AN EFFICIENT HUMAN ACTION RECOGNITION SYSTEM USING SINGLE CAMERA AND FEATURE POINTS

Maithili K., Rajeswari K., Mohanapriya R., Krithika D.

Department of Information Technology, Christ College of Engineering and Technology, Puducherry, INDIA
Email: ka_maithu@yahoo.com, rajeswari2009.it@gmail.com, priya.mohanaravi@gmail.com, dkrithikaintech@gmail.com

Abstract

In this paper, an efficient human action recognition system using feature points and single camera method based on neural network representation and recognition is proposed. By now, representation of action videos is based on learning rarely related human body posture method called Self Organizing Maps (SOM). Fuzzy distances from human body posture prototypes are used to produce a time invariant action representation. Multi layer perceptrons are used for action classification. The algorithm is trained using data from a multi-camera setup. An arbitrary number of cameras can be used in order to recognize actions using a Bayesian framework. Due to the growing interest in visual surveillance has led to human action recognition. So we propose a new and efficient method for human action recognition system using single camera and feature points. Our proposed method overcomes the problems in the existing system and recognizes the action of the required human. The system is developed in such a way, first it is trained using the feature extraction and feature tree method and then system will be capable of identifying the action from postures. We prove that our proposed is very efficient and can recognize actions quickly too.

Index Terms Human action recognition, multilayer perceptrons, feature tree, visual surveillance.

I. INTRODUCTION

Human action recognition is an active research field, due to its importance in a wide range of application, such as intelligent surveillance [1], Visual surveillance, Human action recognition, Crowd behavior analysis, tracking of an individual in crowded scenes, etc. The term Artificial Intelligence (AI) is the study of modeling of human mental function by computer Program, where the term action is related with the term activity and movement. Therefore Action is referred to as single period of human motion patterns (like walking step) but Activities consist of a number of action/movements (like dancing).

The objective of the estimation process is to find the most probable action according to the parameters. We have to estimate which posture the current image stands for, then recognize which action the posture sequence means. A critical problem in a recognition system is how to improve the accuracy and speed. There are two classes of estimation approaches. They are learning-based and example-based [2]. The learning based approaches use trained classifiers, while the example based ones search in exemplars. Action recognition aims to recognize the actions from agents and environment conditions. It plays vital role in many different applications and also in different fields such as human-computer interaction, medicine, and sociology etc, Visual (or video) surveillance devices have long been in use to gather information and to monitor people, events and activities. There is an increasing desire and need in video surveillance applications for a proposed solution to be able to analyze human behaviors and identify subjects for standoff threat analysis and determination. The main purpose of this survey is to look at current developments and capabilities of visual surveillance systems and assess the feasibility and challenges of using a visual surveillance system to automatically detect abnormal behavior, detect hostile intent, and identify human subject.

Problem Statements

A critical problem in a recognition system is how to improve the accuracy and speed. There are two classes of estimation approaches. They are learning-based and example-based. The learning based approaches use trained classifiers, while the example based ones search in exemplars.

Recognition of human action approach methods has many drawbacks in practice, which include

1. Less in efficiency to cope up with recognition problems.
2. The requirement of training stage in particular area to obtain good performance is less possible.
3. Difficult to recognize simultaneous multiple actions and incapable in performing recognition frame by frame.
4. The method should allow continuous action recognition over time.
5. The use of multi-camera setups involves the need of camera calibration for a particular camera setting.

