International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Artificial Intelligence based Action Recoginition using Image Processing

V. Jeyaramya ME[1], J. Gowsalya[2,] V.Manasvini[3], R. Mahalakshmi [4]
[1]*Associate Professor,*
[2, 3, 4] *Final year –Department of Electronics and communication Panimalar institute of technology.*

Abstract: Human action recognition system proposed here recognizes the behavior of a person in real-time. The system aims at communicating the recognized gestures with the camera system. The proposed system initiates on identifying the human action provided in the database and it sends the trigger signals to the camera system so as to record and store the video stream within the system. Template matching algorithm is used to recognize and identify actions from the captured video frame directly. The image which we are getting from the video is in turn compare with database and after that the output will obtain. As a last phase we are detecting the action also as well as sending an alert message also. Existing video processing system uses the Hadoop platform in order to perform the distributed computing. In our proposed model we use neural network classifier. It allows back to back propagation and provides better efficient and accurate outcomes.
Keywords: Human action recognition, trigger signals, template matching algorithm.

I. INTRODUCTION

Technological advancement in the 3D depth sensors solved many of online computations to stock multiple simple cameras from point of view to capture depth information of the scene. It changed ground views of many computer vision problems like human activity recognition, HCI, tracking, alert systems are responding to suspicions and so on. Because of variance in illumination, cluttering in the background, occluded human body parts, learning temporal features that generalize for same actions is a difficult task. To address these issues we have used hand-engineered features and proven that remarkable data specific performances are obtained. Most of these models, cannot be generalized for real world scenarios. In recent years, deep architectures had proven promising results by using it to extract and learn the hierarchical discriminative features of input data. Human action recognition is an intricate field since the static object characteristics, time, and motion features are considered. Moreover, due to environmental variations including different viewpoints, moving the backgrounds and the large intra-class variations of different actions, the recognition of human actions is quite difficult. On other hand, with exponential growth of multimedia data and videos from the different origins such as CCTV increases the demand of distributed computing to provide services more efficiently. In instance, in every minute almost 300 hours of video are uploaded. Existing video processing system uses Hadoop platform to perform the distributed computing. Though hadoop shows low efficiency in iterative computation, which is essential in machine learning. Moreover, Hadoop does not support real time computation.

II. EXISTING SYSTEM

A. Threshold Segmentation

The simplest method of image segmentation is thresholding. Thresholding is used to create a binary image from a gray scale. colour images are segmented to produce binary image. Each pixel in the source image is assigned to two or more classes in the segmentation process.

B. Knn Classifier

In the pattern recognition, the k-nearest neighbour algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. k- NN is a type of instance-based learning, or lazy learning where function is only approximated locally and all computation is deferred until the classification.

C. Drawback in Existing System

1) It is sensitive to noise.
2) It is difficult to set threshold.
3) Choosing k may be tricky.

4) Test stage is computationally expensive.
5) No training stage, all the work is done during the test stage.
6) This is actually the opposite of what we want. Usually we can tolerate long time training steps, but we want fast test step.

III. PROPOSED SYSTEM

In existing system, it is difficult to set the threshold value. The accuracy is not up to the mark in the existing system. We overcome these drawbacks in our proposed system. This network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with three neurons and an output layer (on the right) with three neuron. These neural network provides better accuracy. They also enable back to back propagation. The back to back propagation assists to recognize the actions from the input video. The actions are fed as database. Those database are fed into the neural network and then back to back propagation is performed. So that actions are recognized.

A. Block Diagram

The basic block diagram of this proposed model:

