Badminton Shot Recognition with LSTM Network

Paper Id : 040

Presentation by:

Avinav Jain

Problem Statement

- To classify and recognize various strokes played in the game of badminton which could further help identify player’s strategy and weakness.

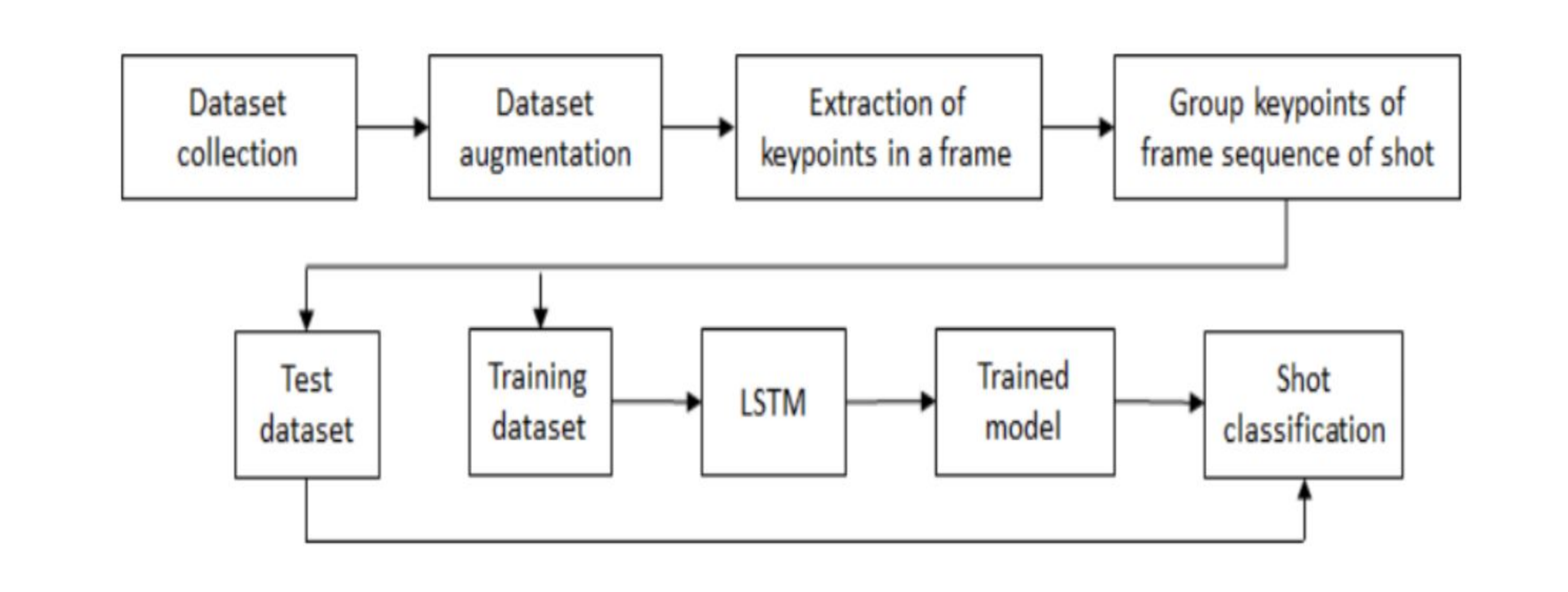
Motivation behind the Problem

- A Player's posture while playing a shot in racket sports such as tennis or badminton speaks a lot about his skill level, so detecting how correct is the player's posture while playing various shots.
- Sports video data is recorded for nearly every major tournament but remains archived and inaccessible to large-scale data mining and analytics. It can only be viewed sequentially or manually tagged, which is time-consuming and prone to errors.
- Applications of sports video analytic systems provide aid in coaching to design the strategy to improve their player's performance, fitness, weaknesses, and strengths assessment and to identify different tactics a player makes.
- In recent times, research on badminton videos has explicitly been rare, and most past research either concentrates on tennis videos or generalizes it to racket sports.

