International Journal of Image Processing and Vision Science

Volume 1 | Issue 3 Article 3

January 2013

SCENE UNDERSTANDING USING BACK PROPAGATION BY NEURAL NETWORK

ARTI TIWARI
Department of Elex & Telecomm. Engg.Chouksey Engg. College,Bilaspur, tiwaryaarti@yahoo.com

JAGVIR VERMA
Department of Elex & Telecomm. Engg.Chouksey Engg. College,Bilaspur, jagvirverma@rediffmail.com

Recommended Citation
TIWARI, ARTI and VERMA, JAGVIR (2013) "SCENE UNDERSTANDING USING BACK PROPAGATION BY NEURAL NETWORK," *International Journal of Image Processing and Vision Science*: Vol. 1 : Iss. 3 , Article 3.
DOI: 10.47893/IJIPVS.2013.1030
Available at: https://www.interscience.in/ijipvs/vol1/iss3/3

SCENE UNDERSTANDING USING BACK PROPAGATION BY NEURAL NETWORK

ARTI TIWARI[1] & JAGVIR VERMA[2]

[1,2]Department of Elex & Telecomm. Engg.Chouksey Engg. College,Bilaspur
Email:tiwaryaarti@yahoo.com, jagvirverma@rediffmail.com

Abstract: In this paper, we proposed an efficient method to address the problem of scene understanding that is based on neural network (NN) and image segmentation. We utilized a multilayer perceptron (MLP) to train the network and features are extracted using pixels in the RGB color space. In this work, object samples in images with varying lighting conditions are used to obtain a wide object color distribution. The training data is generated from positive and negative training patterns in the color planes. Subsequently, training set is fed to an MLP, trained by the back propagation algorithm using these object samples. We apply the above mentioned NN-based object classifier to the test image which is applied to image segmentation and corresponding to the pixel level of the object in test image particular object is determined.

Keywords*: Neural network, Multi layer perceptron, Red green blue color space.*

I. INTRODUCTION

Scene understanding in complex environment is a challenging problem which has a fundamental impact to the object-based video coding, content-based image retrieval, object tracking, and face recognition systems. Real-time object detection is an important and preliminary step to a variety of applications requiring intelligent human-computer interaction. Since there can be number of object present in an image and a particular object can be of various type, so the main challenge in a scene understanding (object detection) process is the amount of variation in visual appearance (such as size, color, shape, surrounding condition, lighting condition, shadows, pose, position, and orientation). A general statement of the object detection problem can be described as the determination of the location and size of the existing objects in images.

II. RELATED WORKS

The detection of several objects in images and video sequences has been regarded as a challenging problem in the field of image processing; hence, many approaches have been proposed to accomplish this task. The multi-module neural network (MMNN) [1] is a hierarchical network with cascade connections, and consists of several modules which can detect specific features. This system consists of several modules in parallel which are trained to respond selectively to human face components: the eyes, the nose, and the mouth. At last, the face area is detected by integrating the outputs of previous cell layer. The dominant factor in the running time of the Rowley system is the number of 20 x 20 pixel windows which the NN must process. Viola and Jones [2] proposed a face detection system based on a multi-classifier cascade and Ada-Boosted perceptron. They used a set of rectangular features to represent face objects which are captured by the Harr basis functions. The task of face detection is one of the general object detection problems; by the same token, Schneiderman and Kanade[3] proposed a complex statistical classifier to detect objects. They used a Bayesian method for detection purposes which represents the statistics of both face and non-face appearances taking advantage of a product of histograms that is used in respect to the joint statistics of a subset of wavelet coefficients capturing local features in space, frequency, orientation, and the position of the aforementioned appearances. Neural networks have also been extensively used for pattern recognition problems, including face detection.
Rowley et al. [4] proposed a connected NN which incorporates face knowledge. The NN is designed to look at 20 x 20 pixel windows. One hidden layer with 26 units looks at different regions based on facial feature knowledge. The dominant factor in the running time of the Rowley system is the number of 20 x 20 pixel windows which the NN must process. Recently, a set of orthogonal, noise robust transformation invariants, distribution sensitive moments called as Eigen moments has been proposed by [5]. The invariance properties of radial Tchebichef of moments and their application pattern recognition tasks have been investigated by R.Mukundan [6].

