VOL. 7, NO. 9, SEPTEMBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

EFFICIENT OBJECT DETECTION AND CLASSIFICATION USING MODIFIED ELM

Balaji N. V.
E-Mail: nvbresearch@gmail.com

Dr. M. Punithavalli
Department of Computer Science,
Sri Ramakrishna Arts College for Women,
Coimbatore, India.

ABSTRACT

Object detection is one of the typical difficulties in computer technology which has its usage to surveillance, robotics, multimedia processing, and HCI. The multi-resolution framework is utilized by the proposed technique for object detection. In this efficient object detection, the lower resolution features are first used to discard the majority of negative windows at comparatively small cost, leaving a relatively small amount of windows to be processed in higher resolutions and this helps to attain better detection accuracy. Then the frameworks on Histograms of Oriented Gradient (HOG) features are used to detect the objects. For training and detection, the classifier used previously is Support Vector Machine (SVM) and Extreme Learning Machine (ELM). Modified ELM is used in the proposed technique to reduce the time for detection and improve the accuracy of classification. The input weights and hidden biases are created with the help of integrated Analytic Network Process (ANP) and Bayesian Network (BN) model. The experimental result shows that the proposed technique achieves better detection rate when compared to the existing techniques.

Keywords: Object Detection, Histograms of Oriented Gradient (HOG), Modified Extreme Learning Machine (ELM), Multi-resolution Framework, Analytic Network Process (ANP), Bayesian Network (BN).

INTRODUCTION

Object Detection [6] and Recognition in noisy and cluttered images is challenging problem in computer vision. The goal of this research is to identify objects in an image accurately. Today, there is an increased need for object detection. There are several problems in detecting and recognizing the objects in an image. It is an important part in many applications such as image search, image auto-annotation and scene understanding; however it is still an open difficulty due to the complexity of object classes and images. [8] The robot usage will reduce the work load for the housekeepers. Some of the works performed by home robots are cleaning, tidying, fetching objects, etc. These tasks need some degree of semantic knowledge about human surroundings so that they can find the way, search for objects and communicate with humans successfully. For example, for a robot to execute the command "bring me an apple", a robot must recognize the term "apple" and to know which places in the surroundings are expected to contain it (e.g., the kitchen).

This paper provides the better object detection technique. Initially, multi-resolution framework is used to eliminate the unnecessary features from the image which will help in speedy processing.[9] The learning machine used in this work is Modified ELM [7]. Finally, Histograms of Oriented Gradient features are applied for accurate detection of object in the image.

RELATED WORKSc

Broussard *et al*., [1] put forth the physiologically motivated image fusion for object detection using a pulse coupled neural network. PCNN are implemented to combine the results of several object detection techniques to improve object detection accuracy. The object detection assets of the obtained image fusion network are demonstrated on mammograms and Forward-Looking Infrared Radar (FLIR) images. This technique exceeded the accuracy obtained by any individual filtering methods or by logical ANDing the individual objects detection technique results.

Kerekes *et al*., [2] had a look on spectral imaging technique analytical model for sub-pixel object detection [3]. To learn the system parameters in the context of land cover classification, an end-to-end remote sensing system modeling approach was previously developed. In this work, the author extends this technique to sub-pixel object detection applications by including a linear mixing model for an unresolved object in a background and using object detection algorithms and Probability of Detection (P_D) versus False Alarm (P_{FA}) curves to characterize performance.

Kubo *et al*., [4] developed an image processing system for direction detection of an object using neural network. The proposed direction detection system was considered to consist of two modules of neural networks; one is an edge detection module and the other is a direction detection module. The edge detection technique

VOL. 7, NO. 9, SEPTEMBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

could detect edges when differences of gray level between the object and the background were larger than thirteen at any gray level of the object.

Wen-jie Wang *et al*., [5] illustrate the Object-oriented multi-feature fusion change detection method for high resolution remote sensing image [10]. The key steps of this technique are segment the image, choose optimized features from spectral features, texture features, shape features of the segment objects, use optimized features to do change detection, make fusion of the preliminary change detection results of the different optimized features to get the final result.

