http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (24): 8275-8280, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.8275.8280

Object Detection Using Edge Direction Histogram Features

[1]Guangwei Wang, [1]Zenggang Xiong, [2]Yihua Lan and [1]Conghuan Ye
[1]School of Computer and Information Science, Hubei Engineering University, 432000, Xiaogan, China
[2]School of Computer and Information Technology, Nanyang Normal University, 473061, Nanyang, China

Abstract: In In this study, we propose an object detection approach using edge direction histogram features. Since edge points are related to shape information closely, Edge Direction Histogram (EDH) is a very simple and direct way to characterize shape information of an object. We divide an object into several parts and employ edge direction histogram method to extract the EDH features. The EDH descriptor is designed to decouple variations of the object due to affine warps and other forms of shape deformations. We further train a support vector machine classifier for each object part and apply a generalized Hough voting scheme to generate object locations and scales. We evaluate the proposed approach on two different kinds of objects: Car and h. Experimental results show that the proposed approach is efficient and robust in object detection.

Key words: Object detection, edge direction histogram, feature descriptor, hough voting

INTRODUCTION

Object detection is an important task in computer vision which has made great use of learning. This is a difficult problem because objects in a category can vary greatly in shape and appearance. Variation arise not only from changes in illumination, occlusion, background clutter and view point but also due to non-rigid deformations and intra-class variation in shape and other visual properties among objects in a rich category.

How do we deal with the variation, especial the intra-class and pose variability of object Most of the current researches have focused on modeling object variability, including several kinds of deformable template models (Matthews and Baker, 2004) and a variety of part-based, fragment-based models (Fergus *et al*., 2005; Agarwal *et al*., 2004; Felzenszwalb and Huttenlocher, 2005; Amit and Trouve, 2007; Lee, 2008; Schneiderman and Kanade, 2004; Opelt *et al*., 2006).

The pictorial structure models (Agarwal *et al*., 2004; Crandall *et al*., 2005) represent an object by a collection of parts arranged in a deformable configuration, where the deformable configuration is represented by spring-like connections between pairs of parts. Crandall *et al*. propose k-fans model (Leibe *et al*., 2008; Agarwal and Roth, 2002) to study the extent to which additional spatial constraints among parts are a ctually helpful in detection and localization. The patchwork of parts model from Amit and Trouve (2007) is similar but it explicitly considers how the appearance model of overlapping parts interacts to define a dense appearance model for images. It is proved that adding spatial constraints gives better performance.

There are several possibilities to represent object classes. A star shape model (Shotton *et al*., 2005; Leibe *et al*., 2005) can be easily trained and evaluated in contrast to the constellation model (Fergus *et al*., 2003) or complex graphical models (Sudderth *et al*., 2005). It allows using as many parts as required, since the complexity scales linearly. Moreover, this model is flexible enough to deal with large variations in object shape and appearance of rigid and articulated structures.

The method of (Leibe *et al*., 2008) give a highly flexible learned representation for object shape that can combine the information observed on different training examples. (Opelt *et al*., 2006) explore a similar geometric representation to that of Leibe *et al*. but use only the boundaries of the object, both internal and external (silhouette) (Dalal and Triggs, 2005).

A number of recent approaches have attempted to learn features weights automatically (Bosch *et al*., 2008; Grauman and Darrell, 2007) using variants of multiple kernel learning. These learning mechanisms, however, only allow identify and weigh the most discriminate features but do not allow to identify and model the interplay between features that may prove important to representing object well.

Our approach has two methods to deal with the variation of object, both local and global. Firstly, multi-cues of object part have been incorporated as a key component of local features. In this study, we introduce EDH descriptor for shape information to describe the

Corresponding Author: Guangwei Wang, School of Computer and Information Science, Hubei Engineering University, 432000, Xiaogan, China

object part. Secondly, we present a voting method based on object part to detect object. The voting method codes the global geometry of generic visual object categories with spatial relations linking object center to object part. The proposed framework can be applied to any object that consists of distinguishable parts arranged in a relatively fixed spatial configuration. Our experiments are performed on images of side views of car and horses the horse will be used as a running example throughout the study to illustrate the ideas and techniques involved.

