A Comprehensive Analysis Of Object Detection And Classification Techniques For The Purpose Of Computer Vision Applications

Jaswinder Singh [#1], Dr.B.K.Sharma [*2]

[#] Research scholar, Dr. A.P.J Abdul Kalam Technical University
Lucknow, India.

[*] Principal Scientific Officer, NITRA Technical Campus
Ghaziabad, India.

Abstract—Recently computer vision system based applications have grown rapidly. In the field of computer vision there are various application present which are used in real-world application scenarios such as medical imaging, security, visual surveillance etc. Among these application visual surveillance is also an important application. In order to obtain better surveillance system, object detection and recognition techniques need to be addressed. In this field, object recognition have gained lots of attraction for real time application various techniques have been proposed for object recognition. In this article we study about the stat-of-art techniques presented for object recognition. Later, a study is presented regarding learning based scheme for object detection system. Finally we have concluded that learning based approaches are robust to obtain the object detection but still the complexity remains a challenging issue for learning based schemes.

Keyword-image classification, object detection, convolution neural network, deep-learning.

I. INTRODUCTION

Recently, Computer Vision Techniques have been proven as a promising technology for real time applications such as surveillance system, medical imaging etc. In real time applications i.e. video surveillance system various occlusions and objects affect the performance of surveillance system. The technique of object detection is a promising and challenging technique in this field of compuover vision system. Object detection is a main component for image processing based application. Main aim of this technique is to identify the objects in any real time natural scene image i.e. if in an image face, dog and building are present then these object are detected using object detection scheme. Object detection techniques perform the desired operation with the help of feature extraction technique and feature learning scheme. Each object belongs to different object class which is known as object classification and carried out with the help of classification techniques. Classification is a process to identify the category of detected object in any given image. Each object or each class of object has its own features and characteristics which is responsible to make difference in each class. This techniques helps to recognize and classify the objects from any given input image. Object recognition techniques

is widely used in various real time applications such as image retrieval, surveillance, security, vehicle tracking and parking systems etc. Various challenges present in this field of object recognition. Challenging issue is considered about robustness of the object recognition system. This technique is used for various real-time applications where guaranteed performance is demanded, this performance is degraded due to image scale variation, different viewpoint, deformations and imaging conditions of image acquisition. Another issue is caused due to huge amount of data in the every object class which causes conflict resulting in performance degradation.

In this context of object detection, various techniques have been proposed recently. These techniques depend on the different techniques of object detection i.e. frame difference based method [2], optical flow computation [3], point detectors [4], temporal difference method [5], background subtraction techniques [6] etc. In figure 1, a general diagram of object recognition system is presented where image is acquired from the image source, this acquired image is denoted as input image, this image is passed through image pre-processing stage where image enhancement , morphological operations are applied. Next stage performs image segmentation and finally object recognition is performed.

Frame differencing method is applied on video sequence to detect the objects. This method computes difference of two consecutive frames or images and detects object. This method shows good performance even if computation environment is dynamic in nature. This method suffers from the issue of complete outline of the moving object which results in performance degradation of object detection. Another method is known as Optical flow based computation method which depends on the computation of optical flow and image clustering. This method is capable to provide information regarding movement of the each object present in image. Compare to background subtraction based method, this method provides 85% better accuracy for object detection. Key challenge in this technique includes various aspects such as computational complexity and noise-sensitivity which makes it more complex to implement for real time application scenarios.

As discussed before, Point detector is also used for object detection in computer vision system based applications. This is a technique which extracts the keypoints from the image considering texture and location information of image. This approach provides better results where illumination and camera viewpoints are varying frequently. In this field, Harris detector [7], KLT detector method [8] and SIFT detector [9].

Next, we discuss about temporal difference method for object detection. Working process of this method is similar to the frame differencing method. According to this method, pixel-wise difference of two-or three frames is computed resulting in the extraction of moving region from input image sequence. If moving object consist of uniform texture pattern then this approach results in better adaptability considering dynamic scene change for object detection. According to this approach, if foreground is not moving relatively to the background then objects cannot be detected by applying this technique. Finally, we describe background subtraction based method for object detection. This method develops a model to represent the scene and estimates the deviation of each frame of input video sequence. According to this method, if any significant change is observed in the background then is marked as object and the pixels which are present in the change marked region are marked as object and remaining pixels are assumed

for further processing of object detection. This process is known as background subtraction approach for object detection. Based on this various benchmark studies such as Hidden markov models [10] and region based methods [11] etc.

