Decision level data fusion in Speech and Image Recognition Systems

A Project Report

submitted by

Gaurav Sanjay Newalkar (12EC43)

Nithin Rao Koluguri(12EC51)

Nikhil Lunavath (12EC68)

under the guidance of

Dr.Ashvini Chaturvedi

in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE - 575025

March 15, 2016

ABSTRACT

We aim to design and implement object recognition in an image using Matlab. Our task is to detect an object in an image pertaining to the command given through speech. We first acquire the image in real-time then segment out the objects and locate where each object is located in the image. We used Histogram of Gradient(HOG) features and Support Vector Machine (SVM) for training and testing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Problem definition

Object recognition and pattern analysis is one of the active research topics in digital image processing. The technology is leaping into so much advancement that image recognition will become part and parcel of our daily lives. Applications such as Ultra sound, MRI use image processing to detect broken bones, tissues, Tumors and various kind of diseases and are used for various other industrial applications Contemporary it is used for detecting airport luggage scanning and for detecting the quality of food grains to detect fungi and other micro diseases.. So now in modern era image processing is used for security reasons through thumb print recognition, eye retina detection and then for crime detection it is also used for face recognition.[2]
satellite imagery is used to detect crop growing patterns, their cultivated area, advance warning to farmers in Australia is given for pesticide and disease. One of the state of the art applications is the cruise missile guidance system developed by US Defense to Map the territory for possible accurate target selection by using GIS and image processing. Even in Pakistan Sugar mills are using satellite imagery for planning their sugar can purchasing campaign and procurement planning. Imaging for medical reasons CT- Scan, MRI and Ultra Sound is now a days a common content in the patients medical history, detection of fungi is good topic to research. Researchers have put lot of effort in image processing.
For this project, we used Otsu's method and flood fill algorithm for segmenting and separating objects in image.Then we used HOG for feature extraction and SVM classifier for training and testing.

1.2 Previous work

Papageorgiou et al describe a pedestrian detector based on a polynomial SVM using rectifed Haar wavelets as input descriptors,with a parts (subwindow) based variant in [1]m.
Depoortere et al gave an optimized version of this [2].
Gavrila Philomen take a more direct approach,extracting edge images and matching them to a set of learned exemplars using chamfer distance.This has been used in a practical real-time pedestrian detection system [3].
Viola et al [4] build an efficient moving person detector,using AdaBoost to train a chain of progressively more complex region rejection rules based on Haar-like wavelets and space-time differences.
Ronfard et al build an articulated body detector by incorporating SVM based limb classifiers over 1st and 2nd order Gaussian filters in a dynamic programming frame work similar to those of Felzenszw alb Huttenlocher and Iof fe Forsyth[5].

1.3 Objectives

1. To recognize 4 different words through automatic speech recognition and convert to 3 digit binary word.
2. Based on the 3 digit binary word, A Image Recognition system to be developed to recognize the object pertaining to the command given.

CHAPTER 2

Description

Object recognition is a process for identifying a specific object in a digital image or video. Object recognition algorithms rely on matching, learning, or pattern recognition algorithms using appearance-based or feature-based techniques. Common techniques include edges, gradients, Histogram of Oriented Gradients (HOG), Haar wavelets, and linear binary patterns. Object recognition is useful in applications such as video stabilization, automated vehicle parking systems, and cell counting in bio-imaging.

It is basically done by two methods
1.Correlation Method
2.Supervised Learning Method

2.1 Methods

In Correlation method,the acquired image in real time is taken and starting from the top-left corner the required object prototype image(window) is correlated with the acquired image to right bottom. The location where we get maximum correlation coefficient is said be the location where the required object is found. But by the above lines we can say that it is time consuming and takes to much calculations. Hence we adopt Learning method.

