A review and an approach for object detection in images

Article *in* International Journal of Computational Vision and Robotics · January 2017

DOI: 10.1504/IJCVR.2017.081234

2 authors:

Kartik Umesh Sharma
Prof Ram Meghe College of Engineering and Management

Nileshsingh V. Thakur
Nagpur Institute of Technology, Nagpur, India

A review and an approach for object detection in images

Kartik Umesh Sharma* and
Nileshsingh V. Thakur

Department of PG Studies (Computer Science and Engineering),
Prof. Ram Meghe College of Engineering and Management,
New Express Highway Road,
Badnera-Amravati, Maharashtra, PIN-444701, India
Email: karthik8777@gmail.com
Email: thakurnisvis@rediffmail.com
*Corresponding author

Abstract: An object detection system finds objects of the real world present either in a digital image or a video, where the object can belong to any class of objects namely humans, cars, etc. In order to detect an object in an image or a video the system needs to have a few components in order to complete the task of detecting an object, they are a model database, a feature detector, a hypothesiser and a hypothesiser verifier. This paper presents a review of the various techniques that are used to detect an object, localise an object, categorise an object, extract features, appearance information, and many more, in images and videos. The comments are drawn based on the studied literature and key issues are also identified relevant to the object detection. Information about the source codes and online datasets is provided to facilitate the new researcher in object detection area. An idea about the possible solution for the multi class object detection is also presented. This paper is suitable for the researchers who are the beginners in this domain.

Keywords: object detection; localisation; categorisation; object recognition.

Reference to this paper should be made as follows: Sharma, K.U. and Thakur, N.V. (2017) 'A review and an approach for object detection in images', *Int. J. Computational Vision and Robotics*, Vol. 7, Nos. 1/2, pp.196–237.

Biographical notes: Kartik Umesh Sharma received his Bachelor of Engineering degree in Information Technology from Sipna College of Engineering and Technology, Amravati, India in 2012. He received his Master of Engineering degree in Computer Science and Engineering from Prof. Ram Meghe College of Engineering and Management, Badnera, India under Sant Gadge Baba Amravati University, Amravati, India in 2014. His research interest includes image processing, video processing, language processing and algorithms. Currently, he is an Assistant Professor in Department of Computer Science and Engineering at Prof. Ram Meghe College of Engineering and Management, Badnera, India.

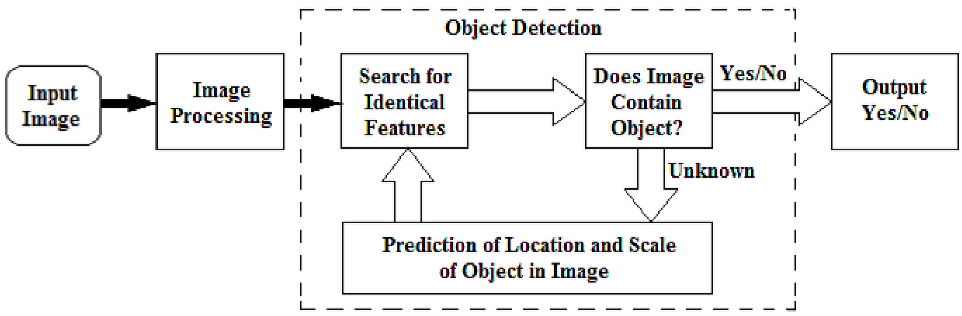
Nileshsingh V. Thakur received his BE (CSE) and ME (CSE) degrees from Government COE, Amravati and COE, Badnera under SGBAU in 1992 and 2005 respectively. He received his PhD in CSE from VNIT, Nagpur, India on 2010. His research interest includes image processing and advanced computing. He is having over 23 years of teaching and research experience. Presently, he is a Professor and Dean at PRMCEAM, Badnera, India. He is the author or co-author of more than 60 scientific publications in international journals and conferences. He is an editorial board member of eight international journals and worked as reviewer for international journals and conferences.

1 Introduction

Object detection (OD) system finds objects in the real world by making use of the object models which is known a priori. This task is comparatively difficult to perform for the machines as compared to Humans who perform OD very effortlessly and instantaneously. In this paper we will give a review of the various techniques and approaches that are used to detect objects in images and videos.

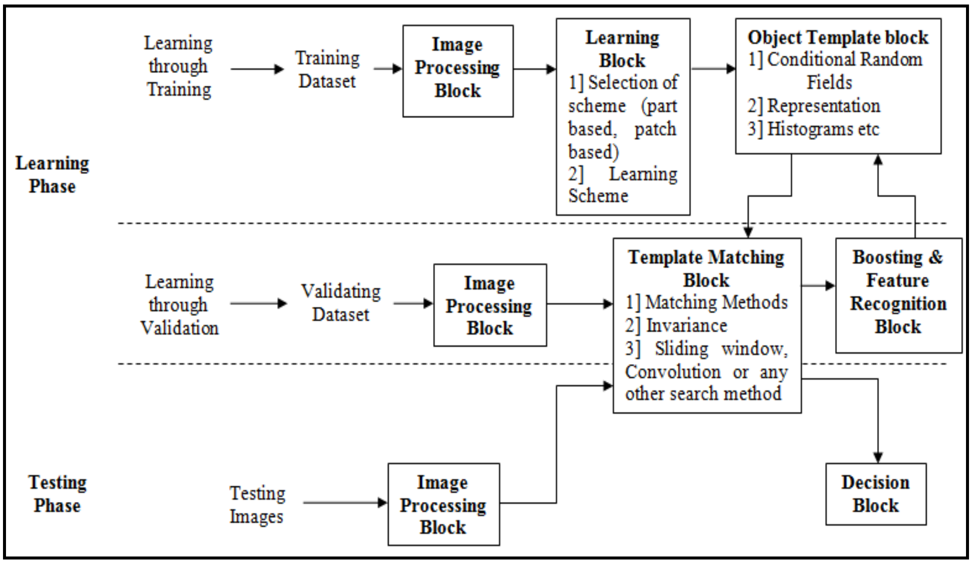
Basically an OD system can be described easily by seeing Figure 1 which shows the basic stages that are involved in the process of OD. The basic input to the OD system can be an image or a scene in case of videos. The basic aim of this system is to detect objects that are present in the image or scene or simply in other words the system needs to categorise the various objects into respective object classes.

Figure 1 Basic OD model



The OD problem can be defined as a labelling problem based on models of known objects. Given an image containing one or more objects of interest and a set of labels corresponding to a set of models known to the system, the system is expected to assign correct labels to regions in the image. The OD problem cannot be solved until the image is segmented and without at least a partial detection, segmentation process cannot be applied. The term detection has been used to refer to many different visual abilities including identification, categorisation and discrimination.

Figure 2 Working of the OD system



2 General methodology

The OD system basically comprises of two main phases namely: the learning phase and the testing phase which are shown in Figure 2 that shows the normal working of the OD system. Learning phase is mainly meant for the classifier so that it recognises the objects present in the image that is given as input to the system. Learning phase can be further classified as learning through training and learning through validation. Learning through training comprises mainly of the learning block where a proper learning scheme is defined, it can be part-based or patch-based, etc. The object template block then makes use of the learning's that were done previously to represent the objects with various representations like histogram representation, random forest representation, etc. Whereas on the other hand, learning through validation block does not require any sort of training as they are validated beforehand. Hence after preprocessing the image, directly template matching is done which produces the features of an object in the image. The main purpose of the testing phase is to decide whether an object is present in the image that is given to the system as input and if yes then to which object class does it belongs to. Here the image is searched for an object by various searching techniques like the sliding window technique, and according to the output of the searching mechanism, a decision is made on the object class.

3 Classification of OD mechanisms

This section classifies the various OD mechanisms based on search, feature classification, template creation and based on matching. We have classified the OD types as sliding window-based, contour-based, graph-based, fuzzy-based, context-based and some other types. Here we will review the work carried out by various authors in the field of OD.

3.1 Sliding window-based OD

Sliding window OD has received remarkable attention as it is considered as a very basic method of detecting objects in an image or video. The sliding window technique basically works by searching through the whole image or scene in order to find the object that is of interest. This is the reason why it failed to meet the criteria of real time applications due to higher execution times and inaccurate localisation. Localisation accuracy is important especially while the OD process is to be followed by object recognition.

Bergboer et al. (2007) have studied the various learning methods which are use for realising context-based OD in paintings, namely the gradient method and the context detection method. The gradient method is used to transform a spatial context into a gradient towards an object, whereas the context detection method makes use of the sliding window approach to search the image regions that are likely to contain the object of interest. Basically the gradient method works totally on assumptions which may lead to higher timing constraints when there is only single object that is to be detected in an image. On the other hand, the context detection works based on sliding window which again introduces the timing constraint as it searches each window for the presence of an object. Clearly the issue of inaccurate localisation has been a concern while using the sliding window-based OD technique; Segvic et al. (2011) have explained how localisation accuracy could be achieved by removing the need for spatial clustering of the nearby detection responses. This leads to three main goals namely high recall, high precision and accurate localisation. Spatial clustering could be used to suppress the number of false positives but at the price of localisation uncertainty.

Sliding window technique initially fixes the size of the window in which it will be searching for the object, but in order to increase the rate at which the detections must happen, Comaschi et al. (2013) have proposed a sliding window approach that decides on the step size of the window at run time, which helps to apply this technique of sliding window to real time applications. They have also demonstrated that how this technique improves the performance of Viola Jones OD, and also claimed to have achieved a speedup of 2.03x in frames per second without compromising the accuracy factor. The main issue being the space utilised. Divvala (2012) has studied the two factors that influence the performance of sliding window technique for OD, namely context and subcategories. The use of the first factor that is context shows how the performance of the sliding window approach could be improved. As the siding window approach searches the total image for the presence of an object in it, while the use of context can be made in order to know whether a particular object is present in that region or not. The subcategories factor is where the information within the sliding window is used to split the training data into smaller groups which have reduced appearance diversity which leads to simpler classification. Here he has discussed only about the two factors, there could be many other factors too such as contour which could affect the performance to a great deal. Subburaman et al. (2010) have presented a technique which is used to reduce the number of miss detections while increasing the grid spacing while the sliding window approach is used for OD. They have achieved it by using a patch to predict the bounding box of an object within the search area, and in order to improve the speed of estimating the bounding boxes the authors make use of the decision tree with simple binary test at each node. Although they claim that their proposed system works on a wide variety of images, still we feel that an occluded image could remain a challenge.

Gualdi et al. (2011) have presented statistical-based search for sliding window technique which makes use of Monte Carlo sampling for estimating the likelihood density functions with Gaussian kernels, which is a multi stage strategy where the proposal distribution is progressively refined by taking into account the feedback of the classifier. Their proposed method exploits the presence of a basin of attraction around true positives to drive an efficient exploration of the state space, using a multi-stage sampling-based strategy. We feel that the concept of multi-stage particle windows can be extended for the training phase as well. An object is salient if it differs from its surroundings or if the object contains rare or outstanding details. Yanulevskaya et al. (2013) have proposed an approach to detect salient object. Here proto-objects are considered as the units of analysis, where a proto-object is a connected image region which can be converted to an object or an object part. The process proceeds by segmenting a complex image into proto-objects and then finding the saliency of each proto-object, the most salient proto-object is considered as being the salient object. Sudowe and Leibe (2011) have investigated how geometric constraints can be used for efficient sliding window OD; here they derived a general algorithm for incorporating ground plane constraints directly into the detector computation. They claimed that their approach allows to effortlessly combining multiple different detectors which automatically compute the region of interest (ROI). As we have seen earlier in Lampert et al. (2008) that object localisation is an important issue that needs to be seen into, but most OD systems rely on binary classification for locating the objects but not giving out the information about the location of the object. Lampert et al. (2008) have tried to achieve the same goal as that of the sliding window approach that is object localisation and retrieving the localised object by proposing a simple yet powerful branch and bound scheme. Table 1 gives an overview of the various approaches for detecting objects using the sliding window approach. The parameters like, the dataset used, the concept that has been made use of, the performance evaluation parameter, issues addressed, author's remarks and our findings are summarised in the Tables 1(a) and 1(b).

