Fight and Aggression Recognition using Depth and Motion Data

Anne Veenendaal, Eddie Jones, Zhao Gang, Elliot Daly, Sumalini Vartak, Rahul Patwardhan

Abstract:

This paper examines human action recognition using color and depth data (RGB-D) and motion analysis of image frames. Specifically, it explores recognition of aggressive actions such as throwing, kicking, punching, threatening by using 3D depth sensors. The RGB-D data was obtained from video recordings of actions and infrared sensing device. 23 participants were asked to enact aggressive actions. The SVM classifiers were trained using the features extracted from the series of frames and tested against violent actions simulated on the basis of real life fight scenes. The results indicated that the performance for actions containing single individuals was better but the system performed poorly for group activity detection in terms of accuracy.

Introduction:

Many studies have focused on human activity [1], [2], [3], [4] through [11] recognition in the last decade. The studies have used techniques such as motion analysis [1], temporal analysis [2], graphical models [3], shape analysis [4] and probabilistic classification [5]. There have been studies that have examined real time surveillance [6], canonical space [7], time sequence analysis [8], view invariance [9], rule based techniques [10] for human activity detection. In the last decade emotion recognition using multiple modalities [11] to [31] and various software based implementation of automated systems have been done. Most of these studies have either used single person for enacting the activity or emotion of have considered only full frontal postures. Our research focusses on recognition of anger, aggression and fights from a side-view. It also uses non marker based recognition and extracts features using human blob extractions from color and depth data obtained from the infrared sensors.

Method:

A total of 23 participants enacted aggressive actions. 20 were male and 3 were female. The age group was between 25 to 32. The participants posed at a distance of 1.5 to 4 meters from the sensor. The data was captured from side view. Edge detection and shape analysis was performed to extract the human shape and the Depth data was used for blob detection and extraction of human shapes from the blob. The reference skeleton was used to extract the shape from the color channel. The features used were the movement of 4 points on the face, 4 points on each hand and 3 points on body, 2 points on hip and 3 points on each leg. The distance between each point and x,y,z co-ordinates were used as features.

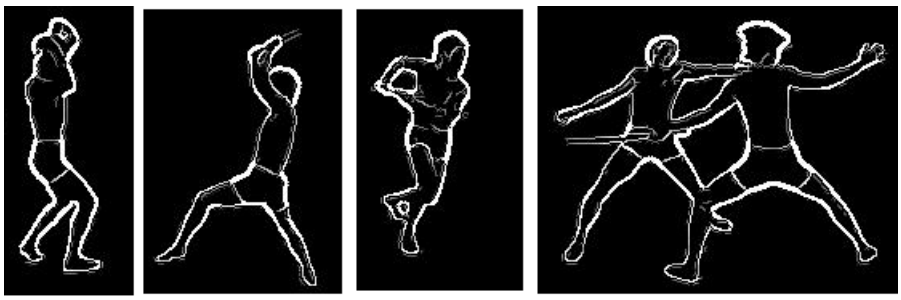
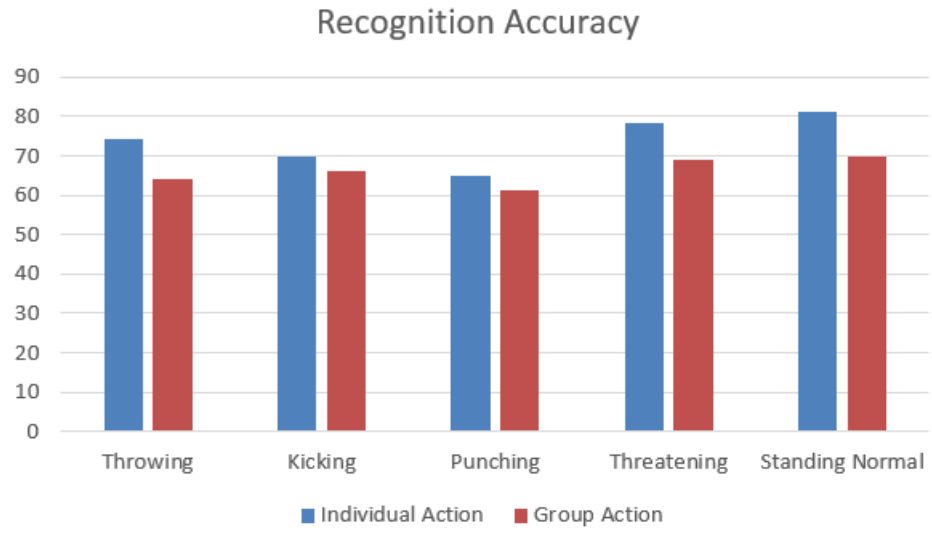


Fig. 1. Actions used by participants for enacting aggression.

