Faculty of Engineering and Information Technology
University of Technology, Sydney

Detecting Text in Clutter Scene

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

by

Xia Cui

November 2014

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Abstract

We often encounter cluttered visual scenes and need to identify objects correctly to navigate and interact with the world. As text takes the typical form of a human-designed informative visual object, retrieving texts in both indoor and outdoor environments is an important step towards providing contextual clues for a wide variety of vision tasks. Furthermore, it plays an invaluable role for multimedia retrieval and location based services.

Text detection from clutter background is nevertheless a challenging task because the text, being figures in image, can be presented in various ways with lots of room for uncertainty such as size, scale, font type, font texture and colour, unpredicted decorative elements put on the text, etc. The situation will be even more complicated if the text is presented in a clutter background where non-text objects possess similar low-level features to text. Further, all these objects are composed of distinct geometric shapes and they are similar with the essential composition elements of text objects. Pursuing a robust text feature descriptor is therefore always difficult because special feature descriptor is only a fragment of text existence. It needs the completely understanding of text.

Regarding the design, understanding, representation and calculating of text as one unitary process of text perceiving, we deal with the completely understanding and representation of text in image with many kinds of aspects in different levels. Without following the legend feature based solution, this research is motivated by perceptual image processing and the observation of painting masters. It will explore a brand new solution by investigating the spatial structure of text and the compositional complexity of the visual object (i.e. text) in image. The research will present the composition granularity indicator and expose novel discriminable attributes embedded inside text objects, which can successfully differentiate text regions and non-text regions on clutter backgrounds.

As figures in image with the clutter scene, it is merely the physical appearance of text which provides the perceptual content and plays a central role for text detection, i.e. location and coarse identification. During the view-construction of text, properties of individual character and textual organization of characters build up the physical appear-

ance. When observers see text appearance in clutter scene, they describe their feelings in terms of crowding effect and clutter. However, the appearance of text still has enough saliency to reveal an informative message. Accordingly, text not only has the characteristics of crowding effect and clutter but also follows the principles of saliency.

Significantly, the crowding effect of text is derived from the space regularity of in-built neighbouring letters which have commonalities beside their distinctiveness. In addition, low-level features of individual letters contribute to the commonalities and distinctiveness from the moment that the font is designed.

Therefore, the computational model of text appearance is built up to integrate the three-level properties, including features of individual characters (low-level features), properties for spatial regularity (i.e. neighbourhood, appearance similarity), and the crowding statistics property of space averaged over pooling regions.

In terms of image processing, if we consider the view construction of text, the features of individual characters in image processing are obtained on the basis of the properties of construction, including mean intensity, local RMS contrast, shape, pixel density, edge density, stroke width, straight line ratio, height to width ratio, stroke width to height ratio, etc.

For the purpose of calculating the properties of space regularity and the crowding space averaging property, the spatial elements and relations are quantified and these involve space granularity and composition rules.

If we examine the works of painters, especially impressionists, they use directional brushstroke or colour patches as space granularity to represent “formless” visual objects in space regularity instead of clear contour shape sketches. The space regularity of patches, i.e. repetitive patterns, can offer a compositional format to express an artist’s feelings about an object rather than to simply describe it. Secondly, it is the harmonious proportions among component parts that bridle component space patches into objects. If we consider the painter’s harmonious proportions, the component parts of an object can be said to react simultaneously so that they can be seen at one and the same time both together and separately.

Similarly, image is described by a set of grey space patches in multi-grey levels. In addition, each space patch groups pixels in position proximity and similarity, in just the same way as the colour patch is used by impressionists. The space organisation of them is also quantified as the measurement of space relations, especially in terms of the

neighbourhood and proportions among component parts. Moreover, the harmonious proportions among space patches are captured by the mathematical tool of geometric mean. Geometric mean (i.e., GM) is calculated over those space patches which possess the same grey level, and considered as the space granularity to form objects. Grey patches with the same GM are composed of GM regions, which are enlarged, extended kinds of pooling regions. Regions given by clusters which have resulted from similarity and neighbourhood are direct, compact pooling regions. Therefore, the statistical properties of space averaging are calculated over GM regions and image is represented as a set of GM regions over which text and other visual objects are analysed by GM indication.

