT descriptors.

R. Saravanakumar et al. [9] represented the objects using the properties of the HSV color space. Adaptive k-means clustering algorithm was applied to cluster objects centroids color values and co-ordinates were sent to next frame for clustering. A three dimensional representation of the HSV color space is a hexacone, with the central vertical axis representing intensity.

Xiaobai Liu et al. [10] proposed hybrid online templates for object detection which uses different features such as flatness, texture, or edge/corner. The template consists of multiple types of features, including sketches/edges, texture regions, and flatness regions. Sketch/edge regions usually consist of various links, ridges, such as corners and junctions. Texture regions are a large number of objects that are either too small or too distant to the camera. In contrast, flatness regions are always filled with homogeneous color or intensity. The limitation of this method is; as the discriminative power of features change along with the object movements, the hybrid template should be adaptively updated by either adjusting the feature confidences, or substituting the old features with the newly discovered ones from the currently observed frames.

Liu et al. [11] presented a novel semiautomatic segmentation method for single video object extraction. Proposed method formulates the separation of the video objects from the background as a classification problem. Each frame was divided into small blocks of uniform size, which are called object blocks if the centering pixels belong to the object, or background blocks otherwise. After a manual segmentation of the first frame, the blocks of this frame were used as the training samples for the object-background classifier.

Johnsen et al. [12] used Approximated median filter to perform background modeling. For the implementation, better results were obtained by scaling the increment and decrement by a step factor if the absolute difference between the current pixel and the median-modeled background pixel is bigger than a threshold.

Victoria Yanulevskaya et al [13] proposed an object-based visual attention theory for the task of salient object detection. They assume proto-object being a unit of attention and argue that notion of an object should be taken into account while assessing object saliency.

Carlos Cuevas and Narciso García [14], proposed the background modeling algorithm to detect real time moving object. The proposed method combined the background model and foreground model to detect an object from complex image very quality taken by non-completely static cameras. It finds the bandwidth matrices for the kernels which are used in background modeling. This proposed method updating the background model for reducing the misdetections. Bahadir Karasulu and Serdar Korukoglu [13], proposed Moving object detection and tracking by using annealed background subtraction method in videos. Current frame is subtracted from image and it is used to classify the pixel either foreground or background by comparing the difference with the threshold. Simulated annealing (SA) technique is used to rectify the p – median problem. The total weighted distance between demand points (nodes) and the closest facilities to demand points are used to find the p number. SA-based hybrid method is developed for performance optimization of back, which is used to detect and track object(s) in videos.

Ling CAI Et Al [15], proposed a stereo vision-based model for multi-object detection and tracking in surveillance. Illumination variation, shadow interference, and object occlusion problems are overcome by using stereo model. They identified the feature points after they projected into 2D ground plane. For grouping the projected points according to their height values and locations on the plane, a kernel-based clustering algorithm is used.

Bangjun Lei and Li-Qun Xu [16], proposed a detection and tracking of objects in wide range of outdoor surveillance and monitoring scenarios in real-time video analysis system. Adaptive background modeling technique is used to extract the foreground regions. A blob analysis is used for object tracking. It gives better result for non-crowded and the static state of scene.

2.1 Feature Descriptor

Feature point descriptors are widely used in many computer visions tasks such as marker less-augmented reality simultaneous localization and mapping. Some available descriptors are:-

The *SIFT descriptor*, proposed over a decade ago, has been widely adopted as one of the highest quality options. However, it imposes a heavy computation burden. This drawback has drawn extensive efforts for optimizing its speed without compromising too much quality.

SURF is arguably the most noticeable. But recent experiments have shown that the SURF descriptor is still too computationally heavy. Thus only a limited number of points can be handled for real-time applications such as AR, especially for handheld devices such as smart phones and tablets.

BRISK descriptor is a method for key point detection, description and matching. A novel scale-space FAST-based detector reveals the descriptor sped which is combined with the assembly of a bit-string descriptor whose intensity is compared with each neighborhood key point by sampling dedication.

BRIEF descriptor is proposed as an efficient feature point descriptor which uses binary strings. BRIEF using simple intensity difference tests for computation and even when using relative few bits, it is highly discriminative. The descriptor is much faster than other state-of-the-art ones, not only in matching and construction but also yields higher rates of recognition.

ORB descriptor is proposed on the basis of very fast binary BRIEF descriptor, which is resistant to noise and invariant to Rotation. SIFT is at two orders of magnitude slower than ORB, it is an efficient alternative to SIFT or SURF. It is an efficient alternative to SIFT or SURF. Its efficiency is tested on smart phone while patch-tracking and objects detection including several real-world application.

