

© 2006 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Automatically Detecting Road Sign Text from Natural Scene Video

Wenjing Jia¹, Xiangjian He¹, and David Tien²

¹*Computer Vision Research Group, University of Technology, Sydney*

²*School of Information Technology, Charles Sturt University*

¹{wejia, sean}@it.uts.edu.au, ²dtien@csu.edu.au

Abstract

Automatic detection of text on road signs can help drivers keep aware of the traffic situation and surrounding environments by reminding them of the signs ahead. Current systems can only detect constrained road signs or produce unsatisfying performance when dealing with complex scenes in practical use. This paper firstly reviews the existing techniques used for text detection from natural scene. A novel system which detects text on road signs from natural scene video is then proposed. Our detailed approaches and methodology give a promising solution to this problem in order to reduce the running time and improve the recognition rate.

1. Introduction

Information provided by road and traffic signs is essential for road safety. Text on road and traffic signs contains abundant information essential for road safety. For motor vehicles, road signs provide information for navigation, describe the current traffic situation, define right-of-way warnings about potential risks, permit or prohibit certain turn and warn about risky driving situation [1]. For passengers, road signs provide essential information to mark objects, such as buildings, streets and many other places [2, 3]. However, the complex background in city scene may distract drivers' attention from noticing these signs.

Recently, extensive research effort has been put on detection and recognition of road sign text on either still or consecutive images. Even though the research effort, correctly and efficiently detecting and recognising text on road signs under arbitrary environments is still a hard, unsolved, and very challenging problem. Major reasons are complex scenes, uncontrollable illumination conditions, wide varieties of character appearance, and different orientations of text.

As this result, correctly detecting text on road signs imposes great challenges and complexities upon the

problem. Many researchers have put various constraints to simplify the problem when designing algorithms. The performance of most previous research work, when tested with unconstrained natural scene images, is still not good enough for practical use.

This paper proposes a novel system which in real-time detects various road and traffic signs from natural scene video and recognises text or symbols on them. The objective of the system is to help drivers keep aware of the traffic situation and surrounding environments by reminding them of the signs ahead and hence to improve road safety.

Unlike most currently existing research, the system to be developed in this research will detect and recognise not only specific traffic signs under "good" conditions, but also texts with wide varieties on more complex natural scene video under more complex conditions. Furthermore, the system will detect not only the road signs that are immediately in front of a driver but also those to the left- and right-front of the driver. In the proposed program, both low-level image features and high-level character features will be used complementarily to improve the performance of detection. Moreover, since in outdoor scenes, illumination conditions can vary considerably, special attention will be devoted to the robustness of the system to work under various illumination conditions.

The remaining parts of the paper are organised in the following sequence. In Sect. 2, the recent related research work is summarised. The proposed system and approaches are introduced in details in Sect. 3. Sect. 4 summarises the paper.

2. Background

Extensive research effort has been made for the detection and recognition of road sign text. In this section, recent related work is summarised.

In [4], Escalera et al. proposed an algorithm to detect and classify road traffic signs. Their system could only detect signs with red borders and specific shapes.

Similarly, the algorithms in [5] and [6] only dealt with circular road signs. Silapachote's system in [7] could only recognise and identify signs that were within a database of known signs. Some other systems could only deal with signs with specific sizes or aspect ratios, including Kim's algorithm [8] and Ezaki's systems in [2, 3].

More complicated and practical systems have only been studied in recent years. Some of the most significant research include Chen and Yuille's visual aid system [9], Silapachote's Visual Integration and Dissemination of Information (VIDI) system [7], Kim's text extraction system [8], Ezaki's text reading system [2, 3], and Mattar's sign classification algorithm [7], etc. However, most of these systems can only work under "good" conditions. For instance, Piccioli in [10] assumed that there was no perspective distortion of signs in shape. Chen and Yuille [9] assumed that the camera was kept in horizontal level, and there was very little blur in images.

These limitations have obviously constrained the applicability of the existing systems.

In terms of system's performance, there is a trade-off between the processing speed and the recognition rate. The recognition rate reported by some researchers was "almost perfect", but it took too long for their methods to be executed in real-time. For instance, Zin's method in [5] took 3 to 5 seconds to process one input image in order to achieve an "almost perfect" detection result, and Chen and Yuille's visual aid system [9] also required 3 seconds to achieve its 93% recognition rate. Obviously, this speed is unacceptable for a real-time application. Some faster systems, however, could not give a satisfactory recognition rate when tested on complex scenes. For example, Kim's algorithm in [8], when tested on complex natural scene images, gave a 92.6% recall rate, and Ezaki's systems in [2, 3] showed a low precision rate.

3. Approach and Methodology

The application of automatically detecting and recognising text on road signs in real-time requires a smart system like the one to be developed in this research program that operates in real-time and runs in unconstrained scenes without human adjustment while keeping the cost low. Both fast processing and accuracy is essential in order to provide correct information to drivers instantly. In this research, we will set up a high-performance driver assistant system for detecting and recognising text and symbols on road and traffic signs from natural scene video in real-time. The application scenario is that a wide-angle controllable surveillance camera is mounted on a moving vehicle to capture

scene video in front of the vehicle. The algorithm to be developed then detects signs from captured video, recognises text or symbols on road signs, and helps the driver to manoeuvre properly in traffic by reminding him/her of the signs ahead in advance. Major approaches to be used in this program are presented in this section.

