A Hybrid System of SOM and RBF for Recognizing Human Activity from Video

Dagnachew Melesew Alemayehu, Bhabani Shankar D. M.

***Abstract*— Recognizing human activities in fully controlled environment is very simple task but in partially observable and uncontrolled environments recognizing activities from videos is a challenging problem and has many practical applications for instance Security and Surveillance. Computer Vision has been researched for a long period of time to recognize human activities from video on which the environment is uncontrolled, the problem is reduced to activity prediction from unfinished activity streaming, which has been studied by many researchers. There could be uncontrolled environments that may create error for the recognition. In this paper, we try to combine both supervised and unsupervised learning network method that can recognize human activities from both controlled and uncontrolled environments**.

***Index Terms*—** Artificial neural network, Object recognition, SOM, RBF and Computer Vision.

I. INTRODUCTION

Machine learning involves adaptive mechanisms that enable computers to learn from the previous knowledge. Learning capabilities can improve the performance of an intelligent system over time. The most popular approaches to machine learning are artificial neural networks. Artificial neural networks (ANN) are machine learning techniques that are used to model human brain and consist of a number of artificial neurons. Neurons in artificial neural network consist of fewer connections than biological neurons-each neuron in ANN receives a number of inputs and an activation function is applied to these inputs.

Neural networks look like the human brain in the following two ways:

i. A neural network acquires knowledge through learning.

ii. A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights

In artificial neural network, a self-organizing map (SOM) for unsupervised learning and radial basis function network (RBF) for supervised learning can be integrated into one system for a computer vision that are capable of recognizing human activities from video. A learning machine need not be given all of the details of its environments if we provide all the details of the environment it takes time to execute all the instruction and its sensors are responsible for detecting the environments. In this paper we investigate neural network architecture for recognizing human activities that is taken from videos. A learning machine that is capable of recognizing human activities from videos is done by combining two learning paradigms which can be defined as follows:

In supervised learning, the goal is to construct an input and output mapping $Y=F(x)$ that predicts the output $y=(y_1, y_2, \ldots, y_m)$ for an input $x=(x_1, x_2, \ldots, x_n)$, the mapping is found from example of the desired output $\{y_1, y_2, \ldots\}$ at the input data points $(x_1, x_2 \ldots\}$ in order to minimize the expected output error.

In, unsupervised learning, only a set of input data $\{x(1), x(2)\ldots\} \in R$ is given and the goal is to construct a mapping so that the output $\{y(1), y(2)\ldots\} \in R$ fully characterizes the statistical properties of the input.

In general, this requires a nonlinear function approximate with good generalization characteristics. We choose the radial basis function RBF network as the main framework for our study of recognizing human activities from video. Moreover, it is possible to train the RBF network in two stages with the basis function-first being determined by unsupervised learning and then the second layer weights being determined by the supervised learning.

In computer vision for recognizing human activities system presented in this paper, a self-organizing map (SOM) is first used to cluster sensory information into prototypes according to a topographic mapping, then the RBF is used to implement the nonlinear mapping

Manuscript received June 17, 2014

Dagnachew Melesew Alemayehu, School of Computing and Electrical Engineering, Bahir Dar University, Ethiopia

Bhabani Shankar D. M., School of Computing and Electrical Engineering, Bahir Dar University, Ethiopia

from sensory spaces to a motor action space. In this two layer learning structure, the SOM has two functions - one is to compress the large amount of sensory information in order to reduce the computation time and the other is to divide the high dimensional sensor information space into some small areas which are used to generate smoother approximations between the sensing information. The cluster center obtained with the SOM are used to initialize the center of the basis function in the RBF network and corresponding receptive fields are calculated by maximum likeli hood estimation, the second layer weights in the RBF network are then obtained by the supervised learning. By using the well-known least mean square (LMS) algorithm, the output of the RBF Network is transitional.

II. RELATED WORK

Artificial neural network is a popular machine learning approach in recent computer vision and robot system learning and is used for simulating human brain. The main idea is to use the ANN approaches for recognizing human activity which is taken from a video device which has become a part of our everyday life. With video sharing websites experiencing relentless growth, it has become necessary to develop efficient indexing and storage schemes to improve user experience. This requires learning of patterns from raw video and summarizing a video based on its content. Content based video summarization has been gaining renewed interest with corresponding advances in content based image retrieval (CBIR) [6]. Summarization and retrieval of consumer content such as sports videos is one of the most commercially viable applications of this technology [7].

Security and surveillance systems have traditionally relied on a network of video cameras monitored by a human operator who needs to be aware of the activity in the camera's field of view. With recent growth in the number of cameras and deployments, the efficiency and accuracy of human operators has been stretched. Hence, security agencies are seeking vision-based solutions to these tasks which can replace or assist a human operator. Automatic recognition of anomalies in a camera's field of view is one such problem that has attracted attention from vision researchers [[8], [5]]. A related application involves searching for an activity of interest in a large database by learning patterns of activity from long videos can be referred from [9], [10].