Table 1(a) Summary of sliding window-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Bergboer et al.	Self-created	Gradient and context methods	Accuracy and speed	Time complexity	Fast object detection is achieved.	Accuracy is achieved at the cost of speed.
Segvic et al.	Self-created	Binary classification	Accuracy	Localisation	Avoids spatial clustering.	Localisation uncertainty exits.
Comaschi et al.	CMU/MIT face dataset	Adaptive sliding window	Speed	Space utilisation	High speed when compared to OpenCV.	High space utilised.
Divvala	PASCAL VOC dataset	Context modelling	Accuracy	Performance gains and computational tractability	Subcategories used for computational tractability.	Factors like contour, shape, etc., not considered.

Table 1(b) Summary of sliding window-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Subburaman et al.	CMU/MIT face dataset	Use of patch for prediction	Speed	Reduction of miss detections	This method works on a variety of images.	Occluded images could be an issue.
Gualdi et al.	INRIA dataset	Multi stage particle window	Detection rate	Time complexity	This method increases the localisation accuracy also.	MS-PW could be extended to training phase.
Yanulevskaya et al.	Standard dataset	Salient object detection	Accuracy	Saliency of an object	Combination of two types of saliency improves performance.	Highly time consuming.
Sudowe et al.	INRIA dataset	Geometric constraints	Speed	Computation of region of interest	Their approach allows combination of various detectors.	Still detection rates are slow as whole image is searched.
Lampert et al.	UIUC/ PASCAL VOC 2006 datasets	Sub-window search	Robustness and speed	Object localisation	Global optimality is retained.	Different other shapes like circles eclipses not considered.

3.2 Contour-based OD

Stiene et al. (2006) have proposed an OD approach which is based on range images, as range images are well suited to contour extraction. They have used a 3D laser scanner and reliable contour extraction with floor interpretation for the process of OD. Although this process has high performance while working on range images, converting the natural images into range images is the overhead process that needs to be performed every time an object is to be detected. Contour-based OD can be well formulated as a matching problem between model contour parts and image edge fragments, and hence Yang et al. (2012) have used this problem and have treated it as a problem of finding dominant sets in weighted graphs, where the nodes of the graph are pairs composed of contour parts and edge fragments and the weights between nodes are based on shape similarity. The main advantage of this system is that it can detect multiple objects present in an image in one pass. Still the question arises that can this system detect objects in an occluded image or other types of images. Basically the objects in an image can be characterised on the basis of their appearance and by the shape of their contours; Schlecht and Ommer (2005) have investigated a local representation of contours for OD that complements appearance-based information. They have combined contour and appearance information into a general voting-based detection algorithm and have claimed that the combination has significantly improved the performance compared to the other voting methods. The combination of contour representation and appearance descriptor increases an additional step in the OD process which increases the time to detect and localise an object. In order to justify that the selected contour is exactly what the image was being searched for, Zhu et al. (2008) have introduced a shape detection framework and have named it as contour context selection. As the shape-based detection is invariant to changes of object appearance, their approach makes use of salient contours as tokens for matching the shape. They have termed the task of matching the contours as a set-to-set contour matching problem and have claimed that their approach takes linear time to compare the contours. Although they have claimed that their approach is able to detect objects in cluttered images, they have not spoken about an occluded image which would give a hard time to their approach to detect objects.

Arbeláez (2006) has presented an approach to boundary extraction which relies on the problem formulation which occurs in the framework of hierarchical classification that allows for region-based segmentation and edge detection as a single task. He has defined generic ultra metric distances by integrating local contour cues along the region boundaries and combining this information with region attributes. Although this approach extracts boundaries of objects but still the problem of localisation remains an issue which the authors have not addressed. Segmentation is the basic process that is involved while detecting an object in an image; Amine and Farida (2012) have proposed an approach which makes use of a deformable model ‘Snake’ which they have termed as an active contour for segmenting the range images. The process is again restricted to range images; the question still lies about the various types of images. Ferrari et al. (2008) have presented a family of scale invariant local shape features formed by chains of k connected, roughly straight contour segments (kAS), and their use for OD. Here they have demonstrated kAS within a sliding window object detector, where windows are subdivided into tiles, each described by a bag of kAS. The authors have limited the use of kAS to a simple detection framework. Authors have not specified the range of k, whether k could be varied greatly or not.

Table 2(a) Summary of contour-based OD work

Authors	Dataset used	Concept used	Performance evaluation parameter	Issues addressed	Authors remark	Our findings
Stiene et al.	Various datasets used	Insensitivity of range data	Speed (reduction in number of false negatives)	Feature extraction	Performance comparable to MPEG-7 standard.	Converting natural image to range data may consume time.
Yang et al.	ETHZ shape dataset	Dominant set computation	Scale and average detection rates	Mapping for features extraction	Multiple objects at multiple scales can be detected in one pass.	Occluded or cluttered images may affect performance.
Schlecht and Ommer	ETHZ shape dataset	Appearance and shape of an object	Speed (false negatives)	Characterisation of an image	This approach can also find most relevant contours and junctions in object hypothesis.	Combination of contour representation and appearance descriptor increases the time to detect.
Zhu et al.	ETHZ and INRIA datasets	Shape of an object	Precision	Set to set matching problem.	Objects in cluttered images can be detected easily.	Authors have not spoken about how their approach works on occluded images.
Arbeláez	BSDB dataset	Ultra metric distances	Area and quadratic error	Boundary extraction	Higher level approaches can benefit from this low-level representation.	Localisation problem may exist.
Amine and Farida	SASB dataset	Homogeneity	Accuracy and speed	Image segmentation	Multi agent systems are used for additional iterations.	The process is restricted to range images.
Ferrari et al.	INRIA dataset	Grouping of adjacent contours	Feature complexity	Straight contour segments	Contour segments can be reused.	Authors have not spoken on the range of *k*.

Table 2(b) Summary of contour-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Shotton et al.	Several challenging datasets used	Contour features	Orientation specificity	Multi scale categorical objects recognition	Have build class-specific codebook of uncluttered contour fragments.	Number of fragment of contours is not specified.
Ravishankar et al.	ETHZ dataset	Deformation happens at high curvature points	Detection rates (FPPI)	Deformable object detection	Local scale variations, bending deformations handled.	Variation in objects not addressed.
Schindler and Suter	ETHZ dataset	Global shape	Detection rates (FPPI)	Shape similarity	The method id invariant to shape and rotation.	Image deformation not considered.
Lu et al.	ETHZ and PASCAL dataset	Particle filters	Recall and precision	Contour grouping	Global shape explicitly employed.	Edge images have not been addressed.
Maire et al.	BSDS dataset	Local and global cues	Recall and precision	Localising junctions	Idealised line drawings are produced.	Contours' passing through an image is a query.

The system proposed by Shotton et al. (2008) not only recognises objects based on local contour features but also is capable of localising the object in space and scale in the image. Fragments of contours could be a good idea to guess the object but here lies a question that how many fragments could be feasible.

Ravishankar et al. (2008) have proposed an efficient multi stage approach to object recognition in real and cluttered images that is robust to scale and rotation. The system makes use of the shape information to localise objects and detect their contours. They claimed to extract object contours in the matching stage using dynamic programming and also k-segment grouping together with centroid of the object can localise it. They have not spoken about the variations in the objects or occlusion present in the images. Schindler and Suter (2008) have presented a method for OD based on global shape. Although this method applies well for natural images, its performance under occluded images will be a certain issue that has to be considered and also if the image is deformed. Lu et al. (2009) have defined the process of object detection and recognition as a contour fragment grouping and labelling problem. The system works by first performing selection of relevant contour fragments in edge images then grouping of the selected contour fragments and finally their matching with the model contours; but the conclusion of all these steps is reached by making use of particle filters (PF) with static observations, the main advantage of using PF is the fact that global shape similarity can be explicitly employed. The authors have only spoken about the edge images which can be the limitation of this system. Maire et al. (2008) have presented a framework for contour and junction detection. Basically the authors develop a contour detector using the combination of local and global cues. Detecting junctions is a problem as the image intensity surface is a bit confusing in the neighbourhood of a junction, and hence the right approach to junction detection should take the advantage of the contours that are incident on the junctions. Although the junctions are detected by making use of the contours passing through them, but it is not clear that all the junctions in an image have contours passing through it. Tables 2(a) and 2(b) summarise the approaches for OD based on contours.

3.3 Graph-based OD

Model-based methods play a central role to solve different problems in computer vision. A particular important class of such methods relies on graph models where an object is decomposed into a number of parts, each one being represented by a graph vertex. He et al. (2004) have presented a skeleton-based graph matching method for object recognition and object localisation which makes use of the skeleton model and contour segment model for this purpose. The use of these models helps reduce the matching space comparatively. This method has worked with satisfaction in case of biomedical images but still there remains a scope to implement this method on various other types of images and see if we still get satisfactory results. Felzenszwalb and Huttenlocher (2004) have addressed the problem of segmenting an image into regions; this is achieved by defining a predicate in order to measure an evidence for a boundary between two regions by making use of a graph-based representation of the image and by developing an efficient segmentation algorithm based on the predicate defined earlier. However finding a segmentation that is neither too coarse nor too fine is an NP-hard problem, hence there remains a huge scope in redesigning this method of image segmentation and to get good results. Dasigi and Jawahar (2008) have discussed a representation scheme for efficiently modelling parts-based representation and matching them, as graphs can be used for effective representation of images for detection and retrieval of objects, the problem of finding a proper structure which can efficiently describe an image and can be matched in low computational expense remains a problem. They in their discussion have compared two graphical representations namely the nearest-neighbour graphs and the collocation tress, for the goodness of fit and the computational expense involved in matching. A graph model-based tracking algorithm which generate a model for a given frame termed as reference frame was used to track a target object in the subsequent frames. Paixão et al. (2008) have proposed a different method and have claimed to have improved the recent algorithm in many ways, mainly instead of updating the model; each analysed frame is back-mapped to the model space thus providing robustness to the method as model parameters need not be modified each time. The variation in staining, fixation and sectioning procedures during gland segmentation may lead to considerable amount of artefacts and variances in tissue sections, which may result in variances in gland appearance. Gunduz-Demir et al. (2010) have presented a new approach to gland segmentation which decomposes the tissue image into a set of primitive objects and segments glands making use of the organisational properties of these objects, which are quantified with the definition of object-graphs.

