EXTRACTING GENERIC TEXT INFORMATION FROM IMAGES

A Thesis Submitted for the Degree of
Doctor of Philosophy

By

Chao Zeng

in

School of Computing and Communications
UNIVERSITY OF TECHNOLOGY, SYDNEY
AUSTRALIA
SEPTEMBER 2013

UNIVERSITY OF TECHNOLOGY, SYDNEY
SCHOOL OF COMPUTING AND COMMUNICATIONS

The undersigned hereby certify that they have read this thesis entitled "**EXTRACTING GENERIC TEXT INFORMATION FROM IMAGES**" by **Chao Zeng** and that in their opinions it is fully adequate, in scope and in quality, as a thesis for the degree of **Doctor of Philosophy**.

Dated: September 2013

Research Supervisors: _____
Xiangjian He

_____
Wenjing Jia

CERTIFICATE

Date: **September 2013**

Author: **Chao Zeng**

Title: **EXTRACTING GENERIC TEXT INFORMATION FROM IMAGES**

Degree: **Ph.D.**

I certify that this thesis has not already been submitted for any degree and is not being submitted as part of candidature for any other degree.

I also certify that the thesis has been written by me and that any help that I have received in preparing this thesis, and all sources used, have been acknowledged in this thesis.

Signature of Author

Acknowledgements

First and foremost, I sincerely appreciate my principal supervisor Professor Xiangjian He for providing me with this precious opportunity of studying PhD under his supervision at University of Technology, Sydney (UTS). His insightful guidance and his continuous encouragement give me impetus throughout my entire PhD study. I owe my research achievements to his excellent supervision.

I also would like to express my deepest gratitude to my co-supervisor Dr. Wenjing Jia who always offers enlightening suggestions and patiently corrects my paper writing. Her consistent support during my research work and the completion of this thesis will never be forgotten.

I am also much indebted to the staff and my fellow research students in Faculty of Engineering and Information Technology (FEIT), UTS, especially the following people for offering various assistance during the completion of this research work. They are Qiang Wu, Ruo Du, Muhammad Abul Hasan, Sheng Wang, Man To Wong, Massimo Piccardi, Richard Yi Da Xu, Min Xu, Zhiyuan Tan, Aruna Jamdagni, Liangfu Lu, Jie Liang, Ava Bargi and Damith Mudugamuwa.

My special thanks should extend to my father Xiangheng Zeng, mother Likuan Zhang and my wife Yaxin Xu. This thesis could not have been completed successfully without their persistent encouragement and firm support.

Last but not least, I appreciate the financial assistance provided by International Postgraduate Research Scholarship of Australia and UTS President's Scholarship. Furthermore, FEIT, UTS is also acknowledged for offering me a travel fund for attending an international conference.

To My Parents and My Family

Table of Contents

Table of Contents viii

List of Tables ix

List of Figures x

Abstract 1

1 Introduction 3
- 1.1 Applications of Text Information Extraction (TIE) Research 3
- 1.2 Existing Methods of Text Detection and Binarisation 7
  - 1.2.1 Framework of Text Detection 7
  - 1.2.2 Framework of Text Binarisation 8
- 1.3 Unsolved Problems 8
- 1.4 Research Objectives 10
- 1.5 Author's Contributions in This Thesis 10
- 1.6 Thesis Structure Overview 11

2 Review of Some Related Work 13
- 2.1 State-of-the-art Text Detection and Binarisation Methods 14
  - 2.1.1 Existing Text Detection Methods 14
  - 2.1.2 Region Classification 18
  - 2.1.3 Existing Text Binarisation Methods 20
- 2.2 Recent Advances on Graph-based Image Segmentation Techniques 24
  - 2.2.1 Introduction to Image Segmentation 24
  - 2.2.2 Background 25
  - 2.2.3 Supervised Graph-Based Image Segmentation Methods 26
  - 2.2.4 Unsupervised Graph-Based Image Segmentation Methods 36
- 2.3 Summary 39

