Towards Structured Analysis of Broadcast Badminton Videos

Supplementary Material

1. Dataset Details

Following is the list of all available match videos from London Olympics 2012, and we selected the first ten to annotate for our Badminton Olympics Dataset. Also, the label distribution of the strokes is skewed, as we can observe in Fig 1.

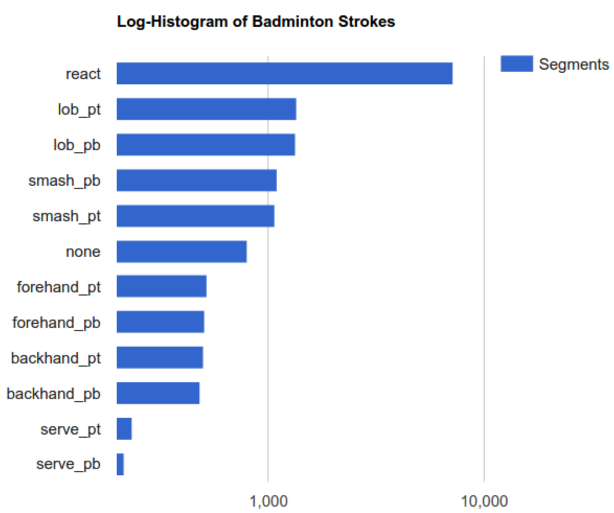

Figure 1: **Distribution of Labels.** Shown here is the log histogram of the classes present in our dataset. We simply count the number of segments belonging to each class to create this histogram and as can be seen, the dataset shows considerable skew.

1. WeiLee-Long-SemiFinals-LondonOlympics-2012 (M)
2. Lee-Dan-SemiFinals-LondonOlympics-2012 (M)
3. Firdasari-Zaitsava-GrpO-LondonOlympics-2012 (W)
4. Baun-Augustyn-GrpG-LondonOlympics-2012 (W)
5. Nguyen-Parupalli-GrpD-LondonOlympics-2012 (M)
6. Yihan-Nehwal-SemiFinals-LondonOlympics-2012 (W)
7. Hidayat-Abian-GrpO-LondonOlympics-2012 (M)
8. Na-Fasungova-GrpD-LondonOlympics-2012 (W)
9. Li-Wang-SemiFinals-LondonOlympics-2012 (W)
10. Chen-Zwiebler-R32-LondonOlympics-2012 (M)
11. WeiLee-Dan-Finals-LondonOlympics-2012 (M)
12. Magee-Hosny-GrpI-LondonOlympics-2012 (W)
13. WeiLee-Parupalli-QtrFinals-LondonOlympics-2012 (M)
14. Karunaratne-Parupalli-R16-LondonOlympics-2012 (M)
15. Nguyen-Tan-GrpD-LondonOlympics-2012.mp4 (M)
16. Chen-Wacha-GrpL-LondonOlympics-2012 (M)
17. WeiLee-Lang-GrpA-LondonOlympics-2012 (M)
18. Sung-Yip-GrpJ-LondonOlympics-2012 (W)
19. Dan-Evans-GrpP-LondonOlympics-2012 (M)
20. Sasaki-Soeroredjo-GrpN-LondonOlympics-2012 (M)
21. Shenck-Gavnholt-GrpN-LondonOlympics-2012 (W)
22. Cordon-Hurskainen-GrpM-LondonOlympics-2012 (M)
23. Sasaki-Cordon-R16-LondonOlympics-2012 (M)
24. Lee-Chen-QtrFinals-LondonOlympics-2012 (M)
25. Wang-Li-Finals-LondonOlympics-2012 (W)
26. Nehwal-Xin-Bronze-LondonOlympics-2012 (W)
27. Lee-Long-Bronze-LondonOlympics-2012 (M)

