International Research Journal of Education and Technology
ISSN 2581-7795

Behaviour Analytical Model for Crowd Attentiveness based on Skeleton Pose Estimation, Person Detection and Oculus Behaviour

Sathish M [1] | Parveen H [2] | Monika R [3]

[1, 2, 3] - SRM Valliammai Engineering College - Kattankulathur, Kanchipuram, Tamil Nadu.

ABSTRACT - Analysis of Human Behaviour has attracted many research attention in Computer Vision. Various applications of Computer Vision such as, education, health care, human-computer interactions and video understanding. Though many research work in this domain, the problems and challenges have remained unsolved. Some of the challenges are Occlusion, Eye Movement Metrics, Interclass variation, etc., In the part of large group, our project aims to extract behavioural information from the input videos or live streams which contains input as Crowd Environment. This method is applied in Classroom Environment to enhance the teaching quality by analysing the student activity. Many studies have focused on the physical activity of a student such as hand raising gestures and sleeping activity by Pose Estimation and Person Detection. So, we are proposing Oculus Behaviour to improve and enhance the accuracy of the existing system. Therefore, this project proposes a Behavioural Analytical Model for Crowd Attentiveness based on Skeleton Pose Estimation, Person Detection and Oculus Behaviour.

Keywords- skeleton pose estimation; crowd behaviour; person detection; eye tracking; deep learning; CNN.

1. INTRODUCTION

The skeleton pose estimation and person detection are the most recently used technique for Human Behaviour Recognition. In this technique, the skeleton data were initially identified, including the joint location and the human poses were identified with Person Detection. This method uses RGB deep images captured by Microsoft Kinect Sensor as an input images for Behaviour Recognition and also it uses OpenPose framework to identify Human Poses. To further improve the existing system, Mediapipe framework is used to detect the activity of the Oculus.The main contributions of this paper are as follows:

- An Analysis of Crowd Behaviour Model was proposed and applied for the Classroom Environment.

- An Error Correction Scheme was proposed for Pose Estimation.

- Person Detection and Oculus Behaviour to decrease the incorrect connections.

The feasibility can be achieved and performance will be improved.

2. METHODOLOGY

2.1.Deep learning

Deep learning is a machine learning and artificial intelligence (AI) technique that mimics how humans acquire knowledge. Data science, which covers statistics and predictive modelling, incorporates deep learning as a key component. Deep learning is highly useful for data scientists who are responsible with gathering, analysing, and interpreting massive amounts of data; it speeds up and simplifies the process.

Deep learning can be regarded of as a means to automate predictive analytics at its most basic level. Deep learning algorithms are built in a hierarchy of increasing complexity and abstraction, unlike typical machine learning algorithms, which are linear.

The following are some of the fields where deep learning is currently being used:

- **Medical Investigation:** Deep learning has begun to be used by cancer researchers as a means of automatically detecting cancer cells.

- **Military and Aerospace:** Deep learning is being used to recognise items from satellites in order to determine regions of interest and safe or risky zones for troops.

- **Text Generation:** Machines are taught the grammar and style of a piece of writing, and then use this model to create an entirely new text that matches the original text's proper spelling, grammar, and style.

- **Computer Vision:** Deep learning has considerably improved computer vision, allowing computers to detect objects and classify, restore, and segment images with extraordinary accuracy.

2.2.Haar Cascade Algorithm

Haar Cascade uses the object detection approach to look for faces. A lot of positive and negative images are used to train the algorithm. Each captured image is converted into grayscale (black and white) which allows us to get a monochrome image hence getting the required area. Setting a threshold area value in the parameter allows us to ignore smaller areas which are not important to us.

The Haar Cascade Algorithm takes the following 4 conditions into consideration:

1. Edges.
2. Curves.
3. Surface area.
4. Corners.

2.3. CNN

CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vison that includes image and video recognition, along with recommender systems and natural language processing (NLP).

A neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain. Traditional neural networks are not ideal for image processing and must be fed images in reduced-resolution pieces. CNN have their “neurons” arranged more like those of the frontal lobe, the area responsible for processing visual stimuli in humans and other animals. The layers of neurons are arranged in such a way as to cover the entire visual field avoiding the piecemeal image processing problem of traditional neural networks.

A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to trains limited for image processing and natural language processing.

3. EXISTING SYSTEM

First, Existing System will use Recorded Video or Live Classroom Camera to capture consecutive frame which will be used as an input image.Then, Skeleton Data will be collected using the OpenPose framework and Person Behaviour Data will be analysed using Haar Cascade Algorithm.The skeleton pose estimation and person detection are the most recently used technique for Human Behaviour Recognition.In this technique, the skeleton data were initially identified, including the joint location and the human poses were identified with Person Detection.

3.1. Drawbacks

However, detecting the Pose Estimation and person Detection gives the features. But the accuracy of the existing system is less.

The existing system will not notice the eye activity of a person. The person may be sleeping without bowing. This problem will not be covered.

4. PROPOSED SYSTEM

The proposed model (i.e., Oculus Behaviour) combined with the existing system such as Skeleton Pose Estimation and Person Detection to produce the result.

Skeleton Pose Estimation detects the Skeleton Structure of each Person to monitor the Sitting, Standing, Hands Up and Hands Down activities.

Person Detection detects how many person are present in the environment.

Oculus Behaviour detects the eyes of a person in the environment to monitor the eyeball activities such as Looking Right, Looking Left,Looking Straight / Listening andCounts the Blink.

The following are the procedure for the proposed system.

- First, the System will use Recorded Video or Live Classroom Camera to capture consecutive frame which will be used as an input image.

- Then, Skeleton Data will be collected using the Mediapipe framework and Person Behaviour Data will be analysed.

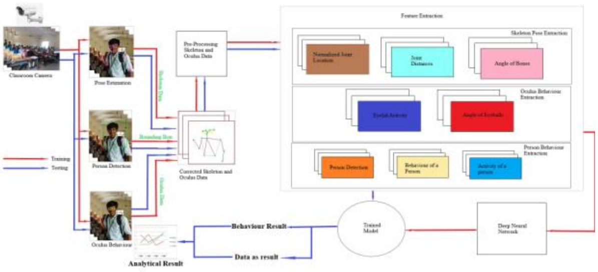
- The System will analyse the Oculus Behaviour such as Activity of the Eyelid and Angle of Eyeballs using Haar Cascade Algorithm.

- Then, Skeleton Data and Oculus Behavioural Data will be Pre-Processed.

- Second, features from Skeleton Pose, Person and Oculus Behaviour will be extracted to generate feature vectors.

Finally, extracted data will be classified to recognize Crowd Behaviours. Classification of Actions will be done using Deep Neural Network (DNN).

4.1. Architecture



4.2.Advantages of Proposed System

The proposed system will give you the exact position of eyes and iris. It detects the eyes and iris and give you the blink counts and position of the eyes in the console. The positions such as Looking Left, Looking Right and looking Straight / Listening will be shown. This should work with the existing system which improves the accuracy from 94.5% to 97%.

The Precision and Recall value will give 10% and 7% higher than the Existing System. The result obtained from this project will also enhance the performance in complex situations.

5. IMPLEMENTATION

5.1. Pseudo code

5.1.1. Live Video Capturing / Recorded Video

1. Run the GUI.
2. Choose anyone of the Option.
3. If option == recorded video
4. Browse the Video in the storage.
5. Choose the appropriate one.
6. Else if option == live stream
7. Open the Webcam.

5.1.2. Pose Estimation

1. Capture the Video or Upload the Video.
2. Detect the Humans.

3. Draw the Skeleton Pose.
4. If Left or Right Elbow Angle is greater than 40 and Left or Right Shoulder Angle is greater than 90
5. Display the Position of the Hand.
6. If Right Knee Angle is greater than 145 and Left Knee Angle is greater than 145
7. Display the Position of the Person.