In Learning method, We segment the image and then separate out each object

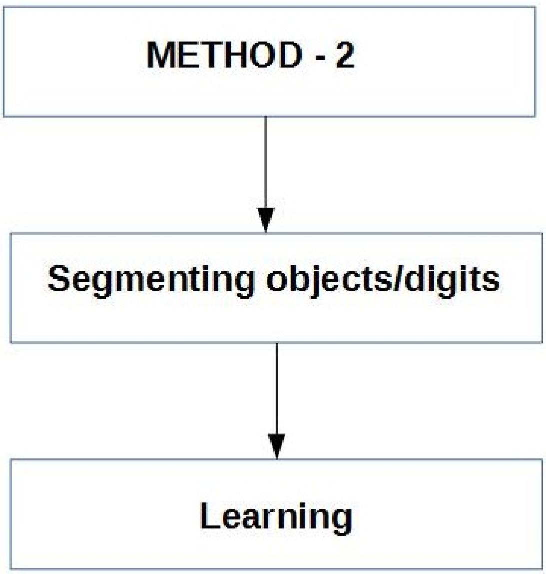

Figure 2.1: Basic Flow of Learning Method

from the image along with the coordinates of the location of the object. Then these individual objects(test images) are then passed to extract features and detect the object using classifier.

2.2 Otsu's Segmentation

In computer vision and image processing, Otsu's method, named after Nobuyuki Otsu, is used to automatically perform clustering-based image thresholding, or, the reduction of a graylevel image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal. Consequently, Otsu's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis.
The extension of the original method to multi-level thresholding is referred to as the Multi Otsu method.Equations are given below

---

The *weighted within-class variance* is:

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

Where the class probabilities are estimated as:

$$q_1(t) = \sum_{i=1}^{t} P(i) \qquad q_2(t) = \sum_{i=t+1}^{I} P(i)$$

And the class means are given by:

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \qquad \mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)}$$

Figure 2.2: Class means and weighted class variance

We show you a test image taken in a room before and after Otsu's Segmentation.
After Otsu's Segmentation we pass the image to flood fill algorithm to separate out each object based on minimum object size. An object(digit 3) is shown after passing through flood fill algorithm in fig 2.4.

