Human Activity Recognition using OpenPose

MUHAMMAD SHIRAZ

MAY 2021

INFORMATION SYSTEMS ENGINEERING DEPARTMENT

KOCAELİ UNIVERSITY

TEACHER: PROF. OĞUZHAN URHAN

Outline

Introduction

- Human Activity Recognition
- Active Learning

Goals

Literature Review

Training Data

- Data Collection and Feature Extraction
- Classification Techniques
- Query Strategies of active learning

Evaluating Model

Results

Conclusions

Goals

Human action recognition based on video data has diverse applications. For example, it can be used for interactive games where a person's movement is recognized by the computer and used as the input for playing games etc.

- Compare and find a model that best fit our system in terms of accuracy and efficiency.
- Reduce the labeling time and labor works using active learning.

Literature review

Use one or multiple camera to do a vision-based recognition.

The result shows that the accuracy of all 5 models are higher than 96%. Furthermore, three of the models achieved an accuracy of more than 99%, including SVM (kernel method), DNN (3 layers, 100x100x100), and Random Forests (100 trees with depth 30).

Recognition accuracy of five models, when using (a) selected features from 4 frames, or (b) raw features from single frame.

Introduction: Activity Recognition

The aim of this project is to create a model that can identify the human actions like running, jogging, walking, clapping etc.

Algorithm

- Detecting human skeleton from image
- Preprocessing features
- Feature Extraction
- Classification

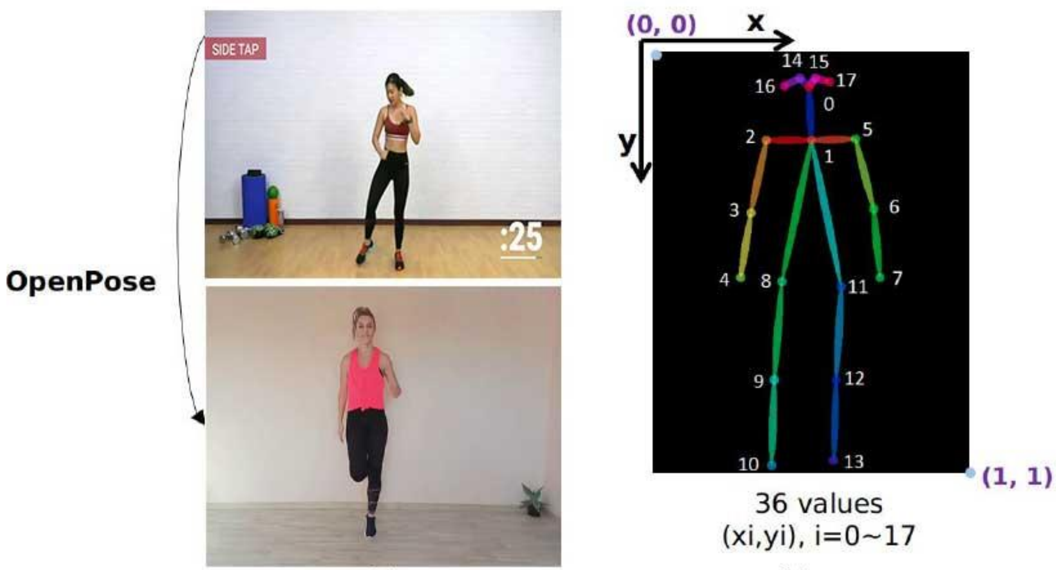
Recognition accuracy

- Speed of algorithms
- Performance on video of actions

The recognized actions are running, jogging, walking, and clapping respectively. They are all kind of reasonable and fits the common sense.

Detecting human skeleton from image

The OpenPose algorithm is adopted to detect human skeleton from the image. The key idea of OpenPose is using Convolutional Neural Network to produce two heapmaps, one for predicting joint positions, and the other for associating the joints into human skeletons.



(a) (b)

Preprocessing features

The raw skeleton data are preprocessed before extracting features. The preprocessing includes 4 steps as summarized:

1. Scale the coordinate

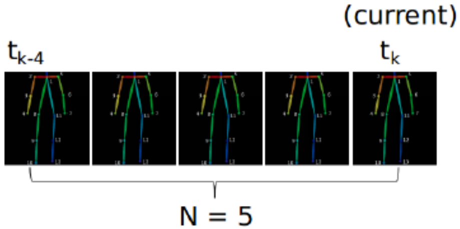
2.Remove all joints on head

3. Discard frames who have no Neck or Thigh

4. Fill in the missing joints

Feature Extraction

After the previous preprocessing step, the joints' positions are complete and good to use. In this section, I have used the joint positions from N=5 frames as raw features, and then manually design and extract more salient features that might be useful for distinguishing the action types.