This field of object detection has grown rapidly in recent years due to its huge demand in real time applications. Existing techniques of object detection are based on the manual designed feature, due to development in technology, now we consider automated feature learning techniques. Dalal et al. [12] discussed about the object recognition technique based on the feature extraction method using linear SVM technique. This work reviews existing model edge and gradient based feature set and shows that the proposed technique outperforms by considering Histograms of Oriented Gradient (HOG) technique. This technique shows that fine orientation, spatial inning, contrast normalization are the important components which affect the object detection performance. These type of techniques are based on the manual feature designing. In order to improve the performance furthermore, Krizhevsky et al. [13] developed a new approach for object recognition using deep convolution neural network, named as ConvNet. To obtain higher and faster rate of object detection, in [14] authors have presented a new technique. According to this technique, initially image integral is obtained which helps to improve the computation time. In next stage, AdaBoost learning technique is applied. This technique collects visual feature set and helps to develop an efficient classifier and finally classifier are combined together with the help of cascading technique resulting in the computation time reduction.

In figure 2, feature extraction based object recognition process is presented. According to this figure, input image is considered during first step, in the second step features are extracted and stored in a matrix form later residue matrix is obtained by applying matrix reconstruction and finally feature matching is applied to detect the object.

Object detection and tracking remains an open research problem even after research of several years in this field. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how one defines the object to be detected and tracked. If only a few visual features (e.g. color) are used as representation of an object, it is not so difficult to identify the all pixels with same color as the object. However, there is always a possibility of existence of another object or background with the same color information. Moreover, the change of illumination in the scene does not guarantee that the color will be same for the same object in all the frames. This leads to inaccurate segmentation based on only visual features (e.g. color). This type of variability changes is quite obvious as video objects generally are moving objects. The images of an object may change drastically as it moves from one frame to another through the field of view of a camera. This variability comes from three principle sources namely variation in target pose or deformations, variation in illumination and partial/full occlusion of the target [15].
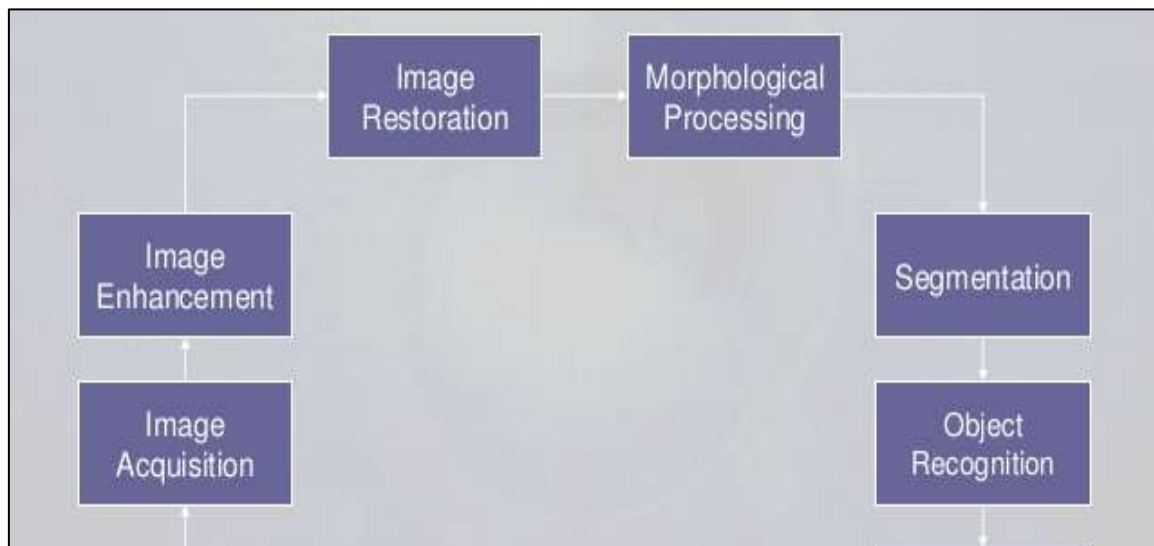
Figure 1.Fig.1. General diagram of Object recognition

The typical challenges of background subtraction in the context of video surveillance have been listed below:

A. Illumination Changes

It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day. Sudden illumination changes can also occur in the scene. This type of change occurs for example with sudden switching on/off a light in an indoor environment. This may also happen in outdoor scenes (fast transition from cloudy to bright sunlight). Illumination strongly affects the appearance of background, and cause false positive detections. The background model should take this into consideration.

B. Dynamic Background

Some parts of the scenery may contain movement (a fountain, movements of clouds, swaying of tree branches, wave of water etc.), but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees). Handling such background dynamics is a challenging task.

C. Occlusion

Occlusion (partial/full) may affect the process of computing the background frame. However, in real life situations, occlusion can occur anytime a subject passes behind an object with respect to a camera.