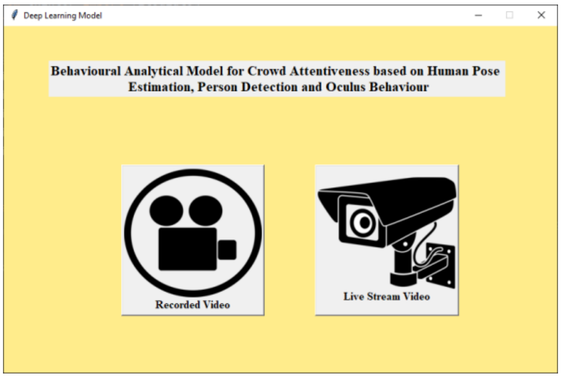
5.1.3. Person Detection

1. Capture the Video or Upload the Video.
2. Detect the Humans.
3. Draw the Bounding Boxes.
4. If person > 1
5. Count the number of person.
6. Display the maximum count.

5.1.4. Oculus Detection

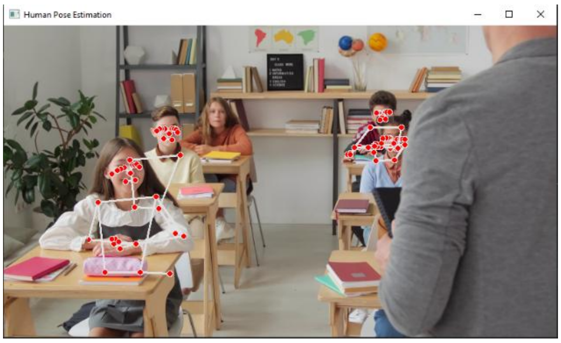
1. Capture the Video or Upload the Video.
2. Detect the eyes.
3. Detect the Iris.
4. Draw the shape of the eyes and iris.
5. If the distance of upper eyelid and lower eyelid is lower than the ratio
6. Count the Blink as 1
7. If the position of the Iris is less than or greater than the ratio
8. Display the Position of the Iris.

5.2.Output

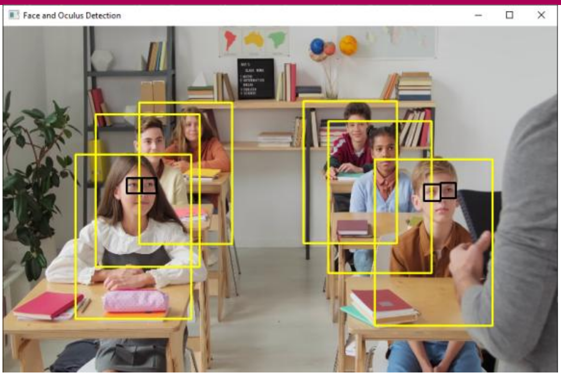


Behavioural Analytical Model for Crowd Attentiveness based on Human Pose Estimation, Person Detection and Oculus Behaviour

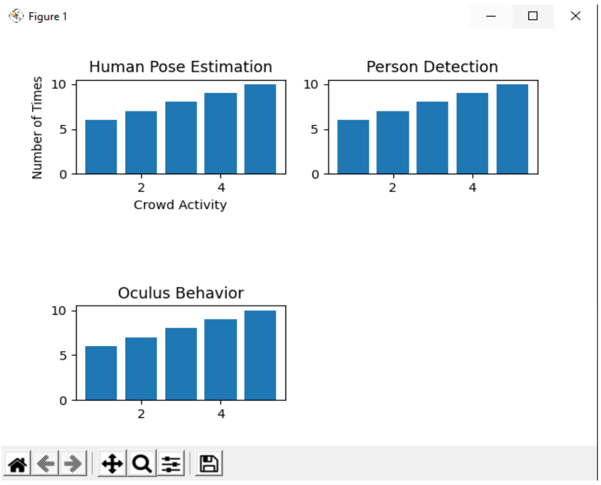
(a). GUI Interface enables you to choose if you want the recorded video or live stream to be processed with the detection



(b). Mediapipe Framework does the Skeleton Pose Estimation of Students in the Classroom.