The rest of the study is organized as follow. Section 2 describes the EDH feature descriptor. Section 3 presents the SVM classifier and voting algorithm. In section 4, experiments on real images show that the feature descriptor is effective for object categorization.

Object part feature descriptor: Based on the observation that for a wide variety of common object categories, shape (Sullivan *et al*., 2008; Ferrari *et al*., 2008; Opelt *et al*., 2006) matters more than other information (Bouchard and Triggs, 2005; Zhu *et al*., 2010). In this study, we use shape information as a key component for object detection.

Object representation and feature extraction are essential to object detection. In this section, we describe a novel object pattern representation combining edge and spatial structures.

Here, we adapt the method of (Gu *et al*., 2009) for describing object part features. An object part is subdivided evenly its bounding box into a n×n grid. Each grid encodes information only inside the part. We capture different cues from the cells and each type of cues is encoded by concatenating grid signal into a histogram. In this study, we consider contour shape cues.

Contour shape, we use shape information as a key component for object detection. Since, edge points are related to shape information closely, Edge Direction Histogram (EDH) is a very simple and direct way to characterize shape information of an object. EDH is computed by grouping the edge pixels which fall into edge directions and counting the number of pixels in each direction.

For any sample image, we compute histogram-based edge direction representation as follows. First we apply variance normalization on the gray image to compensate the effect of different lighting conditions, next we use Canny edge detector to transform the image into an edge image, then we compute the edge direction image and finally we compute the edge direction histogram as object part descriptor.

EDH is computed by grouping the edge pixels which fall into edge directions and counting the number of pixels in each direction. Edge map are extracted by edge detection operator (We use Canny edge detector) and each of edge points can be represented with the vector $\vec{D}_{i,j} = \{dx_{i,j}, dy_{i,j}\}$, where $dx_{i,j}$ and $dy_{i,j}$ are, respectively, horizontal and vertical differences of the point. Each point's edge direction (i.e., gradient direction) is calculated with the Eq.:

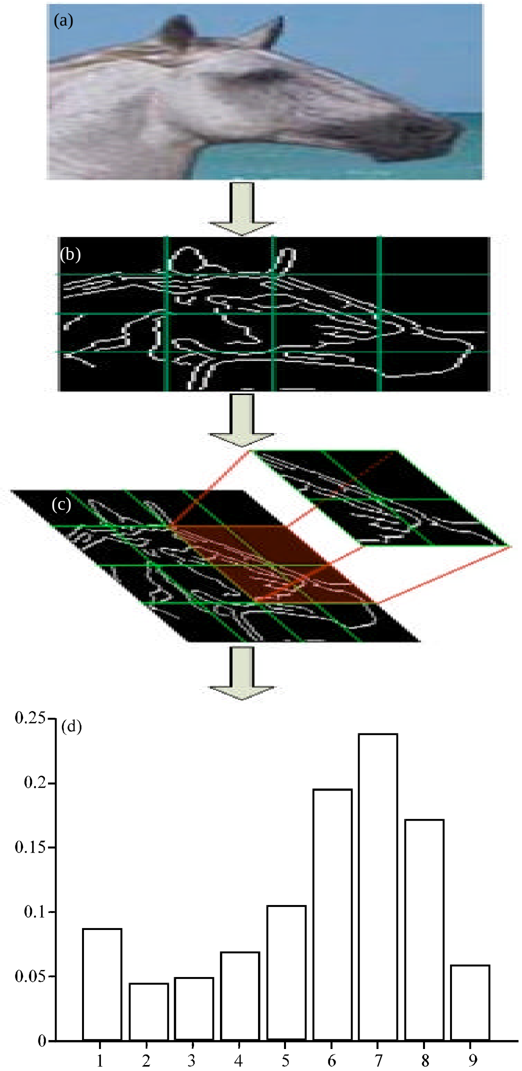

Fig. 1: Contour shape descriptor, (a) Original image, (b) Edge map, (c) 4x4 grid on Edge map and (d) Edge direction histogram of one block containing 4 grids

$$\theta = \arctan\left(\frac{dx_{i,j}}{dy_{i,j}}\right)$$

we then divide direction into bins (e.g., 20° per bin) and calculated the orientation histogram over some region. A global direction histogram of an object part would average too much spatial information to infer pose. We describe an object part window by dividing evenly its bounding box into n×n grid and accumulating a local 1-D histogram of edge pixels within the 2×2 grid, as illustrated

