JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Action Detection System Using Video Processing and Machine Learning

Darshan Deshmukh[1], Mrs.Shafali Gupta [2], Vedant Mahajan[3], Sujit Chaudhari[4], Bhimashankar Dhotre[5]
Department of Computer Engineering,
R.M.D Sinhgad School of Engineering, Pune, India.

Abstract

In compared to its analogue predecessors, the digital IP video system offers a wide range of features and capabilities, but it still faces significant difficulties in the creation and implementation of intelligent and efficient systems. This initiative aims to satisfy the rising need for automated analysis and comprehension of human behaviour. The presentation of a real-time human activity identification system built on global feature extractions. The suggested method calculates an image's characteristics from video sequences. The goal is to combine local and global properties to create a powerful real-time recognition system. We utilize the UT Dataset, which includes the following actions: embracing, handshakes, kicking, punching, and pushing. Convolution neural networks (CNN) are used to instantly identify various human actions. Similar videos based on action recognized in input videos are retrieved from the dataset as output.

Keywords: Analysis, Action recognition, Machine learning, Camera Storage, Object detection.

1. INTRODUCTION

As more digital video cameras are utilized in daily life, an increasing amount of video footage is being produced, shared online, and stored in massive video data sets. Human action recognition is a popular area of study because to its potential uses in content-based video retrieval, human-computer interaction, and sports annotation. As an illustration, the visual system in a sizable public area may automatically extract high-level semantic data from the video with accurate human action recognition. The early attempts at human action recognition used the tracks of a person's body parts as input characteristics.

The majority of current research, however, moves away from the high-level representation of the human body (such as the skeleton) and towards a collection of low-level data (such as local features) since full-body tracking from movies is still a challenging challenge. Microsoft Kinect, one of the most recent developments in depth sensors, provides real-time full-body tracking at an affordable price with sufficient accuracy. This gives us another opportunity to research skeleton-based characteristics for activity identification.

Past research proposed algorithms to classify short videos of simple periodic actions performed by a single person (e.g. ‘walking’ and ‘kicking). In real-world applications, actions and activities are seldom periodic and are often performed by multiple persons (e.g. ‘pushing and ‘hand shaking). Recognition of complex non-periodic activities, especially interactions between multiple persons, will be necessary for a number of applications (e.g. automatic detection of violent activities in smart systems).

The main objectives of this system are:

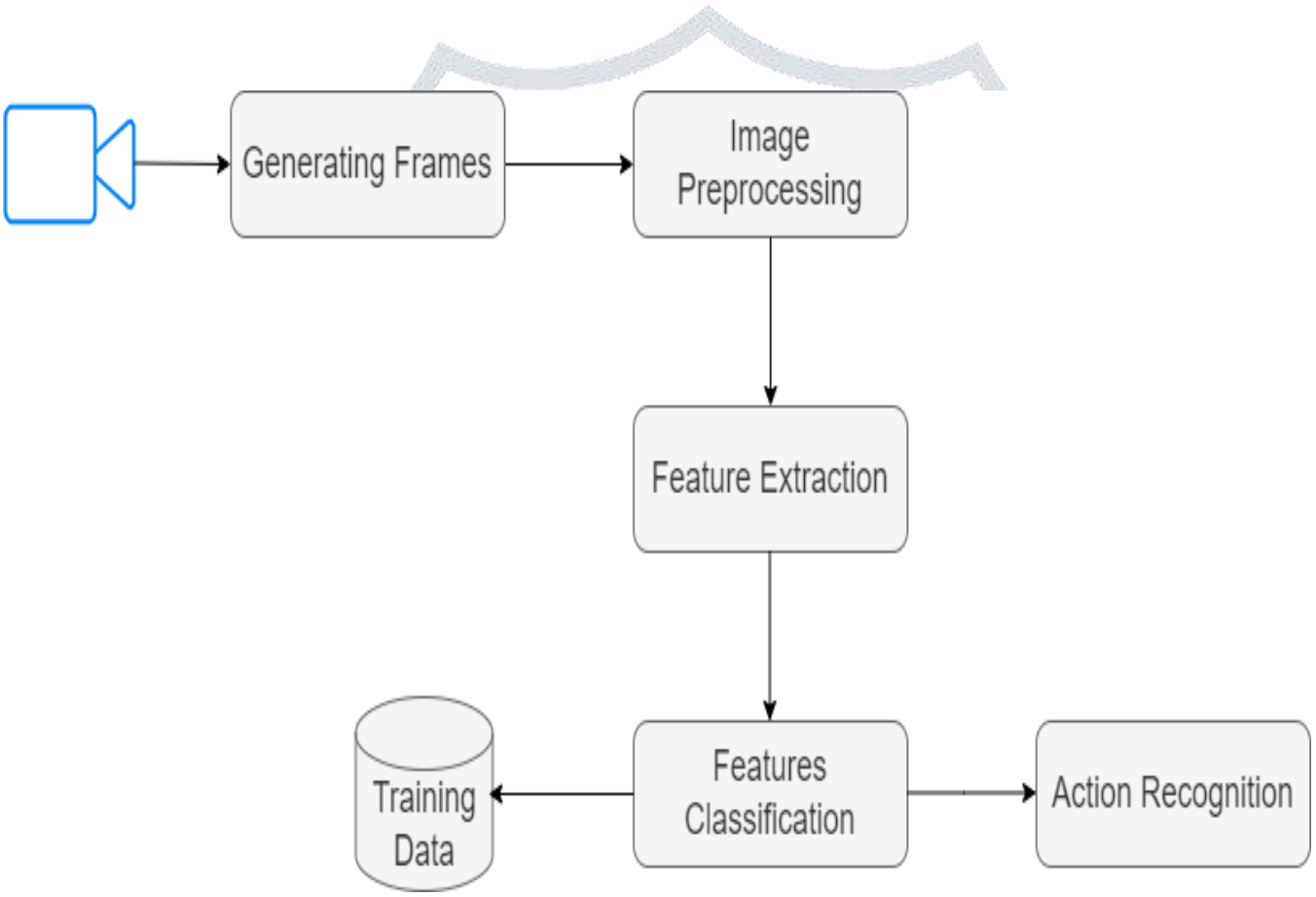
1) Accurate result in recognizing common activities.