FREAK represents retina-inspired and

ISSN : 2456-2734, Vol. 1, Issue 1, August 2016, pp.50-56

coarse-to-fine descriptor to improve the performance of current image descriptors. It performs better than other state-of-the-art key point descriptor while remaining faster to compute with lower memory load and also more robust than SIFT, SURF or BRISK.

Lightweight binary descriptors such as *BRISK*, *FREAK*, *BRIEF* and its variant *rBRIEF* (or ORB descriptor), become increasingly popular as they are very efficient to store and to match. Raw intensities of a subset of pixels within an image patch for binary tests, and thus have low discriminative ability. Lack of distinctiveness incurs an enormous number of false matches when matching against a large database. The basic problem with existing descriptors is high computing complexity, matching and storing the feature point descriptors.

III. PROBLEM STATEMENT

From the existing literature a number of problems are identified as:

1) In multiple objects scene, many occlusions can participate in the scene and can generate false appearance of an object.
2) There a number of feature that increases the computational complexity of the object recognition system.
3) The matching of the object against the large database leads to huge number of false matches, thereby reducing the performance of the system.
4) The change in conditions such as lighting, pose, illuminations also affect the object recognition.

IV. PROPOSED METHODOLOGY

The proposed method is a 2 step process:-

1) Smoothing: The image is smoothened at the corner for the boundary values of the image. It obtains the homogeneous points and corner points of the image and selects the feature vectors.
2) Object Recognition: The process of object recognition is done from the set of features obtained after smoothing of the image. Local Difference Binary (LDB) is done for the object recognition. The image is divided into small blocks and the information from each pixel is extracted. This information describes the efficiency of the process and is the average of the pixel intensity values. The first-order gradient is obtained from the average intensities that confine the variation in the images. The level of block division needs to be fine as more granularity leads to more distinct recognition. For flow of proposed methodology refer Fig 1.

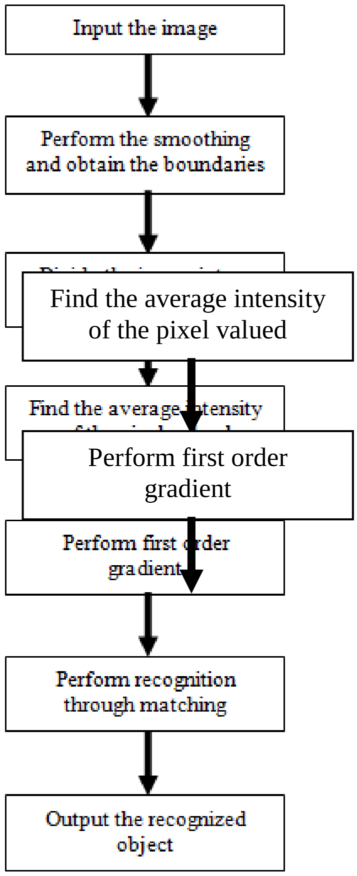


Fig 1: Flow Chart of proposed methodology.

V. RESULT

The proposed work is implemented in Matlab 13 on Windows Platform. In this Paper we have compared our results with the Existing methods. We are using the PASCAL- VOC 2007 dataset to test our Experiments.