D. Clutter

Presence of background clutter makes the task of segmentation difficult. It is hard to model a background that reliably produces the clutter background and separates the moving foreground objects from that.

E. Camouflage

Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications. Camouflage is particularly a problem for temporal differencing methods.

F. Presence of Shadows

Shadows cast by foreground objects often complicate further processing steps subsequent to background subtraction. Overlapping shadows of foreground regions for example hinder their separation and classification. Researchers have proposed different methods for detection of shadows.

G. Motion of the Camera

Video may be captured by unstable (e.g. vibrating) cameras. The jitter magnitude varies from one video to another.

H. Bootstrapping

If initialization data which is free from foreground objects is not available, the background model has to be initialized using a bootstrapping strategy.

I. Video Noise

Video signal is generally superimposed with noise. Background subtraction approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artifacts.

J. Speed of the Moving Objects and Intermittent Object Motion

The speed of the moving object plays an important role in its detection. If the object is moving very slowly, the temporal differencing method will fail to detect the portions of the object preserving uniform region. On the other hand a very fast moving object leaves a trail of ghost region behind it in the detected foreground mask. Intermittent motions of objects cause 'ghosting' artifacts in the detected motion, i.e., objects move, then stop for a short while, after which they start moving again. There may be situations when a video includes still objects that suddenly start moving, e.g., a parked vehicle driving away, and also abandoned objects.

K. Challenging Weather

Detection of moving object becomes a very difficult job when videos are captured in challenging weather conditions (winter weather conditions, i.e., snow storm, snow on the ground, fog), air turbulence etc.
Following are the basic steps for tacking an object, as describe in many literature.

1) Object Detection
Object Detection is to identify objects of interest in the video sequence and to cluster pixels of these objects. Object detection can be done by various techniques such as frame differencing, Optical flow and Background subtraction.

2) Object Classification

Object can be classified as vehicles, birds, floating clouds, swaying tree and other moving objects. The approaches to classify the objects are Shape-based classification, Motion-based classification, Color based classification and texture based classification.

This paper is structured in the following way: Section 1 gives introduction to object detection. Section 2 provides the related work. Section 3 deals with brief explanation on several object detection and object classification methods and section 4 provides conclusions.

II. RELATED WORK

In this section we discuss various state of art techniques in the field of object recognition, object detection and classification. These techniques are based on feature extraction scheme, classifiers and based on the learning features.

Kalogeiton et al [16] reviewed various techniques for object detection and concluded that it is a challenging task in computer vision field. Furthermore, authors have presented that object detection task becomes more challenging in video sequences and video sequences are used widely in real-time applications. However, it is presented in this work that object detection performance depends upon the training process of the feature set and feature extraction technique. According to this article, variation in domain shift factors, spatial location performance and image qualities are the main components which are responsible for object detection performance in video or image datasets.

Choudhury et al. [17] developed a new approach for object detection by considering background subtraction as a benchmark technique for fixed camera view scenarios. However, there are various components present which affect the performance of object recognition or detection. These components are illumination variation, shadow casting, background movement and camera vibration etc. In this article authors presented a study regarding background subtraction technique to improve the accuracy of object detection model. In this work, various mathematical models are considered and categorized based on the working process.