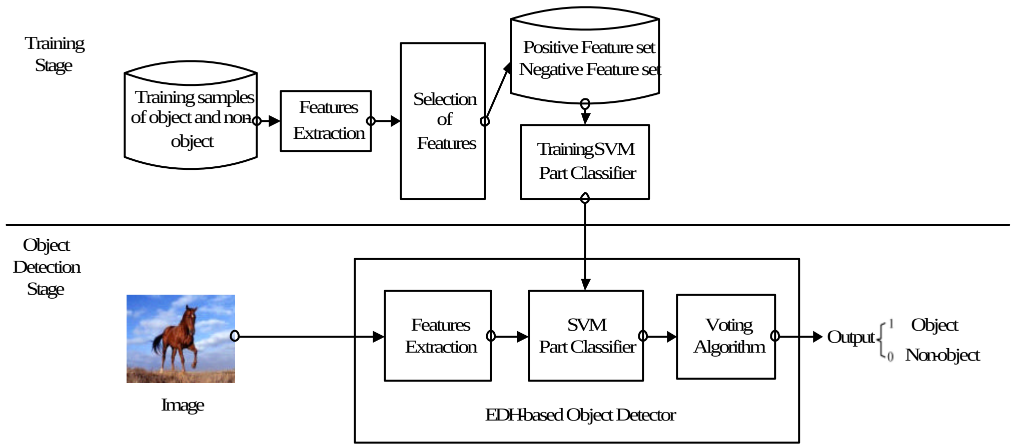

Fig. 2: Architecture of the proposed object detection approach

in Fig. 3. In the experiments reported, we use n = 4. The combination of these histograms then represents the descriptor. Fig. 1 presents the extraction procedure of EDG descriptor.

LEARNING AND DETECTION

Support vector machine for object detection: After running the above feature extraction algorithm, we train some SVM classifiers for object detection in images using the feature set. The task of learning is to establish classifier {Cf(.), Cf (.), ..., Cf (.)} for an object with n parts. Take a classifier for example, given a set of training image windows labeled as positive (object) or negative (nonobjective), each image window is converted into a feature vector as described above. These vectors are then fed as input to a supervised learning algorithm that learns to classify an image window as member or nonmember of the object pattern. In our experiments, we chose linear SVM as classifier.

A SVM performs pattern recognition for a two-class problem by determining the separating hyper plane that maximizes the distance to the closest points of a training set. In our approach, we first adopt an SVM method as the evaluation classifier in selecting informative spatial histogram features and then use the selected feature set to train an SVM for object detection using the Libsvm (Chang and Lin, 2004) software Sliding window classification (Ferrari *et al*., 2008; Ballard, 1981) is a simple, effective technique for object detection. The detection problem is to determine whether the query image contains object part instances and where it is. The classifier Cf (.) is applied to fix-sized windows at various locations in the feature pyramid, each window being represented as a feature vector f(x,l), where x specifies the position of the window in the image and l specifies the level of the image in pyramid. The following expression represents the classifier Cf(.) at one of the sliding windows.

If t_{Pij}, >q, then h_{Pij} = (x,l) and it is a hypothesis position of part P_i Otherwise, it is not. Then we put h_{Pij} into the i^{th} part hypothesis set H_{pi} = {h_{Pi}, hp_{ik} Let H_G = {H_{P1}, H_{P2},..., H_{Pn}} denote the collection of all parts hypothesis position. The architecture of object detection is presented in Fig. 2.

Voting and detection: The goal of this section is to generate bounding boxes of that category in the image. To achieve it, we use a generalized Hough voting scheme based on the transformation between detected parts as well as associated object structure.