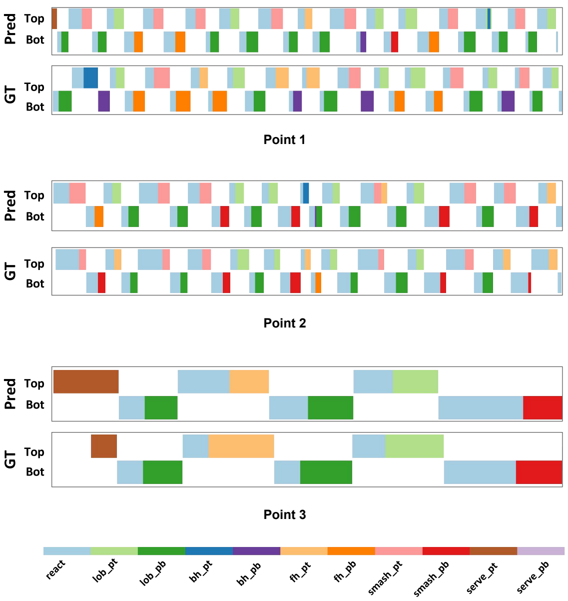

Figure 2: **Additional Stroke Visualizations** Each row represents a for stroke visualization for 3 different points of varying lengths and from different matches. The third row demonstrates that the react class is segmented out adequately well, also, it's important to remember that the human segmentation effort is in itself subjective.

Feature	Metric	d=5	d=10	d=15	d=20
HOG	mAP@0.1	58.95	61.79	60.62	60.02
	mAP@m	56.07	57.67	56.89	55.38
SpatialCNN	mAP@0.1	59.16	59.19	61.23	60.65
	mAP@m	54.51	56.24	57.06	53.59

Table 1: Additional results on different metrics on our dataset using ED-TCN. *mAP@m* corresponds to mAP@mid.

2. Analysis of Stroke Segmentation

We report the results for the mAP based metrics in Table 1 for the ED-TCN. As [2] noted in their analysis, for many fine-grained action detection applications that results are not indicative of real-world performance and this can also be applied to sports video. The key issue is that mAP is very sensitive to a confidence score assigned to each segment prediction, and for badminton actions, where the segments are very small and minute, such a metric fails to capture the performance well, unlike video retrieval tasks. For instance, in the third row of Fig. 2 which presents predictions from SpatialCNN features, it's can be observed that the predicted segments even though aligned, need not have their midpoint within the ground truth segment as the action segments are too small.

Also, we report the confusion matrices for the learnt ED-TCN models in Fig. 3 and Fig. 4. As we note, the accuracy of the bottom player strokes, specially the react, smash, lob and serve classes are better than the top player. This is consistent with earlier works [1, 3] in stroke/shot recognition (both coarse and fine labels) which focused on the bottom player since the player is easier to detect and recognize. It should also be noted the model is confused among the forehand and the backhand classes, which are visually very similar. We would be investigating more robust features in detail in our future work. End-to-end jointl training of the feature extraction and the temporal action segmentation models may correct these mistakes.

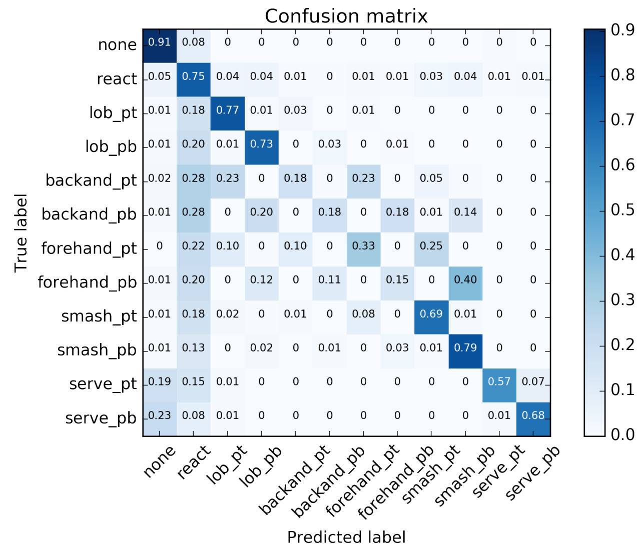

Figure 3: **Confusion Matrix using SpatialCNN features.** The confusion matrix reported is for the ED-TCN with optimal parameters.