2) Recognizes interactions between persons (hand shaking, hugging, pushing, kicking, approaching).

3) Easy to operate.

4) Provides quick and accurate results.

2. SYSTEM MODEL DIAGRAM:



(Figure **1:** System Model Diagram)

3. WORKING PROCESS:

1. Input Image:

Here will upload the Input Image.

2. Image Preprocessing:

In this step will apply the image preprocessing methods like grayscale conversion, image noise removal for further processing.

3. Image Feature Extraction:

In this step will apply the image thresholding and edge detection methods to extract the cell nuclei from leaf image and count that.

4. Image Classification:

In this step will applying the image classification methods like CNN algorithm to classify the diseases.

5. Result:

In this step will show the final action result.

4.PROPOSED SYSTEM:

The proposed work is a two-person interaction-based, video-based system for identifying human activity, Every feature map is a plane, the weight of the neurons in the plane are same. The structure of feature plan uses the sigmoid function as activation function of the convolution network, which makes the feature map have shift in difference. Besides, since the neurons in the same mapping plane share weight, the number of free parameters of the network is decreased. Each convolution layer in the convolution neural network is come after by a computing layer which is used to find the local average and the second extract; this unique two feature extraction structure decreases the resolution.

5.Algorithm:

Convolution Neural Network(CNN)

The structure of CNN includes two layers one is feature extraction layer, the input of each neuron is connected to the local ready fields of the previous layer, and extracts the local feature. Once the local features are extracted, the positional relationship between it and other features also will be displayed. The other is feature map layer; each computing layer of the network is collected of an advantage of feature map. Every feature map is a plane, the weight of the neurons in the plane are same. The structure of feature plan uses the sigmoid function as activation function of the convolution network, which makes the feature map have shift in difference. Besides, since the neurons in the same mapping plane share weight, the number of free parameters of the network is decreased. Each convolution layer in the convolution neural network is come after by a computing layer which is used to find the local average and the second extract; this unique two feature extraction structure decreases the resolution.