III. STATEMENT OF THE PROBLEM

Recognizing human activities from video using RBF for controlled environment is very simple and straight forward, because we feed all the training to the neural network and then, based on the training data set, the network will recognize the activities. But when we apply for uncontrolled environments, the capability of recognizing activities from video becomes minimal and cannot converge easily and also the convergence rate is very low.

IV. METHODOLOGY

The methodology that we use for recognizing human activity from video is a combination of radial basis function (RBF) for recognizing in a fully controlled environment and SOM (Self Organizing Map) for uncontrolled environments.

In RBF, all the training data set is given to the network for training. Once the network is trained using RBF for fully controlled environments, it is a very simple task to recognize the activity. Then the output of this controlled environments or RBF will be taken by SOM for uncontrolled environments. This helps us to take a minimum iteration for choosing an activation value and also provides a higher rate of convergence.

V. THE LEARNING SYSTEM

Recognizing human activities from video through a learning computer vision system is composed of unsupervised learning and supervised learning; there are also input neurons at unsupervised learning. The architecture contains SOM for unsupervised learning and RBF for supervised learning that are integrated into one system for acquiring and recognizing the activities of human which is sensed by a vision system. A generic action or activity recognition system can be viewed as proceeding from a sequence of images to a higher-level interpretation in a series of steps. The major steps involved are the following:

1. Input video or sequence of images
2. Extraction of concise low-level features
3. Mid-level action descriptions from low-level features
4. High-level semantic interpretations from primitive actions

The parameter such as the number of hidden neurons, learning rate and the respective field of RBF have to be carefully chosen in order for the successful implementation of computer vision system, therefore the adjustment of these parameter is crucial for successful application of neural network.

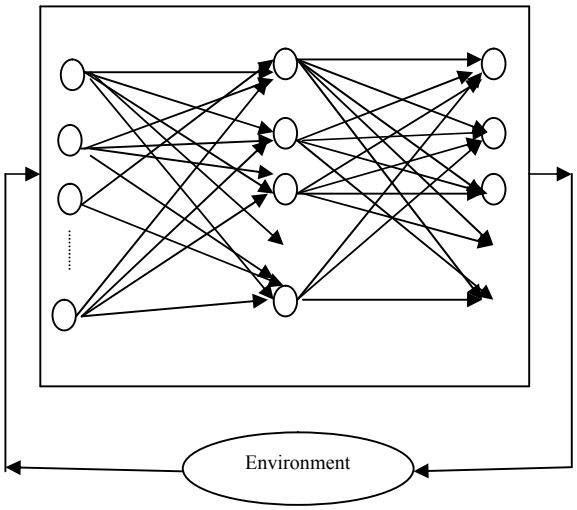


Figure 1: SOM and RBF

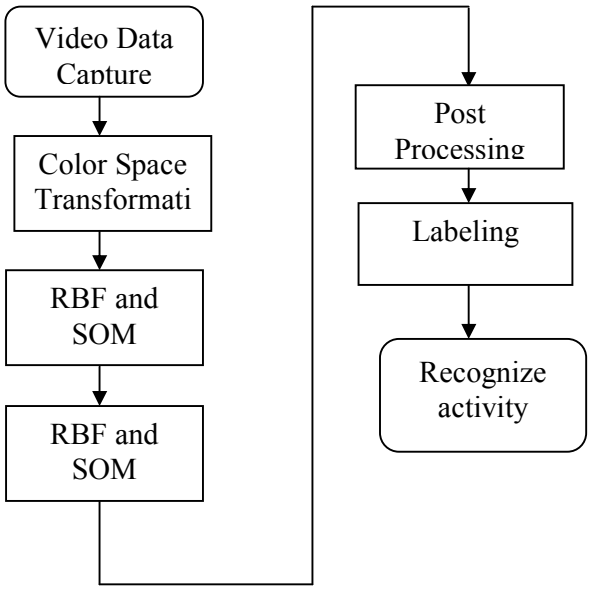
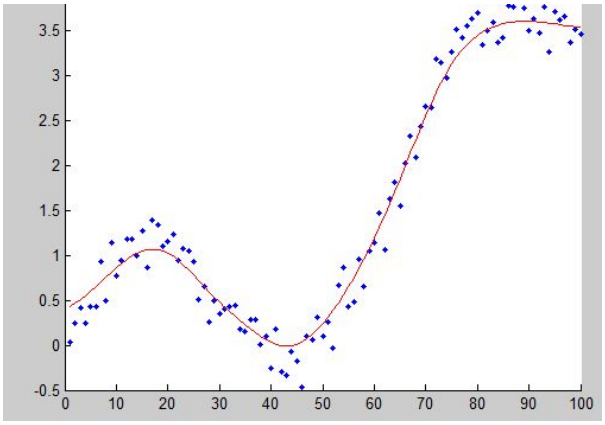


Figure 2: Recognizing human activity

VI.EXPERIMENTS

Intensive experiments have been conducted to address the validity of the sensed videos. First, we have tested the same through the vision program in several stages to check the accuracy. Initially, the achieved results were tested from the vision program and are found to have a remarkable precision although the vision is affected by the camera resolution and lighting. The vision system simulated camera resolution used was 640×480 because this is the average camera resolution in order to sense the visual object.



