Neural Network Techniques Applied on Real time Human Gesture Recognition: A Survey Paper

Sapna Varshney¹, Ritu Tiwari²

¹,² Robotics & Intelligent System Design Lab, Indian Institute of Information Tech. & Mgmt., Gwalior, India

ABSTRACT

Gesture recognition appertains to recognize expressions of action which are very meaningful by a person, that involves the face, arms, hand and body. It is utterly important for designed and systematized intelligent and efficient based human-computer interconnection. Often Self organized Neural Network was employed for determining a topographic record of representative person body gestures. For obtaining compact video representation Fuzzy Vector Quantization is applied to person body gestures emerging in a video that will be use for person classification and action recognition. One feed forward network and another artificial neural networks (ANN) are skilled to identify the person ID and test action video through action class levels. Output combination in network, based on one more feed forward network was performed. In the case of simultaneous several cameras were used for training and detection phases. Gesture recognition are signified various applications. In this survey report the gesture (Real Time Human Action) recognition and beaviour which allocate emphasize particularly on body gestures relate with facial expression that which neural network used. Applications involving Hidden Markov models, PBL(Projection Based Learning Algorithm) approach, Motion Caption Data, Kinect, Histogram 3d joints have been discussed in details. Ongoing presented challenges and upcoming research possibilities have been also high lighted.

KEYWORDS:
Real time Human action, HMM(Hidden markov Model), Convolutional Nerual Networks(CNN), Histogram Techniques etc.

1. INTRODUCTION

Soft computing techniques are widely used for the recognition of human action and behavior. Traditional passive video surveillance system has been replaced with developing an intelligent visual surveillance. The understanding and depiction of human behaviors must be enhanced through visual surveillance enforcement [2], [3]. A person’s routinal activities and determined home care emergency can be monitored by indoor visual behavioral surveillance automatically. Visual surveillance generally sum up with the ongoing stages: person detection, environmental paradigm, tracking, understanding and depicting about behaviours. Understanding behavior signifies motion design sequences or analyzed and classified body postures. Most of avenues has proposed to perform on motion anatomy and division [4] [5] [6]. A real-time motion anatomy by picture features skeletonization was defined application to [4] to adjudge human exercises like that running or walking [5]. A proposed method was paradigm-based, for the analysis of human walking motion. The movement confine information consists of the three-dimensional spot which describe the skeleton of the human. Fig.1 illustrates the representation of person skeleton using the joints. Different types of motion were also apprehended by the (HMM) in study. Support Vector Machine (SVMs) is proposed for human gait discerned which was depend upon motion analysis [6]. A few in number the explorations for body stance classification emphasize on different standing stances [7], [8], [9]. Classified standing posture consist of distinctive human arms masquerade by SVM’s is applied in [7] the model [8] [9]. Human body silhouette is detected through proposed complex 2-D model, and acquiring posture recognition which is basis on a multi-scale morphological tactics [8]
and genetic algorithm [9] is also approached. Dynamic time warping (DTW) was implemented for Dynamic gesture recognition. The manifold neural networks are used for action recognition vastly.

Firstly the Neural Network initiated with hidden markov model (HMMs), neural networks are proposed continuously [10], [11], [12]. [10] The authors used to follow DTW to recognize human gestures. Proposed HMMs model used to recognized Sign language. [11]. Hand pantomime identification through time-delay network and recurrent fuzzy neural network was implemented [12] [13].

Here we describe second neural network, a proposed method of recognition for real-time human being action based on a modified Deep Belief Network (DBN) model. The process can recognize real-time performance, and didn't need extra segmentation process. For actions with related movements and poses, the correctness indeed obviously not low.

Third neural network is, based on the idea, [14] propounded neural dynamic paradigm pertained to as a multiple time scale recurrent neural network (MTRNN) is imitating the regulate system of person motor. An MTRNN is an prolongation of continuous time scale recurrent neural network (CTRNN) [15]. continuous signal can be used as prediction. A global coherent structure appearant system which is based on the phenomenon of self-organization possessed by MTRNN model. Convolutional neural networks is stands on fourth (CNNs) [17] it is a type of broad paradigm which could perform sinuously on raw intake, thus excavate automated execution of feature construction. However,

such paradigms bounded to manage 2-dimensional intake correctly. In this survey paper, a strange 3-dimensional CNN paradigm is developed for action savvy. [17], a novel three-dimensional CNN model is developed for action identification. The model based on feature extraction in temporal and spatial dimension is performed by 3-Dimensional convolutions, by means of that motion captured information is encrypted in multiple adhered frames.