In recognition techniques based on Hough voting (Ballard, 1981), the main idea is to represented the object as a collection of parts and have each part to cast votes in a discrete voting-space. Each vote corresponds to a hypothesis of object location h_G and object Γ. The object is identified by the conglomeration of votes in a small neighborhood of the voting space $V(h_G,G).V(h_G,G)$ is typically defined as the sum of independent votes $\hat{p}_i(v,x,l,d,w,f)$ from each part i, where x is the location of the part in query image, l is the scale of the part and f is the part feature. The final Hough image is computed by combining the votes from all parts for the image in question.

The first step is to generate the voting score S_i at a centroid circle location of the object G by applying a transformation T(.) to h_{Pi} of G. The transformation is a voting procedure which is characterized by:

$$S_i(C_i) = T(h_{Pi},G)$$

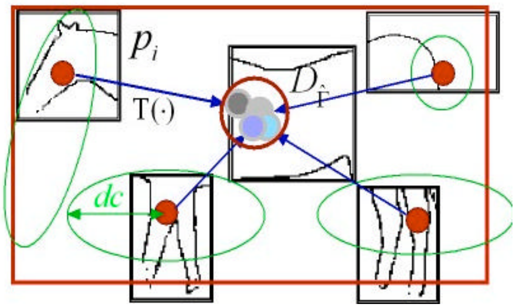

Fig. 3: Voting model. Object part $P_i(i \le i \le m_i)$ (black box) are arranged within the bounding box (red). Small red circles show the optimal position of sub-patterns and green ellipses the spatial uncertainty d_c. Blue arrows represent the voting transformation. Big red circle is the collection of object part voting

where, C_i is the voting location of sub-patterns h_{Pi} . This transformation provides not only the voting score, location C_i but also deviation d_i and weight w_i of sub-patterns in the object pattern. In Fig. 3, blue arrows show the transformation procedure. The transformation T(.) exploits the rough localization provided by the spatial relationship between the sub-patterns and the object pattern:

$$t_{PI},j = Cf\,(f(x,l))$$

If t_{Pij},>q, then$h_{PI,j}$ and it is a hypothesis position of part P_i. Otherwise, it is not. Then we put $h_{Pi,j}$ into the i^{th} part hypothesis set $H_{pi} = \{h_{Pi,1}, h_{P2}, ..., h_{pi,Ki}$ Let $H_G = \{H_{P1}, H_{P2},...,H_{Pn}\}$ denote the collection of all parts hypothesis position. The architecture of object detection is presented in Fig. 2.

Next, we want to find the local max voting score S_G in the Hough image. The object location is obtained by choosing local maxim in the Hough image that has responses above a certain threshold η. The $S_{det}(G)$ can be obtained by:

$$S_{det}(\hat{G}) = \sum_{i=1, c_i \in D_{\hat{\Gamma}}}^{n} (s_i(c_i) \cdot w_i - d_i)$$

The overall detection score $S_{det}(G)$ for object pattern G is a combination of the detected part voting score in the domain D_G. Where W_i the discriminative weight of part is P_i, d_i is the deviation of sub-pattern from the optimal position. If S_{det}, (G)>h, D_G is a centroid of detected object G. Otherwise, it is not.

Table 1: Category recognition rates (at equal error rate) on weizmann and INRIA horse datasets

Dataset	Previous work (%)	Previous work (%)	Marius *et al.*, (%)	Ours (%)
Horses (Weizmann)	92.67	93.12	92.02	94.28
Horse (INRIA)	89.72	91.76	87.14	93.15

Table 2: Comparison of car detection results on test set A

Method	No. of correct detections	No. of false detections	Detection rate (%)	Precision (%)
Zhang *et al.* (2006)	193	45	96.50	81.10
Our approach	196	33	98.00	85.58

Table 3: Comparison of car detection results on test set B

Method	No. of correct detections	No. of false detections	Detection rate (%)	Precision (%)
Zhang *et al.* (2006)	120	37	86.33	76.43
Our approach	129	31	92.80	80.63

EXPERIMENT RESULT

In order to evaluate the effectiveness of the proposed approaches, we conduct experiments of two different object detection tasks. One is to detect side-view car which has non-rigid structure with special componential configuration. The other is horse detection in images. Some performance measures are used to evaluate object detection systems: (1) Detection rate is defined as the number of correct detections over the total number of positives in data set; (2) False positive rate is the number of false positives over the total number of negatives in data set; (3) Precision is the number of correct detections over the sum of correct detections and false positives; (4) Detection is considered correct, if the area of overlap between the predicted bounding box and ground truthbounding box exceeds 50% (Everingham *et al.*, 2007; Ferrari *et al.*, 2006).