In [18], a novel framework is developed for object detection and recognition. This method mitigates the performance issue of existing saliency detection method. In this approach, salient region extraction is performed by taking existing saliency detection technique along with distance weighting, binarization and morphological methods. Later, superpixel segmentation model is applied where Bayesian model is used for saliency refinement resulting accurate saliency mapping. An iterative optimization method is designed to obtain better saliency results by exploiting the characteristics of the output saliency map each time. Through the iterative optimization process, the rough saliency map is updated step by step with better and better performance until an optimal saliency map is obtained. Zhu et al. [19] proposed a new approach for object detection by extending the hierarchical model resulting in sequential frame updating considering dynamic scenes. This method utilizes dictionary learning based scheme developed using two-layer model. This model is developed using three-stage modelling for improving the sparse representation resulting in significant improvement in object detection and classification.
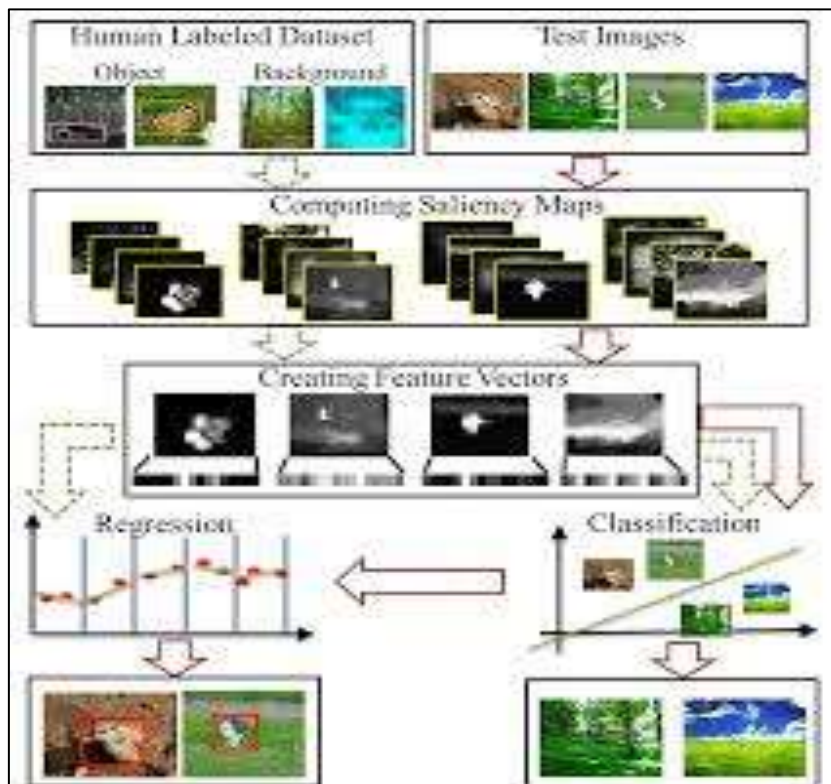
Figure 2. Saliency based object detection and classification

As we have discussed in previous section that feature extraction based techniques are also promising technique in the field of object recognition and detection. By considering this as promising technique, Naqvi et al. [20], developed feature extraction based model for object detection. This approach also utilizes saliency technique as a base model. In this work, authors have developed a hybrid feature modelling to improve the detection performance. Since, images are have diverse nature which shows that images have different context significant nature. In order to deal with diversity, here a feature detector model is developed by introducing feature quality measurement. This approach shows significant performance as tested with benchmark datasets. Similarly, another saliency based technique is developed in [20] which is named as DeepSaliency. According to this article, authors discussed that object detection performance depends on the semantic properties of salient objects. By taking this into account, authors developed a saliency model using convolutional network where inputs are given as raw images and output is obtained in the form of saliency map. According to this approach, data-driven encoding technique is used for information extraction and then multi-tasking based learning model is developed to obtain the correlation estimation between saliency map and segmented image. Finally, feature collaboration is applied which produces effective feature set to analyse visual perception of detected object.

As we have seen the significant improvement in object detection technique for still images or image frames but it remains a challenging task for various scenarios where background varies continuously, illumination change occurs, occlusions and complex backgrounds etc. however, various techniques have been proposed for object detection but due to performance issues there is a need to develop an efficient technique for moving object detection. Probability distribution pixel based techniques also have been presented where image samples are collected

using temporal and spatial domain. To improve the performance of this mode, Fan et al. [22] developed a confidence relationship model for moving object detection technique.

When dealing with salient object that contains several regions with different appearances, salient object detection can be a difficult task as often only parts of the salient object are highlighted and consistency between the salient regions is poor. This study tackles this problem by introducing objectness to assist the salient object detection. Rather than treating objectness in the same manner as other low-level cues (e.g. uniqueness, location etc.) for the determination of regional saliency values, the authors emphasise that objectness should also play a significant role in tuning the consistency between salient regions. The authors integrate objectness, uniqueness and centre bias to find potential salient regions and then enforce consistency between these regions using a full-connected Gaussian Markov random field with the weights determined by the objectness score [23].

