Abstract

Right from the dawn of human civilization, people have started migrating from places to places for various reasons. Our Country; India being a nation with diverse culture and languages, people on moving away from their native may find it very much difficult to understand different languages. Hence, we have designed a system to read Kannada texts present in natural scenes with the aim to provide assistance to the non-native speakers of Kannada language. The area of natural scene text recognition aims on the problem to recognize random texts in images like street signs, name of shops, grocery item labels, and name plates etc. Here, we have tried to focus on recognition of characters in such cases that are not handled effectively by traditional OCR (Optical Character Recognition) techniques. We design a database of images which contains annotated Kannada characters inclusive of alphabets, digits and other symbols.

The problem is depicted as an object compartmentalization skeleton comprising of a database having characters as image. We fetch the potential of different attribute based on Support Vector Machine taxonomy and nearest neighborhood. The project mainly focuses on the areas of scene text recognition, each with a decreasing number of prior conjectures. Firstly, character recognition, where word and character bounding boxes will be used. Next is word recognition, where only word bounding boxes will be used. Lastly removing unwanted objects present in the image is done. For fragmenting text, the images, the bilateral regression segmentation algorithm is used. The system has been tested by submitting 200 different natural scene images and tabulated the results.