Multiple mediums of information from the intake frames are developed by such models. by associating all facts from different channels are used for features representation.

In fifth neural network category the Artificial neural network (ANN) applying motion identification solution was consist of two (NN). Both neural
networks were systematized as in a cataract framework. The very first ANN model has required to recognize the body position feature from the related pictures: and the succeed ANN model is contrived basis on the attributes that are discriminated by the first ANN paradigm is to analyze the direction of body through actions. Concluded, they implemented the pragmatic result of this work is all about demonstrated in this survey.

The format of the survey has followed. Given Section 2 introduction of Gesture Recognition Approaches and literature Review. In section 3 descriptions of Implementation Tools. Section 4 described Proficiency. Section 5 describes Research Gap. Finally concluded in Section 6.

2. GESTURE RECOGNITION APPROACHES
There are numerous methods to sense human action in the acquire picture later than pre-processing as publicized below, we separated these methods in two parts.

2.1. Techniques/Models Based Approaches
Lots of researchers have used these techniques to detect human action in images and videos. These techniques are described below.

Junji Yamato ,et al.(1992) proposed a model of HMMs(Hidden Markov Model) for recognize human actions in valid instance imagery. They used a method of bottom up approach with HMMs. Their experiment used tennis action. They categorized this experiment in six tennis stokes: forehand stroke, backend stroke, backend volley, forehand volley, service and smash. This method deal with 2d images. Experimental results of real time images recognize rate more than 90%, and they confirm result rate increase if number of human want to produce the training data. This method can be extended to 3D object action recognition using HMM model [18].

Hongying Meng, et al.(2006) used the method of motion information and SVM for recognize human action. They had chosen three main techniques to build a method that can work in real time. First they choose motion based history images and related features. Second they choose template matching method instead of state space method. Final they choose linear classifier support vector machine(SVM). The main default feature of method second and third was it could work on in order and motionless background. Data recorded in feature vector if background was motion or noise that reduced the performance classification. [19].

Hongying Meng,et al.(2008) proposed a model of an embedded, re-configurable video processing design. Here they proposed a real-time, embedded image resolution for person action detection, applied on an FPGA-based universally tool. There were three main contribution. Firstly they had evolved a quick person action identification structure with easy motion feature and a SVM. Secondly, they had developed a re-configurable, FPGA based video processing architecture. Finally, they had fruitfully implemented a action identification method on this re-configurable design. They gave 80% average respect rate using some degree of training data [20].

planned method had two steps: attribute extraction and classification. Firstly to extract features from the arm movement data. Extended kalman filter (EKF) was useful to take out the feature from the eating action data in a 3d space in real time and secondly was classification. The hierarchical temporal memory (HTM) network was adopted to organize took out features of the eating action. In this paper result was average triumphant rate of the eating and drinking was 100% [21]. Dirk Gehrig, et al.(2009) proposed a model of HMM based with optical flow of data to recognize human action. For every HMM state they were using 30 dimensional feature vector and 16 Gaussian vector then recognition result fault rate was 13.1%. In this paper person and camera in stable state (not movable state) to avoid ego motion. With the help of stable state they got reliable motion gradient histogram [22].

Hong-Bo Zhangl, et al.(2010) proposed a method of shape based combined with motion feature for action identification. The motion data was used to locate the area of interesting region and for this region well-organized shape and motion depiction was found. To conclude SVM classifier was skilled for the event detection and result detection on KTH person action datasets as well as a variety of human and action demonstrate the accuracy enhanced than Jhuang and Dollars, and the execution time was better to Jhuang system. Their method gave 94.5% accuracy compare to Jhuang’s method. A regular identification method includes two parts: learning part and classification part. These two parts of their identification scheme were illustrated independently in figure 3. [23]. Shih-Yao Lin, et al.(2012) proposed a method of Action Trait Code for action recognition. Action trade code represented a set of velocity types resulting by the standard velocity of the all part of human. Their approached on Cornell Kinect Activity Database compared by a hierarchical utmost entropy Markov model (MEMM). The proposed method achieved higher recognition accuracy than other approaches. They projected an approach called Action Trait Code (Fig. 4) which uses the standard velocity of body parts to acquiesce a code recounting the actions. [24].

