ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Human Action Recognition using Neural Network

[1] S.Maheswari, [2] P.Arockia Jansi Rani,
[1] Research Scholar, M.S.University, [2] Associate Professor, M.S.University

Abstract: - **Recognizing human action from video still remains a problem. Robust solutions to this problem have applications in various domains such as surveillance systems, human computer interaction, smart home health-care systems and control free gaming systems etc. In this paper, an approach for human action recognition based on neural network is presented. In this neural network approach, motion vector feature is used for action recognition. The features are extracted directly from the video without any pre-processing. Since preprocessing is not necessary for the proposed approach computation time has been minimized as compared to that of the existing system. Then the network is trained using scaled conjugate gradient back propagation method. Neural network is trained with the motion vector features. The training is based on the number of hidden neurons, the percentage of sample data taken for training, validation and testing. This system provides an efficient result with a minimum number of hidden neurons and training data. The proposed approach is tested on Weizmann dataset that consists of 10 actions providing 9 videos per activity.**

Keywords: - Feature extraction, motion vectors, neural network, machine learning.

I. INTRODUCTION

Human activity recognition is an important area of computer research. It includes various applications such as surveillance systems, Health care systems and humans and electronic devices interaction. Human action recognition have invariance in scale, position, occlusion of objects, angle of camera through which it usually capture videos.
Learning from a set of examples is an important attribute needed for most pattern recognition systems. Artificial neural network is an adaptive system that changes its structure based on external or internal information that flows through the network and are being widely used in pattern recognition systems [6]. They have a better performance in nonlinear applications. Artificial neural networks are trained, so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the difference between the target and the output reduces to the minimum value [1]. The training of the neural network continues until the mean squared error of the weights lies below a certain threshold or until the maximum number of iterations is reached. Once training is completed, the network can be applied for the actual classification of the data [3]. The classification technique used may be one of the following: 1) Supervised classification - in which the input pattern fall as a member of a predefined class. 2) Unsupervised classification - in which the pattern falls into an unknown class as there are no predefined classes. The recognition problem is a classification or categorization problem, where the classes are either predefined by the system designer or are learned based on the similarity of patterns. It is important to note that learning or training takes place only during the design phase of a pattern recognition system [2].
Human action recognition using neural network [3, 4] show stability and reduce time complexity. The main feature for using neural network in human action recognition is fast learning about actions. A more effective learning algorithm Scaled Conjugate Gradient than the standard back propagation (BP) has been used. SCG belongs to the class of Conjugate Gradient Methods, which shows super linear convergence on most problems. Through several experiments its proven that SCG is at least an order of magnitude faster than BP. The speed-up depends on the convergence criteria, i.e., the bigger demand for reduction in error, the bigger the speed-up. By using a step size scaling mechanism SCG avoids a time consuming line-search per learning iteration, which makes the algorithm faster than other second order algorithms. Scaled conjugate gradient back propogation [4] method is used for recognizing human actions. Ravine phenomena is difficult to handle using back propagation algorithm but Scaled Conjugate gradient back propagation handles the ravine phenomena efficiently and solve the problem without any difficulties. SCG uses second order technique and finds better way to a local minimum than a first order technique but the computational cost is high. The optical flow features (i.e., motion vectors) both the horizontal and vertical flow values are normalized using mean, standard deviation and variance are fed as input to the neural network for recognizing actions.

II. PROPOSED ARCHITECTURE

This section presents the detailed analysis of the performance of the proposed work using neural network. In this approach motion detection is used as the feature for recognizing the action. The optical flow features (i.e., motion vectors) both the horizontal and vertical flow values are normalized using mean, standard deviation and variance are fed as input to the neural network. The standard back propagation algorithm evaluates the gradient of global error function with respect to the weights and update the weights. The optimized method is conjugate gradient algorithm which avoids zigzag approach to the minimum. Moller's scaled conjugate gradient algorithm is employed in this system. It is used because learning speed is faster compared with above said other methods and it eliminates the dependence on user selected parameters. This method is generalized backpropagation algorithm.