II. RELATED WORK

Human action recognition received considerable attention in recent years. Action recognition was also addressed in several ways. We can classify the different actions based on the type of input that had been given to the system. The input was a set of manually tracked points on several parts of the body performing the action. Action Recognition was done by finding the most similar vector in the training set using the canny edge detector algorithm. The goal of our project is an automated analysis of the ongoing events and it is used to set it with the video data. The main application is surveillance system. We propose a generative model approach to learn by means of unsupervised learning approach and recognize human action. In the context of our problem, unsupervised learning is achieved by obtaining action model parameters from unsegmented and unlabeled video sequences, which contain a known number of human action classes. Our project is to permits the computer to learn models for human actions. Then, given a novel video, the algorithm should be able to decide which human action is present in the sequence. The task of automatic categorization and localization of human actions in video sequences is highly interesting for a variety of applications: detecting relevant activities in surveillance video, summarizing and indexing video sequences, organizing a digital video library according to relevant actions, etc. Our algorithm can recognize multiple actions from the video. The levels of understanding the actions from the video are of three different types namely Object-level understanding, tracking level understanding, pose level understanding, and activity level understanding[3]. Human action recognition is essential for surveillance and other monitoring systems in public places. This surveillance can be achieved successful by means of promoting Ubiquitous cameras in public places (e.g. CCTVs). This can monitor suspicious activities for real time actions such as stealing. The main challenges of our paper are environment variations such as the backgrounds (ex., tree), person movement variations (because each person has his/her own style of executing an activity). The ultimate goal is to make computers recognize all of them reliably and levels of human activities such as gestures, actions interactions.

The human beings used to create a continuous sequence of postures in the real world movement. Thus, a human action movement can be represented by a sequence of frames taken from video. These frames resemble each action of the human. The posture recognition can be found out in our project by means of using centroid context [4]. Thus after finding out the postures we can check the behavioral analysis by means of string representation. Then consequently the behavioral analysis is meant for demonstration, actions are viewed from different view angles, and behavioral analysis can be viewed by arbitrary views. The major real time application of our project is to find out the robbery events which can be detected by means of different human beings actions as one could differ from the other.

One of the representations for the human action analysis is appearance based recognition. At this recognition, for action and gesture recognition the appearance model of human body or hand are match it explicitly to the images in a target video that had been kept for action and gesture recognition.

This gesture recognition can be made with the help of a model namely Hidden Markov Models (HMM) and by using their different variants. And the human actions may also differ and also found complicated to handle in case of changes in the clothing from one to another.

The human action recognition is used to find out social interaction which is an evolutionary standpoint. We can test our approach by means of human facial motion, at which the human face can be analyzed by means of different scales formed by it. The proposed approach provides an high recognition accuracy and improves the state of art with the help of training and finding out the actions.

III. EXISTING MODEL

The Self-Organizing Map is one of the neural network models and the Map was developed by professor Kohonen [9]. This model comes under the category of competitive learning networks that is trained using unsupervised learning to produce a two-dimensional array of neurons from distinct representation of the input distance of the training samples, called a map [8].

SOM is a two layer neural network, which consists of input and output layer [6].

- Unsupervised learning neural network
- Maps multidimensional data onto a 2 dimensional grid
- Geometric relationships between image points indicate similarity

The Self-Organizing Map is a two-dimensional array of neurons:

$$M = \{ m, \ldots, m_{prn} \}$$

Each neuron has defined itself neighborhood consists of the surrounding neurons.

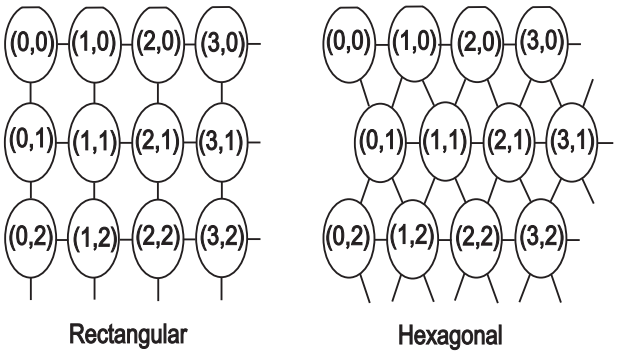

Fig. 1 Different topologies

A self-organizing map comprises of elements or it is made of nodes or neurons. In the input data vector, every node is to a weight vector of same position in the map space as linked with each node.

Generally, the SOM are arranged in a two-dimensional array of nodes. The arrangement is based on regular spacing in a hexagonal or rectangular grid shown in (Fig. 1). The SOM map from higher to lower dimension. This process is made for identifying data space to the vector onto the map is to detect the data space vector from the node with the nearest distance weight vector [7].

One of the main problems with SOMs is:

It is very difficult to determine what input weights to use, that is to getting the right data. So the SOM is sometimes called as missing data. Since, in order to generate a map we need a general solution for each proportion for each member. It is not often possible for all the time because it is very difficult to assume the data so this one of the limitation feature to the use of SOMs.

