Object Detection and Semantic Segmentation using Neural Networks

R.Karthika[#1], S. N Santhalakshmi[*2]

[1]Research Scholar, [2]Assistant Professor, Department of Computer Science, Nandha Arts and Science College, Perundurai, Erode, India

***Abstract* — *Semantic segmentation and object detection are two most common tasks in the field of digital image processing, classification and segmentation. The object detection in repetition domain will be approached to segment objects from foreground with absence of background noise. This work has introduced one automatically detecting an object to increase the accuracy and yield and decrease the diagnosis time. This proposed method represents image Segmentation and Object Detection using NN classifier. The first step for input image segmentation and feature extracted from segmented image using NN classifier. The goal of Classification is to find Object from input ones.*

At the end it is shown the object detected image. The best results can be achieved by this proposed image segmentation and classification image.

Keywords — *Thresholding, GLSM , Probabilistic Neural Networks, Threshold, eigen, Palmprint, vector clustering, kernel tric, semantic segmentation, Down sampling, neural networks, Perceptron, Discrete wavelet, Modeling, simulation, and prototyping, vectors.*

I. Introduction

Identifying particular portion in an image would probably start with image processing techniques such as removing noise, extraction to find lines, regions and areas with certain textures.In a digital image processing, first step is to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. An image may have 512 × 512 or 250,000 pixels, although much larger images are becoming common. There are three basic operations that can be performed on it in the computer for digital images. For a location operation, a pixel value in the output image depends on a single pixel value in the input image. For local operations, several neighboring pixels in the input image determine the value of an output image pixel. In a universal operation, all of the input image pixels contribute to an output image pixel value. These operations, taken separately or in combination, are the means by which the image is enhanced, restored, or compressed.

Recognizing object classes in real-world images is a long standing goal in Computer vision. Conceptually, this is challenging due to large appearance variations of object instances belonging to the same class. Additionally, distinguished from background clutter, scale, and viewpoint variations can render appearances of even the same object instance to be vastly different. Further challenges arise from interclass similarity in which instances from different classes can appear very similar. Consequently, models for object classes must be flexible enough to accommodate class variability, yet discriminative enough to sieve out true object instances in cluttered images. These seemingly paradoxical requirements of an object class model make recognition difficult. This paper addresses two goals of recognition are image classification and object detection.

The task of image classification is to determine if an object class is present in an image, while object detection localizes all instances of that class from an image. Toward these goals, the main contribution in this paper is an approach for object class recognition that employs edge information only. The novelty of our approach is that represent contours by very simple and generic shape primitives of line segments and ellipses, coupled with a flexible method to learn discriminative primitive combinations. These natives are complementary in nature, where line segment models straight curve and ellipse models curved contour. Let's choose an ellipse as it is one of the simplest circular shapes, yet is sufficiently flexible to model curved shapes. These shape primitives possess several attractive properties. First, unlike edge-based descriptors they support abstract and perceptually meaningful reasoning like parallelism and adjacency. Also, unlike contour fragment features, storage demands by these primitives are independent of object size and are efficiently represented with four parameters for a line and five parameters for an ellipse.

The images can be divided into three types:

1.Binary image
2. Gray scale
3. Color image

A. Binary images:

An image is a collection of pixels. Each binary image can have a two values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit (0 or 1).This name black and white, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel, such as grayscale images.

B. Gray scale images:

Gray scale images are always different from theBalck and white images , so it is called as bi-level or binary. Gray scale images are based on the shades of gray color. The another name of gray scale images is monochromatic.The gray scale images are produced from the intensity of each pixel.

C. Color images:

Color images in a digital can consists of information about the colors at each pixel.Each pixel value depends upon the color it appears. A pixel is a combination of red, green and blue.For example: If the valus of R,G,B is 255, it looks like a white.

D. IMAGE SEGMENTATION

Segmentation problems are the bottleneck to achieve object extraction, object specific measurements, and fast object rendering from multi-dimensional image data. Simple segmentation techniques are based on local pixel-neighborhood classification. Such methods fail however to “see” global objects rather than local appearances and require often intensive operator assistance. The reason is that the “logic” of a object does not necessarily follow that of its local image representation. Local properties, such as textures, edginess, ridgeness etc. do not always represent connected features of a given object.