Lu Xia, et al.(2012) proposed a method of Histogram of three dimensional joints for view invariant person action detection. They took out the three dimensional skeletal join location from Kinect depth maps using Rabiner, et al. method [25]. On three dimensional action data set their method established significant view invariance. Their method received higher results on the demanding three dimensional action dataset. Their method provide mean accuracy was 90.92%, standard deviation was 1.74% and the best accuracy was 95.0% [26].

2.2. Soft Computing Based Approach

Soft computing major constituents are Fuzzy System, Probabilistic Reasoning, Evolutionary Computation, Machine Learning, Neural Network, etc and their fusion approaches but we widely focus on neural networks. Neural Network has special capability to obtain important data from convoluted or imprecise data and can be used to take out prototypes and recognize trends that are difficult to be observed by either person or other computer techniques. There are several advantage to use soft computing such as self organization, adaptive learning, real time operation, fault tolerance through unneeded information coding. We describe below many neural network for action recognition such as:

Chung-Hsin Kuo, et al.(2004) proposed a Artificial neural network for snooze motion detection. To identify the sleep motions, they proposed an image-based resolution. Record the video frames, the snooze of the patient by contactless and IR based dark vision camera. ANN structure, were used to instruct the structure by 60 training patterns. These experiment data was 10.79 picture pixcels when evaluated with
the genuine (by inquiry) data. Finally, the results of that effort were also re-defined to assemble the necessities of the clinical assessment and diagnosis [27].

Chia-Feng Juang(2005) selected a fuzzified Takagi Sugeno Kang (TSK)-type recurrent fuzzy network(FTRFN) for managing fuzzy temporal data. Initially there were no regulations in FTRFN. They were made parallel arrangement and parameter culture, where every free constraint in the prerequisite/ result of FTRFN are every tunable. It had been pertained on 1D and 2D fuzzy sequential sequence prophecy. The action identification rate was 92% in this approach [13].

Shuiwan Ji, et al.(2012) proposed a three-dimensional Convolutional Neural Network for Human Action identification. This network (CNNs) was a category of deep models that could perform straightly on the raw input data, thus mechanize the procedure of feature construction. This model extracted features from both temporal and spatial measures by performing three-dimensional convolutions, thereby recording the action information encoded in many nearby frames. They performed tests on the TRECVID 2008 data and the KTH dataset. Result attained in general accuracy of 90.2% as compared with the 91.7% attained by the HMAX model [17].

Haritha Srinivasan, et al.(2012) proposed a method of Artificial Neural Network for estimation of hand force from surface signals. The neural network was skilled with EMG information from wrist exion action as input and energy values from the haptic device as goal. They provided a less no. of errors comparison with other methods.

AO TANG, et al.(2013) proposed a model of Deep neural network for real time hand posture recognition. They projected an efficient algorithm to apply hand recognition and tracking. Secondly, they practical Deep Neural Networks (DNNs) to without human intervention learn features from hand position descriptions, that was insensible to movement, rotation and scaling. The dataset contained not only RGB- Depth- Skeleton data, but audio data as well. Experiments verified that the projected system plant fast and properly, and attained a detection accurateness as high as 98.12% [28].

Pichao Wang1, et al.(2014) proposed a method of Deep Convolutional Neural Networks via Depth Map Sequences for action recognition. Their algorithms could handle view variant cases. Secondly, in order to effectively extracted the body shape and motion information, they generated weighted depth motion maps (DMM) at several temporal scales. The proposed algorithms were evaluated on MSRAction3D, MSR Action3DExt, UT Kinect-Action and MSR Daily Activity 3D datasets respectively. They also combined the last three datasets into a larger one (called Combined Dataset) and tested the proposed method on it. The network recognized the hand posture better than the baseline method with no binarization layer.3-deep ConvNets were skilled on three expected planes of HDMM. Delayed fusion was achieved by merging the softmax class posteriors from the three nets. The overall framework was illustrated in Figure 5. Their algorithms could be separated into three modules: Rotation in 3D Pointclouds, Hierarchical DMM and Networks Training and Class Score Fusion [29].

