Independent Subspace Analysis for Activity Recognition and Fine-grained Classification

A THESIS

submitted by

Munender Kumar
(Roll No : S13002)

for the award of the degree of

Master of Science
(by Research)

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi
India - 175005

April 2017

Declaration

I hereby declare that the entire work embodied in this thesis is the result of the investigations carried out by me in the **School of Computing and Electrical Engineering, Indian Institute of Technology Mandi**, under the supervision of **Dr. Renu M. Rameshan**. This work has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on finding of other investigators. In addition, I certify that no part of this work will, in future, be used for submission in my name, for the award of any other degree at any university.

Place: Mandi

Date:

Munender Kumar

Thesis Certificate

This is to certify that the thesis titled **"Independent Subspace Analysis for Activity Recognition and Fine-grained Classification"**, submitted by **Munender Kumar**, at Indian Institute of Technology Mandi for the award of Master of Science (by research) is a bonafide record of the research work carried out by him under my supervision. The content of this thesis, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Dr. Renu M. Rameshan
(Thesis Supervisor)
School of Computing and Electrical Engineering
Indian Institute of Technology Mandi
Himachal Pradesh - 175005

Dedicated to

My Family, My Friends

and

My Teachers

Acknowledgement

First of all, I would like to express my sincere gratitude to my advisor Dr. Renu M. Rameshan and former advisor Dr. Sukumar Bhattacharya. Their guidance, support and patience have made this research possible. Next, I would like to thank Dr. Arnav Bhavsar for helping and supporting me in tough situations.

I also want to express my sincere thanks to Prof. Timothy A. Gonsalves and to all committee members, Dr. Anil K. Sao, Dr. Prasanth P. Jose, Dr. Arnav Bhavsar and Dr. Padmanabhan Rajan for their support, valuable advice and excellent facilities at MAS Lab for my research work. I would also like to thank Dr. Aditya Nigam for his kind support.

I thank my room-mate, hostel-mates, AMRC and MAS lab - colleagues, institute staff and all well-wishing friends for sharing their time and making this journey unforgettable. With you all it always felt like having another sweet family away from home. Thank you for motivating me to always look forward.

Last but not least, I thank my parents, siblings and my little cousins Gungun, Khushi, Saumya, Arnav, Aditya for their patience, encouragement, love and support.

Munender Kumar

ABSTRACT

Visual recognition is a challenging problem which depends on the discriminative nature and robustness of the features used in recognition techniques. These techniques are mainly focused on adapting hand-designed local features such as SIFT, HOG, k-NN, and SURF etc., which are not scalable to other modalities. Hence there is a paradigm shift from hand-designed local features to unsupervised learning in order to extract features directly from the raw data. Visual signals (images) can be modeled using independent subspace analysis (ISA), an extension to general ICA model, which gives invariant features. ISA has been extended for large dataset to delivers hierarchy of features using convolution and stacking multiple layers of ISA over each other. Albeit performance is good, it takes significant amount of time on large datasets due to high computational complexity and sequential implementation. Two different methods are proposed to speed up feature learning in multilayered ISA. First method for faster feature learning uses parallelization present in the data. MapReduce, a scalable programming model, is used to parametrize ISA model using multiple map-reduce functions over the equal disjoint sets of distributed data. The second method for increasing speed uses spatio-temporal interest point detectors to extract important blocks from video which removes irreverent video blocks. The latter not only enhances the speed but also improves the classification accuracy. Different input level modifications are also proposed which increases the classification performance. A dataset is also created for human-water activities for surveillance purpose near water bodies and the ISA network is applied over it for feature extraction and classification.

Multilayered ISA is used to extract features for fine-grained recognition of similar objects *i.e.,* categorizing various types of leaves, butterflies and birds into their subcategories like breeds and species. This architecture has three ISA layers to extract features from the large image patches. The process convolves learned filters over a large spatial region (image patch) which are learned by applying ISA on small size image patches. Further, discriminative patches are used to train ISA network which correspond to SIFT points and has optimal size based on classification accuracy. Addition of more ISA layers increases the percentage of true-positives

significantly enough while our computational cost is not affected due to the reduction in data size. The proposed approach is tested over leaf, butterfly and bird dataset. Most of the techniques applied on the leaf was focused on structural features since leaves have fine edges. These fine edges are enhanced by applying contrast limited adaptive histogram equalization (CLAHE) on the leaf images. The hybrid technique which work best for leaf dataset is wavelet transform of patches taken around SIFT key points of the enhanced image. It should be hypothesized that adding another ISA layer captures large spatial region and hence gives the complex structure present there. All this together improves percentage of true-positives in the classification by a significant amount.

Features learned from ISA are also used for action recognition in RGB and depth videos where cuboids are extracted around spatio temporal interest points after normalizing the frame size of different videos. The resulting cuboids are concatenated for training multilayered ISA model with two layers. Different dataset are used for testing the framework such as MSR-Action3D, MSRDailyActivity3D, UTD multimodal human action datasets having 20, 16 and 27 activities respectively.

Contents