

# Preface

With the increasing availability of low cost portable digital video recorders, we are witnessing a rapid growth of video data archives today. The need for efficient indexing and retrieval has drawn the attention of researchers towards handling the video databases. However, efficiently handling video content is still a difficult task in the pattern recognition and computer vision community, especially when the size of the database increases dramatically. A lot of ideas have been proposed, but like other frontiers in this community, there is no reliable approach that has theoretical grounding in video content analysis.

Fortunately, studies have shown that humans often pay their first attention to text over other objects in video. It is probably due to the ability of humans to simultaneously process multiple channels of scene context and then focus the attention on texts in video scenes. This fact makes video text detection a feasible and probably the most efficient way for indexing, classifying, retrieving and understanding the visual contents in videos. This can be further used in transportation surveillance, electronic payment, traffic safety detection, sport videos retrieval, and even commercial online advertisements. One example is video-based license plate recognition systems, which are accordingly necessary to help improve the convenience of checking vehicle status at roadside and designated inspection points efficiently. Another example is online video advertising. Driven by the advent of broadband Internet access, today's online video users face a daunting volume of video content from video sharing websites, personal blogs, or from IPTV and mobile TV. Accordingly, how to develop advertising systems especially considering contextual video contents through efficient video text detection techniques has become an urgent need.

Actually, video text detection has not been systematically explored even though people have developed a lot of optical character recognition (OCR) techniques, which are considered as one of the most successful applications in the past decades. For example, to explain a typical Google street video scene view, popular visual understanding methods detect and identify objects such as car, person, tree, road

and sky from the scene successfully. However, regions containing text tends to be ignored. It is probably due to the fact that text from video is sometimes difficult to detect and recognize. The performance of OCR thereby drastically drops when applied to video texts which are either artificially added (*graphic text*) or naturally existing on video scene objects (*scene text*). There are several reasons for this fact. First, the variety of color, font, size and orientation of video text bring difficulties to OCR techniques. Second, video scenes exhibit a wide range of unknown imaging conditions which in general add sensitivity to noises, shadows, occlusion, lights, motion blur and resolution. Finally, the inputs of most of the OCR engines are well segmented texts which have been distinguished from background pixels. Unfortunately, the segmentation of video text is much harder.

This book tries to systematically introduce readers to the recent developments of video text detection for the first time. It covers what we feel a reader who is interested in video text detection ought to know. In our view, video text detection consists of a general introduction to the background of this exploration (Chap. 1), pre-processing techniques (Chap. 2), detection of graphic text from video (Chap. 3), detection of scene text from video (Chap. 4), post-processing techniques such as text line binarization and character reconstruction (Chap. 5), character segmentation and recognition (Chap. 6), video text applications and systems in real-life (Chap. 7), video script identification (Chap. 8), multi-modal techniques which have been proved to be useful for video text detection and video content analysis (Chap. 9), and performance evaluation of the video text detection algorithms and systems (Chap. 10). A reader who goes from cover to cover will hopefully be well informed. However, we also tried to reduce the interdependence between these chapters so that the reader interested in particular topics can avoid wading through the whole book. We present theoretical material in a succinct manner. The reader can easily access to a more detailed up-to-date set of references of the methods discussed for further reading. Thus we are able to maintain a focus on introducing the most important solutions of video text detection in this book.

We are indebted to a number of individuals both from academic circles and industry who have contributed to the preparation of the book. We thank Hao Wang, Zehuan Yuan, Yirui Wu, Run Xin and Trung Quy Phan for collecting materials and their experimental evaluations on particular methods introduced in this book.

Special thanks go to Simon Rees for providing us this chance, and for the commitment to excellence in all aspects of the production of the book. We truly appreciate his creativity, assistance, and patience.

Nanjing, China  
Kuala Lumpur, Malaysia  
Singapore, Singapore  
Kowloon Tong, Hong Kong SAR  
April 04, 2014

Tong Lu  
Shivakumara Palaiahnakote  
Chew Lim Tan  
Wenyin Liu

Video Text Detection

Lu, T.; Palaiahnakote, S.; Tan, C.L.; Liu, W.

2014, X, 264 p. 160 illus., Hardcover

ISBN: 978-1-4471-6514-9