Deenbandhu Singh, et al.(2014) proposed a method of person motion using star skeleton using HMM’s and neural network. The representation was able to find out the mistrustful and nonSuspicious activities. It also detected many humans in video and look after their actions. Motion objects were recognized and their restrictions extracted. In KTH Data Set gave 98% accuracy in test situation, in their mechanism they had given 95% accuracy [30].

Haiting Zhang, et al.(2014) proposed a method of Modified Deep Belief Network Model for person action recognition. Every action was made of a order
of person joint situations. Since the standard DBN could not contract with temporal data, the planned process developed the conditional Restricted Boltzmann Machine (CRBM) to manage the person joint sequence. To validate the efficiency of the future scheme, two skeletal illustration datasets were used for testing. They could show the recognition result using matlab in real time. The recognition accuracy of 1frame was 98.34% [31].

Experimental set up using a humanoid robot. They conducted two experiments with control gestures in an online fashion and for demonstration in a Human-Robot-Interaction(HRI) circumstances with the robot giving audio feedback for the user. The results showed that the mean F-Score for all the gestures was 96.85% [32].

Di wu, et al.(2014) proposed method of Deep dynamic neural network for action identification and recognition. which gave a result of 92.5% and limitation of this method was the heterogeneous inputs from skeletal joints and depth images require different feature learning methods [33].

Poonam Sonwalkar, et al.(2015) proposed method of hand gesticulate detection for real time. Gestures had long been considered as an interaction technique that could potentially deliver more natural. A fast gesticulation detection method was planned to be an border for the human-machine interaction (HMI) of systems. Result of Gaussian filter and the median filter and these noises filters both could shrink most noises and still could stay feature data for hand detection and gesticulation detection [34].

Pichao Wang1, et al.(2015) proposed a method of Deep Convolutional neural Network for action recognition using depth map sequences. Their algorithms could handle view variant cases. Second, in order to effectively extract the body shape and motion information, they generated weighted depth motion maps (DMM) at several temporal scales, referred to as Hierarchical Depth Motion Maps (HDMM). The proposed algorithms were evaluated on MSRAction3D, MSRAction3DExt, UTKinect-Action and MSRDailyActivity3D datasets respectively. The Combined Dataset was a dataset consisting of MSRAction3DExt, UTKinect-Action and MSRDailyActivity3D Datasets gives 90.2% accuracy [35].

3. IMPLEMENTATIONS TOOLS
Mostly researchers used MATLAB with image processing toolbox for human action images and videos recognition and few researchers used C++ also. [36], [37] and [38] used C++ on Windows XP where [36] and [37] they implemented (MFC) which stands for Microsoft Foundation Classes for building user connection and control.

4. PROFICIENCY
Most of the researchers attained zenith accuracy as per their task with respect to techniques and neural networks. All techniques and neural networks gave above 90% accuracy for recognition. Neural network gave 10% more accuracy compare to techniques.

5. RESEARCH GAPS
In this paper we have discussed several techniques and neural network for human action recognition. Now we discuss research gap based on these techniques and neural network, HMM’s, SVM, Histogram method, Action Trade code, 3D Histogram method, deep neural network, convolutional neural network, 3D convolutional neural network, deep belief neural network, deep dynamic neural network, artificial neural network etc. In earlier time mainly simple approaches were used for human action recognition and classification of dataset and then hybrid of these techniques were proposed for the same. There is ongoing research to enhance the accuracy of extracted feature and to reduce number of errors to recognize human action by eliminating unnecessary techniques so that classification accuracy can be improved.

Thus in future, the use of hybrid methods of these approaches with new optimization techniques like deep neural network, convolutional neural network, action trade technique, and histogram method can be experimented to improve the quality of human action.
recognition and error may be reduce to obtain better classification accuracy.