Cyr and Kimia (2004) have presented a method to generate an aspect-graph representation of complex shapes using the dissimilarity between neighbouring views to generate aspects and to select prototypes for complex shape, and then the set of prototypical views obtained from each 3D object is cast in hierarchy. In order to detect an object, the unknown view is compared against the set of prototypes hierarchically. At the coarse levels those prototypes whose distance to query is larger than the radius associated with each prototype are pruned. Lin et al. (2009) have presented a model for representing compositional object categories as an attribute grammar which are embedded in an and-or graph for each compositional object category. The model combines the power of a stochastic context free grammar (SCFG) to express the variability of part configurations, and a Markov random field (MRF) to represent the pictorial spatial relationships between these parts. They have also proposed a recursive inference algorithm which is used to quickly constrain bottom-up detection while testing top-down constraints. Process is time consuming. Yu et al. (2002) have proposed a mechanism based on spectral graph partitioning that readily combine the process of segmentation and recognition into one. Initially the part-based recognition system detects object patches. Patch grouping is used to find set of patches that conform best to the object configuration. This process is integrated with the pixel grouping based on low-level feature similarity, through pixel-patch interactions and patch competition that is encoded as constraints in the solution space. The globally optimal partition is obtained by solving a constrained eigen value problem. Combining the two groupings may lead to extra overhead.

Hori et al. (2012) have made use of graph structural expression in order to develop a method for generic object recognition by embedding a graph into the vector spaces. In order to overcome the drawbacks of the previous methods where the location information of the objects and the relationship between the key points is lost, they have proposed this approach where the graph is constructed by connecting scale-invariant feature transform (SIFT) key points with lines, as a result of this key points maintain their relationship and then the structural representation with location information is achieved. They have not specified on what kind of images this method is feasible. Till now we have seen how are objects detected and localised, but Vajda et al. (2009) have studied the problem of object duplicate detection and localisation. They have proposed a graph-based approach for 3D object duplicate detection which represents the spatial information of the object in order to avoid making an explicit 3D object model. Finite domain constraint satisfaction problem (FDCSP) assumes that the matching problem between regions and labels is bijective. In image interpretation the matching problem is often non-univocal. The non-univocal matching between data and a conceptual graph was not possible until the introduction of arc consistency with bi-level ($FDCSP_{BC}$) constraint is introduced by Deruyver et al. (2009). Lebrun et al. (2011) have addressed the problem of graph matching during the process of OD and localisation using the kernel functions. Basically they have shown in their paper that the similarity between two graphs can be efficiently and effectively computed through a set of walks within the graphs. Firstly they propose kernels on graphs and kernels on walks, and then they propose the solutions for exact and appropriate computations of the kernels. Authors have not specified the range of the walk in a particular graph.

Zhang and Chang (2005) have presented a report in which they present a model named random attributed relational graph (RARG) using which the authors show how part matching and model learning could be achieved by combining variation learning methods with part-based representations. They have tried to solve the part matching problem through the formulation of association graphs that characterises the correspondences between parts in an image and nodes in the object model. For this system to be accurate there is a need for labelling the regions which can be an overhead. Triesch and Eckes (2005) had reviewed object representations based on deformable feature graphs which describe particular views of an object as a spatial constellation of image features; these representations are useful in situations of high clutter and partial occlusions. These representations have a number of advantages namely: it allows recognition without prior segmentation of the object of interest, robustness to small variations in appearance tends to be very good if features are chosen properly and many more. Partial occlusion is dealt with but what if the occlusion is full. Nam and Bao (2012) have presented a method to distinguish the principal objects in an image by making use of graph-based segmentation and normalised histograms (PODSH). Their approach basically focuses on such objects where one might focus while taking images, this approach basically supposes that the position of the main object is located at the centre of an image and the main object holds a large area. Normalised histograms are used to gain the edge and corner information. They claimed that their system has shortcoming due to lack of information of edge and corner information. Liang et al. (2012) have proposed an approach to detect salient objects by making use of adaptive multi-scale colour image neighbourhood hyper-graph representation and spectral hyper graph partitioning methods. Initially their approach extracts a polygonal potential region-of-interest by analysing the edge distribution in an image, then the image is represented by context sensitive hyper graph and then finally an incremental hyper graph partitioning is used to generate the candidate regions for final salient OD. Natural images are only considered, there remains the question whether this approach is feasible on other types of images.

Siddiqi et al. (1999) have applied the theory for the generic representation of 2D shapes, where structural descriptions are derived from shocks (singularities) of a curve evolution process, to the problem of shape matching. The singularities are organised into a directed, acyclic shock graph and the space of all such graphs is highly structured and can be characterised by the rules of a shock graph grammar. They in their paper introduced tree matching algorithm which finds the best set of corresponding nodes between two shock trees in polynomial time. This system has less feasibility when an image is occluded. Serratosa et al. (2003) have presented an article on model function-described graph (FDG), which is a type of compact representation of a set of attributed graphs (AGs) that borrow from random graphs the capability of probabilistic modelling of structural and attribute information. They have defined FDGs, their features and two distance measures between AGs (unclassified patterns) and FDGs (models or classes). Two applications of FDGs are also presented: in the first, FDGs are used for modelling and matching 3D-objects described by multiple views, whereas in the second, they are used for representing and recognising human faces, by several views. Shams et al. (2001) have developed an algorithm which is an extension to the labelled graph matching (LGM) algorithm named LGM1 and have compared the performance of their algorithm by using the state of the art statistical method which is based on mutual information maximisation (MIM). The LGM1 algorithm replaced the pixel values with a Gabor wavelet representation which made it perform superior to the successful version of LGM. Tables 3(a) through 3(c) summarises the approaches for OD based on graphs.

3.4 Fuzzy-based OD

Reyes and Dadios (2004) have developed a logit-logistic fuzzy colour constancy (LLFCC) algorithm for dynamic colour object recognition. This approach focuses on manipulating a colour locus which depicts the colours of an object. A set of adaptive contrast manipulation operators is introduced and utilised in conjunction with a fuzzy inference system and a new perspective in extracting colour descriptors of an object are presented. Again the question here arises about what colour ranges can be detected feasibly. Munoz-Salinas et al. (2004) make use of the information provided by the camera of a robot in order to assign a belief degree on the existence of a door in it; this is done by analysing the segments of the image. Several fuzzy concepts are defined to lead the search process and find different cases in which doors can be seen. Features of the segments like size, direction or the distance between them are measured and analysed using fuzzy logic in order to establish a membership degree of the segments on the defined fuzzy concepts. This work is purely restricted to indoor environments. Bernardin et al. (2007) have presented an automatic system for the monitoring of the indoor environments using the pan-tilt-zoomable cameras. The system makes use of Haar like feature classifier and colour histogram filtering in order to achieve reliable initialisation of person tracks. The system uses a combination of adaptive colour and KLT feature trackers for face and upper body which allows for robust tracking and track recovery in the presence of occlusion. What level of darkness can the system deal with remains a question?

Table 3(a) Summary of graph-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
He et al.	MRI corpus callosum image dataset	Graph matching	Error in matching	Recovery of objects	Works with satisfaction in case of biomedical images.	Variety of images not considered for testing this approach.
Pedro et al.	COIL dataset	Pair wise region comparison	Difference in intensity	Constructing a graph using predicates	Substantial improvement observed due to this method.	This issue is an NP hard problem and hence an improvement still exists.
Dasigi et al.	UKBench dataset	Scheme for modelling	Accuracy	Image matching	Computational time for this process is less compared to others.	Various other types of images not considered.
Paixão et al.	Sequences from CAVIAR project	Back mapping	Time taken to process a video	Tracking an object	Model parameters are not to be modified each time.	An occluded image may cause a concern to this approach.
Demir et al.	Colon biopsy samples	Decomposition of images into primitive objects	Accuracy (FPPI)	Segmentation of colon glands	More tolerance to artefacts and variances in tissues.	Classification of glands not considered.
Cyr et al.	Self-constructed dataset	Similarity bed aspect graph	Sample rate	Representation of complex shapes	Correct object prototypes are always picked.	A 3D cluttered image not considered.
Lin et al.	Caltech dataset	Graph grammar	Accuracy of detection	Compositional object representation	Computational efficiency is relatively slower than other approaches.	Process may be too time consuming.
Yu et al.	Self-constructed dataset	Graph partitioning	Speed	Recognition and segmentation as a single process	Eliminates local false positives.	Combining the two groupings may lead to extra overhead.

Table 3(b) Summary of graph-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Vajda et al.	PASCAL VOC dataset	Graph model	Computational complexity	Object duplicate detection	Robust when using only one or few images for training.	The approach is restricted to still images.
Deruyver et al.	NMR images	Image interpretation	Time	Detection of unexpected objects	Inadequacy may prevail when unexpected objects appear.	Works on assumptions which may not be true always.
Lebrun et al.	Multiple datasets used	Inexact graph matching	Similarity	Object retrieval	Due to weak learning there is no need for the user to indicate the ROI.	Authors have not specified the range of the walk in a graph.
Zhang et al.	Multiple datasets used	Random attributed relational graph	Speed	Part-based OD	Detection accuracy achieved by the single RARG model.	Need for labelling the regions can be an overhead.
Triesch	Multiple datasets used	Deformable feature graphs	Time	Representation of objects	Efficient recognition in the presence of clutter and occlusions.	Partial occlusion is dealt with but what if the occlusion is full.
Nam et al.	DOLSOFT dataset	Normalised histograms	Precision	Distinguishing principle objects	The normalised histogram is added to increase the effect of the system.	System may have shortcomings due to lack of information of edge and corner information.
Liang et al.	MSRA dataset	Context sensitive hyper graph	Precision and recall	Image classification	Precise object boundaries are obtained.	There remains the question whether this approach is feasible on other types of images.
Siddiqi et al.	Self-created database	Singularities of curves	Time	Representation of 2D objects	The grammar permits a reduction of a shock graph to a unique rooted shock tree.	This system may have less feasibility when an image is occluded.

Table 3(c) Summary of graph-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Serratosa et al.	COIL dataset	Function described graphs	Accuracy	Modelling of structural and attribute information	It is more effective and robust to partition a single 3D-object.	Clustered images can be an issue to this approach.
Shams et al.	Self-created database	Graph matching	Accuracy and speed	Pattern recognition	LGM1 algorithm is better than MIM.	Working on occluded images may not be feasible.
Hori et al.	10-class image dataset	Graph expression	Detection rate	Expressing an image as an appearance frequency histogram	Structural representation with location information is achieved.	Authors have not specified on what kind of images this method is feasible.

Malaviya and Malaviya (1993) have discussed various methods for fuzzy logic-based visual object recognition systems. Basically fuzzy logic facilitates the smooth translation of image information into natural language which can be easily processed by fuzzy set theory. They have also commented that in fuzzy type OD, there is a lack of semantic knowledge which raises the question that "What should be actually obtained from a given image?" Elbouz et al. (2011) have proposed and validated a surveillance video system that detects various posture-based events. The system makes use of adapted Vander-Lugt correlator (VLC) and joint-transfer correlator (JTC) techniques in order to make decisions on the identity of a patient and his 3D positions. They also proposed a fuzzy logic technique to get decisions on the objects behaviour and an adapted fuzzy logic control algorithm in order to make a decision based on information given to the system. Again the question arises to the situation where the level of brightness is less. Kaur and Dhir (2013) have proposed the implementation of face detection methods by combining skin detection methods with template matching, where skin detection is carried out using the YCbCr method where as the template matching is performed in edge detected image. This edge detection is performed using a fuzzy edge detection method which is used to detect edges of an image without determining the threshold value. Feasibility of this approach for various kinds of images remains a question.