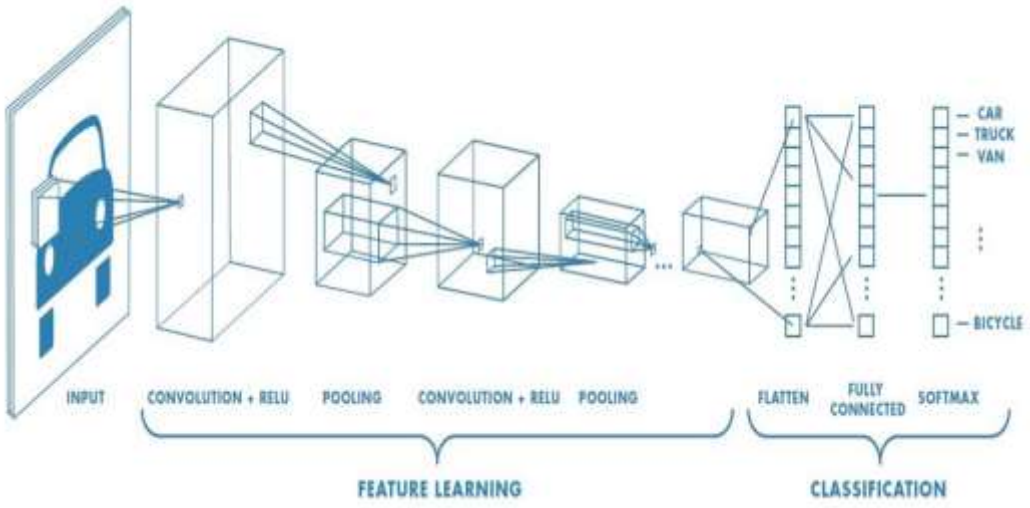


Fig. CNN Layers

Convolution Layer:

Convolution is the first layer to extract features from an input image (leaf image). Convolution preserves the relationship between pixels by learning image features using small squares of input data. Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters i.e. identity filter, edge detection, sharpen, box blur and Gaussian blur filter.

Pooling Layer:

Pooling layers would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains important information.

Fully Connected Layer:

In this layer Feature map matrix will be converted as vector (x1, x2, x3, …). With the fully connected layers, we combined these features together to create a model.

Softmax Classifier:

Finally, we have an activation function such as softmax or sigmoid to classify the outputs i.e. classifiy leaf disease.

6.LITERATURE SURVEY:

A real time human activity recognition system based on Radon transform (RT), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is presented. Artificial Neural Nets (ANN) is used to recognize different human activities[1].

The data extracted using optical flow is converted to binary image. Then Histogram of Oriented Gradient (HOG) descriptor is used to extract feature vector from the binary images. These feature vectors are given as training features to Support Vector Machine (SVM) classifier to prepare a trained model[2].

In this paper, video based action recognition is performed on KTH dataset using four combinations of two feature descriptors and two classifiers. The feature descriptors used are Histogram of Oriented Gradient Descriptor (HOG) and 3- dimensional Scale Invariant Feature Transform (3D SIFT) and classifiers used are Support Vector Machine (SVM) and K Nearest Neighbour (KNN). Features are extracted from frames of training videos using descriptor and clustered to form Bag-of-words model[3].

This approach predicts human actions using temporal images and convolutional neural networks (CNN). CNN is a type of deep learning model that can automatically learn features from training videos. Although the state-of-the-art methods have shown high accuracy, they consume a lot of computational resources. Another problem is that many methods assume that exact knowledge of human positions[4].

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables[5].

Human Activity Recognition Using an Ensemble of Support Vector Machines is employed to improve the classification performance by fusing diverse features from different perspectives. The Dempster-Shafer fusion and product rule from the algebraic combiners have been utilized to combine the outputs of single classifiers[6].

Human motion capture continues to be an increasingly active research area in computer vision with over 350 publications over this period. A number of significant research advances are identified together with novel methodologies for automatic initialization, tracking, pose estimation, and movement recognition. Recent research has addressed reliable tracking and poses estimation in natural scenes. Progress has also been made towards automatic understanding of human actions and behavior[7].

The work presents the study of various famous and unique techniques used for facial feature extraction and Face Recognition. Various algorithms of facial expressions research are compared over the performance parameters like recognition accuracy, number of emotions found, Database used for experimentation, classifier used etc [8]

This work proposes a system that will automatically identify the facial expression from the face image and classify emotions for final decision. The system uses a simplified technique called `Viola Jones Face Detection' technique for face localization. The different feature vectors are club together using a subset feature selection technique to improve the performance of recognition and classification process. Finally the combined features are trained and classified using SVM, Random Forest and KNN classifier technique [9].

The proposed technique use three steps face detection using Haar cascade, features extraction using Active shape Model (ASM) and Adaboost classifier technique for classification of five emotions anger, disgust, happiness, neutral and surprise [10].

In this work implement an efficient technique to create face and emotion feature database and then this will be used for face and emotion recognition of the person. For detecting face from the input image we are using Viola-Jones face detection technique and to evaluate the face and emotion detection KNN classifier technique is used [11].

This paper objective is to display needs and applications of facial expression recognition. Between Verbal & Non-Verbal form of communication facial expression is form of non-verbal connection but it plays pivotal role. It expresses human related or filling & his or her mental situation [12].

