Cathy 07/08~09/08/2017

There are two parts in my presentation today, the first part is trying to understand what the goal is, and the second is the possible directions that I want to try. So please correct me if I am wrong and please advise if you have ideas.

**Part 1. What we are trying to achieve now?**

**Part 2. How to train a model based on the labelled videos?**

What Dr. Kim did is: recognize a movement in the video, mark the process of the movement as, for example, ‘forward’ from the start and end point.

Definition of a movement, can a movement be defined by:

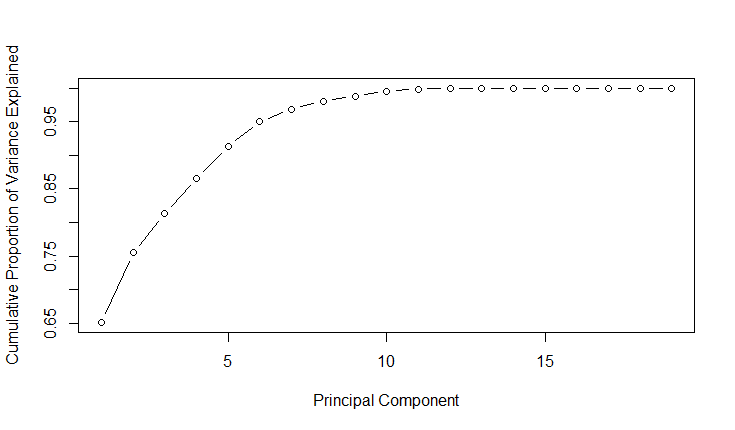
1. One frame/point (Ian’s current work)
2. 2 frames/points

Assumption: If a process is marked as ‘forward’, then the difference between two neighboring frame is also showing a ‘forward’.

I had a quick try: using the N2\_nf4 video, and calculate the difference of  **ith** and ( **i-1)th** frames, 2\_frames\_diff dataset, perform PCA for feature reduction, then classification(RF&SVM) with train/test: 75/25.

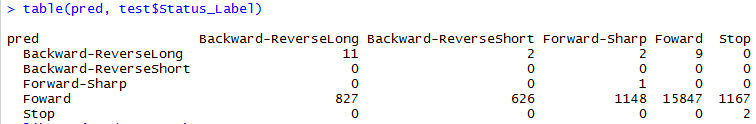
Feature reduction:

Perform PCA on the selected 19 variables (19 variables out of 50, dropping the reference variables and variables with one-value/not-correlated with Status\_Label). Based on the variance explained, 8 components are kept, explaining 98% of the variance in the 19 variables.

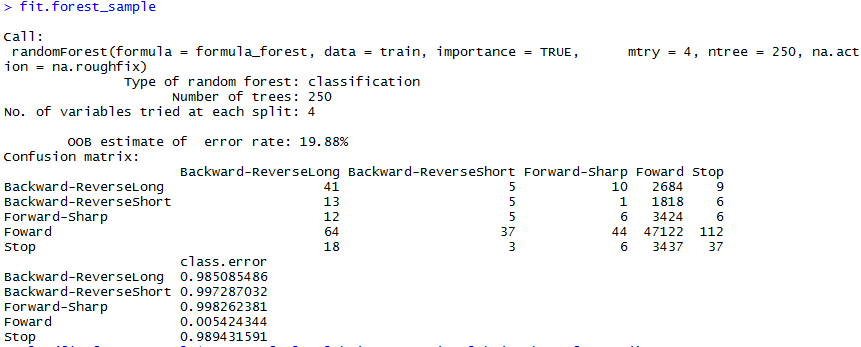


Classification result:

1. Support vector machine:



1. Random forest result:



Summary: the prediction are mainly in the movement ‘Forward’, as ‘Forward’ is the majority movement labelled in the dataset. Also, the reason could be that movement is a process, and cannot be defined by one or two frame.

1. An ordered sequence of frames/points?

This assumption makes sense as a movement process is an ordered data sequence of images/extracted features. And this is also how data scientists define human’s movement like walking, dancing, and sitting.

Possible approach:

1. Feature extraction form the n-frame ordered sequence:

For example, “The 5-s signal segments are used for extracting a total of 26 features: the minimum and maximum values, the mean value, variance, skewness, kurtosis, the first ten values of the autocorrelation sequence, and the maximum five peaks of the discrete Fourier transform of a segment with the corresponding frequencies”[3]. (http://dl.acm.org/citation.cfm?id=1823314)

* **The number of frames to be used: n=?**

How many frames does Dr. Kim use to define a movement?

For example, N2\_nf4, the five-num summary for each movement’s frames number:

"Forward-NTD"



"Forward-Sharp"



"Forward-Shallow"



"Backward-ReverseLong"



"Backward-ReverseShort"



"Stopped-ReverseLong"



"Stopped-Stop"



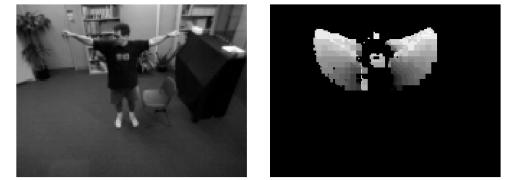
"Stopped-ReverseShort"



* **Which variables to extract value from?**

(other solutions) To be continued….

1. Feature creation: code the **instantaneous** position and velocity of the tracked subject, which is instantaneous position estimate (first sub-group) and **averaging** the velocity and speed estimates over an interval of T frames (second sub-group)[4].( <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.6737&rep=rep1&type=pdf> )
2. Image sequence: Construct a vector-image that can be matched against stored representations of known movements; this image is used as a temporal template[5]. the MEI(Motion energy image) and MHI(motion history image) can be considered as a two component version of a temporal template. (<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=910878>)



1. HMM-based segmentation and recognition of activities from video sequences

References:

[1] human activity classification with miniature inertial and magnetic sensor signals

[2] Scaling to very large corpora for natural language disambiguation

[3] Comparative study on classifying human activities with miniature inertial and magnetic sensors

[4] Human Activity Recognition from Video: modeling, feature selection and classification architecture