Recently, considerable work have been carried out in the field of object detection and classification using learning based schemes. In this field of learning based object detection, Donahue et al. [24] presented a new approach and experimented on PASCAL VOC dataset. These methods are complex to implement and uses both low-level and high-level feature to produce higher-level feature vector. To address this issue, authors proposed an efficient approach for object detection to improve the performance in terms of mean average precision. This techniques provides two main contribution to this research as: (1) this scheme is capable to provide high-capacity neural network using bottom-up approach to localize and segment the objects. (2) This techniques provides boosting the performance by using domain-specific fine tuning. Finally, this method is combined with region proposals and proposed method is described as R-CNN. Another similar technique is proposed in [24], which is based on the region based convolution networks. The key contribution of this work is denoted by its computation speed, hence it is known as Fast R-CNN technique for object detection.
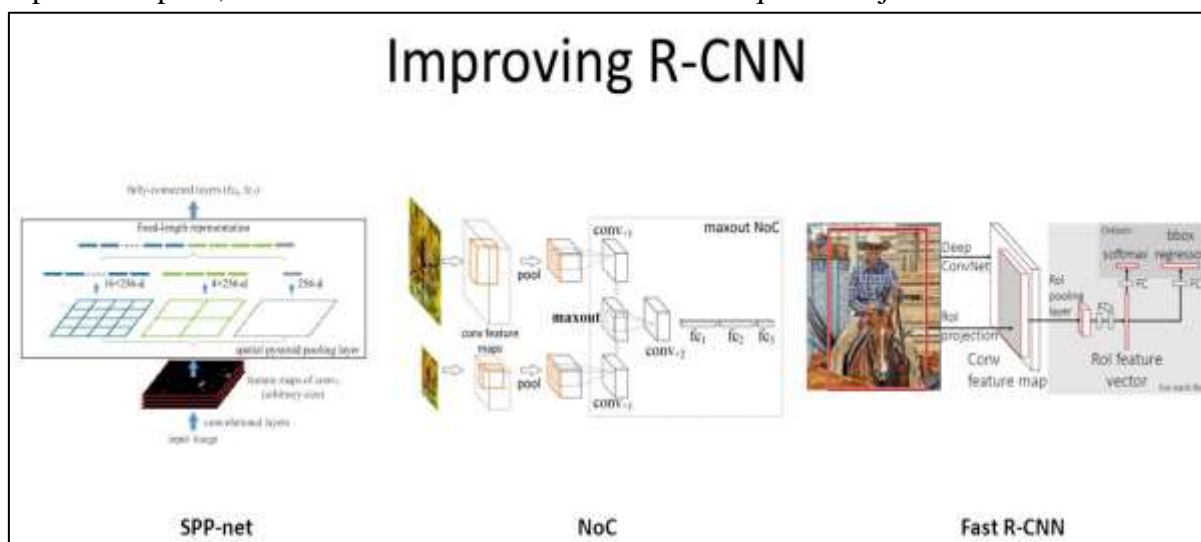

Figure 3. Faster RCNN flow chart

This technique provides better performance when compared to existing deep convolutional network technique. According to this technique, various techniques are used for improving the testing and training process, resulting in improvement in efficiency of the network. Fast R-CNN trains the very deep VGG16 network 9x faster than R-CNN, is 213x faster at test-time,

and achieves a higher mAP on PASCAL VOC 2012. Compared to SPPnet, Fast R-CNN trains VGG16 3x faster, tests 10x faster, and is more accurate.

III. OBJECT RECOGNITION AND DETECTION TECHNIQUES

All In this section, we present the study about state-of-art recognition detection techniques. These techniques are classified as follows:

- (a) Template based object recognition
- (b) Colour based object recognition
- (c) Shape based object recognition
- (d) Local and global feature based object recognition

(a) Template based object recognition

This is a process of finding the smaller parts of any input image to perform matching with the original template image. According to this process, templates are collected from different objects and stored in a pre-defined database. In order to recognize the image object, templates are matched by comparing the input image and test templates. Generally, this approach is used for character recognition, number recognition or vehicle number plate identification etc. by considering both form of image as gray scale and color image. Since it is a pixel-wise matching process so it requires feature extraction model.

Ogunbona et al. [26] presented a new approach for object recognition with the help of improved template matching process. This approach is a combination of spatial and orientation features resulting in performance improvement. In order to obtain the spatial information, generalized distance transform is applied. This transform is used for computing the weights of distance transform by considering orientation map and local gradient information. This technique is tested for benchmark datasets to detect the human, cars etc. In [27], authors have presented a new approach for similarity measurement between visual entities. Internal local self-similarities are known as correlation inside the input image. However, pattern generation for each image is not similar for any of the image. In order to evaluate the self-similarity, a new descriptor is introduced named as self-similarity descriptor. This descriptor provides detailed analysis image or video sequence at different scales. Hence, this model can be implemented for complex visual scenarios such as real time cluttered environment, rough hand-sketches.

(b) Colour based object recognition

This is another technique for object recognition which is widely used in real time application. According to this approach, colour characteristics are used for pattern matching for any given input image. Based on this technique Chunyang et al [28] developed traffic sign detection model to real time driving assistance application. This approach is implemented by considering complex environment. According to this approach, input RGB image is transformed to 8 colour image to reduce the computation overhead. Later, significant colour components are extracted represented as candidate region which is used for further recognition. For feature model, here HOG descriptors are considered for candidate region detection. Finally, Zernike moment is combined with support vector machine for traffic sign recognition.