3 Born-digital Text Detection **41**
3.1 Coarse Detection . . . 43
3.1.1 Maximum Gradient Difference . . . 43
3.1.2 Multiple Layer Image Generation . . . 46
3.1.3 Morphological Operations . . . 48
3.1.4 Cluster Post-processing . . . 50
3.2 Fine Detection . . . 51
3.2.1 T-LBP Descriptor . . . 51
3.2.2 IT-LBP Descriptor . . . 54
3.2.3 SVM-based Text/non-text Classification . . . 56
3.2.4 Bounding Box Integration . . . 62
3.3 Experimental Results . . . 63
3.3.1 Comparison of Different LBP-based Features . . . 63
3.3.2 Results Obtaining by Using Public Dataset . . . 65
3.4 Summary . . . 67

4 Natural Scene Text Detection **69**
4.1 Character MSER Generation . . . 70
4.2 Character MSER Features . . . 74
4.2.1 Geometry-based Features . . . 75
4.2.2 Stroke-Based Features . . . 76
4.2.3 Histogram of Stroke Contour Point Gradient Direction . . . 80
4.2.4 Variance of Local Foreground/Background Colour Difference (VLF-BCD) . . . 81
4.3 Text MSER Retrieval . . . 82
4.3.1 Character MSER Retrieval . . . 84
4.3.2 Text Line MSER Retrieval . . . 85
4.4 Character MSER Grouping . . . 87
4.5 False Alarm Elimination . . . 89
4.6 Experimental Results . . . 91
4.7 Summary . . . 95

5 Text Binarisation **97**
5.1 Gray level-Based Text Binarisation . . . 99
5.1.1 Rationale . . . 104
5.1.2 The Mean-shift Algorithm . . . 107
5.1.3 Mean-Shift Based Channel Image Selection . . . 109
5.1.4 Graph-Based Selected Channel Image Segmentation . . . 113
5.2 Colour-Based Text Binarisation . . . 117

5.2.1 Selective Metric-based Clustering 118
5.3 Experimental Results of the Proposed Methods 123
5.4 Summary 127

6 Conclusions and Future Work **129**
6.1 Conclusions 129
6.1.1 Text Detection 130
6.1.2 Text Binarisation 131
6.2 Future Work 131
6.2.1 Refinement of Pattern Recognition Scheme 132
6.2.2 Parallel Computing for Reducing Running Time 134
6.2.3 Text Recognition 135

Author's Publication list **137**

Bibliography **139**

List of Tables

3.1 Comparison on classification performance of different LBP-based features. 65
3.2 Comparisons between our method and the algorithms in ICDAR2011 Robust Reading Competition Challenge 1 [1]. 66

4.1 Comparisons of classification accuracy rates of MSER classifiers using different sets of features. 92
4.2 Performance comparisons between our method and some state-of-the-art algorithms using the ICDAR2011 Robust Reading Competition Challenge 2 dataset [2]. 95

5.1 Comparison of character recognition rates 126

List of Figures

1.1 Text examples. 4
1.2 Three Steps of the TIE System. 5

2.1 The categorization of graph-based image segmentation techniques 26
2.2 (a) The illustration of constructed min-cut/max-flow model. (b) A cut on the constructed graph. (courtesy of [3]). 29
2.3 (a) Original image. (b) The segmentation result of minimum cut. (courtesy of [3]). 29
2.4 (a) Initialisation. (b) Segmentation result of minimum cut. (c) Segmentation result of topology cut (courtesy of [4]). 32
2.5 Illustration of the random walker algorithm. (a) Initialisation of seed points L1, L2, and L3. (b) Probability of a random walker starting from each node first reaches L1. (c) Probability of a random walker starting from each node first reaches L2. (d) Probability of a random walker starting from each node first reaches L3. (courtesy of [5]). 34
2.6 Results of minimum spanning tree-based image segmentation (courtesy of [6]). 37

3.1 The flow chart of the proposed method. 42
3.2 The computation equations of G_x and G_y. I denotes the source image, G_x and G_y are the horizontal and vertical derivative approximations. $*$ represents the convolution operation. 44