METHODOLOGY

The initial process for the proposed object detection technique is the multi-resolution framework which is applied in different resolution space. The next process is the application of training and detection technique with the help of Modified ELM. Finally, Histogram of Oriented Gradients (HOG) features are applied for detection of objects.

Multiple Resolution Framework

The common approach of object detection using multi-resolution is shown in Figure-1. The framework encodes resolution together with scale spaces in 2D coordinate scheme. Along the vertical axis in scale space, the image gets down sampled in order to use a fixed size detection window to locate objects of larger scales. Along the horizontal axis in resolution space, any detection window is classified hierarchically from its lowest resolution to full resolution. A window rejected in lower resolution will not be passed to any higher resolutions.

Figure-2 shows why multi-resolution is useful for object detection. A Canny edge detector was applied to images from the VOC 2006 at different resolutions. Images in lower resolutions were created by down sampling and then up sampling to the original size for easier comparison. For the motorbikes and cows categories, two samples are picked from each and computed the shape context matching distances shown in the Figure 2.

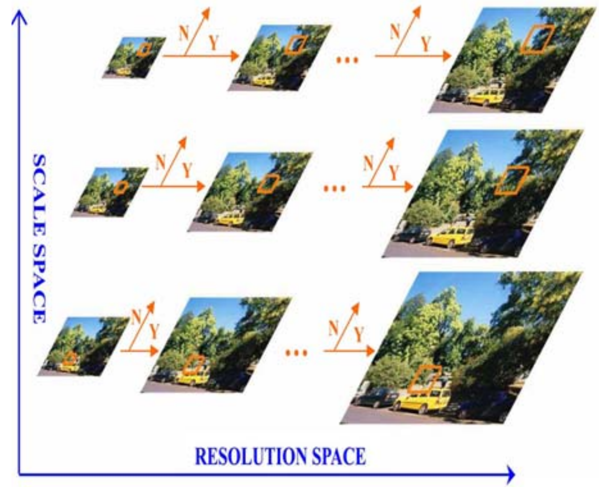

Figure-1. Multiple Resolution Framework for Object Detection.

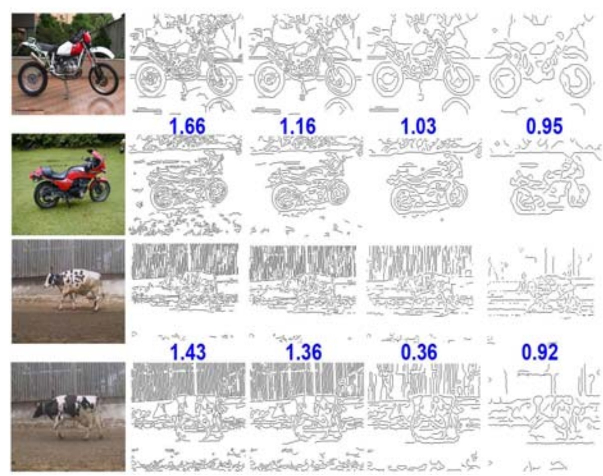

Figure-2. Sample Images from VOC 2006 Database, with Edge Detected Images at Different Resolutions.

Training/Detection Algorithms

For a typical object detection system, the training set is composed of normalized image patches. Negative patches were scanned from images that do not contain any target object instances. Suppose R resolutions are used, where r = 1, ..., R from the lowest resolution to the full resolution. Assume the step of down sampling ratio is α. For each resolution r, the training set is defined as Tr = {(Ir(i), lr(i)), i = 1, ..., Nr}, where Ir(i) is the image patch, lr(i) ϵ {1,-1} is the associated label and Nr is the number of patches. The training starts from the lowest resolution, and then loops between bootstrap and training towards higher resolutions until the full resolution [11]. The output of the training is a hierarchical classifier with components from each resolution [12]. Suppose that the image is down sampled with a ratio of β for each scale up, and that it have S scales s = 1, ..., S, for larger to smaller objects. Suppose the window size at each resolution is (wr, hr) = (w, h)/αR-r, and that the size is fixed across all scales. For each

VOL. 7, NO. 9, SEPTEMBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

resolution r and scale s, a list of status of detection windows γr, s (k) ϵ {1,-1} is kept. The status is initialized to be 1, and set to -1 if the window is rejected by a classifier.