III. OUR APPROACH

Our main contribution consists to introduce a Neural Network to find an optimal combination of this approach. Tests carried on real images show that the neural network improves performance compared to an empirical combination function and to each method used separately. To detect the name

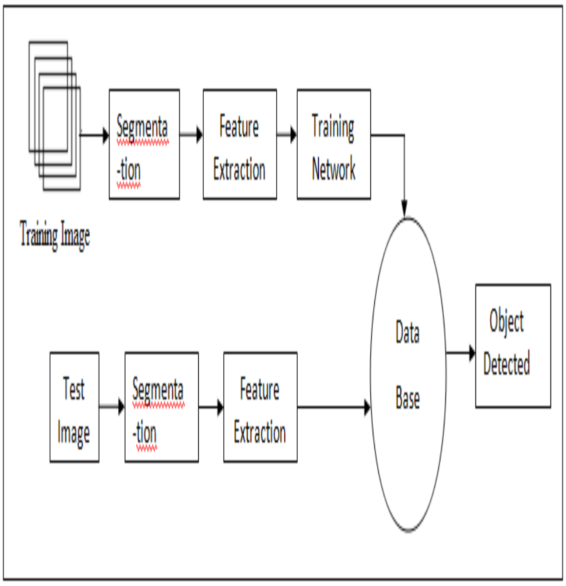

Figure1: Proposed Methodology

of object present in a scene, first we have to train our neural network using any target image by any type of learning available. In our work we will use supervised learning method. After training of the network the name of the object present in the test image can be detected (for example suppose the network is trained for a picture of a computer, than it can easily detect the computer if it is present in the test image).

The Steps to be followed are as follows:
1. Image segmentation.
2. Feature extraction
3. Training of Network
4. Matching with previous data base available.

3.1 Image segmentation:

In computer vision, Segmentation is the process of partitioning an image into multiple segments (sets of pixels, also known as super pixels) [8]. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image Segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Our algorithm is initialized by the clustering method [9, 10]. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire scene, or a set of contours extracted from the scene. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).

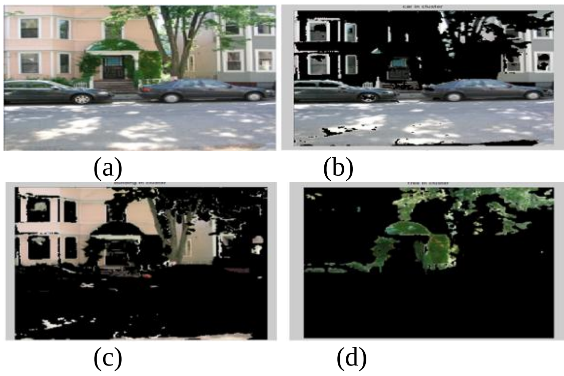

Figure 2 Image segmentation
(a) Original image,
(b) (c) (d) Segmented image

3.3 Feature Extraction Algorithm

Feature is defined as a function of one or more measurements, each of which specifies some quantifiable property of an object, and is computed such that it quantifies some significant characteristics of the object. Feature Extraction Algorithms attempt to reduce a large-dimensionality feature vector into a smaller-dimensionality vector that is easier to work with and encodes less redundancy, using mathematical techniques. Feature selection algorithms, attempt to directly prune out redundant or irrelevant features. The distinction between the two is that the resulting features after feature extraction has taken place are of a different sort than the original features and may not easily be interpretable, while the features left after feature selection are simply a subset of the original features [14, 15]. We classify the various features currently employed as follows:

1. General features: Application independent features such as color, texture, and shape. According to the abstraction level, they can be further divided into:

(a)Pixel-level features: Features calculated at each pixel, e.g. color, location.

(b) Local features: Features calculated over the results of subdivision of the image band on image segmentation or edge detection.

(c) Global features: Features calculated over the entire image or just regular sub-area of an image.

2. **Domain-specific features**: Application dependent features such as human faces, fingerprints, and conceptual features.

These features are often a synthesis of low-level features for a specific domain. On the other hand, all features can be coarsely classified into low-level features and high-level features [16]. Low-level features can be extracted directed from the original images, whereas high-level feature extraction must be based on low-level. As shown in figure some images are taken and their feature is extracted using the property of their shape, position and color. Finally this extracted portion is matched with the previous database stored and the result is obtained when any matching is available.

Figure 3: Feature extraction

3.2 Training of neural network:

In Our Method we have used the back propagation learning algorithm to train our neural network [11, 12]. A feed forward neural network classifier is employed to train on extracted feature vector for each image by back propagation algorithm. The network is optimized for give input data sets [13], for better understanding, the back propagation learning algorithm can be divided into two phases: propagation and weight update.

<u>Phase 1</u>: Propagation - Each propagation involves the following steps:

1 Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.

2 Backward propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

<u>Phase 2</u>: Weight update - For each weight synapse follow the following steps:

1 Multiply its output delta and input activation to get the gradient of the weight.

2 Bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight.

This ratio influences the speed and quality of learning; it is called the learning rate. The sign of the gradient of a weight indicates where the error is increasing; this is why the weight must be updated in the opposite direction.