Literature Review

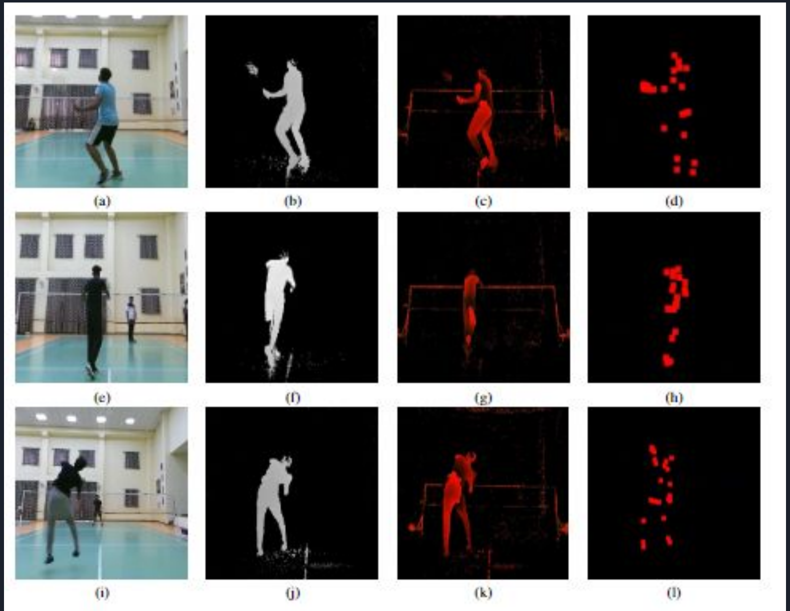
Author	Dataset	Methodology	Result
Ghosh et al.	Olympic Videos	Analysis of racket games by using action recognition, segmentation, and object detection	Player Detection accuracy mAP@0.5 value 97.85%
Chu et al.	Professional Match Videos	Classifying player's strategy using computer vision techniques to process dataset, and predicting using cross validation model	Classification works with 70% accuracy
Mora et al.	THETIS dataset	Classified strokes in the game of tennis using a 3-layered LSTM model	Classification accuracy: 84.10%

Methodology



Dataset Collection

- Dataset is collected in college campus using Microsoft XBox Kinect One Sensor and labelled into three classes, namely, Clear, Serve and Smash.
- Each data item consists of five files, one for RGB video, two for depth maps, 1 video showing joint points of human body, and a json file which has the coordinates of the joint points.
- More than eight players are captured on separate days when capturing the dataset for generalisation purposes.



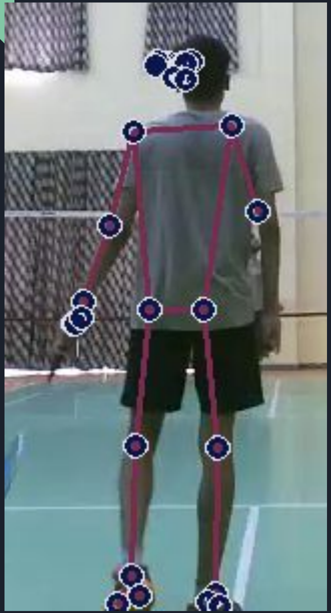
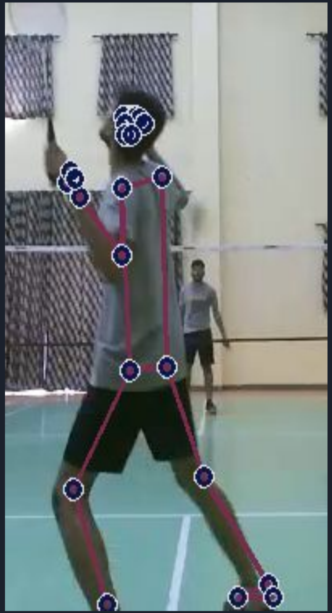
Video Augmentation

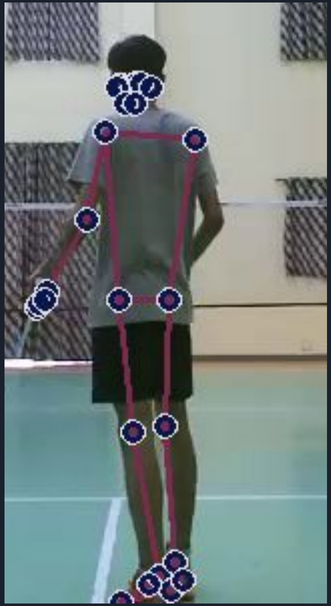
- As a deep learning model requires a large number of training samples, we applied various video augmentation techniques on our collected dataset.
- As part of video augmentation, we extracted frames of videos and applied various small rotations so as to tilt the frames by a small angle.
- Other augmentation techniques include flipping the frames of the video or blurring the frames, normalizing the contrast etc.
- Using this process, we have increased our training data by around 6 times.

Player Body Key points Detection

- The whole frame has a lot of details, which may not correspond to the details needed to predict the type of shot played.
- This may even create hindrances in classification, which may result in bad accuracy.
- Also, using the whole video instead of extracting key points, will require much extra computation power, which may not be feasible.
- To deal with these issues, the next task was to generate the key points for the newly generated dataset.
- We have used MediaPipe holistics for this purpose.

Various frames in a stroke of Clear Shot



Classification Model

- We have built a time-based sequential model(LSTM) to classify these badminton strokes into predefined classes.
- Our dataset is divided into two parts, 90 percent for the training set, and 10 percent for the test set.
- During model training, to further increase the training data, we have used a 5-fold cross-validation method.

