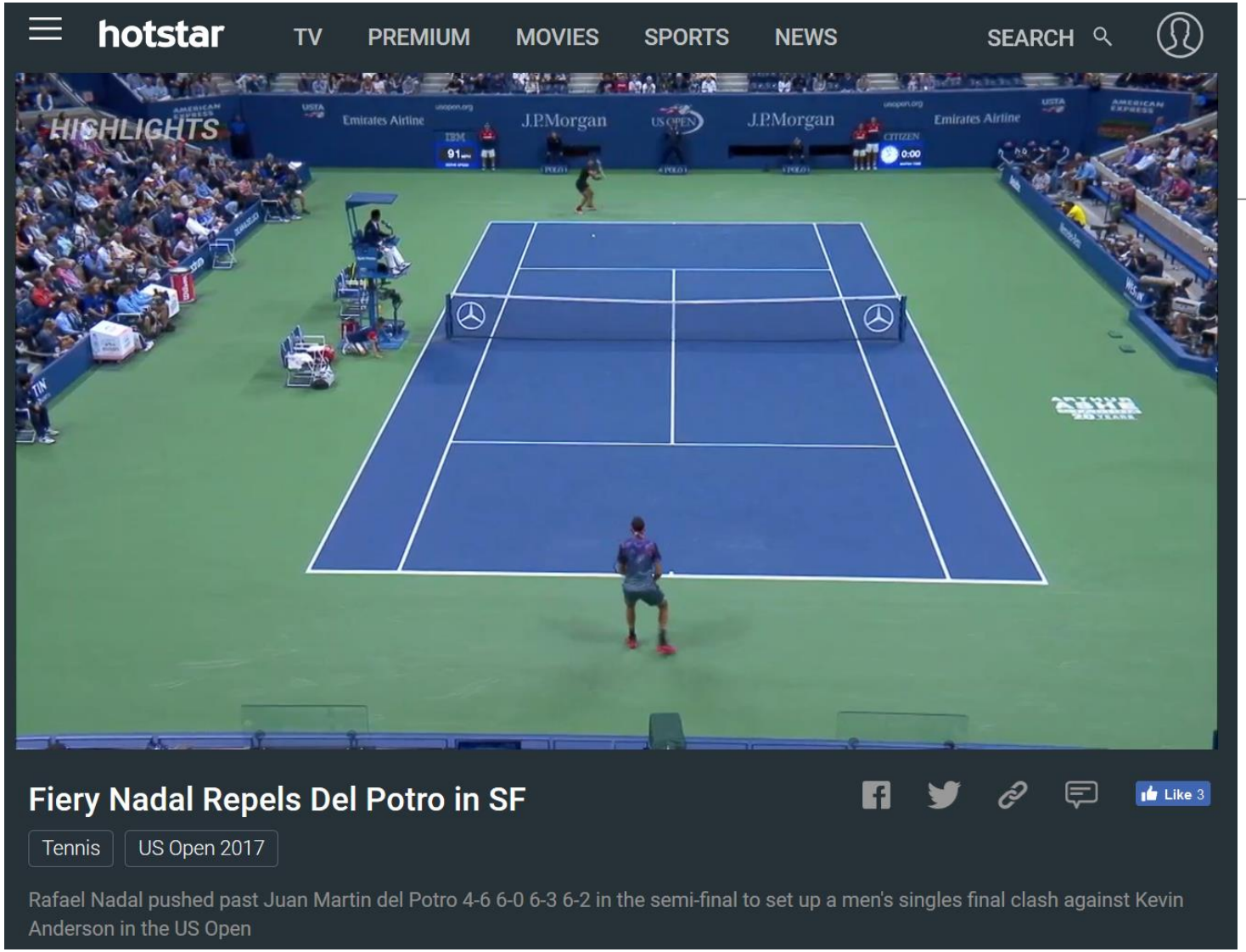
SmartTennisTV

AUTOMATIC INDEXING OF TENNIS VIDEOS

ANURAG GHOSH, C.V. JAWAHAR

CVIT, IIIT HYDERABAD



Motivation

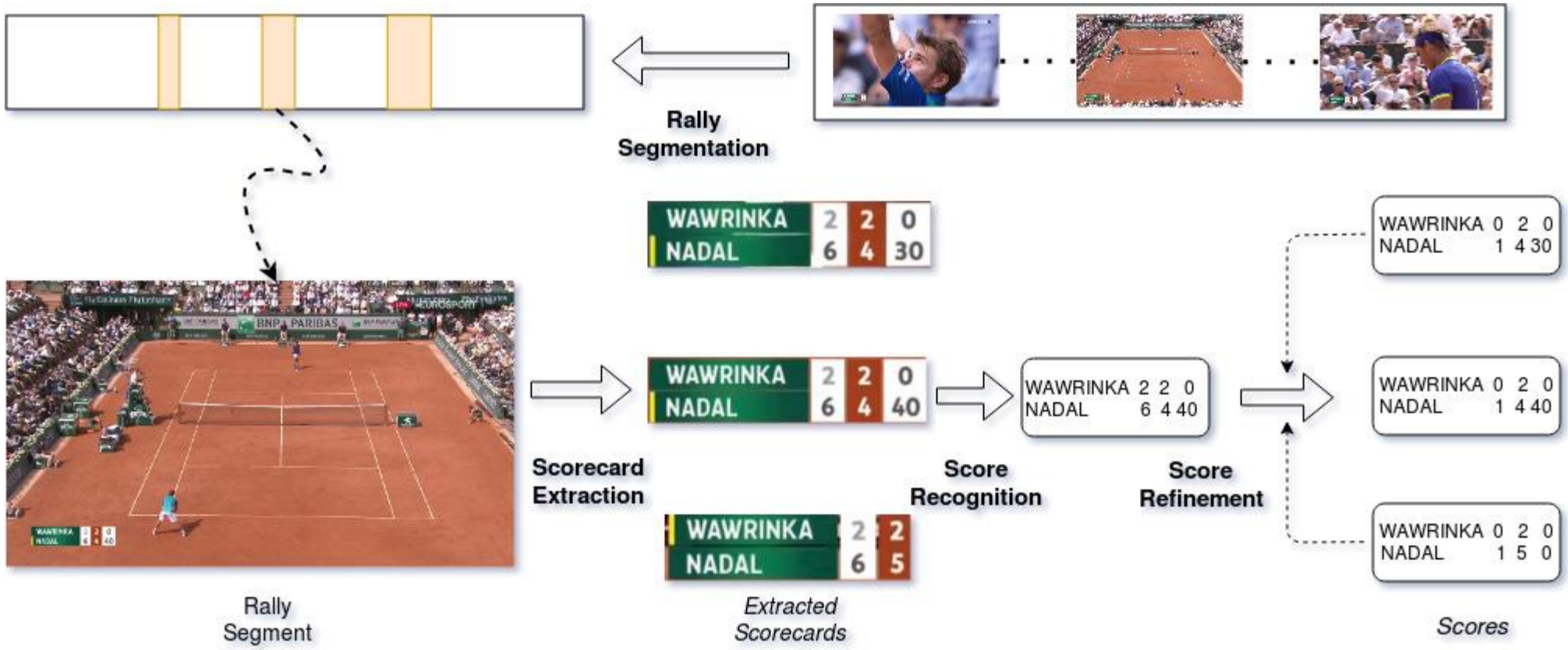
- Streaming Services are plenty
  - Hotstar
  - TennisTV
  - WatchESPN
  - Etc…
- All of them offer hundreds of match replays
- Watching matches from start to end
  - Cumbersome

Motivation

- For enthusiasts and coaches:
  - Watch the real action – the rallies, games and sets
  - Fast forward
  - Revert and revisit any time
- So, what would be neat?
  - Direct/Randomized access to rallies using an index
  - Tagging with semantic labels, like deuce

Approach

Automatically index a tennis match by finding rally segments and extracting the scores.



Approach

- Rally Segmentation
- Scorecard Extraction
- Score Recognition
- Score Refinement
- Event Labelling

Segmenting Rallies

- The play is usually captured from behind the baseline camera view
- Utilize the observation
  - Extract HOG features from every frame
  - Learn a classifier to say game in play or not
  - Apply a Kalman filter on the obtained labels
- Works reasonably well, 95.41% precision in detecting rallies

Extracting the Scorecard

- Scorecards are dynamic in nature
  - Size and shape changes to show more or less information
  - Can't use a common template
  - Differences across tournaments etc.
  - Extract for each rally, i.e. video segment, separately.