3. Additional Analysis

We present additional results for our computed metrics in Fig. 5 and Fig. 6. As can be seen in Fig. 5, the first player (marked green) lost the first set and it's apparent that they were slower than the other player in that set. In the second set, they won comfortably. They seem to have caught up in speed and can be seen to have higher react time to carry their shots, meaning they were in total control of the points. The third set is closer in play, however, the last few points have higher react time for the winning player, meaning they regained control of the match.

Similarly, in Fig. 6, we can see from the dominance sequence that the match was close, even though it ended in straight sets. The average speed and react times are well matched, however, the second player (in blue) has a higher react time to carry their shots in the second set when they also started dominating.

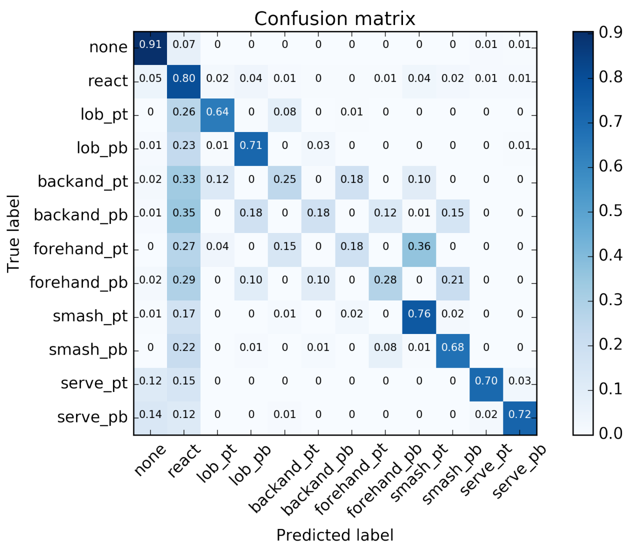

Figure 4: **Confusion Matrix using HOG features.** The confusion matrix reported is for the ED-TCN with optimal parameters.

Fig. 7 shows the footwork for a point played[1]. The footwork visualization shows a common badminton strategy of playing strokes near court's borders. The bottom player wins this point when the top player willingly let the shuttle drop, thinking it would drop outside the court's border. On contrary, the shuttle drop very close but inside the court's border leading to a win for top player.

4. Failure Cases

Badminton is marred by a lot of occlusions and fast paced actions. Fig 8 presents a successful case of player detection, while Fig. 9 depicts the most common failure cases.

Fig. **??** shows the failure cases with respect to top and bottom players for temporal action segmentation. A common case for confusions corresponds to fast paced strokes due to extreme rapid body deformation. Other failures are been due to the dataset bias towards right-handed players, and visually identical strokes (deception strategy).

References

[1] W.-T. Chu and S. Situmeang. Badminton Video Analysis based on Spatiotemporal and Stroke Features. In *Proc. ICMR*, 2017. 2

[2] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. Temporal convolutional networks for action segmentation and detection. 2017. 2

[3] H. Shah, P. Chokalingam, B. Paluri, N. Pradeep, and B. Raman. Automated stroke classification in tennis. *Image Analysis and Recognition*, 2007. 2

