Walking, Lifting, Standing Activity Recognition using Probabilistic Networks

Amol S Patwardhan[1]

[1]Senior Researcher, VIT, University of Mumbai, 400037, India

---***---

Abstract - *This paper examines use of dynamic probabilistic networks (DPN) for human action recognition. The actions of lifting objects and walking in the room, sitting in the room and neutral standing pose were used for testing the classification. The research used the dynamic interrelation between various different regions of interest (ROI) on the human body (face, body, arms, legs) and the time series based events related to these ROIs. These dynamic links are then used to recognize the human behavioral aspects in the scene. First a model is developed to identify the human activities in an indoor scene and this model is dependent on the key features and interlinks between the various dynamic events using DPNs. The sub ROI are classified with DPN to associate the combined interlink with a specific human activity. The recognition accuracy performance between indoor (controlled lighting conditions) is compared with the outdoor lighting conditions. The accuracy in outdoor scenes was lower than the controlled environment.*

Key Words: Negative emotion detection, closed spaces, confined spaces, low lit lighting, sensor, multiple input fusion.

1.INTRODUCTION

Probabilistic networks are very useful to establish the causal relationship between two inter-related events. Such networks are suitable for linking dependencies between variables in the problem domain which could vary from speech recognition, vision based systems, medicine or automotive industries. The dynamic probabilistic networks need a large amount of training data to be effective. Co-incidentally, the video image frames contain an abundance of dynamic motion based images and movement of features which can be exploited for training the network. Several studies [1], [2], [3], [4] to [10] in the last decade have concentration on using the temporal aspects of features from video and image sequences for face recognition, gesture recognition and human activity. As a result, this study examines the use of DPN to detect specific actions inside a controlled environment such as walking, sitting, lifting objects and also tested the performance of the system under outdoor lighting to measure the generalizability of the recognition system. Moreover, the focus is shifting from static image analysis to research on temporal changes in series of images for detecting human activities. The

performance of recognition systems under low lighting or outdoor conditions and in case of noisy background and chaotic and uncontrolled conditions is still poor. Human activity recognition is also closely related to more complex tasks such as emotion recognition. Several studies [11], [12], [13], [14] to [28] have examined the multimodal emotion recognition techniques and the software implementation strategies. Studies [29], [30], [31], [32] to [44] have used joint probability distributions, modelling of activities using manifold and learning their transition probabilities. Hidden Markov model (HMM) based recognition has also gained interest. These methods are practically infeasible and difficult to implement because of the computationally intensive nature of model training process and dependency on the availability of large amount of training data. Additionally, there is a higher degree of co-relation between the features in the data and the network making the accuracy confined and over-fitted to the intra-corpus feature properties. Another strategy currently being explored is the exemplar based approach such as view specific manifolds and spatio-temporal embedded classification using statistical learning and Bayesian inference techniques. But these methods have the same interdependency between the statistical model and the exemplar classes.

2. METHOD

The video of 6 participants walking in a room and lifting objects was recorded. The resolution of the video was set to very low to minimize the image frame size. Each image frame was set to a size of 640 x 480 pixel window. The video of participants resulted in 24 action scenes. Each video consisted of 1500 to 2000 frames and 5-6 minute videos. First each action sequence was manually identified and annotated to establish the ground truth. For the purpose of training the DBN, the data was split into 10 small sets containing a sequence of lifting an object and keeping it down randomly chosen from the master list. Feature extraction was performed on each sequence and the feature vector was used to train the DBN. For the feature vector a bounding box was used with the corner co-ordinates of the box, the average intensity value of the change in pixel intensities for pixels within the box was calculated and the filling ratio and first order moments around the centroid

were used. These features formed a feature vector with 8 components. The feature vectors were set as the observational input for the DBN training. The priors and transition matrices defaults were chosen with random automated event detection.