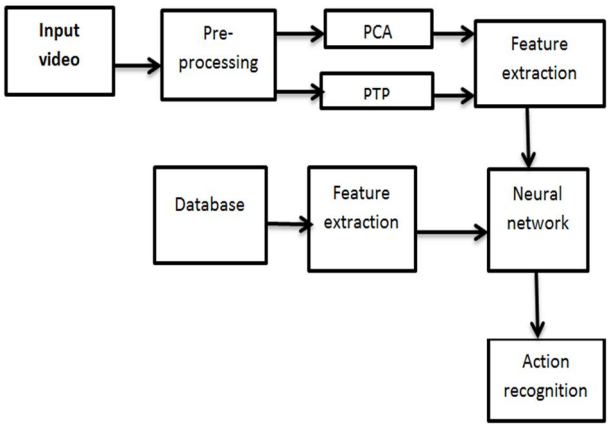


Fig. 1: Basic block diagram of proposed system

1) *Preprocessing:* Pre-processing is a common name for operations with images at the lowest level of abstraction -- both input and output are intensity images. The aim of pre-processing is to improve the image data that suppresses unwanted distortions or enhances some image features important for the further processing.
2) *PCA:* Principal Component Analysis (PCA) is a dimension-reduction tool. It can reduce a large set of variables to a small set such that it still contains most of the information in the large set. The principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible PCA reduces attribute space from a larger number of variables to a smaller number of factors and as such is a "non-dependent" procedure (that is, it does not assume a dependent variable is specified). PCA is a dimensionality reduction or data compression method. The goal is dimension reduction and there is no guarantee that the dimensions are interpretable (a fact often not appreciated by (amateur) statisticians). To select a subset of variables from a larger set, based on which original variables have the highest correlations with the principal component.
3) *PTP:*The proposed Positional Ternary Pattern (PTP) assigns eight bit binary code to each pixel of an image. Initially it computes the edge response of eight neighborhood pixels. Then, we select the primary and secondary direction from those edge responses. Here, we take a further step to select the secondary direction in such a way that, it can represent better corner structure of that pixel. At last, we introduce a ternary pattern of the primary direction, which distinguishes the flat and edge-based region. The coding scheme is described in this section.We apply Kirsch compass masks over the entire image. It convolves in eight different orientations centered on its own position, which obtains its corresponding edge response value in eight directions:Ri = Mi *I where 0 <I <7 (1)

here, I is the entire image, Mi is the ith Kirsh mask, and Ri is the ith response image. Since, shape can be represented by directional information, rather than response magnitude [14], we select the mask index with highest responses as directional information, defined as:

$Zj=\arg \max_{j}\{ Ri(x,y):0<i<7 \}$ (2)

where, Zj holds the two maximum response values from eight Ri Kirsch-responses. Here, j ∈ {1, 2} holds the principle and secondary direction positions. We further check the position of secondary direction.

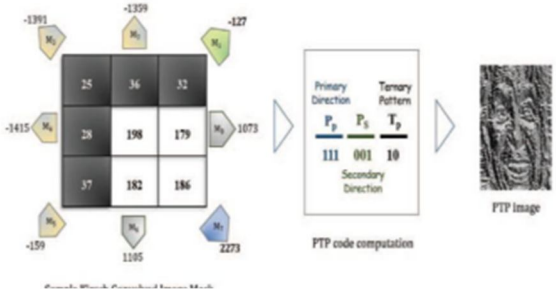


Fig.2: PTP code computation steps

$$\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\begin{bmatrix}5 & 5 & -3\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}$$
$$\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}$$

Fig.3:Kirsch compass masks

If it appears within two neighbors of principle direction, then we search for the maximum response outside the neighborhood of primary direction, and select it as the secondary direction, like this:

$$Z_S = \begin{cases} Z_{Sch}, & \text{if } S_{ch} = \{ P-1, P+1\} \\ Z_{Sch} = \text{otherwise} \end{cases}$$

where, Zs denotes the secondary direction value, already computed from Eq. 2. If its position, s appears within two neighboring pixels of primary direction, p, then Zs changes to Zsch . Otherwise it remains same. We compute Zsch in the following way where, Zsch denotes the new value of secondary direction. Here, arg max returns the maximum Ri Kirsch response value, where i does not include two neighboring position primary direction along with its own position, p. A specific example of this is shown in
In addition, we employ a threshold based ternary coding schema of primary direction to differentiate the smooth region from edge-based region. We encode it using following condition

$$T_P(x,y) = \begin{cases} 1, \text{ if } , Z_p < -\partial \\ 2, \text{if } , Z_p > \partial \\ 0, \text{ if, } -\partial <= Z_p <= \partial \end{cases}$$

where Tj computes the ternary pattern value for primary direction, Zp. Here, ∂ is the threshold value (∂ = 15 used in this paper) to differentiate between smooth and edge region. For instance, in case of edge-based region, we set the ternary value as 2 or 1. Ternary value, 2 represents strong positive response, while, 1 represents strong negative response. Conversely, a weak edge response having ternary value 0, represents the smoother region. With this differentiation, we can generate identical pattern for both smooth and edge-based region. In consequence, the PTP code is created by concatenating the binary form of two principle direction position, along with the ternary pattern. The concatenation is represented by the following equation:
$PTP(x,y)=2^5*P_P(x,y)+2^2*P_s(x,y)+T_P(x,y)$ where (x, y) denotes the current pixel position, while Pp and Ps represent the binary position number of primary and secondary direction respectively. As mentioned earlier, Tp denotes the ternary pattern value for primary direction.