The training and detection is further improved by Modified Extreme Learning Machine (ELM). Previous approaches have used SVM and ELM for the purpose of training and detection.

SVM Algorithm

SVM is used to find an optimal separating hyper plane (OSH) which generates a maximum margin between the categories of data. To build an SVM classifier, a kernel function and its parameters need to be chosen. So far, no analytical or empirical studies have established the superiority of one kernel over another conclusively [13]. The following three kernel functions have been applied to build SVM classifiers:

a) Linear kernel function, $K(x, z) = \langle x, z \rangle$
b) Polynomial kernel function $K(x, z) = (\langle x, z \rangle + 1)^d$ d is the degree of polynomial
c) Radial basis function $K(x,z) = \exp\left\{-\frac{\|x-z\|^2}{2\sigma^2}\right\}, \sigma$ is the width of the function

The usage of SVM will reduce the time requirement for classification and also the accuracy of classification but it can be improved using ELM.

Extreme Learning Machine

Extreme Learning Machine (ELM) meant for Single Hidden Layer Feed-Forward Neural Networks (SLFNs) that will randomly select the input weights and analytically determines the output weights of SLFNs. The ELM has several interesting and significant features different from traditional popular learning algorithms for feed forward neural networks [14].

Extreme Learning Machine Training Algorithm

If there are N samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2} \ldots x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^n$, then the standard SLFN with N hidden neurons and activation function g(x) is defined as:

$$\sum_{i=1}^{\tilde{N}} \beta_i\, g(w_i . x_j + b_i) = \mathbf{0}_j, j = 1, \ldots\ldots, N.,$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ represents the weight vector that links the ith hidden neuron and the input neurons, $ß_i = [ß_{i1}, ß^{i2}, \ldots, ß^{im}]^T$ represents weight vector that links the ith neuron and the output neurons, and b_i represents the threshold of the ith hidden neuron. The "." in $w_i . x_j$ indicates the inner product of w_i and x_j. The SLFN try to reduce the difference between o_j and t_j.

More in a matrix format as H ß = T, where

$$H(a_1, \ldots, a_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}, x_1, \ldots, x_N) = \begin{bmatrix} g(w_1 . x_1 + b_1) & \cdots & g(w_{\tilde{N}} . x_{\tilde{N}} + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 . x_1 + b_1) & \cdots & g(w_{\tilde{N}} . x_{\tilde{N}} + b_{\tilde{N}}) \end{bmatrix}_{N * \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} * m} \text{ and } \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}_{N * m}$$

The result reduces the norm of this least squares equation is:

$$\hat{\beta} = H^+ T$$

where $H^\dagger$ is known as Moore-Penrose generalized inverse.

The usage of Modified ELM will reduce the time requirement for classification and also the accuracy of classification can be improved over the previous techniques [15].

Modified Extreme Learning Machine

A modified ELM technique which uses ELM and LM technique can be described as below:

Initially, the input weights and hidden biases are created with the help of integrated Analytic Network Process (ANP) and Bayesian Network (BN) model. In the Analytic Hierarchy Process (AHP), each element in AHP is considered to be independent of all the others and hence the decision condition are considered to be independent of one another, and the alternatives are considered to be independent of the decision criteria and of each other. But in our case, there is interdependence between the items and the alternatives. ANP does not necessitate independence among elements, so it can be exploited as an effective tool.

By integrating ANP and Bayesian Network in such a way, that the weights, the output of ANP will be given as input to BN. Bayesian Network can obtain inputs from several points, but in the combination BN adopts the weights, as inputs in leafs of the graph; and these weights are structured in vector format. The serial of comparative weights of computed by ANP method can be written in a vector format, this vector will be given as input to BN.

Next, the corresponding output weights are analytically determined with the help of ELM algorithm during the first iteration and randomly generate the output hidden biases. Then, the parameters (all weights and biases) are restructured with the help of LM algorithm [16].

The process for the Hybrid Extreme Learning Machine is described below:

Provided a training set $\aleph = \{x_{\downarrow}i, t_{\downarrow}i\} | x_{\downarrow}i \in R^{\uparrow}n, t_{\downarrow}i \in R^{\uparrow}m, i = \vdash 1, 2, \ldots, N\}$ activation functions $f_1(x) \& f_2(x)$, and hidden nodes namely $\tilde{N} \& K$ of first and second hidden layer.