E. EXISTING SYSTEM

- Threshold segmentation
- PCA
- Support Vector Machine (SVM)
- Fourier transform

THRESHOLDING

The simple method of image segmentation is called the thresholding method. This method is based on a clip-level to turn a gray-scale image into a binary image. Threshold method is used to the find values when multiple-levels are selected. Several popular methods are used in industry including the maximum entropy method, Otsu's method (maximum variance), and k-means clustering. There are lot of recent development methods for thresholding computed tomography (CT) images. The key idea is that, unlike Otsu's method, the thresholds are derived from the radiographs instead of the (reconstructed) image.

Design Steps

(1) Set the initial threshold T= (the maximum value of the image brightness + the minimum value of the image brightness)/2.

(2) Using T segment the image to get two sets of pixels B (all the pixel values are less than T) and N (all the pixel values are greater than T);

(3) Calculate the average value of B and N separately, mean ub and un.

(4) Calculate the new threshold: T= (ub+un)/2

(5) Repeat Second steps to fourth steps upto iterative conditions are met and get necessary region from the brain image.

PCA

PCA has been widely used for dimensionality reduction in computer vision. Result shows that PCA also performs well in various recognition tasks. The basis vectors, bi(x,y) generated from a set of palmprint images are called eigenpalm, as they have the same dimension as the original images Recognition is performed by projecting a new image into the subspace spanned by the eigenpalms and then classifying the palm by comparing its position in palm space with the positions of known individuals.

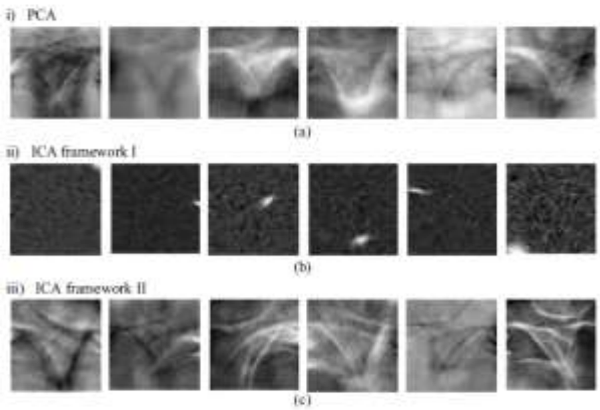


Figure 1: Vector basis generated by each technique.

SVM

During machine learning, Support Vector Machine (SVM) regulates learning models with associated learning algorithms. These algorithms analyze data used for classification and regression analysis. Given a set of training examples with two categories, SVM training algorithm builds a model which assigns the given data to any of the given category. A SVM model is a portrayal of the examples as focuses in space, mapped so that the cases of the different classes are separated by a gap. New data are then mapped into that same space based on the gap.

