# Learning Hierarchical Invariant Spatio-Temporal Feature for Action Recognition with Independent Subspace Analysis

## Original Paper authors:

Quoc V. Le, Will Y Zou, Serena Y. Yeung and Andrew Y. Ng

## Original Paper link:

<http://ieeexplore.ieee.org/document/5995496/>

## Summary of Paper:

This report uses an unsupervised method of feature learning with the use of independent subspace analysis (ISA) and principle component analysis (PCA). The model takes an input of unlabeled raw data and then uses deep learning methods such as stacking and convolution to perform a twostep process of classification.

Initially, ISA is trained on small input patches using both ISA and PCA to provide sequences. These sequences are then convolved with the larger region of the input data, the output is then fed into the layer above. This convolutional stacking enables hierarchical interpretation of the data suitable for recognition. The stacked model is trained greedily, more specifically, the first layer is trained until convergence before training layer two. Using this idea, the training time required is 1-2 hours.

The model uses ISA due its invariance property. Finally, it does not require the use of hand-designed features such as SIFT and HOG.

The use of two-step convolutional stacking, provided results that are competitive to the other models with an accuracy of 93.9% with the KTH dataset.

## Relevance of Paper to our Project:

For application to the project at hand, this method would be suitable, as it’s unsupervised and performs on raw unlabeled data. These properties lend themselves to the task of a non-expert user, performing the machine learning.

This method considers spatio-temporal features, which are required for our chosen application. Allowing for classification of time dependent events, such as reaching for a door, or eating food. The use of twostep convolution reduces the time required to process the data, which is key for a user processing this data at home.

Overall this model performs the image classification within a suitable time frame, takes into consideration the correct dimension of data and provides a fully automated model acceptable for a user.

# Sequential Deep Learning for Human Action Recognition

## Original Paper authors:

Moez Baccouche, Franck Mamlet, Christian Wolf, Cgristophe Garcia and Atilla Baskurt

## Original Paper link:

<https://link.springer.com/chapter/10.1007/978-3-642-25446-8_4>

## Summary of Paper:

This report uses a fully automated deep learning model that extends convolutional neural networks (ConvNet) to 3D domains to process data that has no prior modelling. A recurrent neural network is trained to classify sequences based off temporal evolutions. It does this by using the KTH dataset. It uses the identified features to classify the full sequence.

This model uses a twostep process to classify raw data inputs. Initially it uses a 3D extension of ConvNets to identify features of the data. It then uses the learned features to classify the entire sequence using a recurrent neural network. Furthermore, to capture “real world” sequences the commonly used short term memory becomes insufficient, thus, they use long short term memory (LSTM) to categorise the action sequences.

The use of 3D ConvNets, recurrent neural networks and LSTM, provide a model that can accurately classify the KTH1 dataset to an accuracy of 94.39%. This is competitive among other work.

## Relevance of Paper to our Project:

Through the extension of 2D convnets to 3D, this model has the capabilities to process spacio-temporal sequences, such as the ones required by our project. Furthermore, with the addition of LTSM, this model has the capability to handle “real world” action sequences, unlike some other alternatives.

The model uses a fully automated deep learning process which doesn’t require supervision, this is a necessity for our project as it will be handle by a non-expert. It also has the advantage of taking raw data as an input which is required for the video feed from the chosen camera.

Overall this model performs all the required tasks, however it takes a longer amount of time then some of the other competitors. The report mentions the possibility of reducing it to a single step model which could considerably reduce computational time, the main difficulty will be adaption of the training algorithm, especially when calculating the retro-propagated error.

# Change-Point Detection with Feature Selection in High-Dimension Time-Series Data

## Original Paper authors:

Makoto Yamada, Akisato Kimura, Futoshi Naya and Hiroshi Sawara

## Original Paper link:

<https://pdfs.semanticscholar.org/b25e/19fcb056c2346e3f84875f1b99de519920ad.pdf>

## Summary of Paper:

The report proposes a supervised change-point detection model using the additive Hilbert Schmidt independence criterion (aHSIC). The aHSIC allows for feature selection, this provides for applications in high-dimensional time-series change point detection, as well as providing a more robust environment for dealing with noisy features.

This model is the base of which many other models are designed.

## Relevance of Paper to our Project:

This model has applications to high-dimensional time-series data however, produces false positives if there is an independent feature with high separability.

This model would perform well in situations where a lot of noise is present. Other than this the other options seem the better options.

Overall this model seems to be the base of other models, which has been refined in other papers.