Joints from previous 0.5s
(Vidoe framerate = 10 fps)

Possible Selections of Features

Symbol	Meaning
Xs	Concatinatation of joints' pos of N frames
H	in Xs: Average skeleton height
V_body	in Xs: (Velocity of neck) / H
X	Normalized pos: X = (Xs - mean(Xs)) / H
V_joints	in X: Velocity of all joints, $\{X[t_k]-X[t_{k-1}]\}$
JointAngles	in X: convert pos to joint angles
LimbLens	in X: convert pos to length of limbs

Classification

The total training data are split into two sets: 70% for training, and 30% for testing.

Five different classifiers are experimented, including kNN, SVM, SVM with kernel method, Deep Neural Network, and Random Forests.

The implementation of these methods is from the Python library "sklearn". I tuned the major parameters of each method in order to obtain a better result.

```
classes: ['running', 'jogging', 'walking', 'clapping']
```

Recognition accuracy

The final recognition (classification) accuracy is shown in Fig. 1.1(a), where the features are extracted as described in last section.

The result shows that the accuracy of all 5 models are higher than 96%. Furthermore, three of the models achieved an accuracy of more than 99%, including SVM (kernel method), DNN (3 layers, 100x100x100), and Random Forests (100 trees with depth 30).

		Features of prev 5 frames: X: Normalized joint pos, V_body, V_joints		Features of 1 frame: Xs: Original pos	
Method	**Main settings**	**Accu Train**	**Accu Test**	**Accu Train**	**Accu Test**
kNN	k=5	98.7%	97.4%	95.7%	92.4%
SVM	Linear	97.8%	96.1%	87.9%	86.7%
SVM	Kernal	99.6%	99.1%	90.7%	90.4%
DNN	100x100x100	1	99.4%	1	97.3%
Random Forests	depth 30 trees 100	1	99.2%	1	96.2%

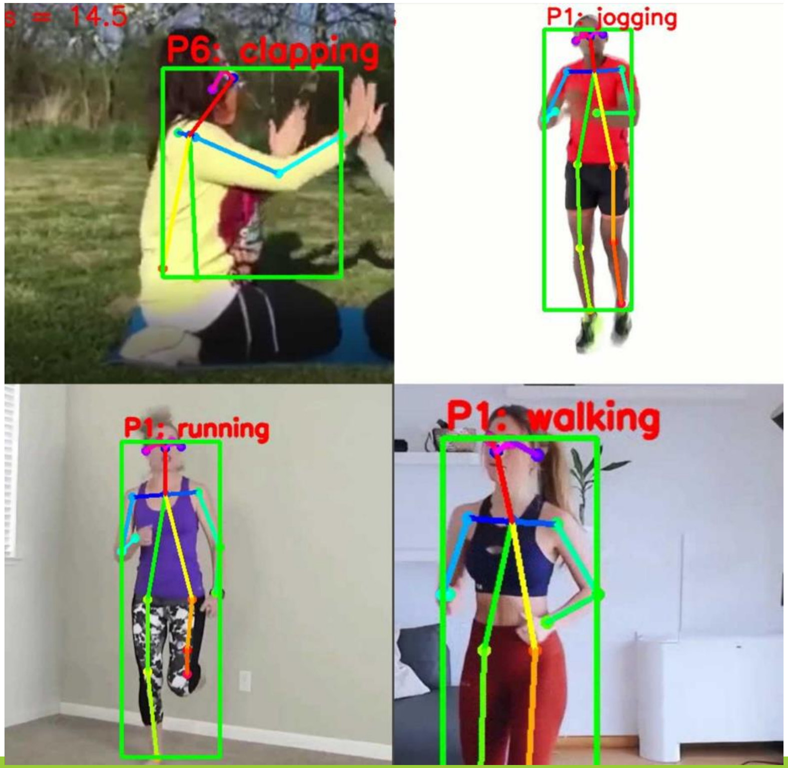
Speed of algorithms

The speed of the algorithms was tested on a laptop with an Intel Core i5 CPU and a GTX-1650 GPU. The OpenPose runs at about 7 fps when the image is resized to 432x368. The time cost for feature extraction and classification is less than 0.01s per frame for all classifiers.

With some future optimization of the program and using better hardware, this action recognition system should be able to run in real time with a framerate of 10 fps.

Performance on video of actions

I tested the performance of the proposed action recognition system on a real video.