Kim et al. (2009) have proposed an object recognition processor which lightens the workload by estimating the global ROI. This estimation of ROI is performed by a neuro-fuzzy controller and this controller also manages the processors overall pipeline stages by using workload aware task scheduling. As pipelining is introduced here raises a question of parallel pipelining. Lopes et al. (2013) have introduced an object tracking approach which is based on fuzzy concepts. The tracking task is performed through the fusion of these fuzzy models by means of an inference engine. Here the object properties considered are very basic, the properties like shape and textures, etc., have not been considered. Rajakumar et al. (2011) have proposed a fuzzy filtering technique for contour detection; the fuzzy logic is basically applied to extract value for an image which is used for edge detection. In their approach, the threshold parameter values are obtained from the fuzzy histograms of an input image, and the fuzzy inference method selects the complete information about the border of the object. Their proposed system works for grey images, but the question whether this system is feasible under occlusion or cluttered image remains a question. Maddalena and Petrosino (2010) have adopted existing approach to background subtraction which is based on self-organisation through artificial neural networks, they have proposed a spatial coherence variant to such approach in order to enhance robustness against false detections and have formulated a fuzzy model just to deal with decision problems. Various video sequences are used. Ma et al. (2012) have combined a fuzzy support vector machine (FSVM) with template matching process to improve the computational efficiency of the process of OD; and also have parallelised the process of template matching on a multi core platform with OpenMP. The system works by initially classifying the samples by template matching and then they are refined by the FSVM classifier. Although the use of multi core platforms increases the computational efficiency, but the use of it also increases the cost and space required. Tables 4(a) and 4(b) summarises the approaches for OD based on fuzzy logic.

Table 4(a) Summary of fuzzy-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Reyes and Dadios	Self-created dataset	Fuzzy colour constancy	Hits and misses	Multi-channel colour imaging	Proposed scheme tremendously cuts processing time.	The question of the colour ranges that could be detected remains.
Oz-Salinas et al.	Self-created dataset	Fuzzy logic	Accuracy	Navigation in indoor environments	Proved to successfully detect doors under strong perspective deformations.	Restricted to indoor environment.
Bernardin et al.	Self-created dataset	Boosted cascades	False positives per image (FPPI)	Monitoring of indoor environments	Quickly acquire and track subjects.	Level of darkness not addressed.
Malaviya and Malaviya	Not applicable	Discussed	Not applicable	Fuzzy logic-based visual object recognition	In fuzzy type object detection, there is a lack of semantic knowledge.	Hence what should be actually obtained from a given image remains a query.
Elbouz et al.	Self-created dataset	Optical correlation	Speed	Patient monitoring	This method is applicable to a very wide range of situations and is robust enough.	The question arises to the situation where the level of brightness is less
Kaur and Dhir	GTVA dataset	Fuzzy edge detection	Detection rate	Detection in still coloured images	The computation time taken by this method is more than classical methods.	Feasibility of this approach for various kinds of images remains a question.
Kim et al.	Any standard dataset	Neuro fuzzy-based pipelining	Execution time	Reducing the workload	Workload aware management is performed to reduce power consumption.	Parallel pipelining not considered.
Lopes et al.	PETS dataset	Fusion of fuzzy models	Accuracy	Single and multiple tracking	The proposed method is robust.	Properties like shape and textures, etc., have not been considered.

Table 4(b) Summary of fuzzy-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Rajakumar et al.	Grey images used	Fuzzy filtering	PSNR value	Shape detection	Filter mask is used for all kinds of images.	Cluttered and occluded images not considered.
Maddalena et al.	Not applicable	Fuzzy spatial coherence	False detection	Background foreground separation	Robust against false detections.	Cluttered image scenes not considered.
Ma et al.	UIUC dataset	Multi-core computation	Accuracy	Template matching	High computational efficiency is achieved.	Cost and space increases.

3.5 Context-based OD

Wolf and Bileschi (2006) have designed a detector for object context. By using context, authors have demonstrated detection of locations that are likely to contain the object of interest. Mainly they have shown that context may be determined from basic visual features like colour and texture. Occlusion remains a point. Perko and Leonardis (2010) have presented a framework for visual-context aware OD; authors have tried to extract visual contextual information from images which can be used prior to the process of OD. In addition, bottom-up saliency and object cooccurrences are used in order to define auxiliary visual context. Finally all the individual contextual cues are integrated with local appearance-based object detector by using a fully probabilistic framework. This system is tested on still images, can it work on other types of images remains an issue.

Kumar and Hebert (2005) had presented a two layer hierarchical formulation to exploit different levels of contextual information in images for robust classification. He has claimed that their approach has two main advantages namely: first, it encodes the short-range interactions as well as the long-range interactions and second being that it can be applied to different domains. Although this system has these advantages, the authors have assumed that the images are labelled, which could not be a case always. Peralta et al. (2012) have presented a method which learns adaptive conditional relationships that depend on the type of scene being analysed. Basically they have proposed a model-based on a conditional mixture of trees which is able to capture contextual relationships among objects using global information about an image. Relationships between objects in an image could be formed only when the image is clear enough but what if the image is occluded. Object categorisation makes use of appearance information and context information in order to improve the object recognition accuracy. Galleguillos and Belongie (2010) have addressed the problem of incorporating different types of contextual information for object categorisation and have also reviewed the different ways of using contextual information for object categorisation. Contextual information would be accurate, once the images are labelled which will not be the case always hence efficiency of this approach could be an issue.

Bergboer et al. (2006) have presented a dual stage context-based (COBA) OD process; in the first stage an object descriptor based on visual features id used to find the object candidates present in the image while in the second stage the identified object candidates are assigned a confidence value based on contextual information. If the image is occluded or cluttered then in that case the descriptor may have an issue while finding an object in an image. Torralba et al. (2003) have presented a context-based vision system for place and object recognition, firstly the locations are identified in order to categorise environments and then use this information in order to provide contextual priors for object recognition. The main advantage of using this system is that it cuts down the number of possible objects that are needed to be considered. But what if the image is cluttered with various different classes of objects remains an issue. Kontschieder et al. (2012) have presented context sensitive decision forests which is used to exploit contextual information that would be helpful for solving the OD problem. Their system has the ability to access the information about each of the samples in the training set which is helpful to learn the contextual information throughout the growing process. In addition to this, they have also introduced a novel split criterion which in combination with a priority-based way of constructing the trees, allows more accurate regression mode selection and hence improves the current context information. Problem may arise when there are a number of object classes present in an image.

Torralba et al. (2004) have presented a method for both detecting and segmenting the objects present in the image. Boosted random fields (BRFs) have been introduced in order to exploit the contextual information, this BRF algorithm combines boosting and conditional random field (CRF) which eases the task of training and inference. Basically they have tried to show hoe contextual information could be helpful for OD. But the question regarding the various kinds of images remains an issue. Torralba (2003) had introduced a simple framework for modelling the relationship between context and object properties which is based on the correlation between the statistics of low level features across the entire image and the objects that it contains. But when the number of object classes in an image increases then the working of this system may come under question. Song et al. (2011) have proposed an iterative contextualisation scheme in order to mutually boost the performance of both OD and classification tasks. For this scheme to work efficiently the authors have initially proposed a contextualised support vector machine (context-SVM) through which a context adaptive classifier is achieved; and then this context-SVM was made use of to boost performance of OD and classification tasks. The scheme boosts up the performance but its efficiency on various kinds of images still remains a challenge.

Chen and Tian have presented a part-based hierarchical compositional model (HCM) which makes use of context information from signage for door detection and classification of the doors that are detected. The basic advantage of using their system is that it can handle partially captured objects as well as large variations in object classes. What if the image is totally occluded or cluttered? Wang et al. (2011) have presented a method named feature context (FC) which is an extension to shape context (SC) method; where FC is used to encode the spatial information of local image features. SC computes histogram of the points that belong to the target shape but on the contrary FC can be applied to the entire image. They have also introduced radial basis coding (RBC) in order to encode the local image features. Applying this process to the entire image may increase the computational cost. Parikh et al. (2008) have proposed a model for context which includes relative location and scale information along with cooccurrence information. Basically when given a segmentation of an image, this model assigns each segment to an object category based on the appearance and contextual information of the segment. Their method may not perform well in case of a cluttered image or an occluded image.

Rutishauser et al. (2004) have tried to investigate to what extent pure bottom-up approach can extract useful information about the location, size and shape of objects from images and how this information is useful in learning objects in an unlabelled image. And have commented that bottom-up attention is useful for variety of applications. He et al. (2004) have proposed an approach which makes use of contextual features for labelling images; these features are incorporated into a probabilistic framework which combines the outputs of many components. Their model basically is a combination of individual models where each model provides labelling information, a classifier that looks at local image statistics, regional label features and global image features that look at the label patterns at local and global levels. Using contextual features for labelling a cluttered image or an occluded image could be a challenge. Fink and Perona have proposed a method which basically generalises the efficient features suggested by Viola and Jones. They have tried to justify the power of a single AdaBoost learner could be augmented using mutual boosting. The concept of mutual boosting could cause computational overhead.

Shi and Malik (2000) have proposed an approach to solve the grouping problem in vision; this is achieved by extracting the global features of an image rather than focusing on the local features of an image. They have treated the grouping problem as a graph partitioning problem and based on it they have proposed the normalised cut criteria for segmenting the graph. The basic purpose of this criterion is that it measures both the total dissimilarity between the different groups as well as the total similarity within the groups. When the global features are considered, there are objects which are not of our interest that are needed to be processed hence increasing the computational cost. Galleguillos et al. (2008) have developed an approach for object categorisation which uses context cooccurrences, location and appearance of an object for that purpose, and hence have named their approach as for cooccurrence, location and appearance (CoLA). CoLA makes use of CRF to maximise object label agreement in the image according to spatial and cooccurrence constraints. Working on different kinds of images could be an issue.

Ramström and Christensen (2004) have proposed a basic model for context-based OD where context was extracted as coherent regions with the help of a distributed segmentation scheme; this helped finding conspicuous visual cues and salient regions. By guiding the foveated part of an active vision system to salient regions and conspicuous visual cues the complexity of the visual search was reduced. What if the background was occluded? Liu et al. (2011) have addressed the problem when using photogrammetric constraints in OD when camera poses are unknown, basically photogrammetric context captures the relationship between object heights and camera viewpoint. They also proposed a branch-bound and cut algorithm to solve the NP-hard problem in structured predictions when cuts of latent variable are embedded into the branch and bound process. What if the objects are cluttered and what if the image is occluded? Sun et al. (2011) have argued that feature selection is an important problem in OD and have demonstrated that genetic algorithms (GAs) provide a simple, general, and powerful framework for selecting good subsets of features, which lead to improved detection rates. They have considered PCA for feature extraction and support vector machines (SVMs) for classi1cation. The goal is searching the PCA space using GAs to select a subset of eigenvectors encoding important information about the target object of interest. Genetic algorithms may increase the time to detect objects.

Sun et al. (2012) have presented a framework which jointly detects objects, estimates the scene layout and segments the supporting surfaces holding these objects. The object detector module is capable of adaptively changing its confidence in establishing whether a certain ROI contains an object (or not) and is-based on iterative estimation procedure by which the object detector becomes more and more accurate. The worst case running time of this algorithm can be high which may increase the computational cost. Russell et al. (2007) have build a system which recognises and localises various object categories in complex images and have achieved it by matching the input image with the images present in the large training set of labelled images. Since there are regularities in object identities across similar images, the retrieved matches provide hypotheses for object identities and location and therefore they use a probabilistic model to transfer the labels from the retrieval set to the input image. Again the worst case to get all the objects matched would have a high execution time.

