Human Body Gesture Recognition System using concepts of Neural Network

[1]Brajesh Kumar, [2]Santosh Kumar
[12]PhD Scholar
[12]B.RA.Bihar University, Muzaffarpur

Abstract: --Human Body Gesture Recognition System using concepts of Neural Network. This is for developing and implemented an experimental setup consisting of a humanoid robot/android able to recognize and execute in real time all the arm gestures of the Dynamic Gesture Language (DGL) in similar way as humans do. Our DGLR system comprises two main subsystems: an image processing (IP) module and a linguistic recognition system (LRS) module. The IP module enables recognizing individual DGL gestures. In this module, we use the bag-of-features (BOFs) and a local part model approach for dynamic gesture recognition from images. Dynamic gesture classification is conducted using the BOFs and nonlinear support-vector-machine (SVM) methods. The multi scale local part model preserves the temporal context. The IP module was tested using two databases, one consisting of images of a human performing a series of dynamic arm gestures under different environmental conditions and a second database consisting of images of an android performing the same series of arm gestures. The linguistic recognition system (LRS) module uses a novel formal grammar approach to accept DGL-wise valid sequences of dynamic gestures and reject invalid ones. LRS consists of two subsystems: one using a Linear Formal Grammar (LFG) to derive the valid sequence of dynamic gestures and another using a Stochastic Linear Formal Grammar (SLFG) to occasionally recover gestures that were unrecognized by the IP module. Experimental results have shown that the DGLR system had a slightly better overall performance when recognizing gestures made by a human subject (98.92% recognition rate) than those made by the android (97.42% recognition rate).

***Keywords:* --Arm Gesture, Neural Network, Pattern Recognition, Ten-fold Cross Validation etc.**

Introduction

Arm gestures represent a powerful natural communication modality between humans, providing a major information transfer channel in our everyday life, transcending language barriers. Hand signs and dynamic arm gestures are an easy to use non-verbal communication modality of humans with other humans and even with some animals. For example, sign languages have already been used extensively among speech or hear-disabled people. Even people who can speak and hear also use many kinds of arm gestures and hand signs to help their communication in audio noisy environments or too far away for hearing applications (e.g. on board of ships or aircraft carriers). More recently, arm gesturing also became a major communication modality for human-computer interaction (HCI). Dynamic gesture communication is a very natural and human-like mode of communication with computers, which complements the static hand signs. As a result of arm being able to move in any direction and to bend to almost any angle in all coordinates, dynamic arm gesturing adds a new dynamic dimension complementing the static hand gestures are limited to significantly less meaning set of hand postures.Visual arm movement recognition is done in both spatial and temporal domains. It requires high accuracy in terms of recognition and time, as well as level of perfection against a cluttered background, variable light condition and variable distance. Visual recognition of dynamic arm gestures has recently been adopted by a number of applications, like smart homes, video surveillance, human-robot communication in smart homes, healthcare and eldercare applications.

A new generation of humanoid robots is currently under

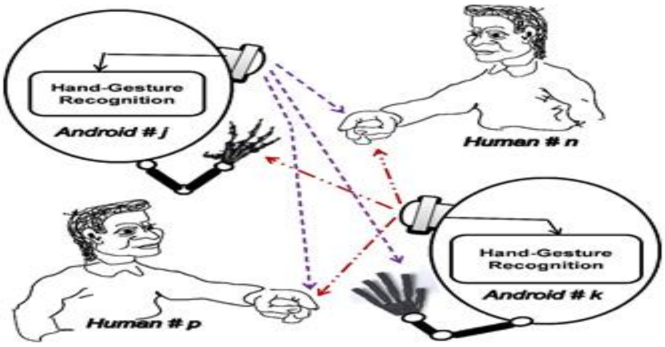


Figure1: Visual recognition of gestures made by androids communicating with humans and other androids.

Ten-Fold Cross Validation

In the ten-fold CV method the data which consists of 100 sequences of dynamic gestures was divided into ten equal and non-overlapping folds. One of the folds is held out each time and the other nine folds are used to train the SLFG. Then the unseen held-out fold is used to test the system. This process is repeated for every fold (i.e. ten times). The fold boundary parameters were set by the number of sequences. That is, for each training and testing pair, the data set was split into 90 sequences of dynamic gestures for training and 10 remaining separate sequences for testing. presents the results of the tenfold CV with 100 sentences (i.e. sequences of dynamic gestures) dataset.

Table 1 below shows the results for each iteration of the 10-fold cross validation, as well as the aggregate accuracy of the 10-fold cross validation iterations with 100 sentences of dataset.

