Human Action Recognition: An Overview

Pravin Dhulekar, S.T. Gandhe, Harshada Chitte and Komal Pardeshi

Abstract In this paper, an approach for detection of human actions from video is given. It has many applications, such as surveillance, human healthcare systems, border security, virtual reality, HMI, etc. This paper includes the detection of human actions and labeling them on an LCD. The frames are made from video and then median filter is applied. The feature extraction is done by applying PCA. The segmentation process is done using Gabor filter. Finally, KNN classifier is applied for classification. Different datasets are also provided.

Keywords Action recognition · Surveillance · Median filter · KNN classifier · Gabor filter · PCA

1 Introduction

Human action recognition is one of the most important topics in computer vision. Action recognition aims to recognize the actions and goals of one or more agents from a series of observations on the agent's action and environmental conditions. The goal of action recognition is to automatically detect the activities from a video and label them on an LCD display. Human action recognition has a wide range of applications, such as security, surveillance, entertainment, video annotation &

Pravin Dhulekar (✉) · S.T. Gandhe · Harshada Chitte · Komal Pardeshi
Department of Electronics and Telecommunication Engineering, Sandip Institute of Technology and Research Center, Nashik, Savitribai Phule Pune University, Pune, India
e-mail: pravin.dhulekar@sitrc.org

S.T. Gandhe
e-mail: stgandhe@gmail.com

Harshada Chitte
e-mail: chitte.harshada12@gmail.com

Komal Pardeshi
e-mail: komalpardeshi30@gmail.com

S.C. Satapathy et al. (eds.), *Proceedings of the International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 468, DOI 10.1007/978-981-10-1675-2_48

retrieval and human computer interaction, border security. A security system and surveillance were operated by human years ago. It will be beneficial to use human action recognition. Video surveillance is an effective tool for today's applications. As this topic is growing very fast, human action detection has naturally become a step to be followed. The applications for human action detection are as follows.

1. Security systems: Many years ago, the security and surveillance systems were monitored by a person and was very time consuming. Our method will be very useful in this field.
2. Healthcare systems: In a healthcare system the rehabilitation process can be monitored by the action recognition.

Our proposed method involves the recognition of human action which involves the preprocessing, segmentation, feature extraction, and classification. Our method will recognize the action through these steps and will display it on LCD display. For the process of preprocessing median filter is applied. Median filter removes the noise. In the process of segmentation, we will take out the silhouettes and will apply the Gabor filter. After the process of segmentation, features are extracted by using PCA. PCA reduces the large dimensions of data space. The classification process is applied for which KNN classifiers are used.

In Sect. 2, we have given some related work. We present our action representation technique and algorithm in Sect. 3, and then we talk about the result in Sect. 4. Finally, we bring a close to this paper in Sect. 5.

2 Related Work

The paper [1] focuses on challenges and gives the overview of current advances in the field. It discusses the features that are extracted by using the global and local representations. Some surveys provide an overview of existing methods to handle the challenges. The author in [1] focuses on the research challenges that can be faced in action recognition, such as variation in viewpoint, occlusions, execution rate, and camera motion. The human Markov models (HMM) is described here. In [2] a method for human action recognition from multi-view image sequences is presented. The method uses the combined motion and shape flow information with variable considerations. Author used a combined local–global optic flow for extracting motion. Flow feature and invariant moments with flow deviations which extracts the global shape flow feature from the image sequences. Ronald Poppe [3] provided a detailed overview of current advances in the field. The limitations for the state of the art and the outline promising directions of research are also discussed. Masoud el. [4] described the article that deals with the problem of classification of human actions from video and uses the motion features. These motion features are computed very easily and they are projected into a lower dimensional space and here matching is performed. The suggested method for classification is very accurate. The article also gives the recovery of two-dimensional and

three-dimensional properties of the person. The various levels of action recognition are described. The levels are split into three categories as low level, mid-level, and high level. Low level consists of core technology. Mid-level consists of human action recognition. High level consists of applications. In paper [5] author concentrated on the approaches that aim on classification of full body motions, such as kicking, punching, and waving. In paper [6], author presents a method for human action recognition from multi-view image sequences that uses the combined motion and shape flow information with variability consideration. A combined local–global (CLG) optic flow is used to extract motion flow feature and invariant moments with flow deviations are used to extract the global shape flow feature from the image sequences. In paper [7] the sampling strategies for detecting the actions are described. It gave the recent trend for sampling for the better performance. The sampling with the high density on action recognition is also explored. They have also investigated the impact of random sampling over dense grid for computational efficiency. Real time action recognition is also given. Vemulapalli et al. [8] represented the 3d skeleton detection. In this paper, a new body part-based skeletal representation for action recognition is given. Inspired by the observation that for human actions, the relative geometry between various body parts provides a more meaningful description than their absolute locations we explicitly model the relative 3D geometry between different body parts in our skeletal representation.