In [29], authors developed a colour attribute based approach for object detection. This approach utilizes two existing methods as part-based object detection and subwindow search approach. In this approach, three main scenarios are considered such as feature combination, invariance in photometric nature, and image feature compactness. In this work, authors present a combined technique of colour and shape based technique. In this technique part-based detection technique and subwindow search technique is used for object localization.

(c) Shape based object recognition

In this section, we present recent studies in the field of object recognition based on the shape feature analysis. Shape analysis is also widely used in real world scenarios for object recognition. Shape feature are more promising when compared to local features such as SIFT because generally these categories can be well-defined by using shape feature instead of using texture feature. Shape feature are the replacement of local features. A. Berg, et.al. [30], have proposed another calculation to discover correspondences between highlight focuses for protest acknowledgment in the system of deformable shape coordinating. The fundamental subroutine fit as a fiddle coordinating takes as information a picture with an obscure question (shape) and thinks about it to a model by taking care of the correspondence issue between the model and the protest. At that point it performs adjusting change and figures a comparability in light of both the adjusting change and the remaining in the wake of applying the adjusting change. The Authors have considered different reasons like Intra-class variety, Occlusion and mess, 3D posture changes that makes correspondence issues more troublesome. Three sorts of requirements to tackle the correspondence issue between shapes are Corresponding focuses on the two shapes ought to have comparative nearby descriptors, Minimizing geometric bending, Smoothness of the change from one shape to the next.

In [31], shape based object detection technique is presented which utilizes edge based clustering and gradient vector girding (GVG) to improve the performance of object detection. In this approach, a pixel-level model is applied to determine the connectivity among pixels. Based on this a graph is built and the images are overlapped using equal size grading. For each cell, multiple graph nodes are computed and then connectivity of each cell is estimated by considering 8-neighbourhood pixels. Finally, the maximum curvature of the result paths is adjusted. Schindler et al. [32] introduced a new approach for object recognition and classification based on the shape features. In this method super-pixel segmentation is applied to obtain the closed edge chain of input image. Later, a probabilistic measurement is performed by considering two different contours of the input image. The main advantage of this work is that it requires only a single object temple to perform the object matching. The basic idea used is to define a distance measures between shapes and then try to find minimum distance. The Authors discussed different deformable template matching techniques listed chamfer matching and spline-based shape matching and compares to elastic matching method and finally proposed a search strategy which relieves the problem of false local minima by re-evaluating the shape distance at each search step.

(d) Local and global feature based object recognition

The most widely recognized way to deal with bland question location is to slide a window over the picture and to characterize each such neighborhood window as containing the objective or foundation. This approach has been effectively used to identify inflexible protests, for example, appearances and autos in [33]. In [34], a technique for protest acknowledgment and division utilizing Scale-Invariant Feature Transform (SIFT) and Graph Cuts is introduced. Filter highlight is invariant for pivots, scale changes, and light changes. By brushing SIFT and Graph Cuts, the presence of items is perceived first by vote handling of SIFT keypoints. At that point the question area is removed by Graph Cuts utilizing SIFT keypoints as seeds. Both acknowledgment and division are performed consequently under jumbled foundations including impediment. Authors in [35], show a strategy for question acknowledgment with full limit identification by joining relative scale invariant component change (ASIFT) and an area blending calculation. The calculation is invariant to six relative parameters in particular interpretation (2 parameters), zoom, turn and two camera hub introductions. The components give solid keypoints that can be utilized for coordinating between various pictures of a protest. They prepared a question in a few pictures with various angles for discovering best keypoints of it. At that point, a vigorous district blending calculation is utilized to perceive and recognize the question with full limit in alternate pictures in light of ASIFT keypoints and a similitude measure for consolidating locales in the picture.

In [36], Histogram of Gradients (HOG) based multistage approach for object detection and object pose recognition for service robots is used. It makes use of the merits of both multi-class and bi-class HOG-based detectors to form a three-stage algorithm at low computing cost. In the first stage, the multi-class classifier with coarse features is employed to estimate the orientation of a potential target object in the image; in the second stage, a bi-class detector corresponding to the detected orientation with intermediate level features is used to filter out most of false positives; and in the third stage, a bi-class detector corresponding to the detected orientation using fine features is used to achieve accurate detection with low rate of false positives. The training of multi-class and bi-class support vector machine (SVM) with their respective features in different levels is described. Antonio Monroy, Angela Eigenstetter and Bjorn Ommer [37], have presented an approach that directly uses curvature cues in a discriminative way to perform object recognition. Integrating curvature information substantially improves detection results over descriptors that solely rely upon histograms of orientated gradients (HoG). The joint descriptor is refered as HoGC. Because of the histogram-nature of the feature vectors, SVM with histogram intersection kernel is used as a classifier. An innate extension of these local approaches is to use sliding window to detect object parts, and then assemble the parts into a whole object. Problem with local features is that recognition may fail because of insufficient local information. This can be solved by using the context of the image as a whole i.e., global features.