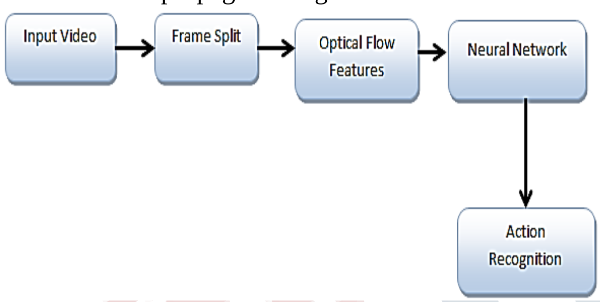

Figure 1: Proposed Action Recognition System using Neural Network

Input Data and Frame Extraction

To process the proposed model Weizmann dataset is taken which is shown in the figure.2. The actual dataset is available in the form of videos so frame extraction is done to extract the frames from these videos. The database contains ten natural human actions (walking, running, skipping,jumping,waving two hands, waving one hand,galloping,jumping jack and bending) performed by nine persons. It contains a total of 93 sequences. All sequences are taken with a static camera with 25fps frame rate, downsampled to spatial resolution of 180x144 pixels.

Motion Feature Extraction Using Optical Flow

There are dense and sparse techniques in optical flow. The sparse technique processes only some pixels from the whole image, and generally executes feature tracking and it is usually used in time-critical applications. Lucas kanade is a sparse optical flow technique which is employed in this system for extracting motion information. The optical flow is the pattern of apparent motion of objects, surfaces and edges in a visual scene caused by the relative motion between the camera and the scene. The optical flow is a vector which represents the object velocity in the images and it is used to detect the direction of the moving objects.

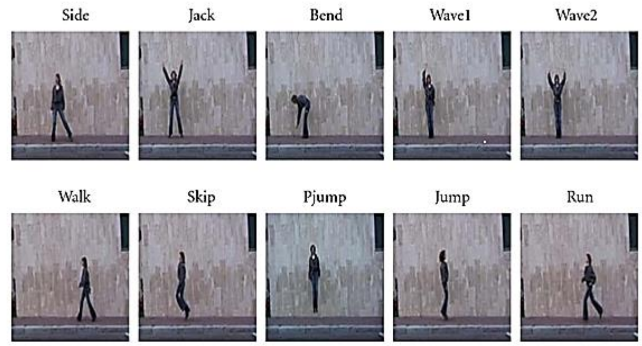

Figure 2: Weizmann Dataset

Neural Network training and testing

Training neural network is equivalent to minimizing a multivariable global error function of the network weights, which can be standard mean square error or any other appropriate error function. Training and testing are carried on the basis of user defined data and for each run the normalized mean square error on the training set after convergence and on the testing set were recorded. The average normalized mean square error is shown in table 1. The minimum value of '0' for nmse denotes the perfect prediction and maximum value of '1' indicates that the error is high. In the training process the neural network requires input variable, optimization algorithm, and the target which is to be achieved. In this system the motion vector features extracted using optical flow are fed as input to the neural network. The input data given to the neural network is splitted into three ways. Ex: If Forty percent of the total input samples are used for training ,then 30 percent for testing and 30 percent for validation. Feed forward neural network with scaled conjugate gradient backpropagation algorithm is used. Supervised learning training pattern is employed in this method. Two hidden layers are used .To select the number of hidden neurons for performing training , neuron number range from 1 to 100.The output of each network with varying number of hidden neurons is evaluated.

The output of network with 10 neurons is efficient. The target is limited between 0 and 1 because of the activation function used. To check the validity and performance the neural network is trained for specified iterations . Training stops when any of these conditions occur: 1. The maximum number of iterations is reached. 2. The performance gradient falls below minimum. 3. Validation performance has increased more than max times since the last time it decreased 4.The achievement of the performance goal.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Performance Analysis

The performance of the system is evaluated on the basis of percent sample data taken for training, number of hidden neurons and mean squared error. The number of iteration is the iteration used for training the input data. If the number of iteration is lower then the time complexity is also reduced. Table 1 shows that most of the cases the desired output is with ten neurons. Lesser number of neurons the complexity is lower. Mean Squared Error (MSE) is the performance metric that determines the network performance.MSE is the average squared difference between outputs and targets. Lower values of MSE shows the better performance of the network as zero indicates no error.