First, we test our algorithm on the INRIA (170 positive and 170 negative images) and Weizmann H (328 images) databases. The INRIA and Weizmann are very challenging datasets of horse images, containing different breeds, colors and textures, with varied articulations, lighting conditions and scales We also compare our method with Zhang *et al.* (2006) on the UIUC image database (Agarwal *et al.*, 2004). Side-view cars consist of distinguishable parts such as wheels, car doors and car windows. These parts are arranged in a relatively fixed spatial configuration. Side-view cars have enormous changes in configurations because of various design styles.

The first set (Test Set A) consists of 170 images containing 200 cars with roughly the same size as in the training images. The second set (Test Set B) consists of 108 images containing 139 cars with different sizes. The test sets are difficult for detection since they contain partially occluded cars and cars that have low contrast with backgrounds. Table 2 and Table 3 show theWe compare our method with (Leordeanu *et al*., 2007) and our previous work (Wang and Chen, 2010; Lan *et al*., 2011) on the INRIA and Weizmann Horse databases. Table 1 shows the results of recognition accuracy at the equal error rate points by comparing the algorithm to Marius *et al*. and our previous work. It can be seen that the performance of the algorithm is superior to other two methods. Table 1 show the detection result experimental results of detection rate and precision on the test set A and B.

CONCLUSIONS

In this study, we have presented an edge direction histogram feature-based object detection approach. This method effectively describes the shape features and trains a SVM classifier. Based on this feature descriptor and part detection classifier, Hough voting method was used to detect object. Extensive object detection experiments show high detection rates with relatively low numbers of false detections. These results illustrate the high discriminant power of edge direction histogram features and the effectiveness and robustness of voting object detection approach.

The proposed framework can be applied to any object category that consists of distinguishable parts arranged in a relatively fixed spatial configuration. In summary, the results show that the object representation using edge direction histogram features is general to different kinds of object classes and the feature descriptor are efficient to extract informative shape features for object detection.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the National Natural Science Funds of China under Grant No. 61370092, Natural Science Funds of Hubei Province under Grant No. 2013CFC005 and Young talent program in Hubei Education Department (Grant No. Q20132709).

REFERENCES

Agarwal, S. and D. Roth, 2002. Learning a sparse representation for object detection. Proceedings of the 7th European Conference on Computer Vision Copenhagen, Denmark, Volume 2353, May 28-31, 2002, Springer, Berlin/Heidelberg, pp: 113-127.

Agarwal, S., A. Awan and D. Roth, 2004. Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell., 26: 1475-1490.

Agarwal, S., A. Awan and D. Roth, 2004. UIUC image database for car detection. http://cogcomp.cs.illinois.edu/Data/Car/.

Amit, Y. and A. Trouve, 2007. Pop: Patchwork of parts models for object recognition. Int. J. Comput. Vision, 75: 267-282.

Ballard, D.H., 1981. Generalizing the hough transform to detect arbitrary shapes. Patt. Recog., 13: 111-122.

Bosch, A., A. Zisserman and X. Muoz, 2008. Image classification using ROIs and multiple kernel learning. Int. J. Comput. Vision.

Bouchard, G. and B. Triggs, 2005. Hierarchical part-based visual object categorization. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 20-25, 2005, IEEE Press, pp: 710-715.

Chang, C.C. and C.J. Lin, 2007. LIBSVM-A library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html.

Crandall, D., P. Felzenszwalb and D. Huttenlocher, 2005. Spatial priors for part-based recognition using statistical models. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 20-25, 2005, IEEE Press, pp: 10-17.

Dalal, N. and B. Triggs 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 25, 2005, San Diego, CA., USA., pp: 886-893.

