Building Text Extraction using OCR

Mayur Zagade[1], Shivani More[2], Manish Pasalkar[3], Anand Narute[4], Prof. Anuradha Thorat[5]

Students, Department of Information Technology[1,2,3,4]

Assistant Professor, Department of Information Technology[5]

Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract: *Image recognition and optical character recognition technologies have become an integral part of our everyday life due in part to the ever-increasing power of computing and the ubiquity of scanning devices. Printed documents can be quickly converted into digital text files through optical character recognition and then be edited by the user. Consequently, minimal time is required to digitize documents; this is particularly helpful when archiving volumes of printed materials. This study demonstrates how image-processing technologies can be used in combination with optical character recognition to improve recognition accuracy and to improve the efficiency of extracting text from images. Two software systems are developed and tested during this study: a character recognition system applied to cosmetic-related advertising images and a text detection and recognition system for natural scenes. The results of the experiment demonstrate that the proposed systems can accurately recognize text in images.*

Keywords: Cosmetic, Ubiquity, Optical