3 Action Representation

The action recognition has following stages:

1. Input Video 2. Image Conversion 3. Preprocessing 4. Segmentation 5. Feature Extraction 6. Classification 7. Action Detection.

1. Input Video:

In this step, the video is given as an input (Fig. 1).

2. Image conversion:

The video which is given as an input is then converted into number of images. Theseimages are then considered as test images. These images are then processed through the further stages.

3. Preprocessing:

Preprocessing enhances the visual appearance of images. It improves the manipulation of datasets. If the enhancement techniques are not used correctly, it can emphasize image artifacts or it can even lead to loss of information. In this paper, we have converted the colored image into gray image. After that we have resized the image. We will get a threshold image by adjusting the threshold value. Preprocessing involves image resampling, grayscale contrast enhancement, noise removal, mathematical operations, manual correction.

Where, image resampling reduces or increases the number of pixels of the dataset. Grayscale contrast enhancement improves the visualization by brightening

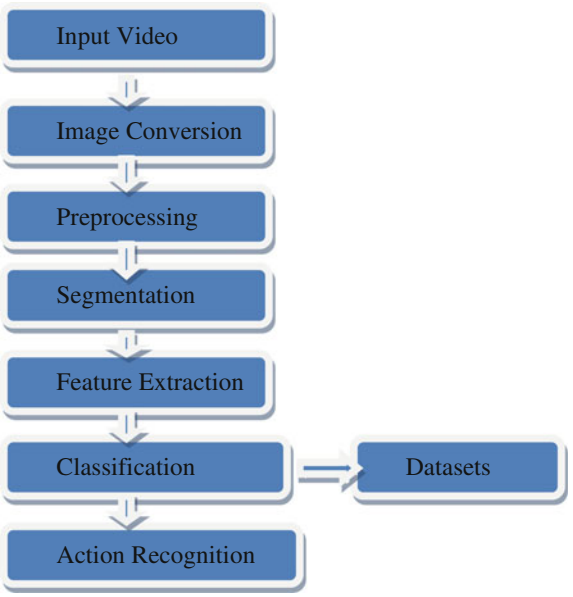

Fig. 1 Work flow of action recognition process

the dataset. Noise removal is typical preprocessing step to improve the result of later preprocessing. It involves some techniques like low-pass filtering, high-pass filtering, band-pass filtering, mean filtering, median filtering.

Median filtering is very widely used in digital image processing because under certain condition, it preserves edges while removing noise. In this paper, we are using median filtering. In median filtering, the 3×3 subregion is scanned over the entire image. At each position, the center pixel is replaced by the median value. Low-pass filtering, high-pass filtering, and band-pass filtering are efficient only in some cases. Most of the times, they blur the image. Median filtering is slower to compute than mean filtering. It prevents edges. It can remove noise.

Mathematical operators such as dilation and erosion are also applied. Dilation is used to connect feature in an image whereas erosion is used to disconnect features in an image and remove small ones. Manual correction tunes an image by editing it.

4. Segmentation:

Image segmentation is an important part in the action recognition system. Segmentation removes unwanted regions as well as finds the boundaries between the regions. Segmentation involves partitioning of an image into the nonoverlapping regions in a meaningful way. It identifies separate objects within an image. Silhouettes are taken out from the images. The Gabor filter can be used for this process of segmentation. Gabor filter can serve as excellent band-pass filter for unidirectional signals. Gabor filters [9] is linear and local. Its convolution kernel is a product of a Gaussian and a cosine function. The filter is characterized by a preferred orientation and a preferred spatial frequency. A 2-D Gabor filter acts as a

local band-pass filter with certain optimal joint localization properties in the spatial domain and in the spatial frequency domain.

5. Feature Extraction:

Feature extraction involves reducing the amount of resources required to describe a large set of data. In this paper, we have used the Principle Component Analysis (PCA) for extracting the features. PCA is sensitive to the relative scaling of original variables. The PCA are orthogonal because they are the Eigen vectors of the co variant matrix which is symmetrical. The purpose of PCA is to reduce the large dimensionality of the data space. The advantages of PCA are that it is a powerful method in image formation, data patterns, and dimension will be reduced by avoiding redundant information without any loss. PCA in signal processing can be described as a transform of a given set of 'n' input vector with the same length 'k' formed in the 'n' dimensional vector.

$$x = [\, x1, x2, \ldots\ldots, xn]^T \tag{1}$$

into a vector y according to

$$y = A(x - mx) \tag{2}$$

The vector mx in above equation is the vector of mean values of all input variables defined by relation,

$$mx = E\{x\} = 1/k \sum xk \tag{3}$$

The rows in matrix A are formed from the Eigen vectors of C_x matrix is possible according to relation

$$C_x = E\left\{(x - mx)(x - mx)^T\right\} \tag{4}$$

The element $C_x(i, i)$ lying in its main diagonal are the variances of x.

$$C_x(i, i) = E\{(x_i - m_i)^2\} \tag{5}$$

The other values $C_x(i, i)$ determine the covariance between input variable x_i and x_j.