Number of hidden neurons	Normal mean square Error					
	Training (70%)	Validation (15%)	Testing (15%)	Training (40%)	Validation (30%)	Testing (30%)
30	0.0045	0.2345	0.2677	0.0046	0.2344	0.2654
20	0.0034	0.2156	0.2578	0.0038	0.2200	0.2574
10	0.0025	0.2132	0.2434	**0.0028**	**0.2138**	**0.2440**

Table1: Normal Mean Square Error for Varying hidden Neurons

Recognition Rate

Recognition rate is calculated on the basis of number of true matches and false matches.

Number of correctly matched actions=87

Total number of tested actions=99

Actions	Recognition Rate
Walk	90
Run	80
Bend	100
Jump	100
Wave1	90
Wave2	90
Jack	80
Pjump	90
Skip	80
Side	90

Table 2:Experimental Results for Recognition Rate

$$\text{Recognition Rate (\%)} = \frac{\text{Number of correct match}}{\text{total Number of tested actions}} \text{X100}$$

The confusion matrix is shown for 10 hidden neurons, 40 percent training , 30 percent testing and 30 Percent validation data. To assess the accuracy of an image classification, it is necessary to create a confusion matrix. In a confusion matrix, the classification results are compared to truth information. The strength of a confusion matrix is that it identifies the nature of the classification errors.

Actions	Walk	Run	Bend	Jump	Wave1	Wave2	Jack	Pjump	Skip	Side
Walk	9	1	0	0	0	0	0	0	0	0
Run	1	8	0	0	0	0	0	0	1	0
Bend	0	0	10	0	0	0	0	0	0	0
Jump	0	0	0	10	0	0	0	0	0	0
Wave1	0	0	0	0	9	0	1	0	0	0
Wave2	0	0	1	0	0	9	0	0	0	0
Jack	1	1	0	0	0	0	8	0	0	0
Pjump	0	0	0	0	0	0	0	9	1	0
Skip	0	0	0	0	0	0	1	1	8	0
Side	0	1	0	0	0	0	0	0	0	8

Table 3:Confusion Matrix

III. CONCLUSION

In this system human action recognition is performed with the optical flow features using neural network. The optical flow is used to extract the motion features. Then using the motion features neural network is trained using scaled conjugate gradient backpropagation method. Then the action is recognized. The videos are processed without any preprocessing and therefore the computation time has been reduced. The results are presented for the Weizmann dataset. Other datasets like KTH dataset, HOHa dataset can also be used. This system provides an efficient result with minimum number of neurons and therefore the complexity has been reduced

REFERENCES

1. Lacher R. C, Hruska S. I, and Kuncicky D. C, "Backpropagation Learning in Expert Networks", IEEE Transactions on Neural Networks 3, pp. 62–72, 1992.

2. Battiti R, "First- and second-order methods for learning; between steepest descent and Newton's method", Neural Computation, Vol. 4, pp. 141–166, 1992

3. Baldi P. F and Hornik K, "Learning in linear neural networks: A survey", IEEE Transactions on Neural Networks, Vol. 6, pp. 837–858, 1995.

4. Weber M, Welling M, and Perona P, "Unsupervised learning of models for recognition", Lecture Notes in Computer Science, Jan 2000.

5. D.Pham, S. Sagiroglu, "Training multilayered perceptrons for pattern recognition: a comparative study of

four training algorithms", International Journal of Machine Tools and Manufacture, vol.41, 2001, pp. 419–430.

6. Christopher M. Bishop, "Neural Networks for pattern recognition", Oxford University Press, published in 2005.

7. W.W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods", Pacific of Journal Optimization, 2:35, 2006, pp. 35–58.

8. Rafael Uetz and Sven Behnke, "Large-scale Object Recognition with CUDA-accelerated Hierarchical Neural Networks", In Proceedings of the 1st IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009.

9. Hafiz T. Hassan, Muhammad U. Khalid and Kashif Imran, "Intelligent Object and Pattern Recognition using Ensembles in Back Propagation Neural Network", International Journal of Electrical & Computer Sciences IJECS-IJENS, Vol. 10, No. 06, pp. 52-59, December 2010

10. Youngbin Park, Sungphill Moon, and Il Hong Suh. Tracking human-like natural motion using deep recurrent neural networks. arXiv preprint arXiv:1604.04528, 2016.