3.3 An example of Canny edge maps obtained with different thresholds. (a) The original colour image. (b) The gray-level image of (a). (c) The Canny edge map of (b) with high thresholds. (d) The Canny edge map of (b) with low thresholds. 45
3.4 An example of MGD map clustering. (a) An original image. (b) MGD map of (a). (c) the four clusters of the MGD map of (a). 47
3.5 The layer images $\textit{LayerImg}_i (i = 1, \cdots, 6)$ generated from Figure 3.13(c). . 48
3.6 (a) a 3×3 square-shape structuring element. (b) a 3×7 cross-shape structuring element. 49
3.7 Neighbour assignment for T-LBP computation. The shadowed pixels represent horizontal neighbourhood pixels of the central pixel P_c. 54
3.8 The local neighbourhood of IT-LBP at four directions. 55
3.9 Possible hyperplanes in a linearly separable case. The red and blue points represent training samples belonging to C1 and C2 respectively. The straight lines L1, L2 and L3 are capable to separate the points into two groups. . . . 58
3.10 Optimal SVM hyperplane having the maximum margin. 60
3.11 Linearly inseparable training data. Any straight lines cannot separate all of the red and blue points into the group they belong to. 61
3.12 Examples of training samples. (a) Positive samples. (b) Negative samples. . 64
3.13 Some text detection results by the proposed method (the detection results are shown in green bounding boxes). 68

4.1 The framework of the proposed natural scene text detection algorithm. . . . 71
4.2 MSER extraction results. MSER regions are marked in white and the remaining regions are marked in black. (a) Original images. (b) Bright-on-dark MSERs. (c) Dark-on-bright MSERs. 73
4.3 Character MSER samples. 74
4.4 Non-character MSER samples. 74
4.5 Stroke contour point gradient direction. (a) A pair of stroke edge points with opposite gradient directions. (b) Quantised gradient directions. 81

4.6 Character/non-character MSER classification. (a) Original scene images. (b) Dark-on-bright MSERs. (c) Bright-on-dark MSERs. (d) and (e) are the MSER classification results in (b) and (c) respectively. The MSERs that are classified as character are marked in white colour and the MSERs that are classified as non-character are marked in red. Best viewed in colour. . . 83
4.7 Single character MSER retrieval. (a) Classification result of MSERs where the MSERs classified as non-character are marked in red. (b) Single character MSER retrieval of (a). The retrieved MSERs are marked in green. Best viewed in colour. 86
4.8 Text line MSER retrieval. (a) Classification result of MSERs where the non-character MSERs are marked in red. (b) Text line MSER retrieval of (a). The retrieved text line MSER is marked in orange. Best viewed in colour. 87
4.9 Text line candidates false alarm elimination. (a) Text line candidates (enclosed by yellow bounding boxes) obtained from dark-on-bright MSER map. (b) Text line candidates (enclosed by yellow bounding boxes) obtained from bright-on-dark MSER map. (c) Final detected text lines (enclosed by green bounding boxes) after false alarm elimination. Best viewed in colour. 90
4.10 The procedure of bootstrap classifier training (courtesy Wei [7]). 91
4.11 Some scene text detection results using ICDAR2011 Robust Reading Competition Challenge 2 dataset by our algorithm. 94

5.1 An example of image segmentation. (a) The original image. (b) The segmentation result of (a). Sub-regions are represented by different colours (courtesy Zhang [8]). 99
5.2 An example of the histogram of a gray level text image. The figure on the right is the histogram of the gray level image on the left. The range of intensity values of a gray level is from 0 to 255. The horizontal axis of the histogram is the intensity values. The vertical axis of the histogram is the number of pixels belonging to a certain intensity value. 101