The various components of the human body were detected using cascade transform to identify the region of interest (ROI). Once the bounding box was detected, the features were extracted using projection methods followed by principle component analysis (PCA). After the dimensionality reduction the bounding box was used to obtain features for the face, body, hands and legs (4 components). Thus the feature vector consisted of 8 x 4 = 32 features representing an activity ranging from walking, sitting, lifting objects and standing in a neutral position. According to the model for Bayesian inference, the detection of activity from an image sequence is done by the maximum a posteriori (MAP) decision rule of probability distribution. This probability is dependent on the prior probability of each class and the likelihood of image sequences given each class. Additionally, the value is normalized overall the observations such that sum of probabilities over all classes is 1. To reduce the computational processing of the model training, every third image sequence frame was considered from a frame rate of 24 per second as a result only 8 frames were considered. The prediction of human activity recognition was implemented using the majority vote based classification scheme. The prediction of each class for the 4 components was also calculated separately and the classifiers for each component (face, hand, body, leg) were used for voting for a class. In case of a tie, the probability of hand was used because of the more dynamic nature of the feature movement while performing an activity and the lack of course facial features from a long distance between the camera and the subject. The performance of the classifier for the various human activity classes is discussed in the results section.

A classifier was trained for each input channel (face, head, hand, body, leg). Additionally, the classifier was trained for each viewing angle. This resulted in a total of 5x4 = 20 classifiers. The result from each classifier was chained to form an ensemble of classifiers and the decision level fusion was implemented to predict the final emotional state. The various viewing angles from the 4 sensors compensated for occlusion. The various input channels provided a comprehensive emotion profile resulting in better emotion accuracy. The participants were asked to wear orange, white and grow beards to mimic realistic conditions in a prison confinement. Some participants were shown reference videos

to assist in enacting the emotion. The participants were also asked to enact emotions while sitting in one spot as well as moving around in the room to depict feeling of frustration and restlessness.

3. RESULT

The recognition accuracy was better in case of non-noisy video sequences. For the noisy sequences the recognition accuracy dropped from 78% to 74% for picking up action and dropped from 79% to 71% for keeping object down on the floor.
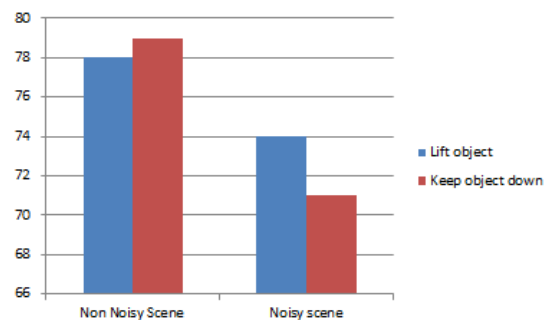
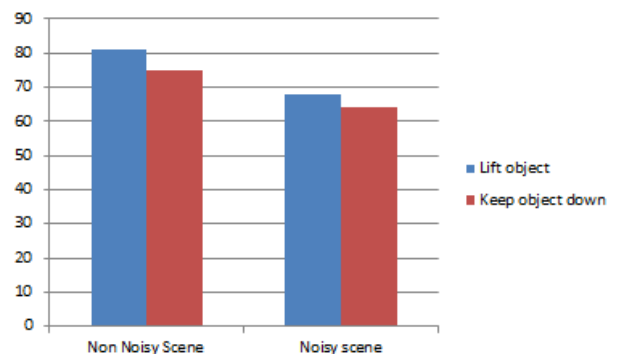

Chart -1: DBN recognition accuracy

Chart -2: HMM recognition accuracy

A comparison with HMM based human activity recognition indicated that HMM was recognition rate was better for lifting action but dropped by a higher degree in case of noisy scene as compared to DBN technique.