[1] https://youtu.be/-yF5pMWafp8?t=831

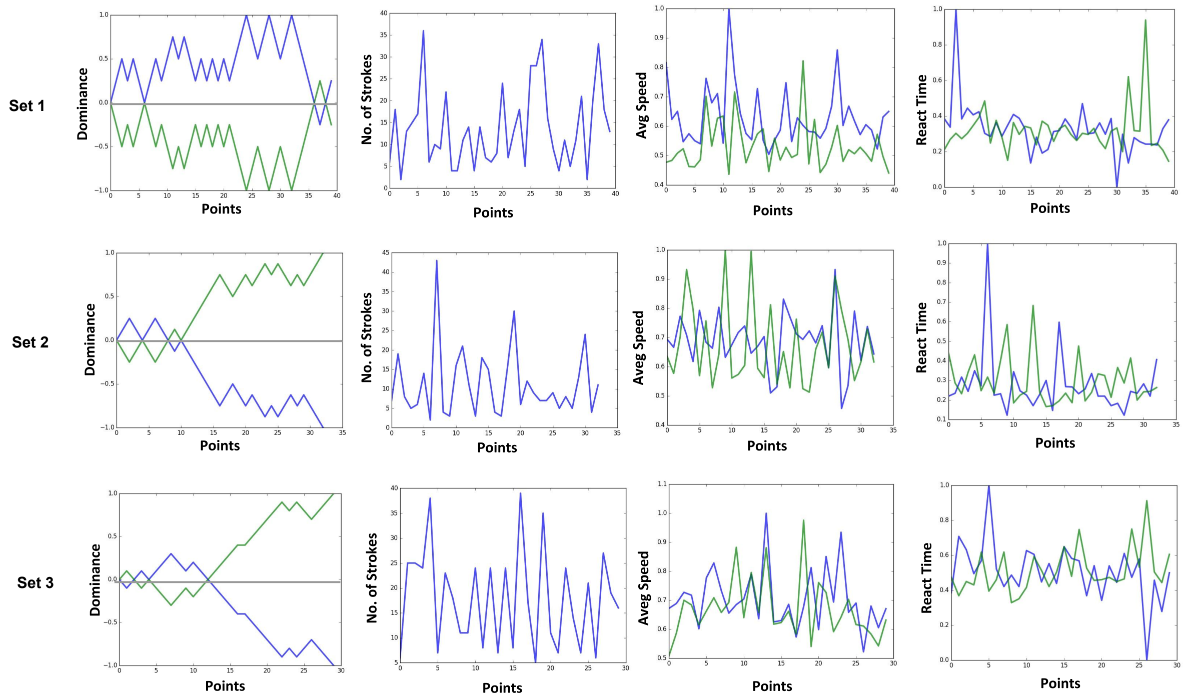

Figure 5: The computed statistics for a match, where each row corresponds to a set. The player marked as green won the match, after losing the first set. *(Best viewed in color)*

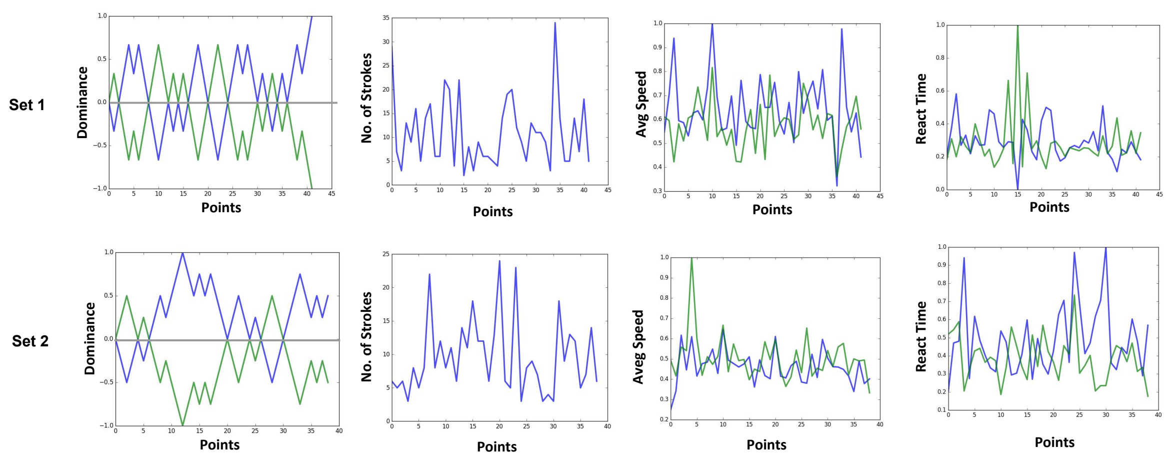

Figure 6: The computed statistics for a match, where each row corresponds to a set. The player marked as blue won the match in straight sets. *(Best viewed in color)*