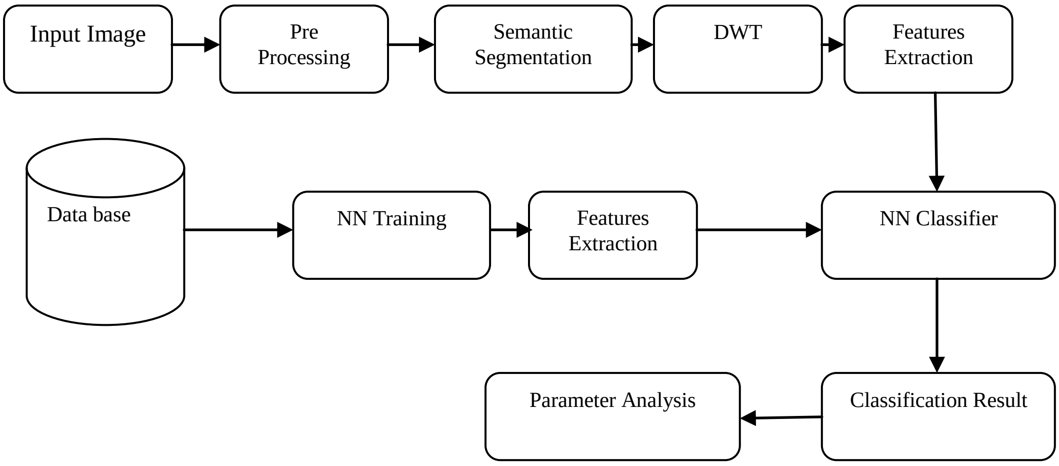


Figure 2: Block diagram for finding an image using NN and sematic segmentation

In addition to linear classification, using a non-linear classification named kernel trick, SVMs can efficiently map their inputs into high dimensional feature spaces.

Supervised learning is not possible, when data are not labeled. An unsupervised learning approach is required to find clustering of these data into groups. Then new data is mapped to form groups. Support vector clustering is a clustering algorithm which provides improvement to support

Vector machines. It is frequently used in industrial applications when data are not labeled or when only some data are labeled as pre-processing for a classification pass.

"Support Vector Machine" (SVM) is a machine learning algorithm. Classification and regression challenges can be efficiently done by using this supervised machine learning algorithm. But this algorithm is commonly used in classification problems. In this algorithm, each data item is plotted as a point in n-dimensional space (n represents the number of features). Value of the particular coordinate is represented by the value of each feature. Now, Classification can be done by finding the hyper-plane which differentiates the two classes.

II PROPOSED METHOD

- NN Network for classification
- Semantic Segmentation for object detection and structural analysis.

A. DISCRETE WAVELET TRANSFORM

A wavelet arrangement is a portrayal of a square-integrable work by a specific orthonormal arrangement created by a wavelet. Here it gives a formal, numerical meaning of an orthonormal wavelet and of the essential wavelet change.

Definition

A function $\psi \in L^2(\mathbb{R})$ is called an orthonormal wavelet if it can be used to define a Hilbert basis, that is a complete orthonormal system, for the Hilbert $L^2(\mathbb{R})$ of square integrable functions. The Hilbert basis is constructed as the family of functions

$\{\psi_{jk} : j, k \in \mathbb{Z}\}$ by means of dyadic translations and dilations of ψ,

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$$

for integers $j, k \in \mathbb{Z}$. This family is an orthonormal system if it is orthonormal under the inner product

$$\langle \psi_{jk}, \psi_{lm} \rangle = \delta_{jl}\delta_{km}$$

where δ_{jl} is the Kronecker delta and $\langle f, g \rangle$ is the standard inner product $\langle f, g \rangle = \int_{-\infty}^{\infty} \overline{f(x)} g(x) dx$ on

$L^2(\mathbb{R})$. The requirement of completeness is that every function $f \in L^2(\mathbb{R})$ may be expanded in the basis as

$$f(x) = \sum_{j,k=-\infty}^{\infty} c_{jk}\psi_{jk}(x)$$

with convergence of the series understood to be convergence in norm. Such a representation of a function *f* is known as a wavelet series. This implies that an orthonormal wavelet is self-dual.

B. GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM)

The co-occurrence matrix explores the gray level in the spatial dependence of texture. Given a position operator P (i, j), let A be the n x n matrix. A[i][j] represents the number of times that points with gray level g[i] occur, in the position specified by P, relative to points with gray level g[j]. By dividing A by the total number of point pairs that satisfy P, matrix C is constructed. Measure of the joint probability that a pair of points satisfying P is represented as C[i][j]. As specified by the point P, C is called a co-occurrence matrix.

Let t be a translation, for every grey-level (a, b), co-occurrence matrix Ct of a region is defined as $C_t(a,b) = card\{(s, s+t) \in R^2 | A[s]=a, A[s+t]=b\}$

Here, $C_t(a, b)$ is the number of site-couples, separated by a translation vector t, where a represents the grey-level of s, and b represents the grey-level of s + t.