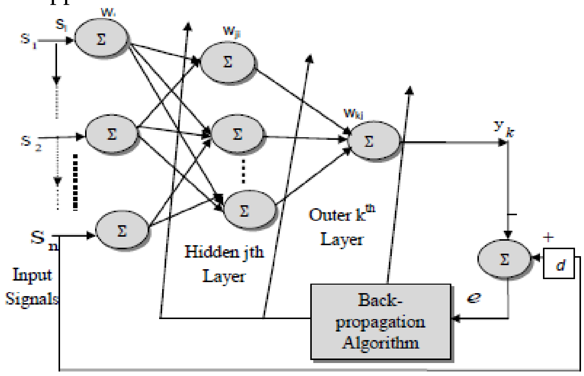

Figure4: Training of Neural Network

3.4 Matching with Database

Finally after the feature extraction this test image is matched with the entire database available in the system. If any matching is obtained between the test image and any of the training images, then the result is reflected according to the training image.The NNs are trained using back Propagation algorithm which is slow in working but its accuracy is good. Its speed can be improved better by using segmentation as we have done in our work.each time when a new image with a new scene will come then we have to perform all operation in that image to detect object.

Figure 5: Qualitative results from our experiments. Top row shows original image, annotated regions and objects, region boundaries, and predicted horizon. Other examples show original image (left) and overlay colored by semantic class and detected objects (right).

IV. DISCUSSION

In this paper we have presented a hierarchical model for joint object detection and image segmentation for scene understanding. Our novel approach overcomes many of the problems associated with trying to combine related vision tasks. Importantly, our method explains every pixel in the image and enforces consistency between random variables from different tasks. Furthermore, our model is encapsulated in a modular energy function which can be easily analyzed and improved as new computer vision technologies become available. One of the difficulties in our model is learning the trade-off between energy terms—too strong a boundary penalty and all regions will be merged together, while too weak a penalty and the scene will be split into too many segments. The Matlab [17] software is used for the experiments on Pentium-IV, 1.76 GHz, 2 MB Cache, 1 GB RAM machine. The colored images for simulation purpose are downloaded from standard dataset [18]. Our work suggests a number of interesting directions for future work. First, our procedure is based on learning of a neural network so it can be replaced with a new object based on a specific application that makes ore global steps. More importantly, our region-based model has the potential for providing holistic unified understanding of an entire scene. This has the benefit of eliminating many of the implausible hypotheses that plague current computer vision algorithms but with a limitation that its data base must be highly strong. Our work can be enhanced by using OHTA technique in future for more color analysis. If back propagation will be replaced by optical back propagation algorithm than may be it will be improved more. Furthermore, by clearly delineating what is recognized, our framework directly present hypotheses for objects that are currently unknown providing the potential for increasing our library of characterized objects using a combination of supervised techniques with image segmentation.

REFERENCES

[1] Tetsuya Taannai and Masafumih Agiwara, **"**A Visual Nervous System Based Multi-Module Neural Network for Object Recognition," in proc of IEEE conference, 1998.

[2] P. Viola and M. Jones, "Robust real-time object detection," in 2nd International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, Vancouver, 2001.

[3] H. Schneiderman and T. Kanade, "A statistical model for 3d object detection system applied to faces and cars," in Proc. of IEEE Conf Computor Vision and Pattern Recognition, 2002.

[4] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20(1), pp 23-28, 1998.

[5] Yap.P.T, Raveendran.P, "Eigen moments, Pattern Recognition," vol.40, pp 1234-1244, 2007.

[6] Mukundan.R, "Radial Tchebichef Invariants for Pattern Recognition," IEEE Conference on image processing, TENCON-2005.

[7] Bardia Mohabbati and Shohreh Kasaei "An Efficient Wavelet/Neural Networks-Based Face Detection Algorithm," 2005 IEEE.

[8] Deva Ramanan "Learning to parse images of articulated bodies," pp 40-45, 2008.

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," Int. J. Computer Vision, 61(1), January 2005.

[10] R. Ronfard, C. Schmid, and B. Triggs, "Learning to parse picture of people," in ECCV, 2002.

[11] Haykin, Simon, "*Neural Networks*, Pearson Education Inc., 2004.

[12] Azzabou N. and Likforman Sulem, "Neural network-based proper names extraction in fax images**,"** Pattern Recognition, Proceedings of the 17th International Conference on 421 - 424 vol.1, 23-26 Aug, ICPR 2004. .

[13] Wenming Cao, Feng hao, Shoujue Wang, "The Application of RBF Neural Networks for Object Recognition," Information Sciences 55, pp.153-160, 2004.

[14]S.Belongie, J.Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Analysis and Machine Intell. 24 (4), pp 509-522, April 2002.

[15] A. Thayananthan, and Stenger, B. and Torr, P.H.S. and Cipolla R., "Shape context and chamfer matching in cluttered scenes," CVPR03. 2003.

[16] Amitabh Wahi, F Mohamed Athiq, C Palanisamy**, "**A hybrid feature extraction method-based object recognition by neural network**,"** in 2008 International Conference on Computing Communication and Networking 2008.

[17]. H. Demuth and H. Mark eds. 2002, "The Matlab version 6.5: User Guide," The Math Works Inc., USA.

[18]. http://staff.science.uva.nl
.

◈◈◈