Everingham, M., L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman, 2007. The PASCAL visual object classes challenge 2007. http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/

Felzenszwalb, P.F. and D.P. Huttenlocher, 2005. Pictorial structures for object recognition. Int. J. Comput. Vision, 61: 55-79.

Fergus, R., P. Perona and A. Zisserman, 2003. Object class recognition by unsupervised scale-invariant learning. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2, June 18-20, 2003, IEEE Press, pp: II-264-II-271.

Fergus, R., P. Perona and A. Zisserman, 2005. A sparse object category model for efficient learning and exhaustive recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 20-25, 2005, IEEE Press, pp: 380-387.

Ferrari, V., L. Fevrier, F. Jurie and C. Schmid, 2008. Groups of adjacent contour segments for object detection. IEEE Trans. Pattern Anal. Mach. Intell., 30: 36-51.

Ferrari, V., T. Tuytelaars and L. Van Gool, 2006. Object detection by contour segment networks. Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, Volume 3953, May 7-13, 2006, Springer, Berlin/Heidelberg, pp: 14-28.

Grauman, K. and T. Darrell, 2007. The pyramid match kernel: Efficient learning with sets of features. J. Mach. Learn. Res., 8: 725-760.

Gu, C., J.J. Lim, P. Arbelaez and J. Malik, 2009. Recognition using regions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, IEEE Press, Miami, FL., pp: 1030-1037.

Lan, Y., H. Ren, Y. Zhang, H. Yu and G. Wang, 2011. A novel object representation model for object detection. Proceedings of the 4th International Workshop on Advanced Computational Intelligence, October 19-21, 2011, IEEE Press, Wuhan, pp: 59-63.

Lee, K.M., 2008. Component-based face detection and verification. Pattern Recognition Lett., 29: 200-214.

Leibe, B., A. Leonardis and B. Schiele, 2008. Robust object detection with interleaved categorization and segmentation. Int. J. Comput. Vision, 77: 259-289.

Leibe, B., E. Seemann and B. Schiele, 2005. Pedestrian detection in crowded scenes. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 20-25, 2005, IEEE Press, pp: 878-885.

Leordeanu, M., M. Hebert and R. Sukthankar, 2007. Beyond local appearance: Category recognition from pairwise interactions of simple features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, IEEE Press, Minneapolis, MN., pp: 1-8.

Matthews, I. and S. Baker, 2004. Active appearance models revisited. Int. J. Comput. Vision, 60: 135-164.

Opelt, A., A. Pinz and A. Zisserman, 2006. A boundary-fragment-model for object detection. Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Springer, Berlin/Heidelberg, pp: 575-588.

Schneiderman, H. and T. Kanade, 2004. Object detection using the s t a t istics of parts. Int. J. Comput. Vision, 56: 151-177.

Shotton, J., A. Blake and R. Cipolla, 2005. Contour-based learning for object detection. Proceedings of the 10th IEEE International Conference on Computer Vision, Volume 1, October 17-21, 2005, IEEE Press, pp: 503-510.

Sudderth, E.B., A. Torralba, W.T. Freeman and A.S. Willsky, 2005. Learning hierarchical models of scenes, objects and parts. Proceedings of the 10th IEEE International Conference on Computer Vision, Volume 2, October 17-21, 2005, IEEE Press, Beijing, pp: 1331-1338.

Sullivan, J., O. Danielsson and S. Carlsson, 2008. Exploiting part-based models and edge boundaries for object detection. Proceedings of the Computing: Techniques and Applications, Digital Image, December 1-3, 2008, IEEE Press,Canberra, ACT., pp: 199-206.

Wang, G. and C. Chen, 2010. Using non-symmetry and anti-packing representation model for object detection. Proceedings of the International Conference on Computational Intelligence and Software Engineering, December 10-12, 2010, IEEE Press, Wuhan, pp: 1-4.

Zhang, H., W. Gao, X. Chen and D. Zhao, 2006. Object detection using spatial histogram features. Image Vision Comput., 24: 327-341.

Zhu, L., Y. Chen, A. Yuille and W. Freeman, 2010. Latent hierarchical structural learning for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, IEEE Press, San Francisco, CA., pp: 1062-1069.