The feature vector was then used to train the SVM classifier with radial basis as the kernel function with a slack value of 0.1. The same data collection was done with groups of two for a total of 5 different actions. The classifier for individual person and groups was trained separately using 10-cross validation. The data was split into 60% for training and 40% for test. The classifier parameters were optimized using grid search.

Results:



The recognition rate for neutral position was better than aggressive actions. The threatening pose showed better recognition accuracy (78.3%) over other actions. The accuracy decreased when the recognition was tested against aggressive actions in a group setting.

Conclusion:

This study provided bench mark results for further exploration of side view point analysis of fight and aggressive action. It also provided insight into how individual vs group actions compare in terms of recognition accuracy. As a future scope the recognition method needs to be compares with other state of the art techniques and actions under different lighting conditions and more view-points.

REFERENCES

[1] R. Poppe, “Vision-based human motion analysis: an overview,” Comput. Vision and Image Understan., vol. 108, pp. 4–18, 2007.

[2] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,” IEEE Trans. Pattern Anal. Machine Intell., vol. 23, no. 3, 2001.

[3] R. Hamid, Y. Huang, and I. Essa, “ARGMode–Activity recognition using graphical models”, in Proc. Conf. Comput. Vision Pattern Recog., vol. 4, pp. 38–45, Madison, Wisconsin, 2003.

[4] S. Carlsson and J. Sullivan, “Action recognition by shape matching to key frames,”in Proc. IEEE Comput. Soc. Workshop Models versus Exemplars in Comput. Vision, pp. 263–270, Miami, Florida, 2002.

[5] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, “Probabilistic posture classification for human-behavior analysis,” IEEE Trans. Syst. Man, and Cybern. A, vol. 35, no. 1, pp. 42–54, 2005.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: Real-time surveillance of people and their activities,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, pp. 809–830, August 2000.

[7] P. S. Huang, C. J. Harris, and M. S. Nixon, “Canonical space representation for recognizing humans by gait or face,” in Proc. IEEE Southwest Symp. Image Anal. Interpretation, pp. 180–185, 1998.

[8] J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov model,” in Proc. IEEE CVPR, pp. 379–385, 1992.

[9] F. Niu and M. Abdel-Mottaleb, “View-invariant human activity recognition based on shape and motion features,” in Proc. IEEE Sixth Int. Symposium Multimedia Softw. Eng., pp. 546–556, 2004.

[10] L. X. Wang and J. M. Mendel, “Generating fuzzy rules by learning from examples,” IEEE Trans. Syst., Man Cybern., vol. 22, no. 6, pp. 1414–1427, 1992.

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

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

[13] 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.

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

[15] 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.

[16] A. S. Patwardhan, “An Architecture for Adaptive Real Time Communication with Embedded Devices,” LSU, 2006.

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

[18] 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).

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

[20] A. S. Patwardhan, J. Kidd, T. Urena and A. Rajagopalan, “Embracing Agile methodology during DevOps Developer Internship Program”, IEEE Software (in review), 2016.

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

[22] A. S. Patwardhan, “Multimodal Affect Recognition using Kinect”, ACM TIST (in review), 2016.

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

[24] S. D’Mello and A. Graesser, “Multimodal Semi-Automated Affect Detection from Conversational Cues, Gross Body Language, and Facial Features,” User Modeling and User-Adapted Interaction, vol. 10, pp. 147-187, 2010.

[25] T. Baenziger, D. Grandjean, and K.R. Scherer, “Emotion Recognition from Expressions in Face, Voice, and Body. The Multimodal Emotion Recognition Test (MERT),” Emotion, vol. 9, pp. 691-704, 2009.

