

**A
PROJECT REPORT
ON
ILLUMINATION VARIANT VIDEO TEXT
RECOGNITION FOR VISUALLY IMPAIRED
PERSONS**

(SPONSERED BY KSCST, BANGALORE)
Submitted to

**VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELGAUM**



In Partial fulfillment of requirement for the award of the degree

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

By

**ABHISHEK M.P.
DEXEETH V.
PRUTHVI M.R.
SHAMITHA M.**

**USN : 4KV10CS001
USN : 4KV10CS029
USN : 4KV10CS063
USN : 4KV10CS074**

VIII SEMESTER BE

Under the Guidance of

Prof. SMITHA M.L. MCA., (Ph.D)

Associate Professor & Head of the Dept. of MCA



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
K. V. G. COLLEGE OF ENGINEERING
SULLIA, D.K.-574327, KARNATAKA
2013-14**

CHAPTER 1

INTRODUCTION

Embedded text in video sequence provides valuable information of paramount importance. Text usually appears as logs, subtitles, captions, banners etc. Text object embedded in video contain much semantic information evaluated for multimedia content. The variation of text occurs due to difference in size, styles, orientation, and alignment. This is why retrieval of data has an important role.

A visually impaired person faces a number of challenges when interacting with the environment because so much information is encoded visually in daily life. One specific difficulty that a blind person would face is the problem of recognition of text data in the videos. Text data present in images and video contain useful information for automatic annotation, indexing and structuring of images. Extraction of this information involves detection, localization, tracking to supplement the visual and audio content.

1.1 Theoretical Background

Several methods for text extraction from video have been proposed in the last decade. Most of them are divided in the following stages: spatial text detection, temporal text detection, image binarization, segmentation and character recognition. From these stages, we observe that the most crucial is the stage of scene text detection which actually concentrates the focus of most researchers in this field. The performance of the whole text extraction system depends on the accurate localization of text in an image or video frame. Some methods have also been proposed for the binarization of the text image [1], although most researchers use state-of-the-art algorithms from the classic document analysis research area. In general, the existing spatial text detection methods can be roughly divided in two categories: region based and texture-based. Region-based methods group pixels that belong to the same character based on the colour homogeneity, the strong edges between character and background or by using a stroke filter. Then, the detected characters are grouped to form text lines according to colour, size and geometrical rules. Text in video especially the superimposed text is the most reliable information to be considered in video indexing research work. Many