Table 1: Results of the Tenfold CV with 100 sentences dataset

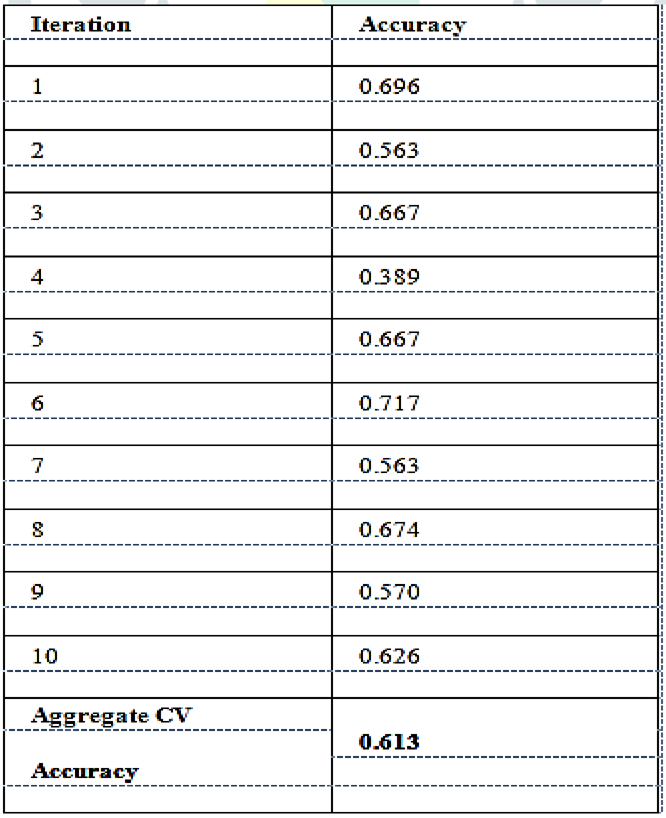
Iteration	Accuracy
1	0.710
2	0.722
3	0.391
4	0.542
5	0.590
6	0.722
7	0.253

8	0.588
9	0.718
10	0.786
Aggregate CV **Accuracy**	**0.602**

We repeated above mentioned tenfold cross validation test with another dataset. In this setup, dataset consists of 200 sentences (i.e. sequences of dynamic gestures) was divided into ten equal and non-overlapping folds. One of the folds is held out each time and the other nine folds are used to train the SLFG. Then the unseen held-out fold is used to test the system. This process is repeated for every fold (i.e. ten times). The fold boundary parameters were set by the number of sentences. That is, for each training and testing pair, the data set was split into 180 sentences (i.e. sequences of gestures) for training and 20 remaining separate sequences for testing presents the results of the tenfold CV with 200 sentences dataset.

Table 2below shows the results for each iteration of the 10-fold cross validation, as well as the aggregate accuracy of the 10-fold cross validation iterations with 200 sentences (i.e. sequences of dynamic gestures) dataset

Table 2:Results of the Tenfold CV with 200 sentences dataset



Iteration	Accuracy
1	0.696
2	0.563
3	0.667
4	0.389
5	0.667
6	0.717
7	0.563
8	0.674
9	0.570
10	0.626
Aggregate CV Accuracy	0.613

The tables above show our stochastic linear formal grammar is a quick learner. With only 100 command sequences it reaches a stable high performance, such that even when the dataset is doubled it shows minimal improvement. It is important to note that without a stochastic linear formal grammar to capture the syntactic patterns of the data, the performance in the table above would drop to 0.08, which is equal to pure chance; since there are 12 dynamic gestures and the chance of predicting the correct one with no syntactic information is 1/12. Therefore the SLFG is performing 0.52 above the baseline, in other words 7.5 times higher in performance than the baseline.

Comparison with Other Approaches

We compare the results of our experiments with comparable previous works. Shiravandi et al. in, used a dynamic Bayesian network method for dynamic hand gesture recognition. They considered 12 gestures for recognition. They achieved an average accuracy of 90%.

Wenjun et al. in proposed an approach based on motion trajectories of hands and hand shapes of the key frames. The hand gesture of the key frame is considered as a static hand gesture. The feature of hand shape is represented with Fourier descriptor and is recognized by the neural network. The combined method of the motion trajectories and key frame is presented to recognize the dynamic hand gesture from unaided video sequences. They consider four dynamic hand gestures for experiment. Their average recognition accuracy is 96%.

Bao et al. in did dynamic hand gesture recognition based on Speeded Up Robust Features (SURF) tracking. The main characteristic is that the dominant movement direction of matched SURF points in adjacent frames is used to help describing a hand trajectory without detecting and segmenting the hand region. They consider 26 alphabetical hand gestures and their average recognition accuracy rate achieved was 84.6%.

Yang et al. in proposed the hidden markov model (HMM) for hand gesture recognition. They consider 18 gestures for recognition. There recognition rate is 96.67%.

In Pisharady et al. used dynamic time wrapping (DTW) and multi-class probability estimates to detect and recognize hand gestures. They used Kinect to get skeletal data. They claimed 96.85% recognition accuracy with 12 gestures. The above mentioned results and our result comparison are given in Table 8. As shown in the table, our experiment achieved the highest accuracy although we had the highest number of aggregate cases, parameters, and scenarios.

Compared to the previous works our work tackles the most complex task as shown by Table 8. As can be observed our task conditions vary among 24 different scenarios whereas the closest previous work tackles only 2.

Future Work

The research will be expanded by a team of student who are developing a new experimental set up using a second-generation life-size android. Further research will also be needed in order to integrate the speech recognition and the dynamic gesture recognition so they complement each other in a single complex recognition task. For instance if some command is unrecognizable to the android in one modality it can be disambiguated using the other modality. It will also provide for a more complex human like way of communication between robots and humans.

