Abstract

Text provides semantic information that may be of a particular interest as it is useful for describing the contents of an image. Therefore, text data present in digital images and videos is an important Region of Interest (ROI) and its detection, localization and extraction is an active research area in the field of computer vision. The wide range of applications and challenges posed by Text Information Extraction (TIE) motivated us to carry out research work in this area. In this research work, we addressed TIE problem and proposed computer vision based techniques in order to extract text information efficiently from images or videos.

There exist a variety of sources where text is occurred in three different forms, and researchers have named as scene text, caption text, and document text (printed or hand written) depend on the source where text is embedded. In this thesis, we attempted to extract these three forms of text from source images or videos. The three forms of text considered in our study, and the corresponding text information extraction approaches proposed in this thesis are described briefly as follows:

Scene text is defined as text existing naturally in images or videos. In order to read and interpret easily, text information is usually written on planar objects like trucks, t-shirts, building walls, billboards, and road signs, etc. This motivated us to detect and extract planar surface, and is followed by extraction of text components within the extracted planar surface. The first phase of the planar surface extraction process involves three steps: a) estimation of disparity map using stereo images or video frames, b) detection of candidate planar surface
from the disparity space and c) segmentation of planar surface by fitting planar model followed by image labeling. Planar surface is segmented from other regions based on the label which is assigned to it. In the second phase, we employed image processing techniques in order to extract linear and curved scene text individually by considering extracted planar surface instead of whole image or video frame.

Caption text defined as text that is artificially overlaid on the image or video after it is captured and is usually done during the editing process. Due to editing process, there is a possibility of inconsistencies in noise levels caused by pixel duplication. This motivated us to propose a technique to extract caption text by finding noise inconsistencies within the image or video frame through multi-resolution analysis, where high-frequency sub-band is used to compute noise variance followed by merging of blocks with similar noise levels. Low level process such as edge generation is used to identify regions belonging to caption text as text regions are rich in edges, and they appear in clusters.

The final and third sub-problem considered in this thesis is extraction of document text (handwritten or printed) which is always contained in document images. In order to extract document text, we initially determine the text-line path in a bid to segment every single text-line separately, this is done by smoothing the image followed by finding the axial line based on a ridge detection algorithm. Each axial line is uniquely labeled, followed by determination of text-line path by fitting a smoothed B-spline curve, where the curvature of every point to the text-line is computed. Finally, text-line is aligned and extracted based on the computed curvature values. We evaluated our technique based on accuracy involved during text-line segmentation in comparison to the ground truth generated through manual annotation.

In order to evaluate the efficiency of our techniques, we conducted several experiments using our own datasets and standard publicly available datasets. The experimental results were compared with the existing methods based on evaluation metrics such as recall, precision and f-measure. The results posted by our techniques are comparable and proven to be better than the existing state-of-the-art methods.