Finally, the representation of an image and the three-level computational text model are put into practice to develop a new-brand algorithm on the public benchmark dataset and to design and implement an automatic processing system on the real big data of the bank cheque. The resulting performance of these tools/processes shows that they are highly competitive and effective.

Acknowledgement

First and foremost I want to thank my supervisor Professor Longbing Cao. It has been an honour to be his Ph.D. student. He has taught me, both consciously and unconsciously, how good big data is done. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am also thankful for the excellent example he has provided as a successful leader scientist and professor.

And I also want to thank my supervisor Dr. Qiang Wu. It has been a great gratitude to be his Ph.D student. He has taught me how good and joyful the thinking ability is done for research as a vigorous and rigorous scientist. I appreciate all his contributions of time and ideas to make my Ph.D full of creative thinking and enjoyment. The endless creative thinking and enthusiasm he has for his research encouraged me to make greater efforts. I am also thankful for the excellent example he has provided as a time manager.

The members of the AAI group have contributed immensely to my personal and professional time at UTS. The group has been a source of friendships as well as good advice and collaboration. I am especially grateful for the fun group of original AAI group members who stuck it out in grad school with me: Ziye Zuo, Wei Wei, Jiahang Chen, Yuming Ou, Zhigang Zheng, XinHua Zhu, Chao Luo, Junfu Yin, Xuhui Fan, and SongYin. I would like to acknowledge honorary group member QiaoYu Sun who was here a couple of years ago. We worked together along with Qiang Wu on the i-Cheque system development, and I am very much appreciated her enthusiasm, intensity, and amazing ability to manipulate text images. And I am also grateful for these honorary group members who have come through the lab: Hongxiu Zhu, Wei Li, Xiaodong Yue. Other past and present group members that I have had the pleasure to work with or alongside of are grad students Jingjiu Li, Mu Li, Can Wang, She Zhong, Chunming Liu, DongYu, JinSong, Pochung Zhang, RenJing, Renhua Song and the numerous visitors and rotation students who have come through the lab.

When I started my PH.D from the cheque document image processing, I have appreciated the camaraderie and local expertise of bank as well as the AAI group early on. In regards to the system design and development, I have appreciated David for his collaboration and the impressive skills.

In my attempted measurements of the crowding structure in an image with clutter scene, I thank the following people for helpful discussions with us: JinSong Xu, Song Yin, DongYu. I thank for their inspirational discussions with us. And I would also like to acknowledge the AAI Facility.

For this dissertation I would like to thank my reading committee members for their time, interest, and helpful comments. I would also like to thank members of my DA defense committee, Jian Zhang, Jinyan Li, Weichang Ye, for their time and insightful questions.

I am grateful to our group's past and present administrative assistant: Lei Zheng, YangBo, Collin Wise and Rayee who kept us organized and was always ready to help. And I am especially great grateful for and touched by the kindness and very generous support from Dr. Dan Luo during tough times.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by the Australian IPRS Scholarship and APA for my first 3 years.

My time at UTS was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for time spent with roommates and friends, especially for elder friends as I made big decision for my life, and for many other people and memories. My time at UTS was also enriched by the AAI sport group, Omma's Church and Campsie Congregation.

Lastly, I would like to thank my family for all their love and encouragement. I am very grateful for my parents who raised me with a love of science and supported me in all my pursuits, especially for the presence of my parents here for three of my years here to take care of my kids. For my kids who safeguard my heart and restore my courage. And most of all for my loving, supportive, encouraging, and patient husband whose faithful support during the final stages of this Ph.D. is so appreciated. Thank you.

Xia Cui

The University of Technology, Sydney

June 2014

Table of Contents