Rabinovich et al. (2007) have proposed an approach which incorporates semantic object context as a post-processing step into any object categorisation model. Authors make use of CRF framework that maximises object label agreement according to contextual relevance. Their approach compares two sources of context: one learned from training data and another queried from Google Sets. Again the running time of this process will be high as matching is done. Singhal et al. (2003) have presented an approach to determine scene content, which is based on a set of individual material detection algorithms, as well as probabilistic spatial context models. As the major limitation to individual material detectors is that the number of misclassifications that occur because of similarities in colour and texture of various materials in an image and to reduce the number of misclassifications, authors have developed a context-aware material detection system. The major challenge to this system will come when the image is cluttered and consists of many materials at a time. Zheng et al. (2009) have proposed a context modelling framework which works without the need of any prior scene segmentation or context annotation. This is achieved by exploring a polar geometric histogram descriptor for context representation. In order to quantify context, authors have formulated a new context risk function and a maximum margin context (MMC) model to solve the minimisation problem of the risk function. Working of this system may be questioned when the image is cluttered or occluded.

Verbeek and Triggs (2007) have introduced a CRF-based scene labelling model which incorporates local features and features aggregated over the whole image. They also introduce a method for learning CRFs from datasets with many unlabelled nodes by marginalising out the unknown labels so that the log-likelihood of the known ones can be maximised by gradient ascent. Considering the features of the total image may cause unnecessary waste of time in certain cases. Belongie et al. (2002) have introduced a shape descriptor named SC which is used for correspondence recovery and shape-based OD. The SC works by capturing the distribution over relative positions of other shape points and finally summarises global shape in a local descriptor. Deforming objects could cause some issues with this shape descriptor. Shotton et al. (2007) have proposed an approach for learning discriminative model of object classes, incorporating texture, layout, and context information efficiently. Basically the learned model is used for automatic visual understanding and semantic segmentation of images. They also proposed a discriminative model that exploits texture-layout filters, features based on textons, which jointly model patterns of texture and their spatial layout. Unary classification and feature selection are achieved using shared boosting to give an efficient classifier which can be applied to a large number of classes. Accurate image segmentation is achieved by incorporating the unary classifier in a CRF. Occlusion remains a problem.

Gepperth et al. (2012) have tried to explore the potential contribution of multimodal context information to OD in an ‘intelligent car’. The used car platform incorporates subsystems for the detection of objects from local visual patterns, as well as for the estimation of global scene properties such as the shape of the road area or the 3D position of the ground plane. In order to quantify the contribution of context information, they have investigated whether it can be used to infer object identity with little or no reference to local patterns of visual appearance. This system is restricted to car detection. Challenges will be given when a cluttered image is given to this system. Oliva and Torralba (2007) have reviewed the role of context in object recognition process. And have presented a computational model of attention guidance that integrates context information with image saliency to determine regions of interest. By comparing scan patterns of different models to those of human observers, authors validate the proposition that top-down information from visual context modulates the saliency of regions during the task of OD. Saliency calculation for an occluded image can be a concern. Kruppa and Schiele (2003) have explored the idea of using local context for face detection purpose; making use of quantitative and qualitative analyses authors claimed that the detection of the local context of faces in greyscale images is feasible, which is in contrast to the traditional object-centred approach to face detection where the role of local context has so far been neglected. This idea has been only performed on grey scale images; there can be an issue while using various kinds of images.

Murphy et al. (2003) have presented a method to combine global and local image features in order to solve the task of OD. Standard approaches to OD focus on local patches of the image, and try to classify them as background or not. They proposed to use the scene context as an extra source of global information, to help resolve local ambiguities. Also a CRF is presented for jointly solving the tasks of OD and scene classification. Large number of object classes may cause problems to this approach. Bar (2004) had proposed a testable model for rapid use of contextual associations in recognition in which an early projection of coarse information can activate expectations about context and identity that, when combined, result in successful object recognition. They also have highlighted some open questions such as how are context frames represented in the cortex, and what triggers their activation? How is contextual information translated into expectations and many more? Tables 5(a) through 5(f) summarises the approaches for OD based on context.

3.6 Other Types of OD

Torrent et al. (2013) have proposed a framework to simultaneously perform OD and segmentation on objects of different nature, which is based on a boosting procedure which automatically decides – according to the object properties – whether it is better to give more weight to the detection or segmentation process to improve both results. Their approach allows information to be crossed from detection to segmentation and vice versa. The timing of this task may increase if initially the object detected is not the one of interest. GE et al. (2009) have studied the use of Asymmetric Adaptive Boosting (AdaBoost) in the OD process. And have commented that Asym-Gentle AdaBoost methods are more robust than Asym-Real AdaBoost and achieve better performance than the previous symmetric and asymmetric AdaBoost algorithms on both face detection and pedestrian detection.

Table 5(a) Summary of context-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Wolf et al.	StreetScenes dataset	Context feature	Detection rates	Context detection	Determination of map of object context in under 10 seconds.	The case of occlusion is not addressed.
Perko et al.	Multiple demanding datasets used	Visual context	Detection rate	Visual context aware OD	Integration is based on modelling.	Variety of images has not been considered for testing.
Kumar	Multiple datasets used	Hierarchical field formulation	Accuracy	Modelling different types of contexts	The formulation is general enough to be applied to different domains.	Authors have assumed that the images are labelled, which could not be a case always.
Galleguillos et al.	Not applicable	Context and appearance information	Not applicable	Object categorisation	Authors review different ways of using contextual information.	Use of appearance information which may improve the result.

Table 5(b) Summary of context-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Peralta et al.	OUTDOOR and SUN09 datasets used	Adaptive context model	Accuracy (FPPI)	Category level recognition	Improves object recognition performance with respect to a single tree model.	Obtaining relationship between objects during occlusion is not explored.
Bergboer et al.	Self-created dataset	COBA approach	Number of false positives	Object validation	COBA performs favourably when compared with current methods.	If the image is cluttered then descriptor may have an issue while finding an object.
Kontschieder et al.	TUD dataset	Context sensitive decision forest	Accuracy	Classification and regression	Allows more accurate regression mode selection and hence improves the current context information.	Problem may arise when there are a number of object classes present in an image.
Torralba et al.	Self-created dataset	Visual context	Accuracy	Place recognition and categorisation	Global image representation that provides relevant information for place recognition.	if the image is cluttered with various different classes of objects remains an issue.
Torralba et al.	Self-created dataset	Boosted random fields	Accuracy and speed	Exploiting image data and contextual information	The BRF algorithm provides a natural extension of the cascade of classifiers.	The question regarding the various kinds of images remains an issue.
Torralba	Self-created dataset	Contextual priming	Accuracy and speed	Modelling a relationship between context and object properties	Object locations and scales can be inferred from a simple holistic representation of context.	Increase in no of classes may be a concern.
Song et al.	VOC 2010 dataset	Iterative contextualisation	Detection rates	Boosting object classification	Context-SVM utilised to iteratively and mutually boost performance of object detection and classification.	Efficiency on various kinds of images remains a challenge.

Table 5(c) Summary of context-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Chen et al.	Self-created dataset	Part-based HCM	Accuracy	Detection and classification of doors	Incorporation of contextual information brings significant improvements.	What if the image is totally occluded or cluttered?
Wang et al.	Caltech dataset	Context feature	Accuracy	Image classification	Using sliding window strategy, FC can outperform more sophisticated object detectors.	Applying this process to the entire image may increase the computational cost.
Parikh et al.	MSRC and Corel datasets used	Appearance and context	Accuracy	Dense scene labelling	Difficult scenes may require the inclusion of more objects.	Cluttered images may be an issue.
Rutishauser et al.	Self-created dataset	Bottom-up attention	Speed	Learning multiple objects from unlabelled images	Other modes of operation, such as learning multiple objects are mentioned.	Bottom-up approach may not be feasible when an image is cluttered.
He et al.	Corel and Sowerby datasets used	Multi scale random fields	Accuracy and speed	Image labelling	The main reasons for our model's success are its direct representation of large-scale interactions.	Using contextual features for labelling a occluded image is not considered.
Fink et al.	CMU/MIT datasets used	Mutual boosting	Accuracy	Information inference	Mutual Boosting could be enhanced by unifying the selection of weak-learners.	The concept of mutual boosting could cause computational overhead.
Shi et al.	Self-created dataset	Graph partitioning	Detection rate	Grouping problem	This approach segments static images and is found that results are encouraging.	Non-ROI are processed thereby increasing cost.
Galleguillos et al.	MSRC and PASCAL datasets used.	Cooccurrence, location and appearance information	Categorisation accuracy	Object categorisation	Spatial interactions among different categories are rather sparse.	Working on different kinds of images could be an issue.

Table 5(d) Summary of context-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Ramström et al.	Self-created dataset	Background context	Accuracy	Staggered recognition	The presented methodology is evaluated in the context of a table top scenario.	What if the background was occluded?
Liu et al.	INRIA dataset	Photogrammetric context	False detections	Object recognition	This model can get significantly better detection performance than models.	What if the objects are cluttered and what if the image is occluded?
Sun et al.	Multiple datasets used	Feature subset selection	Rate of detection	Object detection	Experimental results illustrate significant performance improvements	Genetic algorithms may increase the time to detect.
Sun et al.	In-house dataset used	Context feedback	Accuracy	Detecting objects, estimating scene layout	This approach is built upon an iterative estimation procedure.	The worst case running time of this algorithm can be high.
Rabinovich et al.	PASCAL and MSRC datasets used	Semantic context	False detections	Object categorisation	Authors have incorporated a parts-based generative model for categorisation.	The running time of this process will be high as matching is done.
Singhal et al.	Corel dataset	Spatial context model	Accuracy	Content understanding	This scheme addresses spatial constraints among multiple material types.	The major challenge to this system will come when the image is cluttered.
Zheng et al.	PASCAL VOC2005) and i-LIDS datasets used	Contextual information	False positive detections	Context surrounding of an object	Superior performance of the MMC model through extensive evaluation.	Working of this system may be questioned when the image is cluttered or occluded.
Verbeek et al.	MSRC dataset	Conditional random fields	Accuracy	Scene segmentation	Partially labelled training images could be handled by maximising image segmentations.	Considering the features of the total image may cause unnecessary waste of time.

Table 5(e) Summary of context-based OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Belongie et al.	MNIST dataset	Shape context	False positives	Shape matching and recognition	Shape context leads to a robust score.	Deforming objects could cause some issues.
Shotton et al.	Corel and Sowerby datasets used.	Texton boosting	Accuracy	Image understanding	The proposed algorithm gives competitive and visually pleasing results.	Occlusion remains a problem.
Gepperth et al.	HRI road traffic dataset.	Context information	Accuracy	OD in an intelligent car	Basis function representations allow the simplest learning methods to perform best.	Challenges will be given when a cluttered image is given to this system.
Oliva et al.	Not applicable	Context	Not applicable	Object recognition	A natural way of representing the context of an object is in terms of its relationship to other objects.	Reviewing the role of context in object recognition process is done.
Oliva et al.	Self-created dataset	Top-down information	Accuracy (FPPI)	Determining region of interest	Contextual information provides a shortcut for efficient object detection systems.	Saliency calculation for an occluded image can be a concern.
Kruppa et al.	CMU/MIT datasets used	Local context	Number of correct detections	Locating faces	Using local context yields correct detections.	Various kinds of images are not considered.
Murphy et al.	Self-created dataset	Combining local and global image features	Accuracy and speed	OD and scene recognition	Authors present a conditional random field for jointly solving the tasks of object detection.	Large number of object classes may cause problems to this approach.