Training Data

I collected videos of 4 types of actions, including each activities with 1000 samples

---

(1) Running, (2) Jogging, (3) Walking, (4) Clapping, as shown in Fig. 1.1

Motivation: Fig. 1.2 Snapshots of 4 types of actions in the training data, including: Running, Jogging, Walking, clapping.

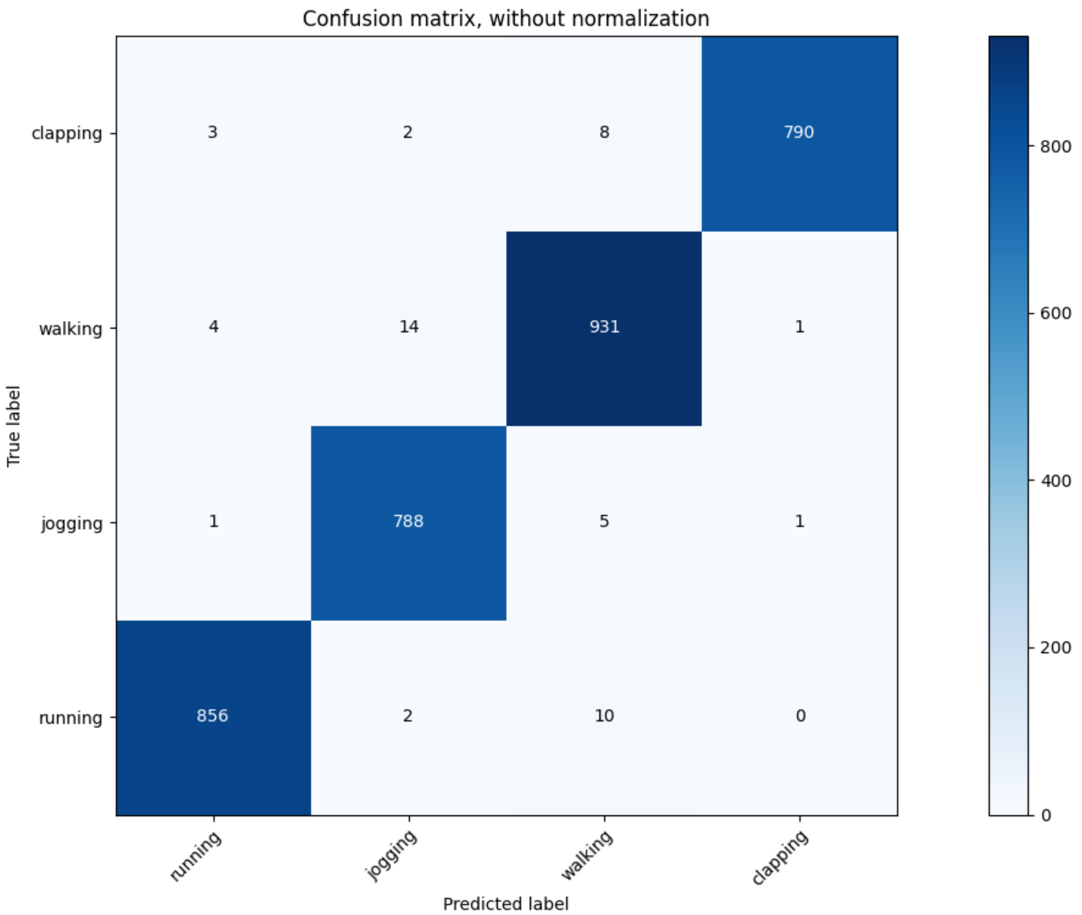
Evaluating Model

These videos were recorded with a size of 640x480 and a framerate of 10 frames/second, so that it's fast enough for capturing the whole movement of an action.