Mapping can result in divided clutters. But we need a lot of map to construct for getting a final good map. For example we taken a sample map, in that we using colors for dividing those data but still we tell that those two groups in comparatively are similar and that they just got divide, but in most of the data of those groups look entirely unconnected. Hence it takes more time [5].

In SOM, it is very difficult to find the different similarities within the sample vectors because every SOM is different. SOMs form try out information for getting final product, commonly encircled by standardized samples, even so similar samples are not forever close to each other.

The final major problem with SOMs is very expensive with regards to computation, which is a major drawback since the proportion dimensions of the data is increase, proportion reducing visualization techniques become more significant, but by the bad luck then time to compute them also increases.

IV. PROPOSED METHOD

In this paper, we propose a framework using feature tree technique, which recognizes the unknown action using the inputs of the single camera [10]. We show that the proposed framework is more accurate and efficient when compared to the existing frameworks in recognizing the actions. The proposed approach does not require the use of the same number of cameras in the training and recognition phases. It is done using feature-tree technique.

Method Description

After image segmentation, the image is decomposed into a number of homogeneous regions.

In Fig. 2, it shows that the image is represented by a two-level tree, where the root node represents the whole image and child nodes represent the region-based objects. The root node is assigned to the global feature, which is the color histogram in this case. Local region-based features, such as color moment, texture, size and shape, are assigned to the child nodes. This enables global and local image features to be integrated through a tree structure. The main advantages in the feature-tree construction stage, local spatiotemporal features are detected and extracted from each labeled video, and then each feature is represented by a pair [*d,l*] where *d* is the feature descriptor and *l* is the class label of the feature. Finally, we index all the labeled features using SR-tree. In the recognition stage, given an unknown action video, we first detect and extract local spatiotemporal features, and then for each feature we launch a query into the feature-tree. A set of nearest neighbor features and their corresponding labels are returned for each feature. Each returned nearest neighbor votes for its label. This process is repeated for each feature, and these votes can be weighted based on the importance of each nearest neighbor. Finally, a video is assigned a label, which receives the most votes. The entire procedure does not require intensive training. Therefore, we can easily apply incremental action recognition using the feature-tree.

A. Preprocessing Phase

For preprocessing, face segmentation should be done based on contrast of the image.

Finding Probability of human Module

To find whether the image is human we should examine that the image is larger connected or not. Then probability to become face should be traced. If the largest connected region is face, then new form should get open. If the module contains noise it should be removed using Digital signal processing filter.

Binary Image Conversion Module

For detecting the face, first we should convert RGB image to binary image. For converting binary image, we should calculate average value of RGB. It should be calculated for each pixel and if the calculated range is less than 110, we should replace it by black pixel else change it to white pixel. By following this we can able to convert the RBG image to binary image.

In database there are two tables. One is person and the other is position."Person" will store the image and "Position" is to store four kinds of index actions. In position table for each index control points are present. There are six different points for lip, 6 control points for right eye, 6 control points for left eye Bezier curve, left eye height and width, lip height and width and also right eye height and width. After by this

