Single and Multiple Instance Learning for Visual Categorisation

Ruo Du

Faculty of Engineering and Information Technology

University of Technology, Sydney

A thesis submitted for the degree of

Doctor of Philosophy

2013

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

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 Author

Production Note:
Signature removed prior to publication.

This thesis is dedicated to my wife and my parents.

For their endless support and encouragement

Acknowledgements

First and foremost I would like to sincerely thank my principal supervisor, Prof. Xiangjian HE, for his guidance, understanding, patience and help on scholarship to make my PhD experience such a rewarding and exciting journey.

I also want to express my utmost gratitude to my co-supervisor, Dr. Qiang Wu, who spent enormous time and energy on me as a mentor and a friend. His guidance was embedded in every step of my studies.

Special thanks also to Dr. Wenjing Jia, Dr. Min Xu, Prof. Massimo Piccardi and Dr. Richard Xu for their great suggestions, knowledge sharing and invaluable assistance.

I wish to thank my fellow research students of our team, Muhammad Hasan, Man Wong, Chao Zeng, Sheng Wang, Minqi Li, Thomas Tan, Aruna Jamdagni, Ying Wan and Mohammed A AmbuSaid, for their assistance and friendships.

Finally, and most importantly, I would like to thank my wife Ting Zhou. Her faith in me, support, encouragement and quite patience made it possible that I could even continue to pursue my PhD after working in IT for ten years. I thank my parents, Changyou Du and Shenghua Liu, and my parents in law, Ruguo Zhou and Shijie Li, for their help and support as always.

Abstract

Nowadays, huge amounts of visual data, e.g., videos and images, have become widely accessible. Therefore, intelligently categorizing the large and growing collections of data for access convenience has been a central goal for modern computer vision research. In this thesis, we describe several newly-developed approaches for visual categorization upon the single and multiple instance learning cases.

In single-instance learning (SIL), each of the training instances has been labeled. Here, we focus on a challenging task of facial expressions recognition where manually labeling each training instance, i.e., face video, is handy. To get the distinct features of expressions, we propose a novel feature representation, Histogram Variances Face (HVF), which integrates dynamic expression information into a static image being invariant to illumination and in-plane rotation. Through HVFs, the facial expression recognition can be cast as a facial recognition problem. We have applied our approach on the well-known Cohn-Kanade AU-Coded Facial Expression database, and then those extracted HVFs are classified by using facial recognition technology, i.e., Eigenfaces and Support Vector Machines (SVMs). The recognition accuracy is very encouraging. We further propose an extension of HVFs, Hexagonal Histogram Variance Faces (HHVFs), which applies HVFs on a hexagonal structure. Comparing to HVFs, HHVFs not only greatly reduce the computation costs but also improve the recognition accuracy.

In multiple-instance learning (MIL), the training instances are divided into groups and the instances in the same group share only one label. MIL arises from many applications where individually labeling training instances is expensive. In this case, we propose a novel algorithm,

multiple-instance learning with a supervised kernel density estimation (MIL-SKDE), to tackle the labeling ambiguity. Our algorithm extends the twin technologies, kernel density estimation (SKDE) and mean shift, to their supervised versions in which the labels of data points will affect the mode seeking. We apply MIL-SKDE in several applications of visual categorization, e.g., image and object categorization, and our algorithm performs superiorly comparing to other state-of-the-art methods. Furthermore, to address the complexity issue of MIL-SKDE, we propose MIL-SS (MIL with speed-up SKDE) to speed up the training process. Experiments shows that it has comparable performances to MIL-SKDE but is much more efficient in training stage.

Finally, we apply MIL-SS in a "bag-of-words" (BoW) system to learn the visual codebook for object categorization on a more comprehensive dataset. Our system consists of four steps: codebook generation, feature coding, feature pooling and classification. Unlike conventional BoW methods that learn codebook from the whole image areas, our method can learn codebook just from the areas of target objects, which significantly improves classification accuracy.

Author's publications for the Ph.D

Journal paper:

1. **Ruo Du**, Qiang Wu, Xiangjian He, and Jie Yang. "MIL-SKDE: Multiple-instance learning with supervised kernel density estimation". *Signal Processing*, Volume 93, Issue 6, June 2013, Pages 1471-1484

Conference papers:

2. **Ruo Du**, Qiang Wu, XiangJian HE, and Jie Yang. "Object categorisation based on a supervised mean shift algorithm". In *12th European Conference on Computer Vision (ECCV) Demos*, Part III, LNCS 7585. Springer, Heidelberg, 2012.

3. **Ruo Du**, Qiang Wu, XiangJian HE, and Jie Yang. "Multi-instance learning with an extended kernel density estimation for object categorisation". In *IEEE International Conference on Multimedia and Expo (ICME) Workshops*, pp.477-482, 9-13 July 2012.

4. Lin Wang, Xiangjian He, **Ruo Du**, Wenjing Jia, Qiang Wu, and Wei-Chang Yeh. "Facial expression recognition on hexagonal structure using lbp-based histogram variances". In *17th International MultiMedia Modeling Conference (MMM)*, pages 35 - 45, 2011.

5. **Ruo Du**, Sheng Wang, Qiang Wu, and Xiangjian He. "Learn concepts in multiple-instance learning with diverse density framework using supervised mean shift". In *Digital Image Computing: Techniques and Applications (DICTA) - Oral presentation*, pages 643-648, 2010.

6. Sheng Wang, **Ruo Du**, Qiang Wu, and Xiangjian He. "Adaptive stick-like features for human detection based on multi-scale feature fusion scheme". In *Digital Image Computing: Techniques and Applications (DICTA)*, pages 375-380, 2010.

7. **Ruo Du**, Qiang Wu, Xiangjian He, Wenjing Jia, and Daming Wei. "Facial expression recognition using histogram variances faces". In *Workshop on Applications of Computer Vision (WACV)*, pages 1-7, 2009.

Contents

List of Tables

List of Figures