Step-1: Randomly choose the starting values of input weight vectors w_1 and bias vector b_1 with the help of

VOL. 7, NO. 9, SEPTEMBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ANP-BN technique and bias vector b_2 without using the ANP-BN technique.

Step-2: Determine the first hidden layer output matrix a_1. With the help of ELM algorithm, determine the output weight
$w_2 = a_1^{-1}.t$

Step-3: Determine the second hidden layer output matrix a_2, errors $e_1 = t - a_2$ and determine the sum of squared errors for overall input.

Step-4: Determine the Jacobian matrix. Calculate the sensitivities with the recurrence relations.

$$s_q^m = f^m(n_q^m)(w^{m+1})^T.s_q^{m+1}$$

after initializing with the following equation

$$s_q^M = -f^m(n_q^m)$$

Augment the individual matrices into the Marquardt sensitivities using the following equation:

$$s^m = [s_1^m, s_2^m, \ldots, s_Q^m]$$

Determine the elements of the Jacobian matrix with the equations

$$[J]_{h,l} = s_{i,h}^m \times s_{j,k}^{m-1} \quad \text{and} \quad [J]_{h,l} = s_{i,h}^m$$

Application to HOG

In Histogram of Oriented Gradients (HOG) feature for pedestrian detection, pixels are first grouped into smaller spatial units called cells. For each cell, a histogram feature on gradients orientations is extracted. The magnitude of the gradient is applied as the weight for voting into the histogram. The descriptor of each block is the concatenation of all cell features. Inside each detection window, densely sampled and overlapping blocks produce redundant descriptors, which is important for better performance. Finally, a linear Support Vector Machine (SVM) is used to classify individual detection windows. In the proposed multi-resolution approach based on the above method, the author used blocks of varying sizes to capture features in different spatial frequencies. An important consequence of the feature hierarchy is that, both more-global (low resolution) and more-local (high resolution) features are encoded.

EXPERIMENTAL RESULTS

The proposed technique for object detection is experimented on PASCAL Visual Object Classes challenge 2006 database (VOC 2006). There are 10 groups of objects are presented in the database such as bicycles, buses, cats, cars, cows, dogs, horses, motorbikes, people and sheep. The multi-resolution approach is applied for the sheep and car categories. The proposed technique obtained the best results among all participants on these two categories.

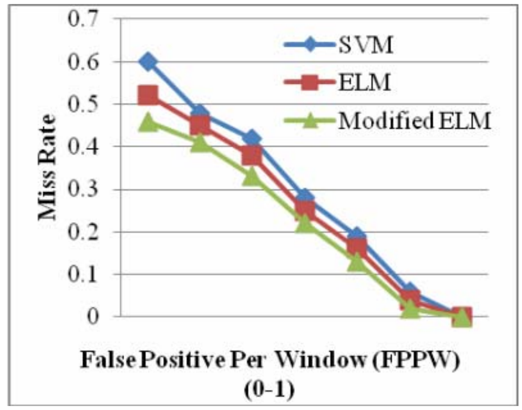

Figure-3. Comparison between proposed and existing methods using detection error trade-off curves on sheep category in voc2006 database.

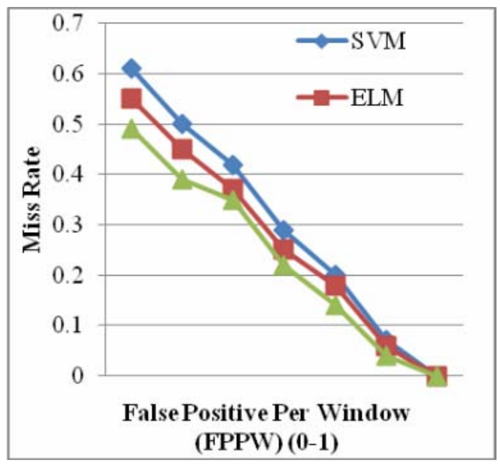

Figure-4. Comparison between proposed and existing methods using detection error trade-off curves on car category in VOC2006 database.

