Chapter 7
Conclusions and Future Directions

In this thesis, we addressed text information extraction problem, which involves two major processes, text localization and text extraction in three different environments and they include scene text, caption text and document text. The processes involved were fully investigated and novel computer vision techniques were developed and presented in this thesis which provide solutions to challenges associated with them. The work presented in this thesis can be adopted in many applications such as page segmentation, transcribing historical documents into textual electronic formats, license plate localization, address block localization, navigation to vehicles, visually impaired persons and robots, and content-based image or video annotation, indexing and retrieval.

7.1 Conclusions

We have investigated the real importance of text information available in both digital videos and images in order propose and develop efficient and effective computer vision techniques that are applicable to a wide range of applications. Through various experiments conducted using benchmark datasets, as well as our own generated datasets and evaluated using available standard evaluation metrics, it is demonstrated that the proposed approaches are more robust and accurately better in most challenging areas containing light variations, oriented text-lines, low contrast, variation in font size and writing style, and script independency. The main contributions of our research work presented in this thesis are summarized as follows:

The selection of high sharpness and high quality video frames are very vital for the accurate extraction of text information and has a direct impact on the
robustness of the text information extraction algorithms. We presented a novel approach to assess the quality of stereo frames based on the analysis of saliency region, in our case, planar surfaces where it is assumed that text information is written on these surfaces. Our approach achieves promising results compared to existing techniques as it is helpful while selecting high quality pairs of frames for the creation of reliable dataset aimed for evaluation of proposed text information extraction algorithms based on stereo frames.

We presented a novel approach for the extraction of scene text where the novelty of our approach lies in the extraction of planar surface, as it is assumed that scene text are always written on planar objects for the purpose of readability. The extraction of planar surface provides an accurate candidate text block segmentation in complex environments such as tree-leaves which have text-like features are removed completely leaving only text features contained in planar surface. Further, text block segmentation reduces our working area for employing other low level processes to extract text within text block. The experimental results show that our approach is capable of handling multi-script text. Another important aspect is that, our technique is invariant to font sizes, noise contrast and or text-line orientation and can be used to extract all manner of text script. Though our approach proved to be computationally costly in terms of the time taken to complete the whole process, it is noteworthy that the same approach produces an output which is most accurate for text information extraction in a highly complex environments which is still a pipe dream for the existing techniques.

We proposed a new approach to extract curved multi-script scene text based on fuzzy curve tracing performed within reduced image space i.e. extracted planar surface. The experimental results obtained using our own dataset demonstrates that our approach achieves highest true positive rate compared to existing methods. Compared to the publicly available datasets, our dataset used for experimentation is characterized by two complexities: a) text does not span from border to border because the text objects are captured away from the camera and b) existence of complex background containing text like features such as tree-leaves, pathways and buildings. The superiority of our technique compared with the
existing methods emanates from its ability to extract text from images with complex background accurately with low false positives. Another advantage of our approach lies on the use of fuzzy curve tracing algorithm. Fuzzy curve tracing algorithm alleviates the problem faced by existing methods which employ bounding boxes when dealing with oriented text strings. Bounding boxes enclose more of non-text area than text area and sometimes overlap each other.

We presented a novel approach to extract caption text in videos based on noise inconsistencies caused by pixel duplication through multi-resolution analysis. Extraction of caption text based on noise inconsistencies require neither a prior information nor reference image but it works independently on a single frame. Another advantage is that proposed technique can work perfectly irrespective of compressed or un-compressed images or videos. Our approach detects the existence of caption text through identifications of various noise levels. However, our approach fails to detect caption text if the image is structured where there may be various isolated regions with totally different variances.

We presented a novel approach for the extraction of arbitrary oriented handwritten text in document images. The main benefit of our approach is that it can handle the presence of multi-oriented text-lines within the same image. This is because usage of rotating smoothing filters, which responds differently to a pixel depending on its orientations. Our technique is also capable to model a local curve based on the adjacency points with potentially similar curvature values within a specified range. It is observed that one of the limitations of the existing methods is deformation of characters while aligning, this leads to low recognition rate during OCR. However, our approach addresses this limitation associated with multi-oriented text-lines by rotating characters and is free from character deformation resulting to a better candidate text for OCR.
Chapter 7. Conclusions and Future Directions

7.2 Future Directions

It is observable from the results that further research is needed to improve the efficiency and effectiveness of the proposed techniques. This is so because of the following identified limitations faced by our approaches: 1) based on the nature of sub-problems addressed, it becomes hard to develop a single technique to address TIE problem for three different forms of text considered in this thesis, 2) the use of stereo frames for scene text extraction becomes computationally expensive due to the number of steps involved, and 3) when occluded or non-connected planar surfaces appear in the same axis, it becomes difficult to extract them, for scene text extraction.

It has been demonstrated that hybrid methods performs better and thus becomes a promising direction. Currently unsupervised algorithms carries the day in all aspects of research, for example machine learning methods that employ unsupervised learning produces good results without human intervention and are increasingly applied to many research areas. Our techniques can be extended by incorporating other newly emerging trends in computer vision and machine learning in order to achieve high accuracy and to overcome the above said limitations.