EfficientPose: Scalable Single-Person Pose Estimation and Classification

03CS6902 Mini Project
Design Report

CHN20CSIP03 PRAVEENA K M
praveenamurali1004@gmail.com
M. Tech. Computer Science (Image Processing)

Department of Computer Engineering
College of Engineering Chengannur
Alappuzha 689121
Phone: +91.479.2165706
http://www.ceconline.edu
hod.cse@ceconline.edu

Abstract

This project is an attempt to implement an approach for human activity recognition and classification using a person's pose skeleton in images. This implemnetation is divided into two parts; a single person poses estimation and activity classification using pose. Pose Estimation consists of the recognition of 18 body key points and joints locations. Using these pose information ,the images are classified into 5 action classes

Contents

Chapter 1

Introduction

The goal of a Human Activity Recognition (HAR) system is to predict the label of a person's action from an image or video.Understanding human behavior in images gives useful information for a large number of computer vision problems and has many applications like scene recognition and pose estimation.One of the popular vision based HAR systems uses pose information. Poses have had remarkable success in human activity recognition, Poses provide useful information about human behavior. Pose estimation is a computer vision task that infers the pose of a person or object in an mage or video.It is a technique used to estimate how a person is physically positioned, such as standing, sitting, or lying down.There are two approaches: a bottom-up approach, and a top-down approach. With a bottom-up approach, the model detects every instance of a particular keypoint (e.g. all left hands) in a given image and then attempts to assemble groups of keypoints into skeletons for distinct objects. A top-down approach is the inverse – the network first uses an object detector to draw a box around each instance of an object, and then estimates the keypoints within each cropped region

1.1 Proposed Project

This project aims to implement an efficient human activity recognition system by extracting the pose information of a single person in the image.

1.1.1 Problem Statement

This project aims to localizing human skeletal keypoints of a person from an image ,and then the classification of the activities into action classes such as walking,standing,sitting and running using the extracted pose key points .

1.1.2 Proposed Solution

The proposed solution for this activity recognition and classification consists of two sequential tasks.The system takes, as input, a color image of size w $\times$ h and produces, the 2D locations of anatomical keypoints for the person in the image.and then the identification of salient actions through the observation of the extracted pose key points with the help of KNN classification algorithm.

Chapter 2

Project Design

2.1 Human Pose Estimation

Human Pose Estimation is the task of extracting the body's skeletal key points and joints locations corresponding to the human body parts. It uses all those key points and joints to associate the two-dimensional structure of the human body.First, feedforward network simultaneously predicts a set of 2D confidence maps S of body part locations and a set of 2D vector fields L of part affinities, which encode the degree of association between parts. Finally, the confidence maps and the affinity fields are parsed by greedy inference to output the 2D keypoints for the person in the image.Successive stages are used to refine the predictions made by each branch. Using the part confidence maps, bipartite graphs are formed between pairs of partsUsing the PAF values, weaker links in the bipartite graphs are pruned .

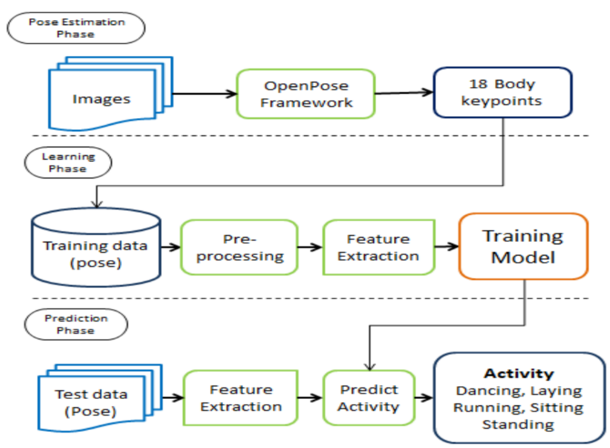

Figure 2.1.1: proposed framework

2.1.1 Part Affinity Field Maps (L)

It contains two-dimensional vectors that encode the body part's positions and orientations in an image. It encrypts your data in the form of a double link be-tween body parts.

$$L = (L1, L2, L3....Lc)(1)$$
$$L_c \varepsilon R^{w*h*2} \ ,c \ \ \varepsilon 1...C$$

, where C is the total number of limbs, R is the real number, L is the set of part affinity field maps, and w x h is the dimension of each map in the set L.

2.1.2 Confidence Map S

It is a two-dimensional representation of the belief that a particular part of the body can be placed on a specific pixel.

$$S = (S1, S2, S3....Sj)(2)$$
$$L_c \varepsilon R^{w*h} \ , j \ \ \varepsilon 1...J$$

, where J is the total number of body parts, R is the real number, and S is the set of confidence maps.
The number of keypoints detected is dependent upon the dataset has been trained.The 18 different body key points are R$_{Ankle}, R_{Knee}, R_{Wrist}, L_{Wrist}, R_{Shoulder}, L_{Shoulder}, L_{Ankle}, L_{ear}, R_{Ear}, R_{Elbow}, L_{Elbow}, L_{Knee}, L_{Eye} R_{Eye}, R_{Hip}, L_{Hip}, Nose, andNeckisused.$

2.2 Activity Classification

The classification algorithm takes 18 body keypoints (x-axis and y-axis values of each point) as input.The k-NN algorithm is used for pose classification .K-nearest neighbor (KNN) is a supervised machine learning algorithm used for classification, and it's a non-parametric, lazy algorithm.To convert pose landmarks to a feature vector, here uses pairwise distances between predefined lists of pose joints, such as distances between wrist and shoulder, ankle and hip, and two wrists. Since the algorithm relies on distances, all poses are normalized to have the same torso size and vertical torso orientation before the conversion.
The distance function used in the algorithm d is given by.

$$d(p, q) = \sqrt{\sum (q_i - p_i)^2}$$

where p, q are vectors containing keypoints of two different images and i=1 . . . n

2.3 Hardware & Software Requirements

Operating System	: Any Operating System
Supporting software	: Python,Opencv
Processor	: Intel Core i5 7th Gen 2.50GHz
RAM	: 8GB
Monitor	: Any colour monitor
Dataset	:COCO Dataset

Chapter 3

Project Progress

Below are the work done so far:

1. Studied the reference paper .
2. Installed Python ,opencv and started learning it.
3. Collected some data sets.
4. Made the design of the project.

3.1 Work Schedule

Schedule of completed work (till august10)

1. Identify suitable project area and topic
2. Study the reference paper well.
3. Check for other papers related to this topic.
4. Prepare for IC and select a guide
5. Analyse various methods which can be used in this project.
6. Install Python and start learning it.
7. Conducted literature survey on related work of project
8. Make the design of the project.

Work scheduled for coming time period 1.

1. Start the implementation.
2. Complete the implementation of each step

References

[1] Zhe Cao; Gines Hidalgo; Tomas Simon; Shih-En Wei; Yaser Sheikh. Openpose: Re-altime multi-person 2d pose estimation using part affinity fields.IEEE Transactionson Pattern Analysis and Machine Intelligence, pages 128–137, july 2019.

[2] Heri Ramampiaro Espen AF Ihlen Daniel Groos. Efficientpose: Scalable single-person pose estimation.Springer journal of Applied Intelligence, 51:2518–2533,6,November 2020.

[3] Gehler P Schiele B Andriluka M, Pishchulin L. 2d human pose estimation: newbenchmark and state of the art analysis. InIEEE Conference on computer visionand pattern recognition (CVPR), 2014.

[4] https://medium.com/dsaid-govtech/human-pose-estimation-and-human-action-recognition-experimenting-for-public-good-dabde16521b3