6. CONCLUSION

Table 1. Literature Review On Realtime Human Actions

<table>
<thead>
<tr>
<th>Year</th>
<th>Paper Name</th>
<th>Results</th>
<th>Limitations/Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>Recognizing human action in time-sequential images Using Hidden Markov Model.</td>
<td>A recognition rate of over 90% was achieved.</td>
<td>HMM’s have never been applied to motion recognition. Or Apply in 2d images Or Not provide a geometric representation of the human body.</td>
</tr>
<tr>
<td>2004</td>
<td>ANN based sleep motion recognition using night vision cameras.</td>
<td>The position center of root mean square error (RMSE) existed in these test figures was 10.79 image pixels when comparing to the authentic (by Investigation) data.</td>
<td>Since related paper did not apply the ANN model to study the case when the patient was in bed, quilt, bed quilt condition is observed through by discussing the pattern matching or vision investigation system can not receive the bed-cover with complex colors simulation.</td>
</tr>
<tr>
<td>2005</td>
<td>A recurrent fuzzy network for fuzzy temporal sequence processing and gesture recognition.</td>
<td>92%</td>
<td>FTRFN to manage real world temporal linguistic troubles or it had been apply on 1D and 2D fuzzy temporal sequence prophecy.</td>
</tr>
<tr>
<td>2006</td>
<td>Recognizing human actions on motion history images(MMHH)</td>
<td>The performance on motion history images(MMHH) is a little bit higher.</td>
<td>Thi implemented technique rely on no correct tracking, since enormously tracking algorithms may acquire an extra computational cost for method.</td>
</tr>
<tr>
<td>2007</td>
<td>Human body posture classification by a NFN and SVM.</td>
<td>Average identification rate was 97.8.</td>
<td>Removes darkness more completely and</td>
</tr>
<tr>
<td>2008</td>
<td>Real-time human action recognition on an embedded, video processing architecture.</td>
<td></td>
<td>They gave 80% overall detection rate using definite training data.</td>
</tr>
<tr>
<td>2008</td>
<td>Detection of Activities for daily life eating and drinking.</td>
<td>Average recognition of g eating and drinking was 100% in that experiment.</td>
<td>This paper only focus on recognition of drinking and eating activities.</td>
</tr>
<tr>
<td>2009</td>
<td>HMM-based human motion recognition with motion feature.</td>
<td>13.1% when using 16 Gaussian for Each HMM state and 30 dimensional feature vectors.</td>
<td>The focus on special regions could not further permit to detect the movement of more than one human.</td>
</tr>
<tr>
<td>2010</td>
<td>Real-time human action recognition based on shape combined with motion feature.</td>
<td>Comparing jhuang’s method the mean gavod r od e 94.5% accuracy.</td>
<td>Proposed approach depends on the frame classifier, so they only will think the continuous dissimilar actions detection in the similar video.</td>
</tr>
<tr>
<td>2012</td>
<td>Convolutional neural network for human action recognition using Action Trait Code.</td>
<td>Hig recognition 97.7/9 tha other of 7.2% accuracy</td>
<td>Such method does not consider the motion data encoded in several contiguous frames.</td>
</tr>
<tr>
<td>2012</td>
<td>Human action recognition using Histogram 3d joint.</td>
<td>By testing, the generally Mean accuracy was 90.92%, the standard deviation is 1.74% and the best accurateness was 95.0%.</td>
<td>Their algorithm utilizes depth information only.</td>
</tr>
<tr>
<td>2013</td>
<td>A real time hand posture recognition system using Deep Neural Network.</td>
<td>As high as 98.12%.</td>
<td>If the color of back- Ground is very close to that of hand skin, those methods are...</td>
</tr>
</tbody>
</table>
### References


AUTHOR’S BIBLIOGRAPHY

Dr. Ritu Tiwari received PhD degree from NIT Raipur. She is currently associate professor for ABV-IIITM Gwalior MP, India. She developed keenly interest in Robotics, Artificial Intelligence, Soft Computing and Applications. She has worked on occasion with publications like in 372 citations, 111 publications along with 5 books, 11 chapters, 69 articles and 28 conference papers aside. She has contributed in the field of Multi-robot navigation and Artificial bee colony algorithm.

Sapna Varshney is currently re-search scholar at ABV-IIITM Gwalior MP, India. She has determined interest in Soft Computing and Neural Networks. She has worked as a lecturer of computer science in Manglayatan University in 2012 to 2013. She fully wants to contribute in the field of real time human action recognition related algorithms.