Layer Type	Output Shape	Param Count
Time Distributed	(None, 15, 225)	0
LSTM	(None, 15, 512)	1511424
LSTM	(None, 15, 256)	787456
LSTM	(None, 128)	197120
Dense	(None, 64)	8256
Dense	(None, 32)	2080
Dense	(None, 3)	99

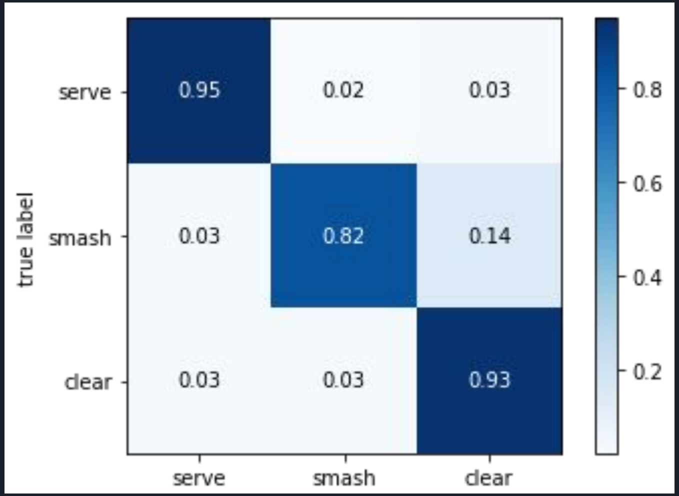
Total Params: 2,56,435

Model Architecture

Results on Test Set

	Precision	Recall	f1-score	Support
Serve	0.95	0.94	0.94	194
Smash	0.82	0.95	0.88	186
Clear	0.93	0.78	0.85	172
accuracy			0.89	552

Classification report



Confusion Matrix

Conclusion

- We have trained a time-based sequential neural network to classify badminton strokes into three types.
- After extracting the key points of the players' bodies using MediaPipe Posture Detection, we trained our sequential network architecture and achieved an accuracy of 89.49 percent using a 5-fold cross-validation technique.
- In the future, this study can be extended further to classify more distinct shots in the game and figure out the strategy of the player.
- This can also help coaches in evaluating one's performance and detecting weak spots.

References

1. Ban, K.W., See, J. Abdullah, J., Loh, Y.P.: BadmintonDB: A Badminton Dataset for Player-specific Match Analysis and Prediction. In: Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports, pp. 47–54.ACM (2022).
2. W. T. Chu and S. Situmeang. 2017. Badminton Video Analysis based on Spatiotemporal and Stroke Features. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR '17). Association for Computing Machinery, New York, NY, USA, 448–451. https://doi.org/10.1145/3078971.3079032
3. T. Tsunoda, Y. Komori, M. Matsugu and T. Harada, "Football Action Recognition Using Hierarchical LSTM," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 155-163, doi: 10.1109/CVPRW.2017.25.
4. Kulkarni, K. M., and Shenoy, S. (2021). "Table Tennis Stroke Recognition Using Two-Dimensional Human Pose Estimation" CVPR Sports Workshop 2021 arXiv. https://doi.org/10.48550/arXiv.2104.09907
5. A. Ghosh, S. Singh and C. V. Jawahar, "Towards Structured Analysis of Broadcast Badminton Videos," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 296-304, doi: 10.1109/WACV.2018.00039.
6. S. V. Mora and W. J. Knottenbelt, "Deep Learning for Domain-Specific Action Recognition in Tennis," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 170-178, doi: 10.1109/CVPRW.2017.27.
7. M. Sharma, M. Lamba, N. Kumar and P. Kumar. (2021). Badminton match outcome prediction model using Naïve Bayes and Feature Weighting technique. Journal of Ambient Intelligence and Humanized Computing. 2020. 10.1007/s12652-020-02578-8.
8. L. Wang, Y. Qiao, X. Tang. (2015). Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors. 10.1109/CVPR.2015.7299059.
9. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

10. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

11. K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In ICLR, 2014 arXiv. https://doi.org/10.48550/arXiv.1409.1556

12. M. Mlakar and M. Luštrek. 2017. Analyzing tennis game through sensor data with machine learning and multi-objective optimization. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp '17). Association for Computing Machinery, New York, NY, USA, 153–156. https://doi.org/10.1145/3123024.3123163

13. V. Renò, N. Mosca, M. Nitti, T. DOrazio, C. Guaragnella, D. Campagnoli, A. Prati, and E. Stella. (2017). A technology platform for automatic high-level tennis game analysis. Computer Vision and Image Understanding. 159. 10.1016/j.cviu.2017.01.002.

14. B. L. Bhatnagar, S. Singh, C. Arora, and C. V. Jawahar. Unsupervised learning of deep feature representation for clustering egocentric actions. In IJCAI, 2017.

15. A. Fathi and J. M. Rehg, "Modeling Actions through State Changes," 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2579-2586, doi: 10.1109/CVPR.2013.333.

16. 16. Lea, C., Reiter, A., Vidal, R., and Hager, G. D. (2016). Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. In ECCVW, 2016. arXiv. https://doi.org/10.48550/arXiv.1602.02995

17. A. Kumar, J. Garg, and A. Mukerjee. (2015). Cricket activity detection. International Image Processing, Applications and Systems Conference, IPAS 2014. 10.1109/IPAS.2014.7043264.

THANK YOU