As discussed in previous section that, recently for real time object detection and classification purpose, learning based approaches are used widely in this research field. In table 1, we present a comparative analysis of learning based object detection schemes.

Table 1. Comparative study of object detection techniques

Technique	Language	OS	CNNs	AE	RBMs	Learning Method	Loss
Deep Learn-Toolbox[38]	Matlab	Lin,Windows	+	(stAE)	+(DBN)	BP	MSE
Caffe [38]	C++, Python, Matlab	Lin, WIn	+	-	-	SGD	1. Euclidean 2. Softmax 3. Logistic 4. Info gain 5. Hinge 6. Sigmoid cross entropy
Torch [39]	Lua,C	Lin, Android,iOS	+	+	-	SGD, LBFGS,CG	1. MSE 2. Cross-entropy 3. Quadratic error
Darch [40]	R	Win,Lin	-	+	DBN	BP,CG	1. MSE 2. Cross-entropy 3. Quadratic error
PyLearn [41]	Python	Lin	+	CAE	Ss RBM	SCG,BGD	1. Cross entropy 2. Log-likelihood
Deepmat[42]	Matlab	Win, Lin	+	St DAE	DBN	GD, LBFGS, CD, PCD	1. Squared error 2. Cross entropy 3. Hinge

IV. CONCLUSION

In this article, we have discussed about object detection techniques which are proposed recently for real time applications. Here we have discussed template based technique, color based object detection, shape based technique local and global feature based technique for object detection. Template matching scheme require huge amount of database for better template creation and feature extraction. Color based technique are depend on the color transformation and feature extraction mode, It provides better results but if it is used for complex scenarios then it suffers from the performance issue and overall performance of the system degrades. Shape and global feature are known as manual feature extraction technique whereas in real-time scenario complex environments are considered for performance evaluation where these techniques are not providing efficient results. From this study we have concluded that for real time scenarios, learning based schemes are useful to perform the object detection. We have discussed various learning based studies also. These studies are providing better performance but suffer from the complexity issue. Hence in future work, complexity reduction of learning based technique can be carried out for object detection and classification.

REFERENCES

[1] Y. Tang; X. Wang; E. Dellandrea; L. Chen, "Weakly Supervised Learning of Deformable Part-Based Models for Object Detection via Region Proposals," in IEEE Transactions on Multimedia , vol.PP, no.99, pp.1-1, 03 October 2016

[2] H. Yin, Y. Chai, S. X. Yang and X. Yang, "Fast-moving target tracking based on mean shift and frame-difference methods," in Journal of Systems Engineering and Electronics, vol. 22, no. 4, pp. 587-592, Aug. 2011.

[3] A. S. Rao, J. Gubbi, S. Marusic and M. Palaniswami, "Crowd Event Detection on Optical Flow Manifolds," in IEEE Transactions on Cybernetics, vol. 46, no. 7, pp. 1524-1537, July 2016.

[4] M. Zandi, A. Mahmoudi-Aznaveh and A. Talebpour, "Iterative Copy-Move Forgery Detection Based on a New Interest Point Detector," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2499-2512, Nov. 2016.

[5] Z. Li et al., "Rotation Feature Extraction for Moving Targets Based on Temporal Differencing and Image Edge Detection," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 10, pp. 1512-1516, Oct. 2016.

[6] W. Kim and Y. Kim, "Background Subtraction Using Illumination-Invariant Structural Complexity," in IEEE Signal Processing Letters, vol. 23, no. 5, pp. 634-638, May 2016.

[7] J. b. Ryu and H. h. Park, "Log-log scaled Harris corner detector," in Electronics Letters, vol. 46, no. 24, pp. 1602-1604, November 25 2010.

[8] C. Zhang, J. Xu, A. Beaugendre and S. Goto, "A KLT-based approach for occlusion handling in human tracking," Picture Coding Symposium (PCS), 2012, Krakow, 2012, pp. 337-340.

[9] S. Luo, W. Mou, K. Althoefer and H. Liu, "Novel Tactile-SIFT Descriptor for Object Shape Recognition," in IEEE Sensors Journal, vol. 15, no. 9, pp. 5001-5009, Sept. 2015.

[10] J. Lei, G. Li, J. Zhang, Q. Guo and D. Tu, "Continuous action segmentation and recognition using hybrid convolutional neural network-hidden Markov model model," in IET Computer Vision, vol. 10, no. 6, pp. 537-544, 9 2016.