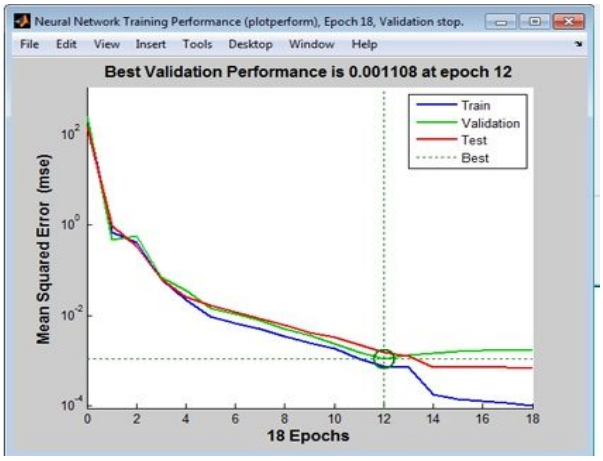


Figure 3 & 4: Performance Evaluation

VII. SAMPLE CODE

```
%clc
clear
close all

n1 = 1:2:200;
x1 = sin(n1*0.1);

n2 = 2:2:200;
x2 = sin(n2*0.1);

xn_train = n1;
dn_train = x1;
xn_test = n2;
dn_test = x2;
%-------------------------------------------------------
switch 1
case 1

P = xn_train;
T = dn_train;
spread = 40;
net = newrbe(P,T,spread);
case 2
P = xn_train;
T = dn_train;
goal = 1e-12;
spread = 40;
MN = size(xn_train,2);
DF = 1;
net = newrb(P,T,goal,spread,MN,DF);
case 3
P = xn_train;
T = dn_train;
spread = 0.5;
net = newgrnn(P,T,spread);

end

err1 = sum((dn_train-sim(net,xn_train)).^2)

X = sim(net,xn_test);
err2 = sum((dn_test-X).^2)
```

VIII. CONCLUSION

In this paper, we combine both RBF and SOM for recognizing human activities from both controlled environments and uncontrolled environments. Partially observed videos providing a machine the ability to see and understand as humans do has long fascinated the scientists, the engineers and even the common man. However, several more technical and intellectual challenges need to be tackled before we get there. The advances made so far need to consolidated, in terms of their robustness to real world conditions and real time performance. This would then provide a concrete ground for further research. Our hybrid neural network has a less performance on recognizing complex environments and also has less performance on recognizing multiple environments at a time. In our future work, we shall focus on providing a better solution for addressing such problems. The data set that we use is also not sufficient i.e. we took 15 videos of different environments.

REFERENCES

[1] S. Haykin, *Neural Networks: A comprehensive Foundation*. 2nd Edition, Prentice Hall.

[2] M. Takizawa, Y. Makihara, N. Shimada, J. Miura and Y. Shirai, *A Service Robot with Interactive Vision- Object Recognition Using Dialog with User,* In Workshop on Language Understanding and Agents for Real World Interaction, 2003

[3.] www.gc.ssr.upm.es/inves/neural/ann1/ anntutorial.html

[4.] Paolo Gaudiano, Eduardo Zalarna, and .Juan Lopez Coronado. *An unsupervised neural network for real time low level control of a mobile robot: noise resistance, stability, and hardware implementation*, Boston University Center for Adaptive Systems, 1994

[5] N. Vaswani, A. K. Roy-Chowdhury, and R. Chellappa, *""Shape Activity": a continuous State HMM for moving/deforming shapes with application to abnormal activity detection,"* IEEE transaction on Image Processing, Volume:14, No. 10, 2005

[6] Y. Rui, T. S. Huang, and S. F. Chang, *"Image retrieval: current techniques, promising directions and open issues,* Journal of Visual Communication and Image Representation, Volume 9, No. 1, 1998

[7] S. F. Chang, "*The holy grail of content-based media analysis*", MultiMedia, IEEE, Vol 9 , Issue 2, 2002

[8] H. Zhong, J. Shi, and M. Visontai, "*Detecting unusual activity in video*", Computer Vision and Pattern Recognition, IEEE Explorer, Vol 2, 2002

[9] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, 2000

[10] W. Hu, D. Xie, T. Tan, and S. Maybank, "*Learning activity patterns using fuzzy self-organizing neural network*," IEEE Transactions on Systems, Man, and Cybernetics, Part B Volume:34 , Issue: 3, 2004

[11] R. R. Elsley. Elsley. *A learning architecture for control based on back-propagation neural networks*. IEEE International Conference on Neural Networks, 1988