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	0	0	0	0	0
2	0	1	0	2	0	0	0	0
3	0	0	1	1	0	0	0	0
4	0	1	0	0	1	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

Classical Co-occurrence matrix

Based on the orientation and distance between image pixels, co-occurrence matrix is constructed. Then, the texture is extracted from the co-occurrence matrix. Commonly proposed texture features are: Energy, Contrast, Correlation, and Homogeneity.

The spatial distribution and the dependence of the grey levels within an area are represented using co-occurrence matrix. Specified a predefined distance and angle, the probability of going from one pixel with a grey level of '*i*' to another with a grey level of '*j* 'is represented using Each ***(i,j)***th entry in the matrices. Sets of statistical measures are computed from these matrices, called feature vectors.

Energy: It is a gray-scale image texture measure of homogeneity changing, reflecting the distribution of image gray-scale uniformity of weight and texture.

$$E = \sum_x \sum_y p(x,y)^2$$

Contrast: Contrast is the main diagonal element, which reflects the image clarity and texture of shadow depth.

Contrast $$I = \sum\sum (x-y)^2 p(x,y)$$

Correlation Coefficient: joint probability occurrence of the specified pixel pairs is measured using correlation coefficient.

Correlation: sum(sum((x- μx)(y-μy)p(x , y)/$\sigma_x\sigma_y$))

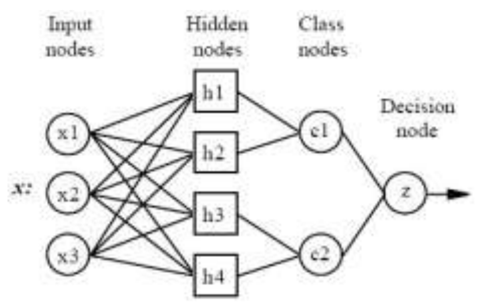
Homogeneity: closeness of the distribution of elements in the GLCM to the GLCM diagonal is measured using homogeneity.

Homogenity = sum(sum(p(x , y)/(1 + [x-y])))

C. PROBABILISTIC NEURAL NETWORKS

There exist a similar architecture for Probabilistic (PNN) and General Regression Neural Network (GRNN), but there exist a fundamental difference: Classification is performed on Probabilistic networks when the target variable is categorical and Classification is performed on General Regression Neural Network when the target variable is continuous. Based on the type of target variable, DTREG will automatically select the correct type of network in both PNN/GRNN network.

ARCHITECTURE OF A PNN:



Input layer — For each predictor variable, there will be one neuron in the input layer. In categorical variables, N-1 neurons are used where N represents the number of categories. Values are standardized by input neurons by subtracting the median and dividing by the interquartile range. Then the values are fed to each of the neurons in the hidden layer, by the input layer.

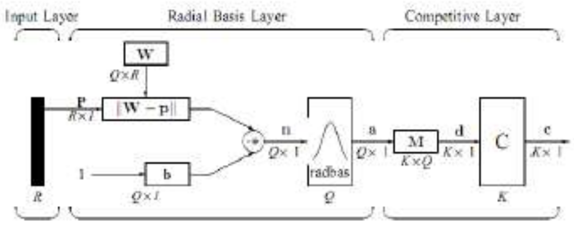
Hidden layer — this layer has one neuron for each case. The values of the predictor variables for the case along with the target value are stored in the neuron. When input layer presents input values, the Euclidean distance of the test case is computed by the hidden neuron from the neuron's center point. Then the neuron applies the RBF kernel function using the sigma values. The resulting value is passed to the Neurons in the next layer.

Pattern layer / Summation layer — This layer is different for PNN and GRNN networks. There is one pattern neuron for each category of the target variable in PNN networks. In each hidden neuron the actual target category of each training case is stored which produces a weighted value output. Pattern neuron is fed with this weighted value which corresponds to the hidden neurons category.

Only two neurons are available in the pattern layer for GRNN networks. One neuron for denominator summation unit and the other is for numerator summation unit. Weighted value output of the hidden neuron is summed up in the denominator summation unit. Weighted value output of the hidden neuron is multiplied by the actual target value in the numerator summation unit.