(c). Haar Cascade and Mediapipe does the Person Detection and Eye Detection.



(d).Values fetched from the above detections are collected and analytical graph is plotted.

6. CONCLUSION

The proposed system uses the detections to get the values and gives the analytical graph. The values obtained from Pose Estimation is Hands Up, Hands Down, Sitting and Standing. The values obtained from Person Detection is total number of person detected. The values obtained from Oculus Behaviour is Looking Right, Looking Left, Looking Straight / Listening. The system is easy to use and no prior experience is needed. By using this system, teachers can improve their way of teaching to get the attention from the students.

7. REFERENCE

1. Yu, M.; Xu, J.; Zhong, J.; Liu, W.; Cheng, W. Behavior Detection and Analysis for Learning Process in Classroom Environment. In Proceedings of the IEEE Frontiers in Education Conference (FIE 2017), Indianapolis, IN, USA, 18–21 October 2017; pp. 1–4.

2. Zheng, R.; Jiang, F.; Shen, R. Intelligent Student Behavior Analysis System for Real Classrooms. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020), Barcelona, Spain, 4–9 May 2020;pp. 9244–9248

3. Zheng, R.; Jiang, F.; Shen, R. GestureDet: Real-time Student Gesture Analysis with Multi-Dimensional Attention-based Detector. In Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), Yokohama, Japan, 11–17 July 2020; pp. 680–686.

4. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtimemultiperson 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. ***2021****, 43, 172–186.*

5. Qiang, B.; Zhang, S.; Zhan, Y.; Xie, W.; Zhao, T. Improved Convolutional Pose Machines for Human Pose Estimation using Image Sensor Data. Sensors ***2019****, 19, 718.*

6. Jin, S.; Liu,W.; Xie, E.;Wang,W.; Qian, C.; Ouyang,W.; Luo, P. Differentiable hierarchical graph grouping for multiperson pose estimation. In Proceedings of the 16th European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 718–734.

7. Dai, Y.;Wang, X.; Gao, L.; Song, J.; Shen, H.T. RSGNet: Relation based skeleton graph network for crowded scenes pose estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; pp. 1193–1200.

8.Cippitelli, E.; Gasparrini, S.; Gambi, E.; Spinsante, S. A human activity recognition system using skeleton data from RGBD sensors. Comput. Intell. Neurosci. ***2016****, 2016, 4351435.*

9. Luvizon, D.C.; Tabia, H.; Picard, D. Learning features combination for human action recognition from skeleton sequences. Pattern Recognit. Lett. ***2017****, 99, 13–20.*

10. Khaire, P.; Kumar, P.; Imran, J. Combining CNN streams of RGB-D and skeletal data for human activity recognition. Pattern Recognit. Lett. ***2018****, 115, 107–116.*

11. Aubry, S.; Laraba, S.; Tilmanne, J.; Dutoit, T. Action recognition based on 2D skeletons extracted from RGB videos. Matec Web Conf. ***2019****, 277, 1–14.*

12.Noori, F.M.;Wallace, B.; Uddin, M.Z.; Torresen, J. A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network. In Proceedings of the Scandinavian Conference on Image Analysis (SCIA 2019), Norrköping, Sweden, 11–13 June 2019; pp. 299–310.

*13. Tzu-Ling Kuo and Chih-Peng Fan, Member, IEEE Department of Electrical Engineering, National Chung Hsing University, Taiwan, R.O.C.****Ab*** *-Deep Learning Based Pupil Tracking Technology for Application of Visible-Light Wearable Eye Tracker. 2020 IEEE International conference on consumer electronics (ICCE).*

14. Akshay, S.; Megha ,Y. J.; ChethanBabu Shetty Department of Computer Science Amrita School of Arts and Sciences, Mysuru Amrita VishwaVidyapeetham,India . Machine Learning Algorithm for to Identify Eye Movement Metrics using Raw Eye Tracking Data. Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020) IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821.