Table 6(a) Summary of other OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Torrent et al.	LabelMe, TUD and Weismann datasets used	Boosting training	Detection rate	Simultaneous segmentation and detection	The approach is valid for performing semiautomatic object labelling.	Timing of this task may increase if ROI is not known a priori.
Jun-Feng et al.	Not applicable	Asymm-etric AdaBoost	Not applicable	Application of AdaBoost to OD	Asymmetric extensions can be derived naturally.	There may be a problem while dealing with cluttered images.
Hsieh et al.	Self-created dataset	Watershed-based transformation	Accuracy	Detecting objects with low contrast	This approach is feasible and effective in detecting objects.	Various types of images not considered.

Table 6(b) Summary of other OD work

Authors	*Dataset used*	*Concept used*	*Performance evaluation parameter*	*Issues addressed*	*Authors remark*	*Our findings*
Nguyen et al.	MIT and INRIA datasets used	Shape-based pattern descriptor	Accuracy in matching	Capturing shape and appearance	This approach performed well when a cluttered background.	Occlusion can be an issue for this approach.
Mimaroglu et al.	Multiple datasets used	Clustering	Execution time	Arbitrary shape OD	ASOD detects the number of objects automatically with respect to the rate.	This approach may not be feasible when image is occluded.
Hussin et al.	Not applicable	Circular hough transform (CHT)	Not applicable	OD from complex background images	This approach automatically detects the desire object.	CHT may not exactly detect the circular object.
Pavani et al.	MIT/CMU datasets used	Haar features	Accuracy and speed	Fast OD	Object detectors based on the proposed features are more accurate and faster.	Creating rectangles can be problematic.
Laptev	VOC2005) dataset	Boosting histograms	Detection rate	Localisation	Validation of the method on recent benchmarks for object recognition shows its superior performance.	Boosting histograms may not be feasible when the image is occluded or cluttered.
Bhanu et al.	SAR dataset	Genetic programming	Accuracy	Synthesising composite operators	GP can synthesise effective composite operators.	It may be time consuming as it finds the features of the total image rather than ROI.
Malagon-Borja et al.	MIT pedestrian dataset	Principal component analysis	Accuracy (FPPI)	Locating pedestrians in a still image	This approach can be generalised to the detection of several different types of objects.	Timing can be an issue because the classifier examines each location in the image.
Zhang et al.	UIUC image dataset	Spatial histogram features	Detection rates and false detections	Detection of distinguishable parts of an object	Feature selection methods are efficient.	There will certainly be an issue if the image is occluded.

Hsieh et al. (2006) have proposed an approach to the detection of small objects which makes use of a watershed-based transformation. Their proposed detection system basically includes two main modules, ROI locating and contour extraction. ROI is generated by making use of an image differencing technique and the watershed-based segmentation algorithm applied on the ROI to extract object contours. Performance of this method would be questioned when applied on various other types of images. Nguyen et al. (2013) have proposed an object descriptor which is able to capture the shape and appearance information of an object in an image. Contours templates which represent object shape are used to derive a set of key points at which the appearance feature named non-redundant local binary pattern (NR-LBP), is computed. Finally an object descriptor is formed that concatenates BR-LBP features and appearance of the object. When an image is occluded then the working of this approach may have some issues. Mimaroglu and Erdil (2011) have considered the problem of detecting arbitrary shape objects as a clustering application by decomposing images into representative data points, and then performing clustering on these points. Their method of ASOD is based on COMUSA which is an algorithm for combining multiple clustering's. Their approach may not work where an image is occluded.

Hussin et al. (2012) have discussed about the various techniques on how to detect the mango from a mango tree. The techniques are colour processing which is used as primary filtering to eliminate the unrelated colour or object in the image. Besides that, shape detection are been used where it will use the edge detection, circular Hough transform (CHT). Pavani et al. (2010) have proposed a method for assigning optimal weights to the rectangles of the Haar-like features so that the weak classifiers constructed based on them give best possible classification performance. The optimal weights were computed in a supervised manner using three different techniques namely: brute-force search, genetic algorithms and Fisher's linear discriminant analysis. Creating rectangles can be problematic when an image is cluttered or highly occluded. Laptev (2009) presented a method for OD that combines AdaBoost learning with local histogram features. He had introduced a weak learner for multi valued histogram features and also analyse various choices of image features. Histogram-based descriptors can be feasible only when the image is natural and clear. It may not be feasible when the image is occluded or cluttered.

Zhao et al. (2011) have proposed a background model named greyscale arranging pairs (GAP) which is based on the statistical reach feature (SRF). This model makes use of the multi point pairs that exhibit a stable statistical intensity relationship as a background model. The intensity difference between pixels of the pair is much more stable than the intensity of a single pixel, especially in varying environments. Occlusion can be an issue for this model. Bhanu and Lin (2004) have made use of genetic programming in order to synthesise composite operators and composite features to detect potential objects in images. Genetic programming can synthesise effective composite operators for OD by running on selected training regions of training images and the synthesised composite operators can be applied to the whole training images and other similar testing images. Malagon-Borja and Fuentes (2009) have presented an OD system which works without assuming any prior knowledge about the image. Their system works as follows: in the first stage a classifier examines each location in the image at different scales. Then in a second stage the system tries to eliminate false detections based on heuristics. The classifier is based on the idea that Principal Component Analysis

(PCA) can compress optimally only the kind of images that were used to compute the principal components (PCs). Thus the classifier performs separately the PCA from the positive examples and from the negative examples; when it needs to classify a new pattern it projects it into both sets of PCs and compares the reconstructions, assigning the example to the class with the smallest reconstruction error. Timing can be an issue because the classifier examines each location in the image.

Zhang et al. (2006) have presented a spatial histogram feature-based OD approach which automatically selects informative spatial histogram features and learns a hierarchical classifier by combining cascade histogram matching and a SVM to detect objects in images. There will certainly be an issue if the image is occluded. Ugolotti et al. (2013) have presented a method for OD in images that is based on deformable models and swarm intelligence algorithms. The task of OD is modelled as an optimisation problem which is tackled using particle swarm optimisation (PSO) and differential evolution (DE). Occlusion and cluttered images may cause an issue to this approach.

Park (2001) had presented a criterion composed of the area variation rate and the compactness of the segmented shape which is based on optimisation that is used to select local optimum thresholds. This method shows to have the shape resolving property in the subtraction image, so that overlapped objects may be resolved into bright and dark evidences characterising each object. His approach may have an issue when an image is cluttered and large number of objects overlaps each other. Nonato et al. (2008) have developed a framework for triangle characterisation in 2D meshes which is applied to the problem of OD and also helps in creating two dimensional models from images. The main advantage of using this is that it removes the unimportant objects present in an image. But if the image is cluttered, then in that case identifying the unimportant object may be an issue with this system. Table 6(a) through 6(c) summarises the various other types of approaches for OD.

4 Open and key issues in OD

Following are the issues in the field of OD:

- Is it necessary to scan the whole image in order to locate the object?, i.e., speed up.
- How to combine the classifiers?, i.e., accuracy.
- Which are good sets of classifiers that are needed to be combined?, i.e., accuracy.
- When and how should the combined classifiers be trained?, i.e., accuracy.
- Should multi-class recognition be performed by detection or by classification?, i.e., speed up and accuracy.
- How to evaluate the performance for an undefined class distribution?, i.e., performance evaluation.

- How can different views of an object be identified as representing a single object?, i.e., accuracy.
- How to handle the occluded objects from detection point of view?, i.e., improvement and efficiency.

5 Online source codes and datasets

The details for the source codes and the standard datasets, generally, used in the process of OD is listed in Tables 7 and 8 respectively. Tables 7 and 8 provide the detail of the authors, the language in which the source code is written and the URL's for those source codes and datasets.

5.1 Online source codes

Table 7 Details of various source codes available

Code for	*Author*	*Language*	*URL*
Colour detection and object tracking	Shemal Fernando	C++	http://opencv-srf.blogspot.in/2010/09/object-detection-using-colour-seperation.html
Face detection using HAAR cascades	Alexander Mordvintsev and Abid K.	Python	http://opencv-python-tutroals.readthedocs.org/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html
Object detection	OpenCV dev Team	OpenCV	http://docs.opencv.org/modules/ocl/doc/object_detection.html
Object detection and tracking	Shamsheer Verma	OpenCV, Visual Studio C++ 2010	http://www.instructables.com/id/OBJECT-DETECTION-AND-TRACKING-USING-OPENCV-VISUAL-/
Training HAAR classifier	Thosten Ball	OpenCV	http://coding-robin.de/2013/07/22/train-your-own-opencv-haar-classifier.html
Real time object tracker	Chesnokov Yuriy	C++	http://www.codeproject.com/Articles/22243/Real-Time-Object-Tracker-in-C
Part-based object detection	Daniel Rodríguez Molina	C++	http://www.uco.es/~in1majim/proyectos/libpabod/
Object detector with boosting	Li Fei-Fei	MATLAB	http://people.csail.mit.edu/torralba/shortCourseRLOC/boosting/boosting.html

5.2 Online datasets

Table 8 Details of various datasets available

Datasets	*URL*
LabelMe toolbox	http://people.csail.mit.edu/torralba/LabelMeToolbox/LabelMeToolbox.zip
CMU/MIT face dataset	http://vasc.ri.cmu.edu/idb/images/face/frontal_images/images.tar
UIUC car dataset	http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/CarData.tar.gz
INRIA person dataset	http://pascal.inrialpes.fr/data/human/INRIAPerson.tar
ETHZ dataset	http://www.vision.ee.ethz.ch/datasets/downloads/ethz_shape_classes_v12.tgz
PASCAL VOC 2010	http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/VOCtrainval_03-May-2010.tar
PASCAL VOC 2011	http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2011/VOCtrainval_25-May-2011.tar
PASCAL VOC 2012	http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2012/VOCtrainval_11-May-2012.tar
BSD dataset	http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300-images.tgz

6 Possible solution

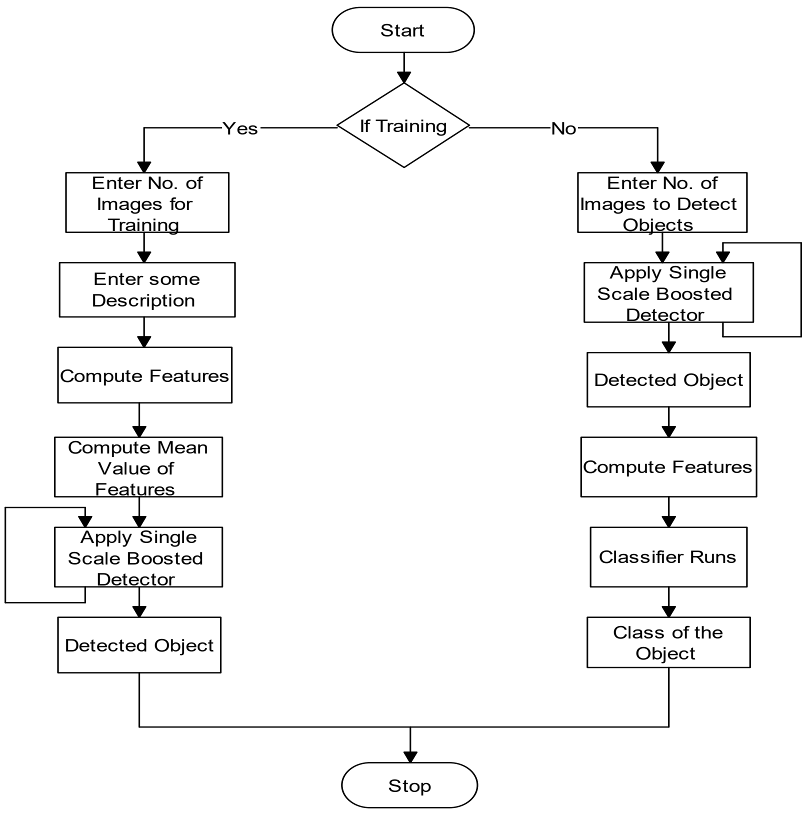
Although for human beings, the recognition of familiar objects of any type and in any kind of environment may be a simple task, but the process of recognition is still a huge difficulty for computers. Especially, the situations when there are changes in light or there is some sort of movements in space, the images of same object look differently. On the other hand, the number of instruments that are able to capture images from day to day life has increased drastically. To make instruments capable to classify such huge amounts of data, OD has become a real challenge for the researchers.

