

Initial Report

1. Get Training Data Set prepared.

The training data would be a set of videos including many different type of action. Since the project aims at classify the action, the training data set has to be large enough otherwise it might lead to a poor classification accuracy.

2. Transfer the data into feature domain.

To transfer the data from spatial domain to feature domain, the initial idea is to apply SIFT. Although it takes a long time to compute the descriptors of an image, it still is a productive way to extract the key-points from a given image.

The algorithm we use here might be changed in the future because of the massive training data. Since adequate training data are required, it is obvious that using SIFT will takes a lot of time on both training and testing. An algorithm which is fast but finding fewer key-points might be applied in actual implementation.

3. Train the Machine.

In industry, various deep learning architectures such as deep neural networks, convolutional deep neural networks and recurrent neural networks have been applied to fields like automatic speech recognition, natural language processing and computer vision.

Therefore, this project chooses to use convolutional deep neural networks for training the classifier. Given that our training data set is a set of videos, convolutional deep neural networks might be the most effective tools for this occasion.

4. Test the Machines by Testing Data Set.

After training the machine, the classifier can be tested by a testing data set and compute the performance of the classifier.

5. Improve the Classification Accuracy.

In this phase, we need to go back to the stage 3 and change the parameters in order to improve the classification accuracy. For example, by changing the parameters in the algorithm can avoid overfitting classifier, thus the performance will be improved.

Also, changing the structure of the network might be necessary in this phase.