3.1. Operating on Colour Natural Scene Video

Most previous research on road sign text detection was based on still images as detecting text from video is more challenging [11]. In practical use, an automatic driver assistant system, once enabled, should be able to capture scene images and run detection algorithms spontaneously and automatically without any other operations. A system working on still images cannot meet such requirement. Consequently, the significance of developing a text detection system that works on video rather than still images stands to reason.

Furthermore, current existing video-based systems rely on the assumptions that there is only a small angle between the object signs and the optical axis, so that there is only slight perspective distortion in captured images. Using these systems, some important signs such as intersected street signs that are not located right in front of the driver will be missed.

In this program, a wide-angle surveillance video camera, which captures not only the scenes in front of the driver, but also scenes to the left-front and the right-front of the driver, will be employed. Due to the potential severe deformation of signs caused by perspective distortion, this has never been attempted before. Text restoration algorithm based on characters' homographic features will be developed to ensure an accurate recognition under such environment.

Colour information is one of the most critical characteristics of signs, which helps to make the detection algorithm faster if properly utilised. Most colour-based features used in previous research, however, were not illumination-independent, thus did not work well under various illumination conditions. We have studied illumination-insensitive features which are based on colour information in [12] and [13]. These ideas, together with a related idea proposed in [14], will be applied in this research.

Besides, the video camera used for this system will be software controllable, of which the focus can be adjusted automatically to get more details of interested areas when needed in order to achieve a more reliable recognition result.

3.2. Using both Global and Local Features

To efficiently and correctly find locations of road signs from a complex background, for instance, city scenes, is a critical task required for an efficient road sign recognition system. Most previous research on locating road signs only utilised limited features and thus could only deal with limited kinds of road signs. This degrades the applicability of their developed systems. Therefore, to develop a practical driver assistant system, information hidden in images and videos must be utilised effectively and efficiently. Using multi-features including both computationally cheap global features and fine local features is significant towards an efficient and robust road sign text detection system.

In this research, we will design an innovative multi-layered hierarchical framework to implement the system, where both global and local features will be appropriately developed and employed to achieve our objectives. The purpose of the lower layers is to rule out those non-text areas in the very early stage of detection using computationally cheap features, usually global statistical features. A wide-range of global features, which are based on multi-scale, oriented band-pass filters, and non-linear grating cells, will be employed for this purpose. These features have been shown to be effective in detecting signs in unconstrained outdoor images [15]. In higher layers, local features, such as scale invariant features (SIFT) [16] and Harr features [17], have been shown to be very effective in object recognition problems [18].

3.3. Detecting Character-Like Areas

Effective features may not produce satisfying classification result unless reliable and robust classifiers have been chosen. Various prior knowledge based thresholding methods obviously cannot make a satisfactory decision adaptable under various environments. Using a learning-based classifier, which gives a stable classification result quickly, is thus critical to the system. Using the AdaBoost machine learning algorithm [19] to train a strong classifier combined with Viola and Jones' cascade approach [20] for text reading from city scene images has been tested by Chen and Yuille in [9], which made the detecting algorithm extremely fast to meet the real-time processing requirement. Our previous research using AdaBoost for number plate detection also demonstrated its effectiveness in detecting character areas with various appearances [21].

Furthermore, in most previous research, focus was put on strings of text. Characters under the same context are assumed to have the same foreground and background colours, same sizes, and be distributed in the same interval roughly in the horizontal direction. This, however, has created lots of troubles under outdoor scene condition because colour can vary greatly

for each character even on the same sign due to unevenly lighting, sizes of each character can be very different due to perspective distortion, and characters may be distributed with significantly different intervals or in different orientations in the same scene. All of these can result in the missing detection of text. Instead, in this research, our focus will be on individual characters in order to cope with text that appears with wide varieties.

In this system, we will use standard AdaBoost training methods to train a strong classifier combined with the Viola and Jones' cascade approach [20]. The cascade approach enables the algorithm to rule out most non-text areas with only a few tests, which can make the algorithm extremely fast. In previous related work, several sets of global features have been used as weak classifiers, which are to be combined by standard AdaBoost techniques. In our system, not only global features but also scale-invariant local features of characters will be developed to produce the strong classifier, which is specially designed for character-like area detection.

Besides, in this research, the above method will be used to detect *character-like* area rather than *text-like* areas to cope with characters that appear with wide varieties under the same context. Extraction of text strings will be implemented by further connecting individual characters after considering various text alignments. This will make the detection algorithm keep a high detection rate and a fast processing speed, even under the situation that there are a couple of characters hard to detect. Our previous experimental results [22] using similar ideas in a related area have demonstrated the promising perspective of this idea.

