CS 6476 Computer Vision

Activity Classification using MHI

Yue Cai

yuecai@gatech.edu

Introduction

In this assignment, I will implement an activity recognition method based on Motion History Images (MHIs) and use this method to perform a knn activity classification. To see the result, we will get a confusion matrix and video in the end.

Methods

Data

In this project, our target is to recognize 6 actions, boxing, handclapping, handwaving, jogging, running, and walking. I have available 6 actions video. These videos are mainly under 1 minute, no more than 500 frames. I obtain 4 sequences on video. We separated the data set with 70 percentage training and 30 percentage testing.

Dataset and sequence file can be obtained via: Recognition of human actions (kth.se)

MHI

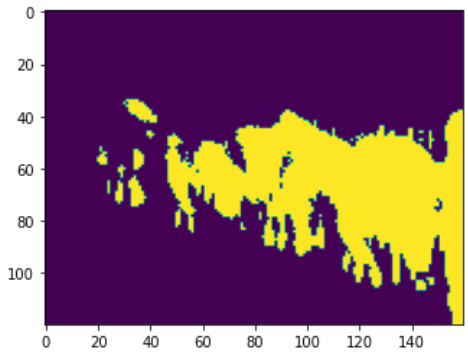
Activity classification has a long history. The creation of MEI (motion energy images) and MHI (motion history images) boosted this field. In this theory, a binary image will be set up first. Based on the binary signal, we will move to the MHI. We control MEI and MHI with theta and tau. Theta decided how much difference between frames we are seeking for. Tau drives the length of motions we recorded. Then, moments will be created. We calculate hu moment. It works via calculating 7 signals with central moments. It boosts recognition.

Classifier

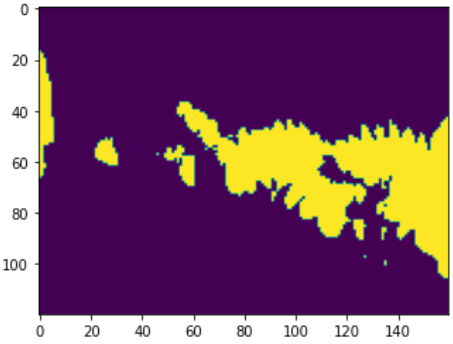
I used K nearest neighbors algorithm to train and predict. The data x is the hu moment and data y is the action index, which means we always get action with the hu. I iterate all samples of data and generally save the model. To proceed with the result, I implemented the model with different videos which contain multiple actions and we will check the performance later.

Implementation

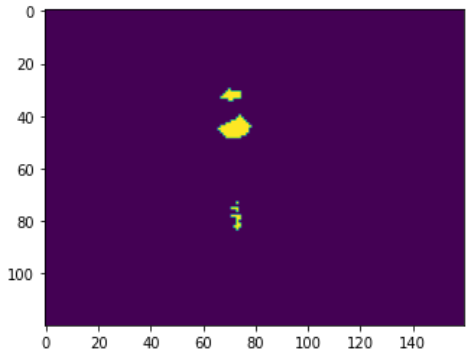
Binary Image



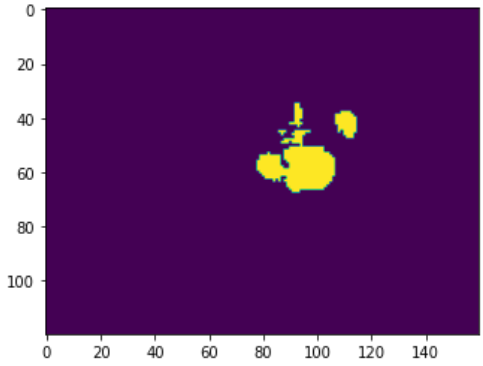
Running



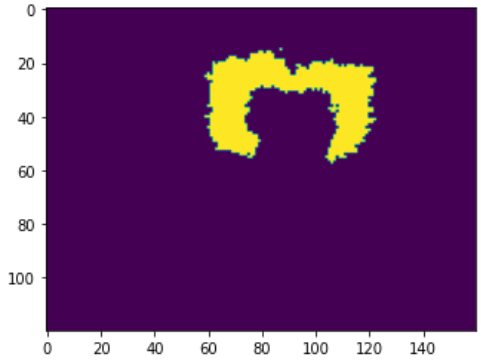
jogging



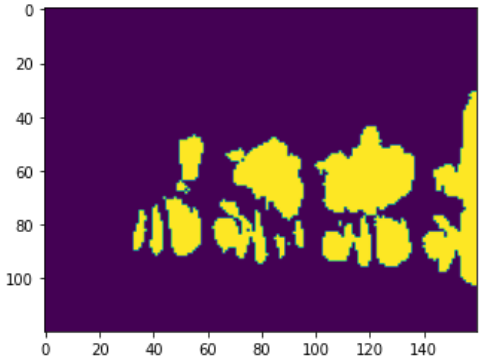
Boxing



Handclapping



Handwaving



Walking

	Running	Jogging	Boxing	Handclapping	Handwaving	Walking
Theta	35	25	15	15	10	30
Tau	13	13	5	5	5	5