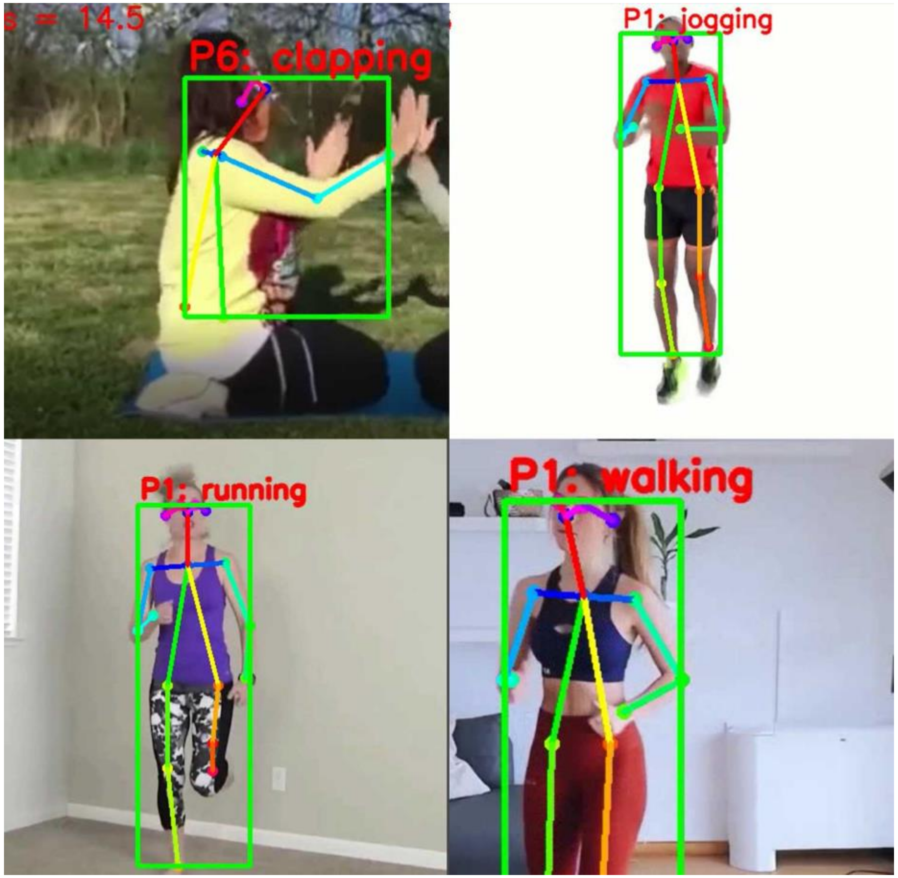
Accuracy report:

	precision	recall	f1-score	support
running	0.99	0.98	0.99	353
jogging	0.97	0.94	0.96	265
walking	0.97	0.99	0.98	372
clapping	0.98	0.99	0.98	85
accuracy			0.98	1075
macro avg	0.98	0.98	0.98	1075
weighted avg	0.98	0.98	0.98	1075

Confusion matrix, without normalization



Results



Action recognition result on a video

Conclusions

In this project, I implemented a human action recognition system that can recognize 4 types of actions. The algorithm is based on a real-time framework by aggregating the skeleton data of a 0.5s window for feature extraction and classification.

The recognition accuracy was up to 99% on the training set composed of more than 10000 samples.

This action recognition system was then tested on real world videos: It achieved stable and accurate recognition performance on a video similar to the training set performed by me, and achieved relatively good result on other videos.

Reference

1)L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” *Pers Comput.,* Lecture Notes in computer Science, vol. 3001, pp. 1–17, 2004.

2)U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, “Location and activity recognition using eWatch: A wearable sensor platform,” *Ambient Intell. Everday Life,* Lecture Notes in Computer Science, vol. 3864, pp. 86–102, 2006.

3)J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, “Activity classification using realistic data from wearable sensors,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

4)N.Wang, E. Ambikairajah,N.H. Lovell, and B.G. Celler, “Accelerometry based classification of walking patterns using time-frequency analysis,” in *Proc. 29th Annu. Conf. IEEE Eng. Med. Biol. Soc.*, Lyon, France, 2007, pp. 4899–4902.

5)T.B.Moeslund,A.Hilton,V.Kr ¨ uger, Asurveyofadvancesinvision-based human motioncaptureandanalysis,Comput.VisionImageUnderstanding 104 (2–3)(2006)90–126.

6)T.B. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Comput. Vision Image Understanding 81 (3) (2001) 231–268.

7)Y. Tao, H. Hu, H. Zhou, Integration of vision and inertial sensors for 3D arm motion tracking in home-based rehabilitation, Int. J. Robotics Res. 26 (6) (2007) 607–624.

8)Preece S J, Goulermas J Y, Kenney L P J and Howard D 2008b A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data *IEEE Trans. Biomed. Eng.* at press

9)N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In AAAI, pages 1541–1546, 2005.

10)S.J. Preece, J.Y. Goulermas, L.P.J. Kenney, D. Howard, K. Meijer and R. Crompton, Activity identification using body-mounted sensors—a review of classification techniques. *Physiol Meas*, **30** (2009), pp. R1–R33.

11)Altun, K., Barshan, B., Tun¸cel, O.: Comparative study on classifying human activities with miniature inertial and magnetic sensors. Pattern Recogn. 43(10), 3605–3620 (2010), doi:10.1016/j.patcog.2010.04.019