C02029: Doctor of Philosophy
CRICOS Code: 00099F
May 2019

Video-based similar gesture action recognition using deep learning and GAN-based approaches

DI WU

School of Computer Science
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

Video-based similar gesture action recognition using deep learning and GAN-based approaches

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

Doctor of Philosophy

by

DI WU

to

School of Computer Science
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

May 2019

Abstract

Human action is not merely a matter of presenting patterns of motion of different parts of the body, in addition, it is also a description of intention, emotion and thoughts of the person. Hence, it has become a crucial component in human behavior analysis and understanding. Human action recognition has a wide variety of applications such as surveillance, robotics, health care, video searching and human-computer interaction. Analysing human actions manually is tedious and easily prone to errors. Therefore, computer scientists have been trying to bring the abilities of cognitive video understanding to human action recognition systems by using computer vision techniques. However, human action recognition is a complex task in computer vision because of the camera motion, occlusion, background cluttering, viewpoint variation, execution rate and similar gestures. These challenges significantly degrade the performance of the human action recognition system. The purpose of this research is to propose solutions based on traditional machine learning methods as well as the state-of-the-art deep learning methods to automatically process video-based human action recognition. This thesis investigates three research areas of video-based human action recognition: traditional human action recognition, similar gesture action recognition, and data augmentation for human action recognition.

To start with, the feature-based methods using classic machine learning algorithms have been studied. Recently, deep convolutional neural networks (CNN) have taken their place in the computer vision and human action recognition research areas and have achieved tremendous success in comparison to traditional machine learning techniques. Current state-of-the-art deep convolutional neural networks were used for the human action recognition task. Furthermore, recurrent neural networks (RNN) and its variation of long-short term memory (LSTM) are used to process the time series features which are handcrafted features or extracted from the CNN. However, these methods suffer from similar gestures, which appear in the human action videos. Thus, a hierarchical classification framework is proposed for similar gesture action recognition, and the performance is improved by the multi-stage classification approach. Additionally, the framework has been modified into an end-to-end system, therefore, the similar gestures can be processed by the system automatically.

In this study, a novel data augmentation framework for action recognition has been proposed, the objective is to generate well learnt video frames from action videos which can enlarge the dataset size as well as the feature bias. It is very important for a human action recognition system to recognize the actions with similar gestures as accurately as possible. For such a system, a generative adversarial net (GAN) is applied to learn the original video datasets and generate video frames by playing an adversarial game. Furthermore, a framework is developed for classifying the original dataset in the first place to obtain the confusion matrix using a CNN. The similar gesture actions will be paired based on the confusion matrix results. The final classification result will be applied on the fusion dataset which contains both original and generated video frames. This study will provide realtime and practical solutions for autonomous human action recognition system. The analysis of similar gesture actions will improve the performance of the existing CNN-based approaches.

In addition, the GAN-based approaches from computer vision have been applied to the graph embedding area, because graph embedding is similar to image embedding but used for different purposes. Unlike the purpose of the GAN in computer vision for generating the images, the GAN in graph embedding can be used to regularize the embedding. So the proposed methods are able to reconstruct both structural characteristics and node features, which naturally possess the interaction between these two sources of information while learning the embedding.

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Di Wu* declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science* and *Centre for Artificial Intelligence*, *Faculty of Engineering and Information Technology* at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

SIGNATURE: Production Note:
Signature removed prior to publication.
[Di Wu]

DATE: 06th May, 2019

PLACE: Sydney, Australia

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my supervisors Prof. Michael Blumenstein (Principal supervisor) and Dr Nabin Sharma (Associate Supervisor) for their fruitful guidance and support throughout my PhD candidature. Their encouragement, interactive regular meetings have made my experience very productive. I am especially thankful to Prof. Michael for his support during the stressful time when I was under highly research pressure and also for my transfer of PhD studies from Griffith University to the University of Technology Sydney. I am also especially thankful Dr Nabin Sharma for taking his time help me to improve my research papers and take care of my research progress.

I am thankful to Griffith University for giving me the opportunity to pursue my PhD studies by granting me scholarships to cover my tuition fee and living expenses. I am thankful to the School of ICT, IIIS and GGRS of Griffith University for their support during the first year of my PhD.

I am thankful to UTS for providing me with the scholarship for the remaining period of my PhD. I cannot forget to mention the state-of-the-art facilities for doing my experiments. I express my gratitude to all staff at the School of Computer Science, CAI, GRS and FEIT staffs of UTS for their assistance. I am lucky to find great friends. I am grateful to Dr Ruiqi Hu, Junjun Chen, Jiale Zhang, Muhammad Saqib, Dr Ranju Mandal, Chandranath Adak and Fujin Zhu for their support and great company.

Finally, I would not have been able to focus on my PhD studies without the constant support, especially the support from my wife, grand father, grand mother and my parents.

List of Publications

Conference :

1. Wu, D., Sharma, N. and Blumenstein, M., 2017, May. Recent advances in video-based human action recognition using deep learning: a review. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 2865-2872). IEEE.

2. Wu, D., Sharma, N. and Blumenstein, M., 2018, December. Similar Gesture Recognition using Hierarchical Classification Approach in RGB Videos. In 2018 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-7). IEEE. **Research Question 1**

3. Wu, D., Sharma, N. and Blumenstein, M., 2018, November. An End-to-End Hierarchical Classification Approach for Similar Gesture Recognition. In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE. **Research Question 2**

4. Wu, D., Chen, J., Sharma, N., Pan. S., Long, G. and Blumenstein, M., 2019. Adversarial Action Data Augmentation for Similar Gesture Action Recognition. In 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. (Accepted) **Research Question 3**

5. Wu, D., Hu, R., Zheng, Y., Jiang. J., Sharma, N. and Blumenstein, M., 2019. Feature-Dependent Graph Convolutional Autoencoders with Adversarial Training Methods. In 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. (Accepted)

6. Zhang, J., Chen, J., Wu, D., Chen, B., and Yu, S., 2019. Achieving Poisoning Attack in Federated Learning using Generative Adversarial Nets. International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE. (Accepted)

7. Zhao, Y., Chen, J., Zhang, J., Wu, D., Teng, J., and Yu, S., 2019. Using Generative Adversarial Network to Defend Poisoning Attacks in Federated Learning. In 2019 International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP). Springer. (Accepted)

JOURNAL :

1. Wu, D., Chen, J., Sharma, N., Zhang, J., Zhang, Q. and Blumenstein, M., 2019. An End-to-End Adversarial Video Data Augmentation Framework for Similar Gesture Action Recognition. Expert Systems with Applications (Under Review), **Research Question 3**

2. Chou, K.P., Prasad, M., Wu, D., Sharma, N., Li, D.L., Lin, Y.F., Blumenstein, M., Lin, W.C. and Lin, C.T., 2018. Robust Feature-Based Automated Multi-View Human Action Recognition System. IEEE Access, 6, pp.15283-15296.

3. Chen, J., Wu, D., Zhao, Y., Sharma, N., Blumenstein, M., and Yu, S., 2019. Fooling Intrusion Detection Systems by Using Adversarially Regularized Autoencoder. Digital Communications and Networks, (Under Review)

4. Zhao, Y., Chen, J., Wu, D., Sharma, N., Sajjanhar, A., and Blumenstein, M., 2019. Network Anomaly Detection by Using Time-decay Closed Frequent Pattern. Information, 10(8), p.262.

Table of Contents

List of Figures

Figure **Page**

LIST OF TABLES

TABLE Page