After studying the referred literature, it is identified that the accuracy and time are the main key parameters to judge any approach of OD. A few of the research papers provide the balancing solution between the accuracy and time. Possible approach to have the good balance between the accuracy and time can be devised by using the concept of Steiner tree. The task of detecting objects and classifying them into a particular class is formulated as a Steiner tree problem. The Steiner tree problem basically deals with finding the minimum path between the given set of vertices. As it is an optimisation problem and NP-hard problem, the scope for research exists.

The basic sliding window approach for OD analyses a large number of image regions (of the order of 50,000) regions for a 640×480 pixel image) to know which of the region contain the object of interest. In many applications, there is a need for recognising multiple object classes, and hence multiple binary classifier are required to run over each region, for instance, if detected object to be classified into one of the ten object classes, the sliding window approach may require to explore 500,000) regions to detect the correct class. Hence there is a need to have the approach which can minimise the search space (number of regions), i.e., the approach should explore only those regions where the probability of occurrence of object of interest is more. Steiner tree-based approach can be developed to address this issue. For this purpose, a Steiner tree-based classifier can be used to classify the object(s) present in a particular image. Multi scale boosted detector can be used for recognising the objects in the image.

Various parameters for a particular image can be identified and the range of values for these parameters should be fixed. Identified parameters can be used to mark the various levels of the Steiner tree. Then, the classifier has to be trained using the range of values for particular class so that the classifier can classify the identified object into particular class. During the evaluation phase the classifier uses the set of values that are obtained for a particular image and matches those values with the range of values store in the database for particular class. When the parameter value obtained for the given image is the same as the fixed value for the parameter, then it follows the path through the node of one level to other, but when the parameter value for the given image does not match the fixed value of the parameter, then the intermediate Steiner node is to be created for this value for the relevant level. For later case, the evaluation of the intermediate Steiner node can be carried out using the Euclidian distance to check the relevance of this intermediate node with the main nodes of that level. Finally, the Steiner tree can be created based on the identified main nodes. Last node of the created Steiner tree is the Steiner node for the class of the object. Execution flow diagram for discussed possible solution is shown in Figure 3.

Figure 3 Flow diagram of the possible solution



7 Conclusions

This paper presents the review of the various methods for detecting objects in images as well as in videos. The process of OD is classified into five major categories namely Sliding window-based, contour-based, graph-based, fuzzy-based and context-based. Apart from this, other approaches that are used for detecting objects like the shape-based detection and Steiner tree-based are also summarised.

A review on the topic of OD has been carried out by Prasad (2012), Madaan and Sharma (2012) and Karasulu (2010). Prasad (2012) has discussed the problem of OD in real images and addressed the various aspects like the feature types, learning model, object templates, matching schemes and boosting methods, where as Madaan and Sharma (2012) have considered the same problem of OD in remote sensing images and explored the concept of pre-segmentation for OD. Karasulu (2010) has reviewed and evaluated the methods for moving OD in videos. Though, the different review papers are available for OD, but, this paper is different from the existing papers. This paper provides the details of the existing approaches based on the key concept which is used as the base for development of the approach. Apart from this, the summary of the available source codes and the datasets used for the evaluation of the OD approach is presented. This paper also provides the idea to solve the multi class OD problem based on the Steiner tree. This paper is useful for the study purpose as well as for the new researchers who want to explore the OD research area.

References

Amine, K. and Farida, M.H. (2012) 'An active contour for range image segmentation', *Signal & Image Processing: An International Journal (SIPIJ)*, Vol. 3, No. 3, pp.17–29, doi: 10.5121/sipij.2012.3302.

Arbeláez, P. (2006) 'Boundary extraction in natural images using ultra metric contour maps', *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop*, p.182.

Bar, M. (2004) 'Visual objects in contexts', *Nature Reviews Neuroscience*, Vol. 5, No. 8, pp.617–629, doi: 10.1038/nrn1476.

Belongie, S., Malik, J. and Puzicha, J. (2002) 'Shape context: a new descriptor for shape matching and object recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp.509–522.

Bergboer, N., Postma, E. and Herik, J. (2007) 'Accuracy versus speed in context based object detection', *Pattern Recognition Letters* Vol. 28, No. 6, pp.686–694, doi: 10.1016/j.patrec.2006.08.004.

Bhanu, B. and Lin, Y. (2004) 'Object detection in multi-modal images using genetic programming', *Journal of Applied Soft Computing*, Vol. 4, No. 2, pp.175–201.

Camp, K. and Stiefelhagen, R. (2007) 'Automatic person detection and tracking using fuzzy controlled active cameras', *IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, C. and Tian, Y. (2010) 'Door detection via signage context-based hierarchical compositional model', *Computer Vision and Pattern Recognition Workshop*, pp.1–6.

Comaschi, F., Stuijk, S., Basten, T. and Corporaal, H. (2013) 'RASW: a run-time adaptive sliding window to improve Viola-Jones object detection', *Proceedings of Seventh International Conference on Distributed Smart Cameras (ICDSC), 2013*, pp.1–6, 29 October to 1 November, doi: 10.1109/ICDSC.2013.6778224.

Cyr, C.M. and Kimia, B.B. (2004) 'A similarity based aspect graph approach to 3D object recognition', *International Journal of Computer Vision*, Vol. 57, No. 1, pp.5–22.

Dasigi, P. and Jawahar, C.V. (2008) 'Efficient graph based image matching for recognition and retrieval', *Proceedings of National Conference on Computer Vision*, *Pattern recognition*.

Deruyver, A., Hodé, Y. and Brun, L. (2009) 'Image interpretation with a conceptual graph-labeling over segmented images and detection of unexpected objects', *Artificial Intelligence*, Vol. 173, No. 14, pp.1245–1265.

Divvala, S.K. (2012) *Thesis on 'Context and Subcategories for Sliding Window Object Recognition*, CMU-RI-TR, pp.12–17.

Elbouz, M., Ayman, A. and Brosseau, C. (2011) 'Fuzzy logic and optical correlation-based face recognition method for patient monitoring application in home video surveillance', *Optical Engineering*, Vol. 50, No. 6, pp.1–13.

Felzenszwalb, P.F and Huttenlocher, D.P (2004) 'Efficient graph based image segmentation', *International Journal of Computer Vision*, Vol. 59, No. 2, pp.167–181.

Ferrari, V., Fevrier, L., Jurie, F. and Schmid, C. (2008) 'Groups of adjacent contour segments for object detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30 No. 1, pp.36–51.

Fink, M. and Perona, P. (2004) 'Mutual boosting for contextual inference', *Proceedings on Advances in Neural Information Processing Systems*.

Galleguillos, C. and Belongie, S. (2010) 'Context based object categorization: a critical survey', *Computer Vision and Image Understanding*, Vol. 114, No. 6, pp.712–722.

GE, J. and Luo, Y. (2009) 'A comprehensive study for asymmetric AdaBoost and its application in object detection', *Journal of Acta Automatica Sinica*, Vol. 35, No. 11, pp.1403–1409.

Gepperth, A., Dittes, B. and Ortiz, M.G. (2012) 'the contribution of context information: a case study of object recognition in an intelligent car', *Journal on Neurocomputing*, Vol. 94, No. 1, pp.77–86, doi: 10.1016/j.neucom.2012.03.008.

Gualdi, G., Prati, A. and Cucchiara, R. (2011) 'Multi-stage particle windows for fast and accurate object detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 8, pp.1589–1604.

Gunduz-Demir, C., Kandemir, M., Tosun, A.B. and Sokmensuer, C. (2010) 'Automatic segmentation of colon glands using object-graphs', *Medical Image Analysis*, Vol. 14, No. 1, pp.1–12.

He, L., Han, C.Y., Everding, B. and Wee, W.G. (2004) 'Graph matching for object recognition and recovery', *Pattern Recognition*, Vol. 37 No. 7, pp.1557–1560.

He, X., Zemel, R.S. and Carreira-Perpinan, M.A. (2004) 'Multiscale conditional random fields for image labeling', *CVPR*, Vol. 2, pp.695–702.

Hori, T., Takiguchi, T. and Ariki, Y. (2012) 'Generic object recognition by graph structural expression', *ICASSP*, pp.1021–1024.

Hsieh, F., Han, C., Wu, N., Chuang, T.C. and Fan, K. (2006) 'A novel approach to the detection of small objects with low contrast', *Journal of Signal Processing*, Vol. 86. No. 1, pp.71–83.

Hussin, R., Juhari, M.R., Kang, N.W., Ismail, R.C. and Kamarudin, A. (2012) 'Digital image processing techniques for object detection from complex background image', in *International Symposium on Robotics and Intelligent Sensors*, Vol. 41, pp.340–344.

Karasulu, B. (2010) 'Review and evaluation of well-known methods for moving object detection and tracking videos', *Journal of Aeronautics and Space Technologies*, Vol. 4, No. 4, pp.11–22.

Kaur, R. and Dhir, V. (2013) 'Fuzzy logic based novel approach for face detection', *International Journal of Latest Research in Science and Technology*, Vol. 2, No. 1, pp.558–566.

Kim, J., Kim, M., Lee, S., Oh, J., Oh, S. and Yoo, H. (2009) 'Real-time object recognition with neuro-fuzzy controlled workload-aware task pipelining', *Micro*, *IEEE*, Vol. 29, No. 6, pp.28–43.

Kontschieder, P., Bul, S., Criminisi, A., Kohli, P., Pelillo, M. and Bischof, H. (2012) 'Context-sensitive decision forests for object detection', *Proceedings on Advances in Neural Information Processing Systems*.

Kruppa, H. and Schiele, B. (2003) 'Using local context to improve face detection', *Proceedings of the BMVC*, pp.3–12.

Kumar, S. and Hebert, M. (2005) 'A hierarchical field framework for unified context-based classification', *Tenth IEEE International Conference on Computer Vision*, pp.1284–1291.

Lampert, C.H., Blaschko, M.B. and Hofmann, T. (2008) 'Beyond sliding windows: object localization by efficient sub-window search', *IEEE Conference on Computer Vision and Pattern Recognition*.

Laptev, I. (2009) 'improving object detection with boosted histograms', *Journal of Vision Computing*, Vol. 27, No. 5, pp.535–544.

Lebrun, J., Gosselin, P. and Philipp-Foliguet, S. (2011) 'Inexact graph matching based on kernels for object retrieval in image databases', *Image and Vision Computing*, Vol. 29, No. 11, pp.716–729.

Liang, Z., Chi, Z., Fu, H. and Feng, D. (2012) 'Salient object detection using context sensitive hyper graph representation and partitioning', *Pattern Recognition*, Vol. 45, No. 11, pp.3886–3901.

Lin, L., Wu, T., Porway, J. and Xu, Z. (2009) 'A stochastic graph grammar for compositional object representation and recognition', *Pattern Recognition*, Vol. 42, No. 7, pp.1297–1307.