3.4. Restoring Text from Perspective Distortion

When the locations of the text areas have been found, correctly extracting the text within the area is nevertheless still a problem. This is mainly because the potential perspective distortion caused during the image capturing process, as well as the various appearance of text caused by the slant and tilt of road signs. Without correctly extracting text from road signs, it is unlikely that the general-purpose OCR technique can read out the text correctly and quickly.

It is noted that, Characters of road signs in the same context all have similar colour distribution. This feature will be used to align characters to form a row-based or column-based structure. Even though the geometry of a sign may be deformed, text on the sign can still keep an aligned layout. So, layout analysis of text will be done to align characters in an optimal way so that characters

belong to the same context will be aligned together. Breaking the constraints that the text on road signs must be aligned in nearly horizontal direction will be another important innovation of this system.

4. Conclusion

In this paper, a novel system is proposed, which can detect text on road signs from outdoor scene video. Based on our study on existing research work, a driver assistant system which operates on colour video is proposed. The algorithms developed in the research then automatically process and analyse input videos in order to detect text on various road signs in real time. Our preliminary experimental results using similar ideas in a related area on number plate detection under various complex conditions show the perspective of the proposed system.

5 References

- [1] V. Kastinaki, M. Zervakis and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, vol. 21, pp. 359-381, 2003.
- [2] N. Ezaki, M. Bulacu and L. Schomaker, "Text detection from natural scene images: towards a system for visually impaired persons," *The 17th International Conference on Pattern Recognition*, vol. 2, pp. 683-686, 2004.
- [3] N. Ezaki, K. Kiyota, B. T. Minh, M. Bulacu and L. Schomaker, "Improved text-detection methods for a camera-based text reading system for blind persons," *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, vol. 1, pp. 257-261, 2005.
- [4] A. de la Escalera, L. E. Moreno, M. A. Salichs and J. M. Armingol, "Road Traffic Sign Detection and Classification," *IEEE Tran. Industrial Electronics*, vol.44, pp.848-859, 1997.
- [5] T. T. Zin and H. Hama, "Robust road sign recognition using standard deviation," *The 7th IEEE Conference on Intelligent Transportation Systems*, pp. 429-434, 2004.
- [6] J. Torresen, J. W. Bakke and L. Sekanina, "Efficient recognition of speed limit signs," *The 7th IEEE Conference on Intelligent Transportation Systems*, pp. 652-656, 2004.
- [7] P. Silapachote, J. Weinman, A. Hanson, M. A. Mattar and R. Weiss, "Automatic Sign Detection and Recognition in Natural Scenes," *The IEEE Conference on Computer Vision and Pattern Recognition*, vol.3, pp.21-27, 2005.
- [8] K. C. Kim, H. R. Byun, Y. J. Song, Y. W. Choi, S. Y. Chi, et al., "Scene text extraction in natural scene images using hierarchical feature combining and verification," *The 17th International Conference on Pattern Recognition*, vol. 2, pp. 679-682, 2004.
- [9] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," *The 2004 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 366-373, 2004.
- [10] G. Piccioli, E. De Micheli, P. Parodi and M. Campani, "Robust method for road sign detection and recognition," *Image and Vision Computing*, vol. 14, pp. 209-223, 1996.
- [11] W. Wu, X. Chen and J. Yang, "Incremental Detection of Text on Road Signs from Video with Application to a Driving Assistant System," *Proceedings of ACM Multimedia*, 2004.
- [12] H. Zhang, W. Jia, X. He and Q. Wu, "Modified Color Ratio Gradient," *The International Workshop on Multimedia Signal Processing (IEEE)*, pp. 317-320, 2005.
- [13] W. Jia, H. Zhang, X. He and Q. Wu, "Symmetric Color Ratio in Spiral Structure," *Lecture Notes in Computer Science*, vol. 3852, pp. 204-213, 2006.
- [14] W. Jia, H. Zhang, X. He and Q. Wu, "Gaussian Weighted Histogram Intersection for License Plate Classification," *The 18th International Conference on Pattern Recognition*, pp. to appear, 2006.
- [15] J. Weinman, A. Hanson and A. McCallum, "Sign detection in natural images with conditional random fields," *The 2004 14th IEEE Workshop on Machine Learning for Signal Processing*, pp. 549-558, 2004.
- [16] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.
- [17] C. P. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection," *The Sixth International Conference on Computer Vision*, pp. 555-562, 1998.
- [18] M. A. Mattar, A. R. Hanson and E. G. Learned-Miller, "Sign Classification using Local and Meta-Features," *The 2005 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 3, pp. 26, 2005.
- [19] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139, 1997.
- [20] P. Viola and M. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade," *Advances in Neural Information Processing System*, 2001.
- [21] P. Viola, M. J. Jones and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *The 9th IEEE Conference on Computer Vision*, vol. 2, pp. 734-741, 2003.
- [22] H. Zhang, W. Jia, X. He and Q. Wu, "Learning-Based License Plate Detection Using Global and Local Features," *The 18th International Conference on Pattern Recognition*, pp. to appear, 2006.