The results from the test video sequence for the dynamic probabilistic network are shown below and it was obtained by the MAP decision rule maximization. The plot of posterior probability Fig. 3 of each class given the sequence of test images showed that the DPN was capable of accurate recognition of various activities such as walking, sitting, lifting objects compared under controlled conditions. A comparison of recognition rates for the various activities

under indoor and outdoor conditions using majority voting scheme is also shown in Fig. 4. In each instance the recognition rates of indoor controlled conditions were better.
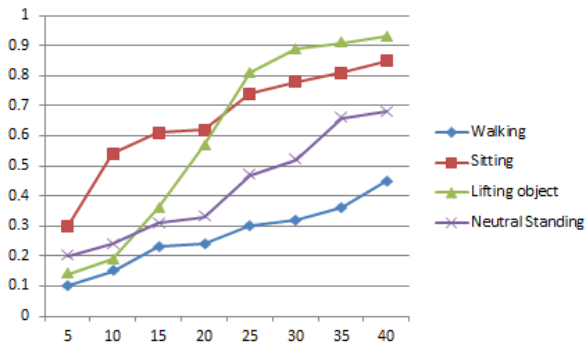

Chart -3: Comparison of posterior probabilities for various

The probability of lifting objects steadily improved as more data was available from subsequent image sequences. The remaining actions probabilities steadily improved before flattening out indicating that the subsequent video frames could not provide discriminating information about the activity.
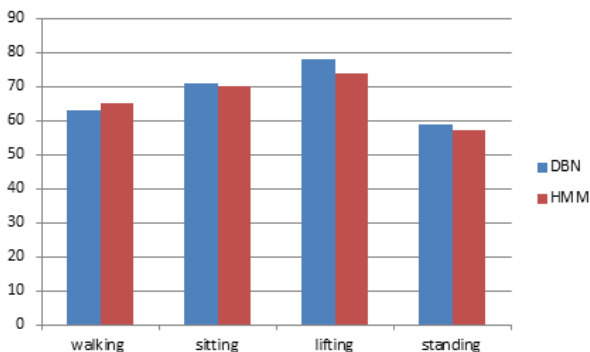

Chart -4: Comparison of accuracies using DBN and HMM.

The DBN performed better for the lifting object activity under controlled lighting, standing, and sitting. In case of the activities representing walking, HMM had a better accuracy over the proposed method.

4. CONCLUSIONS

The DBN based human activity recognition was more robust compared to HMM. The proposed method for using region of interest and then combining the features from face, hand, body and legs to form a feature vector proved useful for recognition under controlled conditions and reducing the computational complexity. But the same strategy was not useful in outdoor scenes. This was evident from the fact that

the recognition rate dropped by a smaller value as compared to that of HMM in case of noisy scene. Additionally, the DBN based technique was more suitable for busy video scenes with several temporal changes and activity as compared to only individual actors in the recorded scene and the corresponding HMM based recognition.

REFERENCES

[1] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Drunken Abnormal Human Gait Detection using Sensors, Computer Science and Emerging Research Journal, vol 1, 2013.

[2] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Fear Detection with Background Subtraction from RGB-D data, Computer Science and Emerging Research Journal, vol 1, 2013.

[3] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Code Definition Analysis for Call Graph Generation, Computer Science and Emerging Research Journal, vol 1, 2013.

[4] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Multi-View Point Drowsiness and Fatigue Detection, Computer Science and Emerging Research Journal, vol 2, 2014.

[5] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Group Emotion Detection using Edge Detecttion Mesh Analysis, Computer Science and Emerging Research Journal, vol 2, 2014.

[6] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Polarity Analysis of Restaurant Review Comment Board, Computer Science and Emerging Research Journal, vol 2, 2014.

[7] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Sentiment Analysis in Code Review Comments, Computer Science and Emerging Research Journal, vol 3, 2015.

[8] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Temporal Analysis of News Feed Using Phrase Position, Computer Science and Emerging Research Journal, vol 3, 2015.

[9] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Decision Rule Driven Human Activity Recognition, Computer Science and Emerging Research Journal, vol 3, 2015.

[10] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Depression and Sadness Recognition in Closed Spaces, Computer Science and Emerging Research Journal, vol 4, 2016.

[11] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Dynamic Probabilistic Network Based Human Action Recognition, Computer Science and Emerging Research Journal, vol 4, 2016.

[12] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Fight and Aggression Recognition using Depth and Motion Data, Computer Science and Emerging Research Journal, vol 4, 2016.