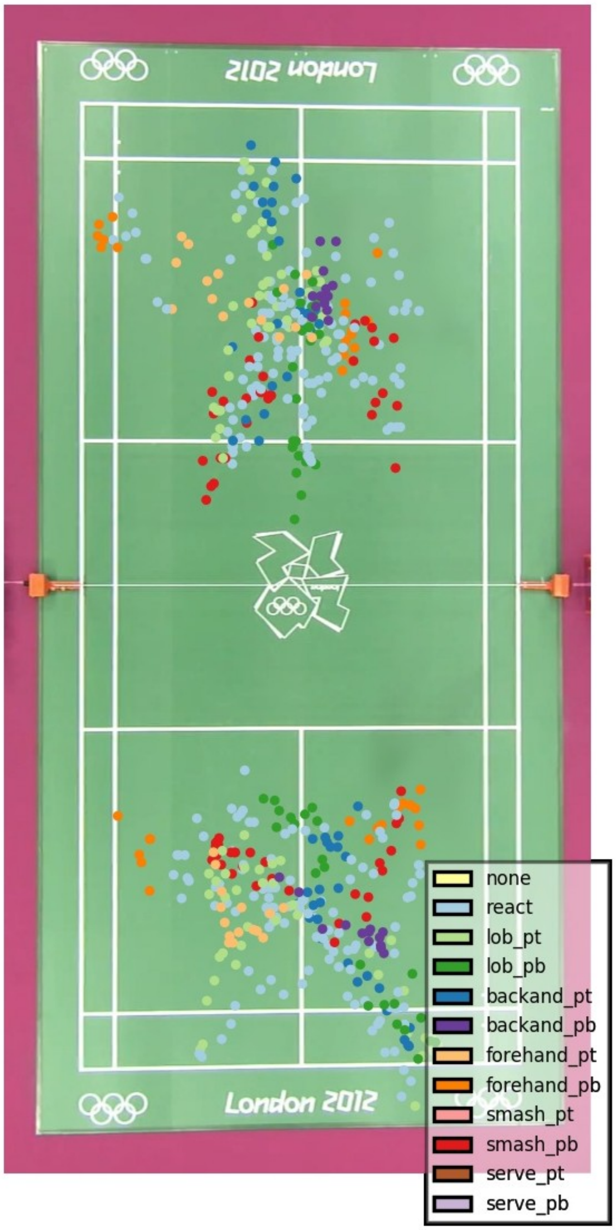

Figure 7: The point summary for the point depicted in Fig. **??**. The bottom player won the point.

Figure 8: The top player is occluded and the detection model is reasonably able to detect him.

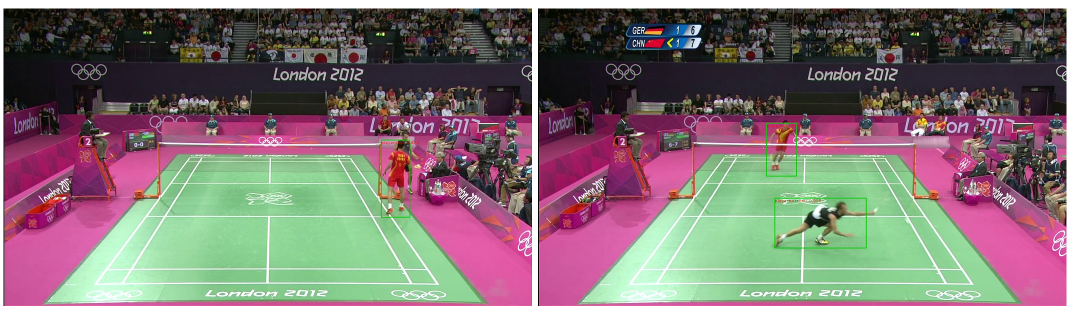

Figure 9: **Left** : The top player is heavily occluded and the detection model failed to detect him. **Right** :he top player is not tightly bounded by the detection box.