In this proposed system it is attention on the human face for recognizing expression. Many techniques are available torecognize the face image. This technique can be adapted to real time system very easily. The system briefly displays the schemes of capturing the image from web cam, detecting the face, processing the image to recognize few results [13].

In this work, adopt the recently introduced SIFT flow technique to register every frame with respect to an Avatar reference face model. Then, an iterative technique is used not only to super-resolve the EAI representation for each video and the Avatar reference, but also to improve the recognition performance. Also extract the features from EAIs using both Local Binary Pattern (LBP) technique and Local Phase Quantization (LPQ) technique [14].

In this study, a frame of emotion recognition system is developed, including face detection, feature extraction and facial expression classification. In part of face detection, a skin detection process is support first to pick up the facial region from a complicated background. Through the feature detection of lip, mouth, and eyes, eyebrow, those feature points are launch [15].

In this work, a new technique for facial emotion recognition is found. The proposal involves the use of Haar transform technique and adaptive AdaBoost technique for face identification and Principal Component Analysis (PCA) technique in conjunction with minimum distance classifier for face recognition. Two techniques have been investigated for facial expression recognition. The former relies on the use of PCA and K-nearest neighbor (KNN) classification technique,while the latter advocates the use of Negative Matrix Factorization (NMF) and KNN technique [16].

7.RESULTS AND DISCUSSION:

8.CONCLUSION:-

In this project, a feature representation and activity recognition system for human activity recognition for video retrieval system is proposed. The key frames selected to represent a sequence of activity, significantly reduced the computational complexity. The feature extraction algorithm is implemented using java to identify the activity of person. This results in increased efficiency for our human activity recognition system. The recognition process is performed by using CNN classifier. CNN classifier shows highest recognition results for human activity recognition. Videos similar to input video are retrieved from the database.

9.AKNOWLEDGMENTS:

It gives us great pleasure in presenting the preliminary project report on ***"Action Detection System Using Video Processing and Machine Learning"****.*

I would like to take this opportunity to thank my internal guide ***Mrs.Shafali Gupta*** *for giving us all the help and guidance we needed. We are really gratefulto them for their kind support. Their valuable suggestions were very helpful.*

I am also grateful to ***Prof. Vina M. Lomte****, Head of Computer Engineering Department, RMD Sinhgad School of Engineering, for her indispensable support, suggestions.*

Darshan Deshmukh.
Vedant Mahajan.
Sujit Chaudhari.
Bhimashankar Dhotre

(BE Computer Engineering.)

4. REFERENCES:

[1] Z.A. Khan, W. Sohn, "Real Time Human Activity Recognition System based on Radon Transform", IJCA Special Issue on Artificial Intelligence Techniques - Novel Approaches Practical Applications, AIT – 2011.

[2]Jagadeesh B, Chandrashekar M Patil, "Video Based Action Detection and Recognition Human using Optical Flow and SVM Classifier", IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016, India

[3]Aishwarya Budhkar, Nikita Patil, "Video-Based Human Action Recognition: Comparative Analysis of Feature Descriptors and Classifiers", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 6, June 2017

[4]ChengbinJin, ShengzheLi, Trung Dung Do, Hakil Kim, "Real-Time Human Action Recognition Using CNN Over Temporal
Images for Static Video Surveillance Cameras", Information and Communication Engineering, InhaUniversity,Incheon,Korea

[5]Jolliffe, I.T., “Principal component analysis”, Springer Series in Statistics, 2nd ed., Springer, 2002.

[6]E. Mohammadi, Q.M. Jonathan Wu, M. Saif, ”Human Activity Recognition Using an Ensemble of Support Vector Machines”,
2016 IEEE International Conference on High Performance Computing Simu- lation (HPCS) , July 2016

[7]T. B., Hilton, A., and Kruger, V., “A survey of advances in vision-based human motion capture and analysis“, Computer Vision
and Image Understanding, vol. 104, pp. 90-126, 2006

5. AUTHORS & MENTOR:

Mr. Darshan P Deshmukh
RMD School of Engineering,
Warje, Pune-58
(Computer Engineering)

Mrs. Shafali Gupta.
RMD School of Engineering,
Warje, Pune-58
(Head of Computer Engineering)

Mr.Vedant Mahajan
RMD School of Engineering,
Warje, Pune-58
(Computer Engineering)

Mr. Sujit Chaudhari
RMD School of Engineering,
Warje, Pune-58
(Computer Engineering)

Mr. Bhimashankar Dhotre
RMD School of Engineering,
Warje, Pune-58
(Computer Engineering)