Fig.2. shows the results of Existing Algorithm

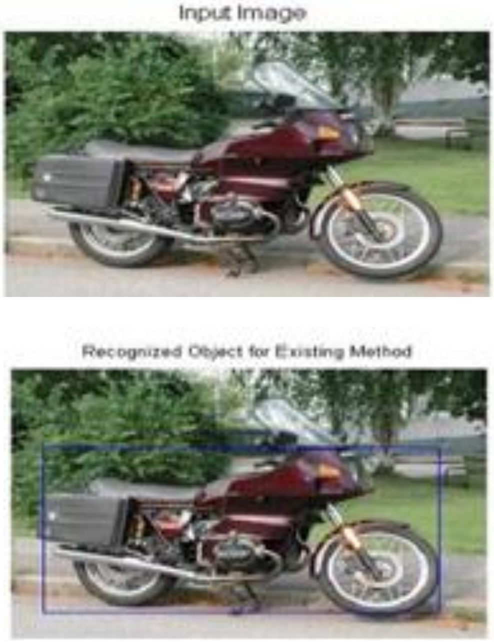


Fig.3. shows the results of Existing Algorithm

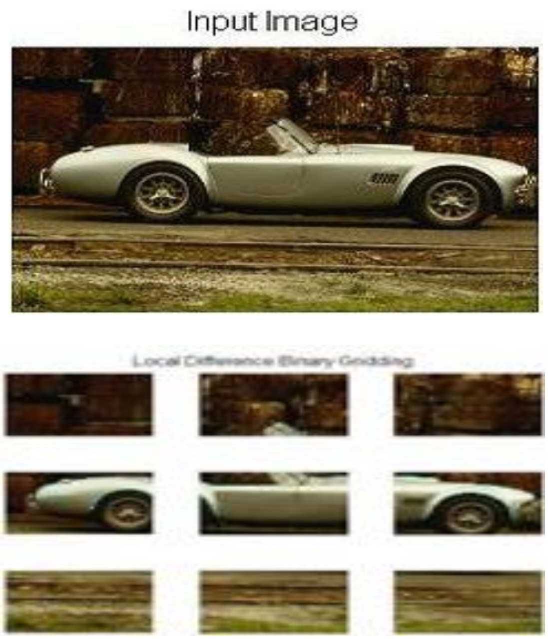


Fig.4. shows the results of Local Difference Binary Gridding of 3*3 Image

Fig.5. shows the results of Local Difference Binary Gridding of 3*3 Image

Fig.6. Final recognized object

ISSN : 2456-2734, Vol. 1, Issue 1, August 2016, pp.50-56

Fig.7. Final recognized object

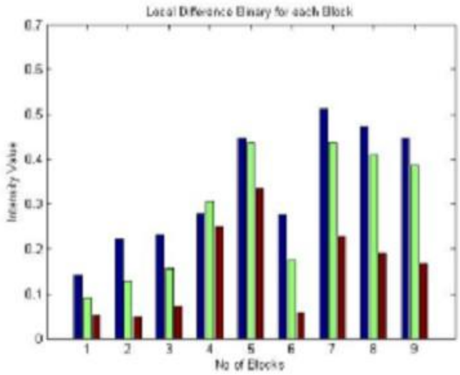


Fig8. Graph shows the intensity value and average value of car image

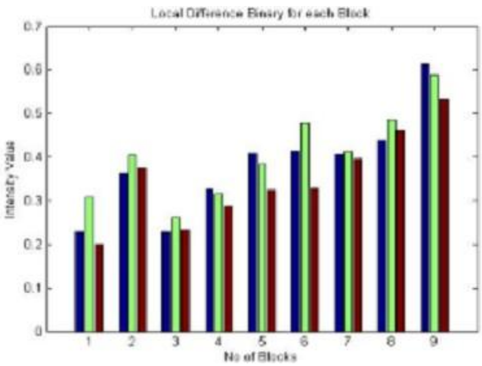
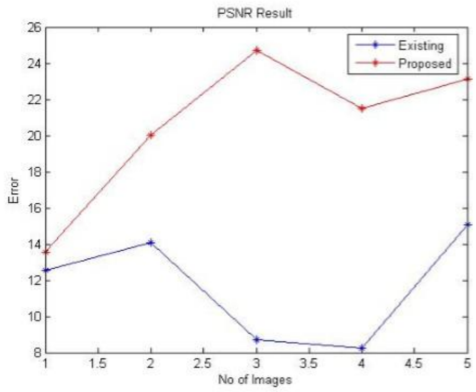
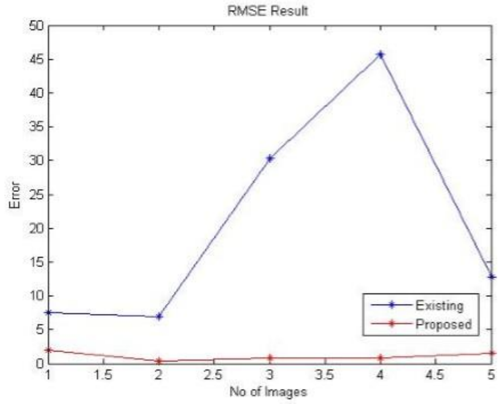