Decision layer — This layer is also different for PNN and GRNN networks. In PNN networks, weighted votes for each target category accumulated in the pattern layer are compared in the decision layer. After comparison, the largest vote is used to predict the target category.

The value accumulated in the numerator summation unit is divided by the value in the denominator summation unit by the decision layer in GRNN networks. The result is used as the predicted target value.Input vector is denoted by p, and is represented as a black vertical bar. Its dimension is R × 1. In this paper, R = 3.



Radial Basis Layer:

The vector distances between input vector p and the weight vector made of each row of weight matrix W are calculated in the radial basis layer. The dot product is used to define the vector distance between two vectors. Assume the dimension of W is Q×R. The dot product between p and the i-th row of W produces the i-th element of the distance vector ||W-p||, whose dimension is Q×1. The distance between vectors is indicated using the minus symbol, "-".

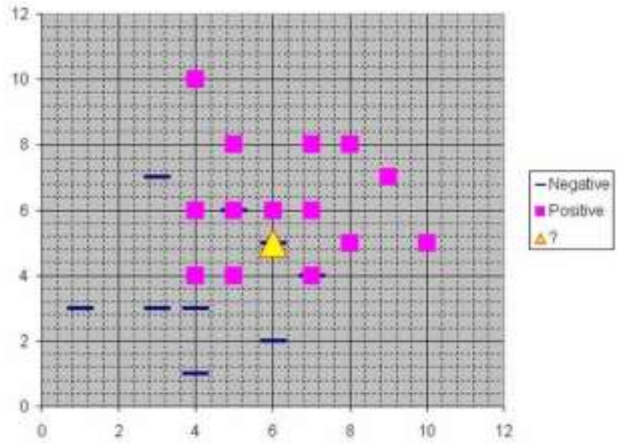
Then, the bias vector b is combined with ||W- p|| by an element-by-element multiplication.The result is denoted as n = ||W- p|| ..p. PNN's transfer function has built into a distance criterion with respect to a center. In this paper, it is defined as radbas (n) = 2 n e- (1). Each element of n is substituted into the equation which produces a, the output vector of Radial Basis Layer. The i-th element of a can be represented as ai = radbas (||Wi - p|| ..bi) where Wi represents vector made of the i-th row of W and the i-th element of bias vector b is bi.

Competitive Layer:

In this Layer, firstly the vector A is multiplied with layer weight matrix M, producing an output vector D. C represents the competitive function, produces a 1 corresponding to the largest element of d, and 0's elsewhere. The index of 1 in C is the number of tumors that the system can classify. The dimension of the output vector, K, is 5 in this paper.

HOW PNN NETWORK WORK:

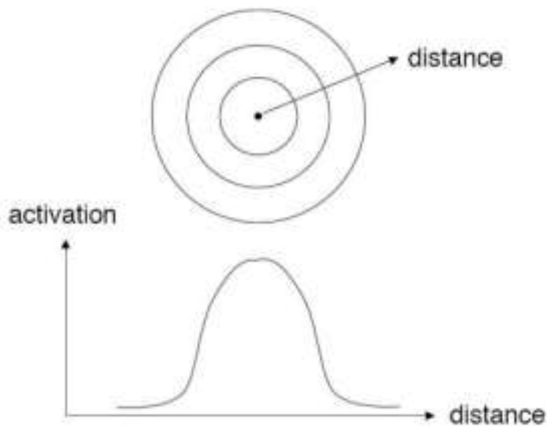
Probabilistic neural networks are similar to K-Nearest Neighbor (k-NN) models, though they vary in implementation. The basic idea is that a predicted target value of an item is likely to be about the same as other items that have close values of the predictor variables. For example considering the neighboring points the nearest neighbor classification is performed. If only one closet point is considered in 1-NN, then the new point will be classified as negative since it is on top of a known negative point. If 9 closest points are considered in 9-NN classification, close negative point may be overbalanced by the surrounding 8 positive points.