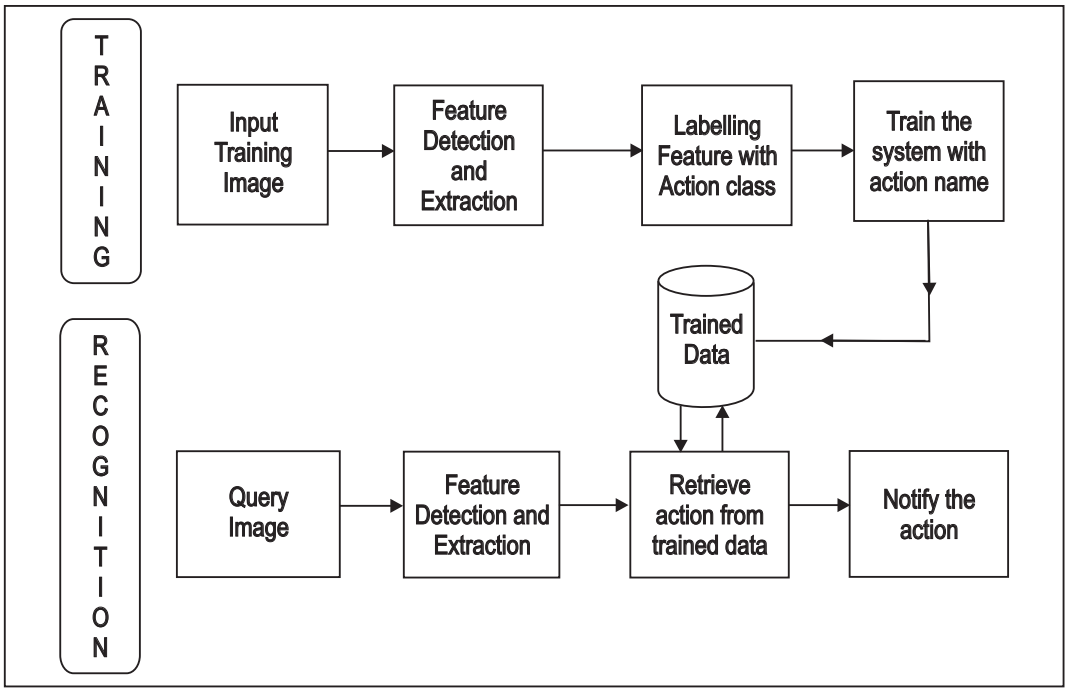

Fig. 2. Feature tree model

approach, the process examines the action of the people.

Movement feature extraction module

From the converted binary image format, we extract the features (such as movement of legs, hands etc.). Features are extracted by applying the edge detection algorithm first; to extract exactly the features of human posture and remove the background / unwanted noises in the picture.

B. Training Phase

In this module, we train the system with the actions depicted in the image. The actions are trained in the system which uses the technique of feature tree. The action representing the image is trained in the system with feature tree model.

C. Testing Phase

For detecting the action of an image, we have to trace the Bezier curve of the right eye, lip and left eye. Then we convert height according to its width and width of the Bezier curve to 100. If the person's action information is already present in the database, then the process will match action height to the possible related present height and the process will give the related action as output. If the information of action is not present in the database, then the process will calculates the mean height for every action in the database for each and every people and then gets a decision according to the average height.

V. CONCLUSION

We illustrate that our method efficiently recognizes the action from the video datasets and also solves problem of inability to cope up the problem of incremental recognition and helps to form a better framework in recognizing actions. Our proposed system using single camera and feature tree will be very efficient in recognizing the actions and it will be very useful in the human surveillances applications.

VI. FUTURE WORK

In proposed system, we recognize the normal human actions like walking, running etc. from images. In future we can recognize the actions from a video sequence. Also in future we can also try to recognize the human action in specific with lip reading etc.

REFERENCES

[1] L. Weilun, H. Jungong, and P. With, "Flexible human behavior analysis framework for video surveillance applications," Int. J. Digital Multimedia Broadcast., vol. 2010, pp. 920121-1-920121-9, Jan. 2010.

[2] R. Poppe, "Vision-based human motion analysis: An overview," Comput. Vis. Image Underst., vol. 108, no. 1-2, pp. 4-18, 2007.

[3] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning," IEEE Trans. Pattern Anal.Mach.Intell., vol.32, no.2, pp. 288-303, Feb. 2010.

[4] L. Weilun, H. Jungong and P. With, "Flexible human behavior analysis framework for video surveillance applications," Int. J. Digital Multime-dia Broadcast., vol. 2010, pp. 920121-1-920121-9, Jan. 2010.

[5] "Self Organizing Maps" Tom Germano, March 23, 1999.

[6] Honkela, T., "Self-Organizing Maps in Natural Language Processing", Espoo 1997.

[7] Self-Organizing Map (SOM), Jaakko Hollmen Fri Mar 8 13:44:32 EET 1996.

[8] Kohonen, Teuvo; Honkela, Timo (2007). "Kohonen Network". Scholarpedia.

[9] Mehotra, K., Mohan, C.K. & Ranka, S. (1997). Elements of Artificial Neural Networks. MIT Press.

[10] L. Weilun, H. Jungong and P. With, "Flexible human behavior analysis framework for video surveillance applications," Int. J. Digital Multimedia Broadcast., vol. 2010, pp. 920121-1-920121-9, Jan. 2010.