Liu, Y., Wu, Y. and Yuan, Z. (2011) 'Object detection using discriminative photogrammetric context', *International Conference on Image Processing*, pp.2405–2408.

Lopes, N.V., Couto, P., Jurio, A. and Melo-Pinto, P. (2013) 'Hierarchical fuzzy logic based approach for object tracking', *Knowledge-Based Systems*, Vol. 54, No. 1, pp.255–268, doi: 10.1016/j.knosys.2013.09.014.

Lu, C., Latecki, L.J., Adluru, N., Yang, X. and Ling, H. (2009) 'Shape guided contour grouping with particle filters', *International Conference on Computer Vision*.

Ma, Y., Wu, W. and He, Q. (2012) 'Algorithm for object detection using multi-core parallel computation', *International Conference on Medical Physics and Biomedical Engineering*, Vol. 33, pp.455–461.

Madaan, T. and Sharma, H. (2012) 'Object detection in remote sensing images: a review', *IJSRP*, Vol. 2, No. 6, pp.1–3.

Maddalena, L. and Petrosino, A. (2010) 'Fuzzy spatial coherence based approach to background foreground separation for moving object detection', *Neural Comput & Applic*, Vol. 19, No. 2, pp.179–186.

Maire, M., Arbelaez, P., Fowlkes, C. and Malik, J. (2008) 'Using contours to detect and localize junctions in natural images', *IEEE Conference on Computer Vision and Pattern Recognition*.

Malagon-Borja, L. and Fuentes, O. (2009) 'Object detection using image reconstruction with PCA', *Journal of Image and Vision Computing*, Vol. 27, Nos. 1–2, pp.2–9.

Malaviya, A. and Malaviya, P. (1993) 'Object recognition using fuzzy set theoretic techniques', *SPIE Proceedings*, Vol. 1962.

Mimaroglu, S. and Erdil, E. (2011) 'ASOD: arbitrary shape object detection', *Journal of Engineering Applications of Artificial Intelligence*, Vol. 24, No. 7, pp.1295–1299.

Munoz-Salinas, R., Aguirre, E., Garcia-Silvente, M. and Gonzalez, A. (2004) 'Door-detection using computer vision and fuzzy logic', *Proceedings of the 6th WSEAS International Conference on Mathematical Methods & Computational Techniques in Electrical Engineering*.

Murphy, K., Torralba, A. and Freeman, W.T. (2003) 'Using the forest to see the trees: a graphical model relating features, objects, and scenes', *Proceedings on Advances in Neural Information Processing Systems*.

Oliva, A., Torralba, A., Castelhano, M.S. and Henderson, J.M. (2003) 'Top-down control of visual attention in object detection', *IEEE Proceedings of the International Conference on Image Processing*, Vol. 1, pp.253–256.

Paixão, T.M., Graciano, A., Cesar Jr., R.M. and Hirata Jr., R. (2008) 'A back-mapping approach for graph based object tracking', *Brazilian Symposium on Computer Graphics and Image Processing*, pp.45–52, doi: 10.1109/SIBGRAPI.2008.32.

Parikh, D., Zitnick, L. and Chen, T. (2008) 'From appearance to context-based recognition: dense labeling in small images', *IEEE Conference on Computer Vision and Pattern Recognition*.

Park, Y. (2001) 'Shape-resolving local thresholding for object detection', *Pattern Recognition Letters*, Vol. 22, No. 8, pp.883–890.

Pavani, K., Gomez, D.D. and Frangi, A.F. (2010) 'Haar-like features with optimally weighted rectangles for rapid object detection', *Journal of Pattern Recognition*, Vol. 43, No. 1, pp.160–172.

Peralta, B., Espinace, P. and Soto, A. (2012) 'Adaptive hierarchical contexts for object recognition with conditional mixture of trees', *Proceedings British Machine Vision Conference*, pp.121.1–121.11.

Perko, R. and Leonardis, A. (2010) 'A framework for visual-context-aware object detection in still images', *Computer Vision and Image Understanding*, Vol. 114, No. 6, pp.700–711.

Prasad, D. (2012) 'Survey of the problem of object detection in real images', *IJIP*, Vol. 6, No. 6, pp.441–466.

Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E. and Belongie, S. (2007) 'Objects in context', *International Conference on Computer Vision*.

Rajakumar, T.C., Perumal, S.A. and Krishnan, N. (2011) 'A fuzzy filtering model for contour detection', *ICTACT Journal on Soft Computing*, Vol. 1, No. 4, pp.197–200.

Ramström, O. and Christensen, H.I. (2004) 'Object detection using background context', *International Conference on Pattern Recognition*, pp.45–48.

Ravishankar, S., Jain, A. and Mittal, A. (2008) 'Multi stage contour based detection of deformable objects', *Computer Vision-European Conference on Computer Vision*, pp.483–496.

Reyes, N.H. and Dadios, E.P. (2004) 'Dynamic color object recognition using fuzzy logic', *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 8, No. 1 pp.29–38.

Russakovsky, O. and Ng, A.Y. (2010) 'A Steiner tree approach to efficient object detection', *Proceeding of: the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1070–1077.

Russell, B.C., Torralba, A., Liu, C., Fergus, R. and Freeman, W.T. (2007) 'Object recognition by scene alignment', *Proceedings on Advances in Neural Information Processing Systems*.

Rutishauser, U., Walther, D., Koch, C. and Perona, P. (2004) 'Is bottom-up attention useful for object recognition?', *Computer Vision and Pattern recognition*, Vol. 2, No. 2, pp.37–44.

Schindler, K. and Suter, D. (2008) 'Object detection by global contour shape', *Pattern Recognition*, Vol. 41, No. 12, pp.3736–3748.

Schlecht, J. and Ommer, B. (2005) 'Contour based object detection', *International Conference on Computer Vision*, Vol. 1, pp.503–510.

Segvic, S., Kalafatic, Z. and Kovaˇcek, I. (2011) 'Sliding window object detection without spatial clustering of raw detection responses', *Proceedings of the Computer Vision Winter Workshop*.

Serratosa, F., Alquezar, R. and Sanfeliu, A. (2003) 'Function-described graphs for modeling objects represented by sets of attributed graphs', *Pattern Recognition*, Vol. 36, No. 3, pp.781–798.

Shams, L.B., Brady, M.J. and Schaal, S. (2001) 'Graph matching vs mutual information maximization for object detection', *Neural Networks*, Vol. 14, No. 3, pp.345–354.

Shi, J. and Malik, J. (2000) 'Normalized cuts and image segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp.888–905.

Shotton, J., Blake, A. and Cipolla, R. (2008) 'Multi scale categorical object recognition using contour fragments', *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 30 No. 7, pp.1270–1281.

Shotton, J., Winn, J., Rother, C. and Criminisi, A. (2007) 'TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context', *International Journal of Computer Vision*, Vol. 81, No. 1, pp.2–23.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. and Zucker, S. (1999) 'Shock graphs and shape matching', *International Journal of Computer Vision*, Vol. 35, No. 1, pp.13–32.

Singhal, A., Luo, J. and Zhu, W. (2003) 'Probabilistic spatial context models for scene content understanding', *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp.235–241.

Song, Z., Chen, Q., Huang, Z., Hua, Y. and Yan, S. (2011) 'Contextualizing object detection and classification', *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1585–1592.

Stiene, S., Lingemann, K., N¨uchter, A. and Hertzberg, J. (2006) 'Contour based object detection in range images', *3D Data Processing*, *Visualization*, *and Transmission*, *Third International Symposium*, pp.168–175.

Subburaman, B., Venkatesh and Marcel, S. (2010) 'Fast bounding box estimation based face detection', *ECCV*, *Workshop on Face Detection*.

Sudowe, P. and Leibe, B. (2011) 'Efficient use of geometric constraints for sliding window object detection in video', *8th International Conference on Computer Vision Systems.*

Sun, M., Bao, S. and Savarese, S. (2012) 'Object detection with geometrical context feedback loop', *IJCV*, Vol. 100, No. 2, pp.154–169.

Suna, Z., Bebisa, G. and Millerb, R. (2011) 'Object detection using feature subset selection', *18th IEEE International Conference on Image Processing*, Vol. 37, pp.2165–2176.

Torralba, A., Murphy, K.P., Freeman, W.T and Rubin, M.A. (2003) 'Context-based vision system for place and object recognition', *Ninth IEEE Conference on Computer Vision*.

Torrent, A., Lladó, X., Freixenet, J. and Torralba, A. (2013) 'A boosting approach for the simultaneous detection and segmentation of generic objects', *Journal of Pattern Recognition Letters*, Vol. 34, No. 13, pp.1490–1498.

Triesch, J. and Eckes, C. (2005) 'Object recognition with deformable feature graphs: faces, hands, and cluttered scenes', *Handbook of Pattern Recognition and Computer Vision*, pp.461–480, doi: 10.1142/9789812775320_0025.

Ugolotti, R., Nashed, Y., Mesejo, P., Ivekovič, S., Mussi, L. and Cagnoni, S. (2013) 'Particle swarm optimization and differential evolution for model based object detection', *Journal of Applies Soft Computing*, Vol. 13, No. 6, pp.3092–3105.

Vajda, P., Dufaux, F., Minh, T.H. and Ebrahimi, T. (2009) 'Graph based approach for 3d object duplicate detection', *Proceedings of the 10th International Workshop on Image Analysis for Multimedia Interactive Services*.

Verbeek, J. and Triggs, B. (2007) 'Scene segmentation with conditional random fields learned from partially labeled images', *Proceedings of Advances in Neural Information Processing Systems*.

Wang, X., Bai, X., Liu, W. and Jan, L.L. (2011) 'Feature context for image classification and object detection', *IEEE Conference on Computer Vision and Pattern Recognition*.

Wolf, L. and Bileschi, S. (2006) 'A critical view of context', *International Journal of Computer Vision*, Vol. 69 No. 2, pp.251–261.

Yang, X., Liu, H. and Latecki, L.J. (2012) 'Contour based object detection as dominant set computation', *Journal on Pattern Recognition*, Vol. 45 No. 5, pp.1927–1936.

Yanulevskaya, V., Uijlings, J. and Geusebroek, J.M. (2013) 'Salient object detection: from pixels to segments', *Image and Vision Computing*, Vol. 31, No. 1, pp.31–42.

Yu, S.X., Gross, R. and Shi, J. (2002) 'Concurrent object recognition and segmentation by graph partitioning', *NIPS*, MIT Press, pp.1383–1390.

Zang, H., Gao, W., Chen, X. and Zhao, D. (2006) 'Object detection using spatial histogram features', *Journal of Image and Vision Computing*, Vol. 24, No. 4, pp.327–341.

Zhang, D.Q. and Chang, S.F. (2005) *Learning Random Attributed Relational Graph for Part Based Object Detection*, Columbia University ADVENT Technical Report.

Zhao, X., Satoh, Y., Takauji, H., Kaneko, S., Iwata, K. and Ozaki, R. (2011) 'Object detection based on a robust and accurate statistical multi-point-pair model', *Journal of Pattern Recognition*, Vol. 44, No. 6, pp.1296–1311.

Zheng, W.S., Gong, S. and Xiang, T. (2009) 'Quantifying contextual information for object detection', *IEEE 12th International Conference*, pp.932–939.

Zhu, Q., Wang, L., Wu, Y. and Shi, J. (2008) 'Contour context selection for object detection: a set-to-set contour matching approach', *Computer Vision – European Conference on Computer Vision*, pp.774–787.