[13] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Sensor Tracked Points and HMM Based Classifier for Human Action

Recognition, Computer Science and Emerging Research Journal, vol 5, 2016.

[14] A. S. Patwardhan, 2016. "Structured Unit Testable Templated Code for Efficient Code Review Process", PeerJ Computer Science (in review), 2016.

[15] A. S. Patwardhan, and R. S. Patwardhan, "XML Entity Architecture for Efficient Software Integration", International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 4, no. 6, June 2016.

[16] A. S. Patwardhan and G. M. Knapp, "Affect Intensity Estimation Using Multiple Modalities," Florida Artificial Intelligence Research Society Conference, May. 2014.

[17] A. S. Patwardhan, R. S. Patwardhan, and S. S. Vartak, "Self-Contained Cross-Cutting Pipeline Software Architecture," International Research Journal of Engineering and Technology (IRJET), vol. 3, no. 5, May. 2016.

[18] A. S. Patwardhan, "An Architecture for Adaptive Real Time Communication with Embedded Devices," LSU, 2006.

[19] A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher, 2013.

[20] A. S. Patwardhan and G. M. Knapp, "Aggressive Action and Anger Detection from Multiple Modalities using Kinect", submitted to ACM Transactions on Intelligent Systems and Technology (ACM TIST) (in review).

[21] A. S. Patwardhan and G. M. Knapp, "EmoFit: Affect Monitoring System for Sedentary Jobs," preprint, arXiv.org, 2016.

[22] A. S. Patwardhan, "Embracing Agile methodology during DevOps Developer Internship Program", IEEE Software (in review), 2016.

[23] A. S. Patwardhan, "Edge Based Grid Super-Imposition for Crowd Emotion Recognition", International Research Journal of Engineering and Technology (IRJET), May. 2010.

[24] A. S. Patwardhan, "Human Activity Recognition Using Temporal Frame Decision Rule Extraction", International Research Journal of Engineering and Technology (IRJET), May. 2010.

[25] A. S. Patwardhan, "Low Morale, Depressed and Sad State Recognition in Confined Spaces", International Research Journal of Engineering and Technology (IRJET), May. 2011.

[26] A. S. Patwardhan, "View Independent Drowsy Behavior and Tiredness Detection", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2011.

[27] A. S. Patwardhan, "Sensor Based Human Gait Recognition for Drunk State", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2012.

[28] A. S. Patwardhan, "Background Removal Using RGB-D data for Fright Recognition", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2012.

[29] A. S. Patwardhan, "Depth and Movement Data Analysis for Fight Detection", International Research Journal of Engineering and Technology (IRJET), May. 2013.

[30] A. S. Patwardhan, "Human Action Recognition Classification using HMM and Movement Tracking", International Research Journal of Engineering and Technology (IRJET), May. 2013.

[31] A. S. Patwardhan, "Feedback and Emotion Polarity Extraction from Online Reviewer sites", International Research Journal of Engineering and Technology (IRJET), May. 2014.

[32] A. S. Patwardhan, "Call Tree Detection Using Source Code Syntax Analysis", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2014.

[33] A. S. Patwardhan, "Walking, Lifting, Standing Activity Recognition using Probabilistic Networks", International Research Journal of Engineering and Technology (IRJET), May. 2015.

[34] A. S. Patwardhan, "Online News Article Temporal Phrase Extraction for Causal Linking", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2015.

[35] A. S. Patwardhan, "Online Comment Processing for Sentiment Extraction", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2016.

[36] A. S. Patwardhan, "Analysis of Software Delivery Process Shortcomings and Architectural Pitfalls", PeerJ Computer Science (in review), 2016.

[37] A. S. Patwardhan, "Multimodal Affect Recognition using Kinect", ACM TIST (in review), 2016.

[38] A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", IEEE TAC (in review), 2016.

[39] A. S. Patwardhan, Jacob Badeaux, Siavash, G. M. Knapp, "Automated Prediction of Temporal Relations", Technical Report. 2014.