The binary images for 6 actions are shown above. In creating these images, frames were compared with the former one and the difference is recorded to track. By adjusting theta, we can control how much

movement between frames would be recorded and what to ignore. It's obvious that the larger movement you made, the larger theta it needs. For some "large" movements like running and jogging, small theta may confuse the results.

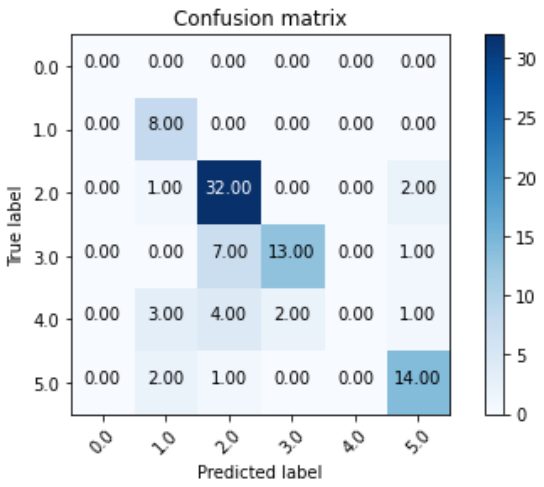
Training and Prediction

A large dataset was used for training. Actually, I tried several algorithms like SVM and NN, however, KNN did the best job. The predicting results are shown below. The accuracy of prediction is 75%. In my opinion, optimization of the index when generating hu moments may improve the prediction.

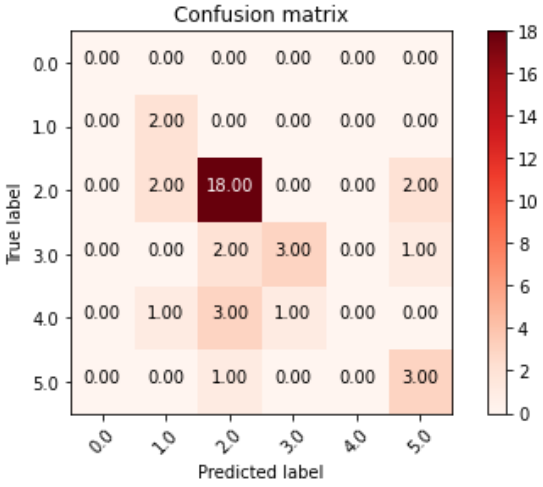
Number of axis is the index of actions.

0: 'boxing', 1:'handclapping',2:'handwaving',3:'jogging',4:'running',5:'walking'

Trained:



Tested:



Application

Knowing a specific frameset influences the result a lot. For a video that we know only contains one action, it is always easy to get the result. When the video has several parts with different actions, if we can not split well, the later HMI will be confusing because it no longer belongs to the former action.

To beat this challenge, I selected frames every second at first. For my video, it is 25 frames per second. Then I found that a shorter period may lead to less trace of the moment. As the result, some long-distance movements may be recognized as a similar short actions. For my experiment, running was always regarded as handclapping. In the end, I selected frames by the minimum of all tau values, since it stands for the length of the action. With this value, we could always get the same piece of movement, then if we know the movement, we will approach another better frame number.

My video results can be found below.

Jogging: https://youtu.be/dA4siGUemXQ

Running: https://youtu.be/GSdKqGpLQ9o

Walking: https://youtu.be/7jlVJQ6x5kM

Boxing: https://youtu.be/96JUzhLsudc

Handwaving: https://youtu.be/73-S-nRlRnI

Handclapping: https://youtu.be/MJFhLTAfQw0

Multiaction: https://youtu.be/jacI_WEjfhw

References

H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic image networks for action recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3034–3042, 2016.

Recognition of human actions. (2005, January 18). Recognition of Human Actions. https://www.csc.kth.se/cvap/actions/

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2, pp. 246–252, IEEE, 1999.

A. Fathi and G. Mori, "Action recognition by learning mid-level motion features," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1–8, IEEE, 2008.

Wikipedia. Image moments | wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LaTeX&oldid=413720397, 2017. [Online; accessed 2-December-2017].

S. Maji, L. Bourdev, and J. Malik. Action recognition from a distributed representation of pose and appearance. In CVPR 2011, *pages 3177-3184, June 2011.*