Based on this foundation, a probabilistic neural network is built and generalized to consider all the other points. From the point being evaluated, the distance is computed to each of the other points. The radial basis function (RBF) is applied to the distance to compute the weight for each point. Since radius distance is passed as an argument to the function, it is named as the radial basis function.

Weight = RBF (distance)

The further some other point is from the new point, the less influence it has.



III. CONCLUSION

Object detection and segmentation is recently growing technology and it is mostly used in video surveillance and retrieval applications. To overcome the threshold based object segmentation issues, semantic segmentation is used and it gives objects detection without any degradation of its shape. Discrete wavelet transforms and gray level co-occurrence matrix both are gives texture and edges of object in good accuracy. After that the neural network which helps to detect and classify the objects. Hence the method mostly used surveillance applications like Military areas, forest area, satellite applications and border areas.

REFERENCES

[1] Ranslational And Rotational Jitter Invariant Incremental Principal Component Pursuit For Video Background Modeling, Paul Rodriguez, Brendit Wohlberg, IEEE,International Conference on Image 2015.

[2] Survey on Techniques Involved in Image Segmentation - Shruti Pardhi,Mrs. K. H.wanjale.

[3] Traditional and Recent Approaches in Background Modeling for Foreground Detection: An Overview - Thierry Bouwmans computer Science Review, Elsevier, 2014, 11, pp.31-66.

[4] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, “Hyperspectral imagery super-resolution by sparse representation and spectral regularization,” *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–10, Oct. 2011.

[5] Special section on background models comparison, Article in Computer Vision Image understanding 122:1–3 ,May 2014 with 89 Reads, Antoine Vacavant, Laure Tougne,Lionel Robinault.

[6] A Robust Approach for the Background Subtraction Based on Multi-Layered Self-Organizing Maps-Article in IEEE Transactions on Image Processing PP(99):1-1August 2016 , Giorgio Gemignani, Alessandro Rozza.

[7] Siqi Bao, Albert C S Chung Feature sensitive label Fusion with random walkerfor Atlas based Image Segmentation, IEEE Transactions on Image processing, April 6,2017.

[8] F. Dufaux –Signal,Image processing and communication - 2013, pp. 483–490.

[9] An Evaluation of Background Subtraction for Object Detection Vis-a-Vis Mitigating Challenging Scenarios - Suman Kumar Choudhury, Pankaj Kumar Sa, Sambit Bakshi and Banshidhar Majhi,IEEE Access, volume 4 2016.

[10] Qingsong,Zhu,Jiaming Mai, Ling Shao, A fast Image Haze Removal algorithm using color Attenuation Prior , July - 07,2015.

[11] M. Aharon, M. Elad, and A. Bruckstein, “The K- SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006

[12] A. S. Charles, B. Olshausen, and C. Rozell, “Learning sparse codes for hyperspectral imagery,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5,pp. 963–978, Sep. 2011.

[13] Garberialia Ghimpeteanue, Thomas Batard, Marcelo Bertalmio,Stacey Levine, A decomposition Framework for Image Denoising Algorithms, dec-08,2015

[14] R.H.Bamberger and M.J.T.S,ith. “A filter bank for the directional decomposition of images: Theory and Design”, IEEE Trans .Signal Processing vol.40,no.4,pp.882-893 ,Apr 1992 .

[15] Xiang-Yang Wang, Yi-Ping Yang and Hong-Ying Yang, “A Novel nonsampled contourlet-domain image watermarking using support vector regression”, Journal of Optics A: Pure and Applied Optics, Volume 11, September 2009.

[16] Kusum Rani, Reecha Sharma, “Study of Different Image fusion Algorithm”, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 5, May 2013.

[17] Apurva Sharma, Anil Saroliya, “A Brief Review of Different Image Fusion Algorithm”, International Journal of Science and Research (IJSR) , Volume 4 Issue 6, June 2015.

[18] Vidya Rajbansi , “Analysis of Image Retrieval Using Linear Transformation Techniques”, International Journal of Scientific and Research Publications, Volume 5, Issue 7, July 2015.