15.RenaldiPrimaswaraPrasetya,FitriUtaminingrum, WayanFirdausMahmudy,Faculty of Computer Science, Brawijaya University, Malang, Indonesia Real Time Eyeball Movement Detection Based on Region Division and Midpoint Position.

16. Ruchika A. Patel and Mr.Sandip R. Panchal PG Student, 2Assistant Professor 1, 2Department of Electronics & Communication SardarvallabhbhaiInstitute of technology Vasad, Gujarat , India- Detected Eye Tracking Techniques: And Method Analysis Survey.

17. Jiang, X.; Xu, K.; Sun, T. Action recognition scheme based on skeleton representation with DS-LSTM network. IEEE Trans. Circuits Syst. Video Technol. ***2020****, 30, 2129–2140.*

18. Agahian, S.; Negin, F.; Köse, C. An efficient human action recognition framework with pose-based spatiotemporal features. Eng. Sci. Technol. Int. J. ***2020****, 23, 196–203.*

19. Liao, W.; Xu, W.; Kong, S.; Ahmad, F.; Liu, W. A two-stage method for hand raising gesture recognition in classroom. In Proceedings of the 8th International Conference on Educational and Information Technology, Cambridge, UK, 2–4 March 2019; pp. 38–44.

20. Mo, L.; Li, F.; Zhu, Y.; Huang, A. Human physical activity recognition based on computer vision with deep learning model. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC 2016), Taipei, Taiwan, 23–26 May 2016; pp. 1–6.

*21. Si, J.; Lin, J.; Jiang, F.; Shen, R. Hand-raising gesture detection in real classrooms using improved R-FCN. Neurocomputing****2019****, 359, 69–76.*

22. Zhou, H.; Jiang, F.; Shen, R. Who are raising their hands? Hand-raiser seeking based on object detection and pose estimation.In Proceedings of the 10th Asian Conference on Machine Learning (ACML 2018), Beijing, China, 14–16 November 2018; pp. 470–485.

23. Wang, Z.; Jiang, F.; Shen, R. An effective yawn behavior detection method in classroom. In Proceedings of the 26th International Conference on Neural Information Processing (ICONIP2019), Sydney, NSW, Australia, 12–15 December 2019; pp. 430–441.

24.Althloothi, S.; Mahoor, M.H.; Zhang, X.; Voyles, R.M. Human activity recognition using *multi-features and multiple kernellearning. Pattern Recognit.* ***2014****, 47, 1800–1812.*

25. Franco, A.; Magnani, A.; Maio, D. A multimodal approach for human activity recognition based on skeleton and RGB data. Pattern Recognit. Lett. ***2020****, 131, 293–299.*

26. Jia, J.G.; Zhou, Y.F.; Hao, X.W.; Li, F.; Desrosiers, C.; Zhang, C.M. Two-stream temporal convolutional networks for skeleton-based human action recognition. J. Comput. Sci. Technol. ***2020****, 35, 538–550.*

27.Negin, F.; Agahian, S.; Köse, C. Improving bag-of-poses with semi-temporal pose descriptors for skeleton-based action recognition. Vis. Comput. ***2019****, 35, 591–607.*

28. Schneider, P.; Memmesheimer, R.; Kramer, I.; Paulus, D. Gesture recognition in RGB videos using human body keypointsanddynamic time warping. In Proceedings of the Robot World Cup XXIII (RoboCup 2019), Sydney, NSW, Australia, 8 July 2019; pp. 281–293.

29. Li, X.; Fan, Z.; Liu, Y.; Li, Y.; Dai, Q. 3D Pose Detection of Closely Interactive Humans using Multiview Cameras. Sensors ***2019****, 19, 1–16.*

30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.