4) Glcm Features

a) *Gray-level co-occurrence matrix:* To create a GLCM, use the gray comatrix function. The gray comatrix function creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i,j) in the resultant GLCM is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image. Because the processing required to calculate a GLCM for the full dynamic range of an image is prohibitive, gray comatrix scales the input image. By default, gray comatrix uses scaling to reduce the number of intensity values in gray scale image from 256 to eight. The number of gray levels determines the size of the GLCM. To control the number of gray levels in the GLCM and the scaling of intensity values, using the Number Levels and the Gray Limits parameters of the gray co- matrix function. See the gray co- matrix reference page for more information.

The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For instance, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the particular offset. To illustrate, the following figure shows how gray co matrix calculates the first three values in a GLCM. In the output GLCM, element (1, 1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. GLCM (1,2)contains the value 2 because there are two instances where two horizontally aligned adjacent pixels have the values 1 and 2. Element (1, 3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the corresponding values 1 and 3. Gray co-matrix continues processing the feeded image, scanning the image for other pixel pairs (i,j) and storing the sums in the corresponding elements of the GLCM.

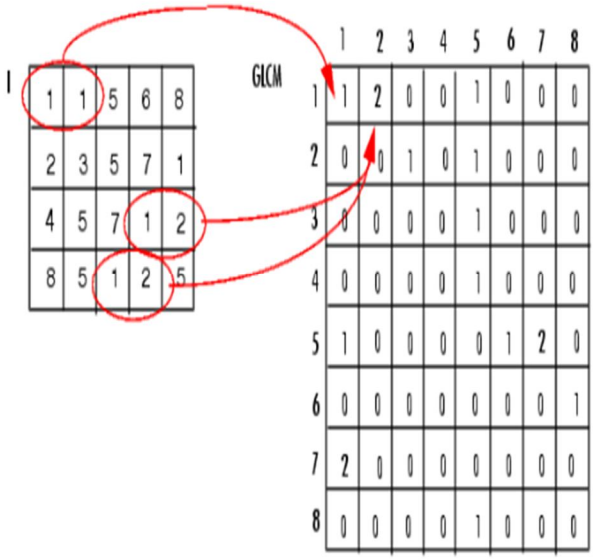


Fig.4:Glcm features

5)Neural Network: Neural networks are the predictive models based on the action of biological neuron. ANN consists of connected nodes. These nodes are termed as artificial neurons. The signals are transmitted between these neurons.

a) *Back Propagation Algorithm* Consider a network with a input x and network function F. The derivative F'(x) is computed in two phases: Feed-forward: the input x is fed into the network. At each node, the derivatives of primitive functions are evaluated and they are stored. Back propagation: The constant 1 is fed into the output unit and the network starts running backwards. The value stored in the left part of the unit gets multiplied with information incoming to the node and this result is transmitted to the left f the unit. The result at the input unit is the derivative of F with respect to x

b) *Steps Of The Algorithm:* The back propagation algorithm is used to make the necessary corrections, after choosing the weights of the network randomly. The algorithm can be done in the following four steps:

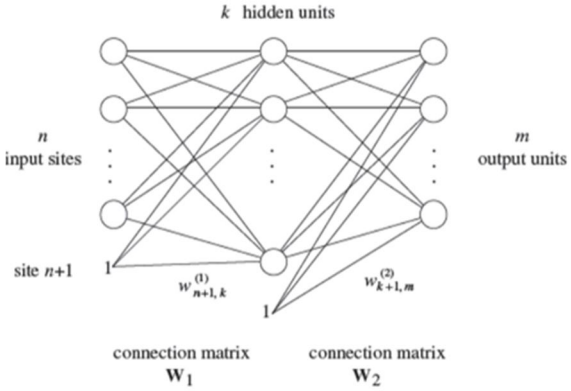


Fig.5:Layers in BPN

Although the implementation is very different, back propagation networks are almost similar to K-Nearest Neighbor (k-NN) models in concept wise. The predicted target value of an item is likely to be same as other items having close values of the predictor variables. Consider this figure:

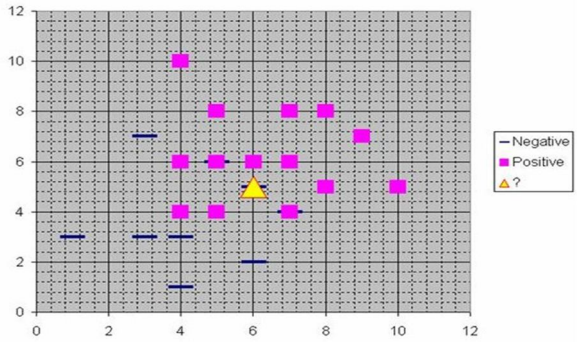


Fig.6:(x,y) coordinates of the model

There are two predictor variables, x and y in training set by assumption. The cases are plotted using coordinates x,y as shown in the figure. Also consider that the target variable has two categories, positive is denoted by a square and negative is denoted by a dash. Now, suppose we are trying to predict the value of a new case represented by the triangle with predictor values x=6, y=5.1, we should predict the target as positive or negative Notice that the triangle is positioned exactly on top of a dash representing a negative value. But that particular dash is in a unusual position compared to the other dashes which are clustered below the squares and left of center. So it could be inferred that the underlying negative value is an odd case. The performance of nearest neighbour classification depends on how many neighboring points are considered. If 1-NN is used , then clearly the new point should be classified as negative since it is on top of a known negative point. If 9-NN classification is used and the neighbouring 9 points are considered, then the effect of the surrounding 8 positive points overbalance the close negative point.