5.3 Different colours can be converted into an identical gray level value. 102
5.4 Colour channel split on RGB colour space. 102
5.5 The flow chart of our colour channel image-based text binarisation method. 105
5.6 Histograms of a clear text image (a) and a degraded text image (b). Sub-figures (c) and (d) are the histograms of the gray level images of (a) and (b) respectively. 106
5.7 The main peaks in different histograms of a colour text image. (a) The original colour image. (b1) The intensity map of (a). (c1), (d1) and (e1) are the images of R channel, G channel and B channel of (a) respectively. (b2), (c2), (d2) and (e2) are the histograms of (b1), (c1), (d1) and (e1) respectively. The red lines in each histogram indicates the locations of peaks (Best view in colour). 111
5.8 The two located local maxima in the histograms of the R, G and B channel images for a sample image (a). The curves shown in (b), (c) and (d) display the estimated density distributions (i.e. histograms) of R, G and B channel images respectively. The blue and red spots (best viewed in colour) are the positions of the two intensity values C_{peak1} and C_{peak2} (indicating the peaks of each histogram). 112
5.9 8-connected neighbourhood of a pixel with a value C_i. If a pixel with a value $C_j(j = 1, \cdots, 8)$ appears in the 8-connected neighbourhood of the pixel with a value C_i, then there is a co-occurrence pair between pixel values C_i and C_j. 115
5.10 The flow chart of our colour-based text binarisation method. 118

5.11 A comparison between the binarisation results obtained using M metric in [9] and that using the proposed M_{norm} metric with Euclidean Distance. (a) Original image. (b) 3-means clustering result with Euclidean Distance. Here, the green, red and blue colours represent the textual foreground cluster, background cluster and the noise cluster respectively. (c) Binarised text (in black) by using M. (d) Binarised text (in black) by using M_{norm}. (Best viewed in colour). 120
5.12 Simple cases. (a) Original image. (b) Otsu's method. (c) The method in [10]. (d) The proposed gray level-based method. (e) The proposed colour-based method. 123
5.13 Uneven lighting cases. (a) Original image. (b) Otsu's method. (c) The method in [10]. (d) The proposed gray level-based method. (e) The proposed colour-based method. 124
5.14 Complex background cases. (a) Original image. (b) Otsu's method. (c) The method in [10]. (d) The proposed gray level-based method. (e) The proposed colour-based method. 125
5.15 Highlight cases. (a) Original image. (b) Otsu's method. (c) The method in [10]. (d) The proposed gray level-based method. (e) The proposed colour-based method. 126
5.16 Other cases. (a) Original image. (b) Otsu's method. (c) The method in [10]. (d) The proposed gray level-based method. (e) The proposed colour-based method. 127

Abstract

As a vast amount of text appears everywhere, including natural scene, web pages and videos, text becomes very important information for different applications. Extracting text information from images and video frames is the first step of applying them to a specific application and this task is completed by a text information extraction (TIE) system. TIE consists of text detection, text binarisation and text recognition. For different applications or projects, one or more of these three TIE components may be embedded. Although many efforts have been made to extract text from images and videos, this problem is far from being solved due to the difficulties existing in different scenarios. This thesis focuses on the research of text detection and text binarisation.

For the work on text detection in born-digital images, a new scheme for coarse text detection and a texture-based feature for fine text detection are proposed. In the coarse detection step, a novel scheme based on Maximum Gradient Difference (MGD) response of text lines is proposed. MGD values are classified into multiple clusters by a clustering algorithm to create multiple layer images. Then, the text line candidates are detected in different layer images. An SVM classifier trained by a novel texture-based feature is utilized to filter out the non-text regions. The superiority of the proposed feature is demonstrated by comparing with other features for text/non-text classification capability.

Another algorithm is designed for detecting texts from natural scene images. Maximally Stable Extremal Regions (MSERs) as character candidates are classified into character MSERs and non-character MSERs based on geometry-based, stroke-based, HOG-based and colour-based features. Two types of misclassified character MSERs are retrieved by two different schemes respectively. A false alarm elimination step is performed for increasing the text detection precision and the bootstrap strategy is used to enhance the power of suppressing false positives. Both promising recall rate and precision rate are achieved.

In the aspect of text binarisation research, the combination of the selected colour channel image and graph-based technique are explored firstly. The colour channel image with the histogram having the biggest distance, estimated by mean-shift procedure, between the two main peaks is selected before the graph model is constructed. Then, Normalised cut is employed on the graph to get the binarisation result. For circumventing the drawbacks of the grayscale-based method, a colour-based text binarisation method is proposed. A modified Connected Component (CC)-based validation measurement and a new objective segmentation evaluation criterion are applied as sequential processing. The experimental results show the effectiveness of our text binarisation algorithms.