[11] M. Xu, L. Zhang and B. Du, "An Image-Based Endmember Bundle Extraction Algorithm Using Both Spatial and Spectral Information," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2607-2617, June 2015.

[12] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. (2005)

[13] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. (2012)

[14] Viola, P., Jones, M.: Robust real-time object detection. IJCV (2001)

[15] Gregory D. Hager and Peter N. Belhumeur; "Efficient Region Tracking With Parametric Models of Geometry and Illumination", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20 (10), pp. 1025–1039, 1998

[16] V. Kalogeiton, V. Ferrari and C. Schmid, "Analysing Domain Shift Factors between Videos and Images for Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 11, pp. 2327-2334, Nov. 1 2016.

[17] S. K. Choudhury; P. K. Sa; S. Bakshi; B. Majhi, "An Evaluation of Background Subtraction for Object Detection Vis-a-Vis Mitigating Challenging Scenarios," in IEEE Access , vol.PP, no.99, pp.1-1, 13 September 2016

[18] J. Lei et al., "A Universal Framework for Salient Object Detection," in IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1783-1795, Sept. 2016.

[19] S. Zhu, J. Du, N. Ren and M. Liang, "Hierarchical-Based Object Detection with Improved Locality Sparse Coding," in Chinese Journal of Electronics, vol. 25, no. 2, pp. 290-295, 3 2016.

[20] S. S. Naqvi, W. N. Browne and C. Hollitt, "Feature Quality-Based Dynamic Feature Selection for Improving Salient Object Detection," in IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4298-4313, Sept. 2016.

[21] X. Li et al., "DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection," in IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3919-3930, Aug. 2016.

[22] Z. Fan and H. Wang, "Robust moving object detection based on spatio-temporal confidence relationship," in Electronics Letters, vol. 52, no. 10, pp. 825-827, 5 12 2016.

[23] X. Sun, A. Su, S. Chen, Q. Yu and X. Liu, "Objectness to assist salient object detection," in IET Image Processing, vol. 10, no. 5, pp. 391-397, 5 2016.

[24] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR. (2014)

[25] Girshick, R.: Fast R-CNN. In: ICCV. (2015)

[26] D. T. Nguyen, W. Li, and P. Ogunbona, "An improved template matching method for object detection," in ACCV, 2009.

[27] E.Shectman, M.Irani, "Matching Local Self-Similarities across Images and Videos," In IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[28] M. Xing, M. Chunyang, W. Yan, W. Xiaolong and C. Xuetao, "Traffic sign detection and recognition using color standardization and Zernike moments," 2016 Chinese Control and Decision Conference (CCDC), Yinchuan, 2016, pp. 5195-5198.

[29] 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

[30] 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.

[31] H. Moballegh, N. Schmude, and R. Rojas, "Gradient Vector Griding: An Approach to Shape-based Object Detection in RoboCup Scenarios," from: www.ais.uni-bonn.de/robocup.de/papers/RS11_Moballegh.pdf

[32] K.Schindler, D. Suter, "Object Detection by Global Countour Shape," Pattern Recognition, 41(12), pp.3736–3748, 2008

[33] C. Papageorgiou and T. Poggio, "A trainable system for object detection," International. Journal of Computer Vision, 38(1), pp.15–33, 2000.

[34] A. Suga, K. Fukuda, T. Takiguchi, Y.Ariki, "Object Recognition and Segmentation Using SIFT and Graph Cuts," In 19th International Conference on Pattern Recognition ,pp. 1-4, 2008

[35] R. Oji, "An Automatic Algorithm for Object Recognition and Detection Based on ASIFT Keypoints," Signal & Image Processing: An International Journal (SIPIJ) Vol.3, No.5, pp.29-39, October 2012

[36] L. Dong, X. Yu ,L. Li, J. Kah Eng Hoe, "HOG based multi-stage object detection and pose recognition for service robot," Control Automation Robotics & Vision (ICARCV), 11th International Conference, pp. 2495 – 2500, Dec. 2010.

[37] A. Monroy, A. Eigenstetter, B. Ommer, "Beyond Straight Lines - Object Detection Using Curvature," In 18th IEEE International conference on Image Processing (ICIP), pp. 3561 - 3564 , 2011.

[38] Caffe – a deep learning framework. http://caffe.berkeleyvision.org

[39] Torch – a scientific computing framework. http://www.torch.ch

[40] Package Darch. http://cran.rproject.org/web/packages/darch/index.html

[41] Pylearn2 – a machine learning library. http://deeplearning.net/software/pylearn2

[42] DeepMat Library. https://github.com/kyunghyuncho/deepmat