$$C_x(i, i) = E\{(x_i - m_i)(\, x_j - m_j)\} \tag{6}$$

PCA is possible according to

$$x = A^T y + mx \tag{7}$$

6. Classification:

For the purpose of classification we are using here KNN classifier. KNN classifier is a minimum distance classifier. KNN algorithm is a nonparametric method used for classification. K-Nearest Neighbors (KNN) algorithm uses a database in which the data points are separated into several separate classes for predicting the classification of new sample point. In KNN classification, the output is a class membership. Here an object is classified by a majority voter of its neighbors. If $K = 1$, then the object is simply assigned to the class of that single nearest neighbor. The algorithm is that a positive integer k is specified, along with a new sample then we select the k entries in our database which are closest to the new sample. We find the most common classification of these entries and this is the classification we give to the new sample.

Classification involves the identification of images. The classification algorithm assumes that the image to be classified has some features and those features belong to one or more classes. The design cycle of the classification involves following steps.

- Collect the data and the labels.
- Choose the features out of them.
- Select a classifier for classification.
- Train the classifier and evaluate it.

4 Experimental Analysis

Experimental analysis gives the comparison between various techniques used for detecting the human actions. After doing the experiment, our method gives us about 90–95 % result (Table 1).

We have also included the graph for ROC (Fig. 2).

Table 1 Comparison of different techniques with proposed techniques

Sr no.	Name of the paper	Result (approx.) (%)	Techniques used
1.	Behavior histograms for action recognition and human detection	86.66	Only classification is done. HOG technique is used
2.	Detecting human action in active video	40–50	Matching of frames is done
3.	Sampling strategies for real time action recognition	83	Dense sampling is used for better performance
4.	Our proposed work	90–95	Median filter, Gabor filter, PCA, KNN classifier

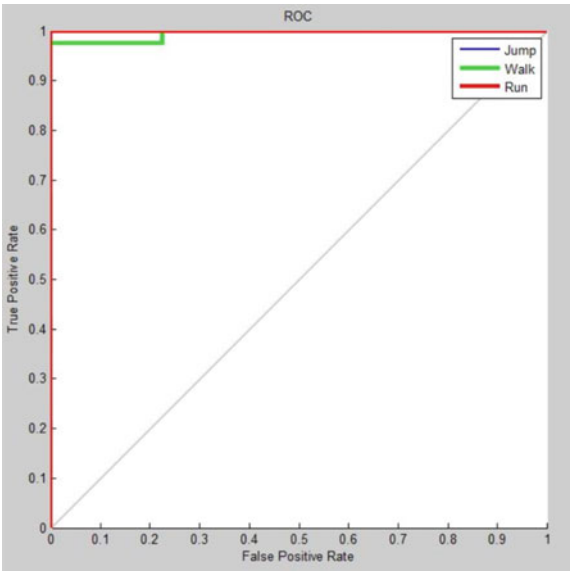

Fig. 2 ROC for some actions like jump, walk and run

5 Conclusion

In this paper, we have explained the methods for feature extraction, segmentation, and classification which are involved in action recognition process. We have provided the methods to recognize human actions from a video. Our method provides the action recognition accurately. The median filter used for preprocessing gives the good result. Other methods also provide good results.

Many more methods can be derived from future work. More methods will be derived to solve the problems involved during detecting actions. The variety of datasets is also discussed.

References

1. Manoj Raman thanz, Student Member, IEEE, Wei-Yun Yau, Senior Member, IEEE and Eam Khwang Teoh, Member IEEE,"Hman Action Recognition with Video Data: Research and Evaluation Challenges", (2014).
2. Hao Jiang, Ze-Nian Liand Mark S. Drew, "Detetcting Human Action on Active Video", IEEE (2006).
3. Ronald Poppe, "A Survey on Vision Based Human Action Recogntion", Elsevier, (2010).
4. Christian Thurau, Czech Technical University, Faculty of Electrical Engineering Department for Cybernetics, Center for Machine Perception 121 35 Prague2, Karlovo n'am'est'$_1$, Czech Republic, "Behavior Histograms for Action Recognition and Human Detection", Springer-Verlag Berlin, (2007).

5. Daniel Weinland a, ft, Remi Ronfard b, Edmond Boyer c “A Survey of Vision Based Methods for Action Representation, Segmentation and Recognition”, Elsevier, (2011).
6. Mohiuddin Ahmad, Seong-Whan Lee, “Human Action Recognition Using Shape and CLG-motion Flow from Multi-view Image Sequences, (2008).
7. Feng shi, Emil Petriu and Robert Lagani’ere, “Sampling Strategies for Real time action Recognition”, CVPR(2013).
8. Raviteja Vemulapalli, Felipe Arrate and Rama Chellappa, “Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group”, Computer Vision Foundation, IEEE (2014).
9. Simona E. Grigorescu, Nicolai Petkov, and Peter Kruizinga, “Comparison of Texture Features Based on Gabor Filters”, IEEE(2002).