Fig9. Graph shows the intensity value and average value of bike image





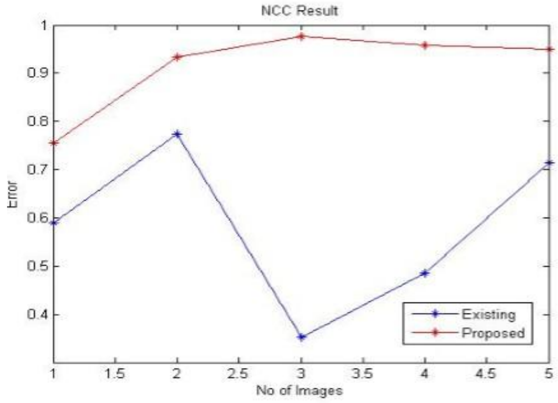


Fig10. Graph shows the comparison of result with the existing and proposed work

VI. CONCLUSION

The proposed method is compared with the localization of objects for the object recognition. The object recognition with the Object localization method performs recognition based on multi-class branch-and-bound formalism. The existing method does the object recognition with bounding box around the object. The proposed method performs the recognition with the boundary of the extraction of the object with the gridding for better recognition. The exact boundary of the object finds the object despite of the background in the image scene. Hence only the object of interest can be recognized better way. The measure Peak to Signal Noise (PSNR), Root Mean Square Error (RMSE), and Normalized Correlation Coefficient (NCC) has been evaluated for the performance of the proposed method over the exiting one. The experimental result shows that the proposed method has a RMSE and PSNR value better that the existing one. The NCC value shows the correlation of the object with the images in the database.

REFERENCE

[1]. Nicolas Pinto, David D. Cox , James J. DiCarlo, Why is Real-World Visual Object Recognition Hard?, PLoS Computational Biology, Vol. 4, Issue 1, pp. 0151-0156, 2008.

[2]. Reza Oji, An Automatic Algorithm for Object Recognition and Detection Based on ASHIFT Keypoints, Signal & Image Processing : An Interna- tional Journal (SIPIJ) Vol.3, No.5, pp. 29-39, 2012.

[3]. V.Subbaroyan, Dr.Selvakumar Raja, Novel Feature Fusion Method of Object Recognition Using Wavelet Transform, International Journal of Scientific & Engineering Research, Vol 3, Issue 9, 2012.10.

[4]. ShulinYang, LiefengBo, JueWang and LindaShapiro, Unsupervised Tem- plate Learning for Fine-Grained Object Recognition, Advances in Neural Information Processing Systems, pp. 31223130, 2012.

[5]. Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua, BRIEF: Binary Robust IndependentElementary Features, ECCV'10 Pro- ceedings of the 11th European conference on Computer vision: Part IV, pp. 778-792, 2010.

[6]. Stefan Leutenegger, Margarita Chliand Roland Y.Siegwart, BRISK: Bi- nary Robust Invariant Scalable Keypoints, ICCV '11 Proceedings of the 2011 International Conference on Computer Vision, pp. 2548-2555, 2011.

[7]. John Chiverton, Xianghua Xie and Majid Mirmehdi,"Automatic Bootstrapping and Tracking of Object Contours ", IEEE Transactions on Image Processing, Vol. 21, No. 3, March 2012.

[8]. Zhenjun Han, Qixiang Ye, Jianbin Jiao," Online feature evaluation for object tracking Using kalman filter "978- 1-4244-2175-6/08/$25.00 ©2008 IEEE.

[9]. S. Saravanakumar, A. Vadivel and C.G. Saneem Ahmed," Human object tracking in video sequences "ICTACT Journal on Image and Video Processing, August 2011, Volume: 02, Issue: 01.

[10]. Xiaobai Liu, Liang Lin, Shuicheng Yan, Hai Jin, and Wenbin Jiang," Adaptive Object Tracking by Learning Hybrid Template Online ",IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 11, November 2011.

[11]. Yi Liu, Yuan F. Zheng, "Video Object Segmentation and Tracking Using ψ - Learning Classification ", IEEE transactions on circuits and systems for video technology, vol. 15, no. 7, July 2005.

[12]. Swantje Johnsen and Ashley Tews," Real-Time Object Tracking and Classification Using a Static Camera "Proceedings of the IEEE ICRA 2009 Workshop on People Detection and Tracking Kobe, Japan, May 2009".

[13]. Victoria Yanulevskaya, Jasper Uijlings, Jan-Mark Geusebroek, "Salient object detection: From pixels to segments", Image and Vision Computing Vol. 31, Pg. No. 31–42, 2013.

[14]. Carlos Cuevas and Narciso García , "Improved background modeling for real-time spatio-temporal nonparametric moving object detection strategies", Image and Vision Computing Vol.31, Pg. No. 616–630,2013.

[15]. Ling CAI et. al., "Multi-object detection and tracking by stereo vision", Pattern Recognition Vol.43, Pg. No. 4028–4041, 2010.

[16]. Bangjun Lei and Li-Qun Xu, "Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management", Pattern Recognition Letters Vol.27, Pg. No. 1816 – 1825, 2006.