Conclusion

The idea of the project got started from a McConnel's idea of orientation histograms. Many researchers found the idea interesting and tried to use it in various applications. From hand recognition to cat recognition and geographical statistics, my supervisor and I had the idea of trying to use this technique in conjunction with Neural Networks. In other approaches of pattern recognition that orientation histograms have been used different ways of comparing and classifying were employed. Euclidean distance is a straight forward approach to it. It is efficient as long as the data sets are small and not further improvement is expected. Another advantage of using neural networks is that you can draw conclusions from the network output. If a vector is not classified correct we can check its output and work out a solution. As far as the orientation algorithm is concerned it can be further improved. The main problem is how good differentiation one can achieve. This of course is dependent upon the images but it comes down to the algorithm as well. Edge detection techniques are keep changing while line detection can solve some problems. One of the ideas that I had lately is the one of tangents but I don't know if it is feasible and there is not time of developing it.To say that I have come to robust conclusions at the end of the project is not safe. This is possible only for the first part of the project. Regardless of how many times you run the program the output vector will always be the same. This is not the case with the perceptron. Apart from not being 100% stable there are so many parameters (e.g. number of layers, number of nodes) that one can play with that finding the optimal settings is not that straight forward. As mentioned earlier it all comes down to the application. If there is a specific noise target for example you can work to fit these specifications.

References

[1] Christopher M. Bishop, “Neural networks for Pattern Recognition” Oxford, 2015.

[2] William T. Freeman, Michael Roth, “Orientation Histograms for Hand Gesture Recognition” IEEE Intl. Workshp. On Automatic Face and Gesture Recognition,

Zurich, June, 2015.

[3] Maria Petrou, PanagiotaBosdogianni, “Image Processing, The Fundamentals”, Wiley

[4] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S Huang, “Visual Interpretation of Hand Gestures for Human-Computer Interaction: A review” IEEE Transactions of pattern analysis and machine intelligence, Vol 19, NO 7, July 2017

[5] Srinivas Gutta, Ibraham F. Imam, Harry Wechsler, “ Hand gesture Recognition Using Ensembles of Radial Basis Functions (RBF) Networks and Decision Trees” International Journal of Pattern Recognition and Artificial Intelligence, Vol 11 No.6 2016.

[6] Simon Haykin, “Neural Networks, A comprehensive Foundation”, Prentice Hall Duane Hanselman, Bruce Littlefield, “Mastering MATLAB, A comprehensive

tutorial and reference”, Prentice Hall

[7] http://www.cs.rug.nl/~peterkr/FACE/face.html

[8] http://www.hav.com/

[9]http://www.dacs.dtic.mil/techs/neural/neural_ToC.html

[10] http://www.tk.uni-linz.ac.at/~schaber/ogr.html

[11]http://vismod.www.media.mit.edu/vismod/classes/mas622/projects/hands/

[12] http://aiintelligence.com/aii-info/techs/nn.html Q. Chen, F. Malric, Y. Zhang, M. Abid, A. Cordeiro, E.M. Petriu, N.D. Georganas, "Interacting with Digital Signage Using Hand Gestures", Proc. ICIAR 2009, Int. Conf. Image Analysis and Recognition, (M. Kamel and A. Campilho - Eds), Lecture

[14] R. H. Liang and M.Ouhyoung, "A Real Time Continuous Gesture Recognition System for Sign Language," in Proc . 3rd International Conference on Automatic Face and Gesture Recognition, 1998, pp. 558-565.

[15] V. I. Pavlovic, R. Sharma, T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, july 1997 pp. 677-695.

[16] M. Alsheakhali, A. Skaik, M. Aldahdouh, M. Alhelou, "Hand Gesture Recognition System," Computer Engineering Department, The Islamic University of Gaza Strip, Palestine, 2011.

[17] S.SidneyFels and G. E. Hinton, "Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer," IEEE Transactions on Neural Networks, Vol. 3, No. 6, November 1992 pp.1-7.

[18] M. Ali Qureshi, A. Aziz, M. AmmarSaeed, M. Hayat, "Implementation of an Efficient Algorithm for Human Hand Gesture Identification," Dept. Electronic Engineering, University College of Engineering & Technology, The Islamia University of Bahawalpur, Bahawalpur Pakistan.

[19] D.J. Sturman and D. Zeltzer, "A Survey of Glove-Based Input, "IEEE Computer Graphics and Applications, vol. 14, pp. 30-39, Jan. 1994.

[20] Kumo, Yoshinori, T. Ishiyama, KH. Jo, N. Shimada and Y. Shirai, "Vision-Based Human Interface System: Selectively Recognizing Intentional Hand Gestures," In Proceedings of the IASTED International Conference on Computer Graphics and Imaging, 219-223, 1998.

[21] Utsumi, Akira, J. Kurumisawa, T. Otsuka, and J. Ohya, "Direct Manipulation Scene Creation in 3D," SIGGRAPH’97 Electronic Garden, 1997.