[26] C. Busso et al., “Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information,” Proc. Int’l Conf. Multimodal Interfaces, T.D.R. Sharma, M.P. Harper, G. Lazzari, and M. Turk, eds., pp. 205-211, 2004.

[27] N. Sebe, I. Cohen, and T.S. Huang, “Multimodal Emotion Recognition,” Handbook of Pattern Recognition and Computer Vision, World Scientific, 2005.

[28] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, “Emotion Recognition in Human-Computer Interaction,” IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, 2001.

[29] Y. Bengio, P. Lamblin D. Popovici, and H. Larochelle. Greedy LayerWise Training of Deep Networks. Advances in Neural Information Processing Systems 19 (NIPS‘06), pp 153-160, MIT Press 2007.

[30] A. Kapoor and R.W. Picard, “Multimodal Affect Recognition in Learning Environments,” Proc. 13th Ann. ACM Int’l Conf. Multimedia, pp. 677-682, 2005.

[31] O. Oreifej and Z. Liu. HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences. In IEEE Conference on Computer Vision and Pattern Recognition, pp 716–723, 2013

[32] L. Xia, C.-C. Chen, and J. K. Aggarwal. View invariant human action recognition using histograms of 3D joints. 2012 IEEE Computer Vision and Pattern Recognition Workshops, pp 20–27.

[33] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In CVPR, pp 1290-97, 2012.

[34] Sijin Li, Weichen Zhang, and Antoni B. Chan. Maximum-Margin Structured Learning with Deep Networks for 3D Human Pose Estimation. Intl. Conf. on Computer Vision (ICCV), Santiago, 2015.

[35] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1):221-231, 2013.

[36] P. Luo, X. Wang, and X. Tang. Pedestrian parsing via deep decompositional neural network. In ICCV, pages 2648-55, 2013.

[37] J. K. Aggarwal and M. S. Ryoo. Human activity analysis: A review. ACM Computing Survey, 43(3):16, 2011.

[38] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In Intl Workshop on Wearable and Implantable Body Sensor Networks, pp 113–116, 2006.

[39] L. Chen, H. Wei, J. M. Ferryman. A survey of human motion analysis using depth imagery. Pattern Recognition Letters, 34:1995, 2013.

[40] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3D points. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp 9-14, 2010.

[41] J. Shotton, T. Sharp, A. Kipman, A. W. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. Communications of the ACM, 56(1):116–124, 2013.

[42] J. Sung, C. Ponce, B. Selman, and A. Saxena. Unstructured human activity detection from RGBD images. In IEEE Intl Conference on Robotics and Automation, pp 842–849, 2012.

[43] C. Zhang, Y. Tian. RGB-D based daily living activity recognition. Journal of Computer Vision and Image Processing, 2(4), Dec. 2012.

[44] D. Gurkaynak and H. Yalcin, Recognition and Classification of Human Activity from RGB-D Videos. IEEE Conference on Signal Processing and Communication Applications (SIU 2015), pp 1642-1646, 2015.

[45] Anne 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.

[46] Anne 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.

[47] Anne 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.

[48] Anne 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.

[49] Anne 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.

[50] Anne 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.

[51] Anne 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.

[52] Anne 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.

[53] Anne 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.

[54] Anne 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.

[55] Anne 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.

[56] Anne 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.

[57] Anne 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.

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

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

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

[61] A. S. Patwardhan, "View Independent Drowsy Behavior and Tiredness Detection", International Research Journal of Engineering and Technology (IRJET), May. 2011.

[62] A. S. Patwardhan, "Sensor Based Human Gait Recognition for Drunk State", International Research Journal of Engineering and Technology (IRJET), May. 2012.

[63] A. S. Patwardhan, "Background Removal Using RGB-D data for Fright Recognition", International Research Journal of Engineering and Technology (IRJET), May. 2012.

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

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

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

[67] A. S. Patwardhan, "Call Tree Detection Using Source Code Syntax Analysis", International Research Journal of Engineering and Technology (IRJET), May. 2014.

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

[69] A. S. Patwardhan, "Online News Article Temporal Phrase Extraction for Causal Linking", International Research Journal of Engineering and Technology (IRJET), May. 2015.

[70] A. S. Patwardhan, "Online Comment Processing for Sentiment Extraction", International Research Journal of Engineering and Technology (IRJET), May. 2016.