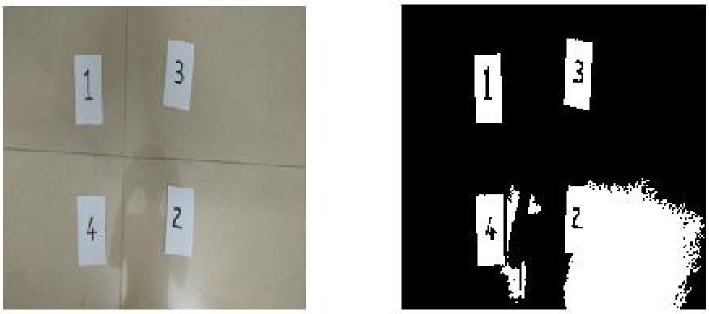

Figure 2.3: Before and after segmentation

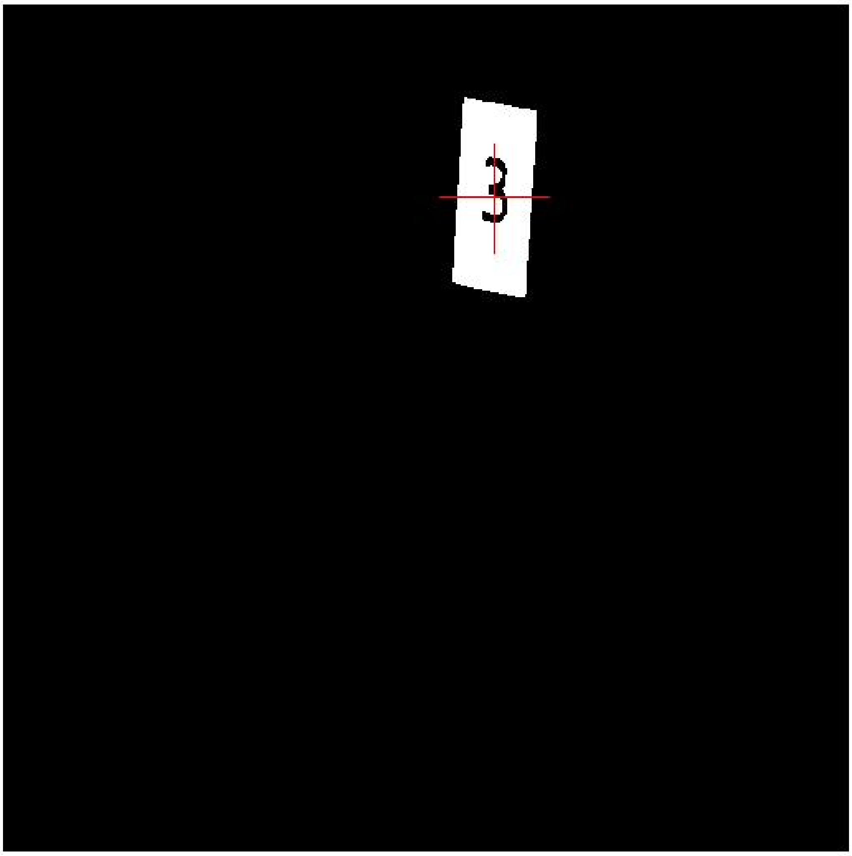

Figure 2.4: Digit 3 with centroid labelled on it

2.3 Feature Extraction :

After extracting each object from image we pass them through another step to extract features of each object.Histogram of Oriented Gradient descriptors, or HOG descriptors, are feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.

The essential thought behind the Histogram of Oriented Gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions.The implementation of these descriptors can be achieved by dividing the image in to small connected regions called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor.

The following image shows the histogram visualisation applied to the above extracted 'digit 3' object.

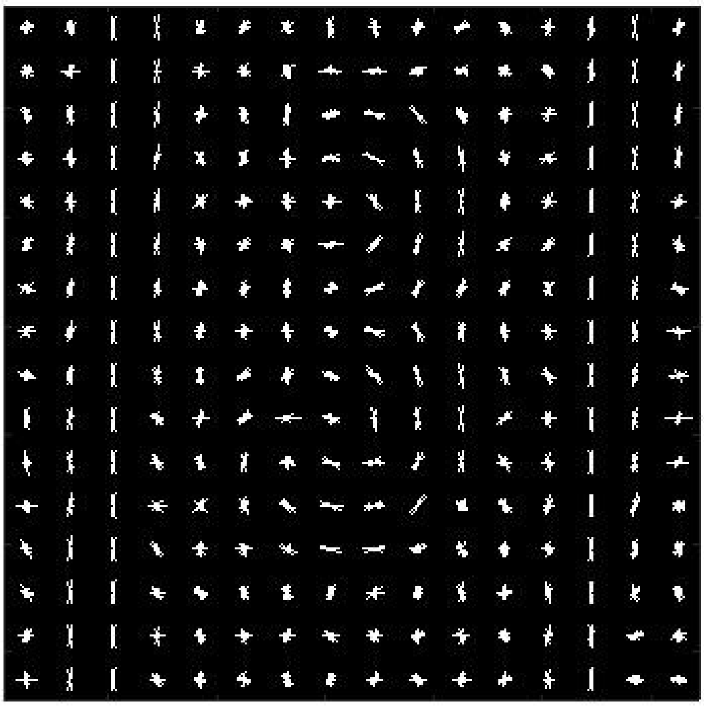

Figure 2.5: HOG visualisation of digit 3

2.4 Future Work

There are copious future applications for speech and image recognition systems. The output evaluated after speech recognition are used in various programs and applications to make physical interaction less systems. If Trained efficiently on many number of objects this project can be used to detect any object in a given image and thereby generalize it.Now a days we see this application in Google Photos where the search results are based on the object detected in image.

REFERENCES

[1] A.Mohan,C.Papageorgiou,and T.Poggio, *"Example-based object detection in images by components",*An International Journal (CSEIJ), Vol. 3, No. 4, April 2001 .

[2] V.dePoortere,J.Cant,B.VandenBosch,J.dePrins,F.Fransens,and L.Van Gool, *"Efficient pedes-trian detection:a test case for svm based cate gorization",*Workshop on Cognitive vision, 2002.

[3] D.M.Gavrila,J.Giebel,and S.Mundere, *" Vision-based pedes-trian detection:the protector+system",* Proc.of the IEEE Intelligent Vehicles Symposium,Parma,Italy,2004.

[4] P.Viola,M.J.Jones,and D.Snow *"Detecting pedestrians using patterns of motion and appearance",* 2003

[5] N. Dalal and B. Triggs, *" Histograms of Oriented Gradients for Human Detection",Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 886-893*,2005

[6] LeCun, Y., Bottou, L., Bengio, Y., and Haffner. P *"Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86, 2278-2324,* 1998

[7] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, *"Reading Digits in Natural Images with Unsupervised Feature Learning",* NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011.

[8] Subransu Maji,Jithendra Malik , *"Fast and Accurate Digit Classification",* University of California at Berkeley.