Figure-3 shows the comparison of proposed method with existing object detection method using DET curves in case of sheep category. From the graph it can be observed that the resulted miss rate for the proposed system with Modified ELM is lesser when compared to the existing system with SVM and ELM. This shows that the modified ELM gives better classification result for the proposed technique. Figure-4 shows the comparison of proposed method with existing object detection method using DET curves in case of car category. From the Figure, it can be clearly seen that the miss rate is lesser for the proposed technique for all levels of false positive per window.

CONCLUSIONS

Object Detection from the image is an important technique applied in many fields such as image browsing, robotics, etc. The existing techniques lack accuracy and take more time detection. To overcome these drawbacks, the multi-resolution based detection system is proposed in this paper. The first step is to discard the negative windows using multi-resolution framework, the next step is learning and detection in which the training samples are supplied for training the system and the detection is performed accordingly. To improve the accuracy of classification, Modified ELM is applied. The input weights and hidden biases are created with the help of integrated Analytic Network Process (ANP) and Bayesian Network (BN) model. Finally, with the help of Histogram of Oriented Gradients (HOG) features, the object is detected from the image. The VOC 2006 database is used for the purpose of experimentation and result shows that this technique yields better accuracy for detection and also miss rate for the proposed technique is lesser when compared to the conventional techniques.

REFERENCES

[1] R.P. Broussard, R.P, S.K. Rogers, M.E. Oxley and G.L. Tarr. 1999. Physiologically motivated image fusion for object detection using a pulse coupled neural network. IEEE Transactions on Neural Networks. 10: 554-563.

[2] J.P Kerekes and J.E Baum. 2002. Spectral imaging system analytical model for subpixel object detection. IEEE Transactions on Geoscience and Remote Sensing. 40: 1088-1101.

[3] http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2006/index.html.

[4] T. Kubo, M. Obuchi, G. Ohashi and Y. Shimodaira. 1998. Image processing system for direction detection of an object using neural network. The 1998 IEEE Asia-Pacific Conference on Circuits and Systems, IEEE APCCAS 1998. pp. 571- 574.

[5] Wen-jie Wang, Zhong-ming Zhao and Hai-qing Zhu. 2009. Object-oriented multi-feature fusion change detection method for high resolution remote sensing image. 17th International Conference on Geoinformatics. pp. 1-6.

[6] Jiqiang Song, Min Cai and M. S Lyu. 2002. Edge color distribution transform: an efficient tool for object detection in images. Conference on Pattern Recognition, Proceedings of 16th International conference. 1: 608-61.

[7] Saad Ali and Mubarak Shah. 2005. A supervised learning framework for generic object detection in images. Tenth IEEE International Conference on Computer Vision, ICCV 2005. 2: 1347-1354.

[8] S. Sheraizin and S. Itzikowitz. 2009. Unmanned object detection for image surveillance systems. IEEE conference on EUROCON'09. pp. 302-307.

[9] A. Steel and D. Brunner. 2008. Detection and Characterization of Urban Objects from VHR Optical Image Data. IEEE International conference on Geoscience and Remote Sensing Symposium, IGARSS 2008. 3: III-1256-III-1259.

[10] S. Belongie, J. Malik and J. Puzicha. 2002. Shape matching and object recognition using shape contexts. PAMI. 24(4): 509-522.

[11] S. C. Brubaker, M. D. Mullin and J. M. Rehg. Towards optimal training of cascaded detectors. In ECCV06. I: 325-337.

[12] J. Fasola and M. Veloso. Real-time object detection using segmented and grayscale images. In ICRA06. pp. 4088-4093.

[13] B. Leibe, E. Seemann and B. Schiele. Pedestrian detection in crowded scenes. In CVPR05. pp. I: 878-885.

[14] S. Mitri, A. Nuchter, K. Pervolz and H. Surmann. Robust object detection in regions of interest with an application in ball recognition. In ICRA05. pp. 125-130.

[15] P. Viola, M. J. Jones and D. Snow. 2005. Detecting pedestrians using patterns of motion and appearance. IJCV. 63(2): 153-161.

[16] V. Ferrari, T. Tuytelaars and L. J. V. Gool. Object detection by contour segment networks. In ECCV06. pp. 14-28.