B. Removing Unnecessary Neurons

One of the limitation of BPN models compared to multilayer perceptron networks is that BPN models are large ,because there is one neuron for each training row. This makes the model to run slower than multilayer perceptron networks when predicting values for new rows.

DTREG provides a solution to remove unnecessary neurons from the model after the model has been constructed.

Removing unnecessary neurons has three benefits

1) The size of the model is reduced.

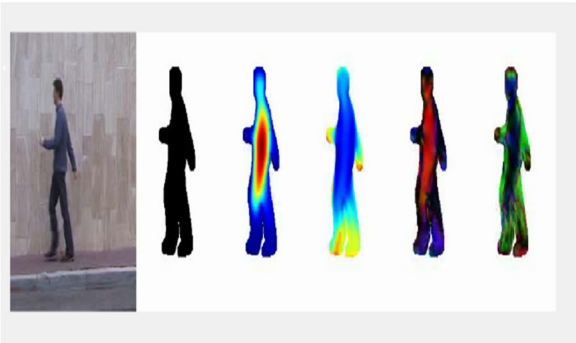
2) The time consumption will be less on processing the model.

3) Removing neurons improves the accuracy of the model.

The process of removing unnecessary neurons is an repetitive process. Leave-one-out validation method is used to measure the amount of error of the model with removal of each neuron . The neuron that causes the least increase in error (or possibly the largest reduction in error) is then removed from the model on further processing. The process is repeated with the remaining neurons

till the stopping criterion is reached.When unnecessary neurons are removed, the "Model Size" section of the analysis report shows how the error changes in accordance with different numbers of neurons. There are three criteria that can be selected to guide the removal of neurons:

C. Frame Conversion



IV. CONCLUSION

Due to the advancement of the Internet, social media video services and intelligent CCTV for video surveillance system, the multimedia data such as video is increasing rapidly. Moreover, understanding video context and identifying video types has an notable significance in the management of massive video data. In order to manage these videos and provide services to the users, it is necessary to understand the human activities from the videos automatically. There are many applications, which focus on the action recognition, such as crowd behavior prediction,robotics, video surveillance, human-machine interaction and sports game analysis.

REFERENCES

[1] Jianguang Zhang,Yahong Han,Jinhui Tang,Qinghua Hu and Jianmin Jiang"Semi supervised image to video adaptation for video action recognition",IEEE Transaction on cybernetics,2016.
[2] Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh, Advances in Human Action Recognition: A Survey, arXiv:1501.05964v1 [cs.CV] 23 Jan 2015.
[3] Chen, Roozbeh Jafari, Nasser Kehtarnavaz, A survey of depth and inertial sensor fusion for human action recognition, ©Springer Science + Business Media New York, 2015
[4] H.Wang et al,"Action recognition using nonnegative component representation and sparse basis selection ,"IEEE Trans.Image Process,vol.23,no.2,pp.570-581,Feb 2014.
[5] B.Yao and L.Fei-Fei,"Grouplet:A structured image representation for recoginizing and human object interactions",in Proc.IEEE Conput.Vis .Pattern Recognition.,San Fransico,CA,USA,2010,pp.9-16
[6] Karl-Friedrich Kraiss, Advanced Man-Machine Interaction Fundamental & Implementation, © Springer-Verlag Berlin Heidelberg,2006
[7] J., Ryoo, M., Human activity analysis: A survey, ACM Computing Surveys 43, 1-43, 2011. Daniel Weinlanda, Remi Ronfardb, Edmond Boyerc, A Survey of Vision-Based Methods for Action Representation, Segmentationand Recognition, volume 115, Issue 2, February 2011, Pages 224-241.
[8] Volker Kruger, Danica Kragic, Ales Ude, Christopher Geib, The Meaning of Action: A review on action recognition and mapping,Taylor and Francis Online, volume 21, Issue 13, 2007.
[9] V.Delaitre,I.Laptev and J.Sivic,"Recognizing human actions in still images;A study of bag-of-features and part based representation" in Proc.brit.Mach.Vis.Conf.,aberystwyth,U.K.,2010,pp.97.1-97.11.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com