Extracting the Scorecard

- Scorecards, however, don't move with the action
  - Find temporally correlated regions from gradient of the frames
  - Coarsely identify the quadrant (summing up region pixels)
  - Apply morphological operators (open, close etc.)
  - Fit a bounding box to identify
- Some of the extracted scorecards here
  - The failure cases in (v) and (vi)

Getting Scores

- We have the scorecard, now what?
  - Extract the scores!
- Can we use OCR directly?
  - Not quite.
  - Scorecards have highly varying contrast regions
  - Text can be white on black, black on white etc.
  - Preprocess needed to create optimal input for each scorecard type

Getting Scores

- Experiment with scene text recognition methods
  - Useful in recognizing text in-the-wild
  - Textspot is SoTA on multiple datasets without finetuning
  - CRNN is trained on the dataset
- We also experimented with Tesseract with manual preprocessing
- Caveats of the approach?
  - Scorecards use widely differing formats – Can't infer game state directly
  - Need to postprocess text based on tournament scorecard style

Getting Scores

- Textspot works the best without any preprocessing!
- Tesseract is okay, also CRNN.

Table 1. Averaged Edit Distance for score recognition (Lower is better)

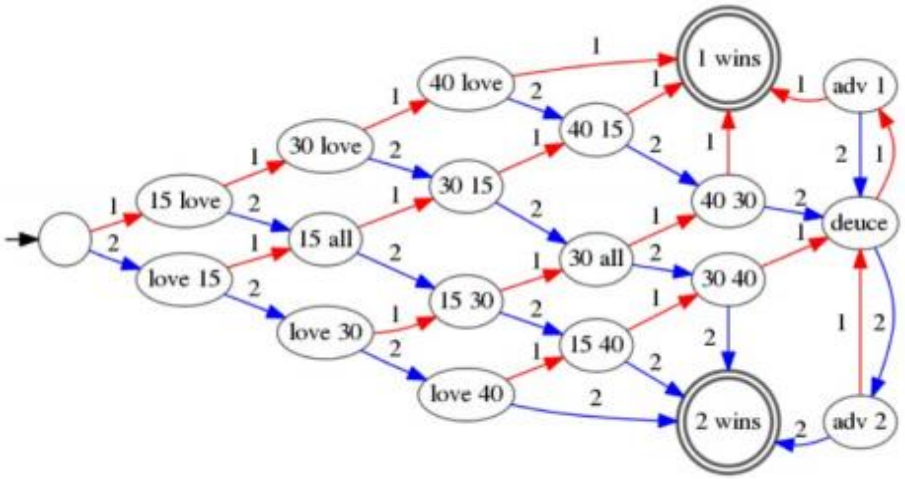
Match	Textspot	CRNN	Tesseract-P
Match 1 (186 rallies)	0.2070	0.4272	0.2612
Match 2 (218 rallies)	0.2178	0.4476	0.3780

Refining Scores

- Textspot is nice!
  - Still, Errors introduced by scene text recognition
  - Sometimes there's no scorecard in the video segment at all.
  - Any way to remove these errors?

Refining Scores

- We know how scores behave
  - Utilize the scoring system as a constraint
  - Refine the captured score sequence using these constraints
  - How to model the scoring system? State Automaton is appropriat
- Assume the score sequence is $s_1, s_2 \ldots s_n$
  - Let's say we wish to refine s_i using the automaton
  - Let P be all possible states reached by s_{i-1} and states that can reach s_{i+1}
  - Assign s_i the *closest possible state in P* w.r.t. current s_i
  - Repeat for the score sequence



Refining Scores

- As we can see, we observe a good increase in accuracy for each of the matches.
- However, the algorithm is no good if the scores recognized are incorrect.

Match	Textspot	Ours
Match 1 (186 rallies)	79.30%	91.66%
Match 2 (218 rallies)	77.90%	80.58%
Match 3 (201 rallies)	92.45%	95.19%
Match 4 (194 rallies)	85.22%	92.18%

Demo?

Interface supports the indexing and retrieval of a match segments (point, game, set) with tagged events

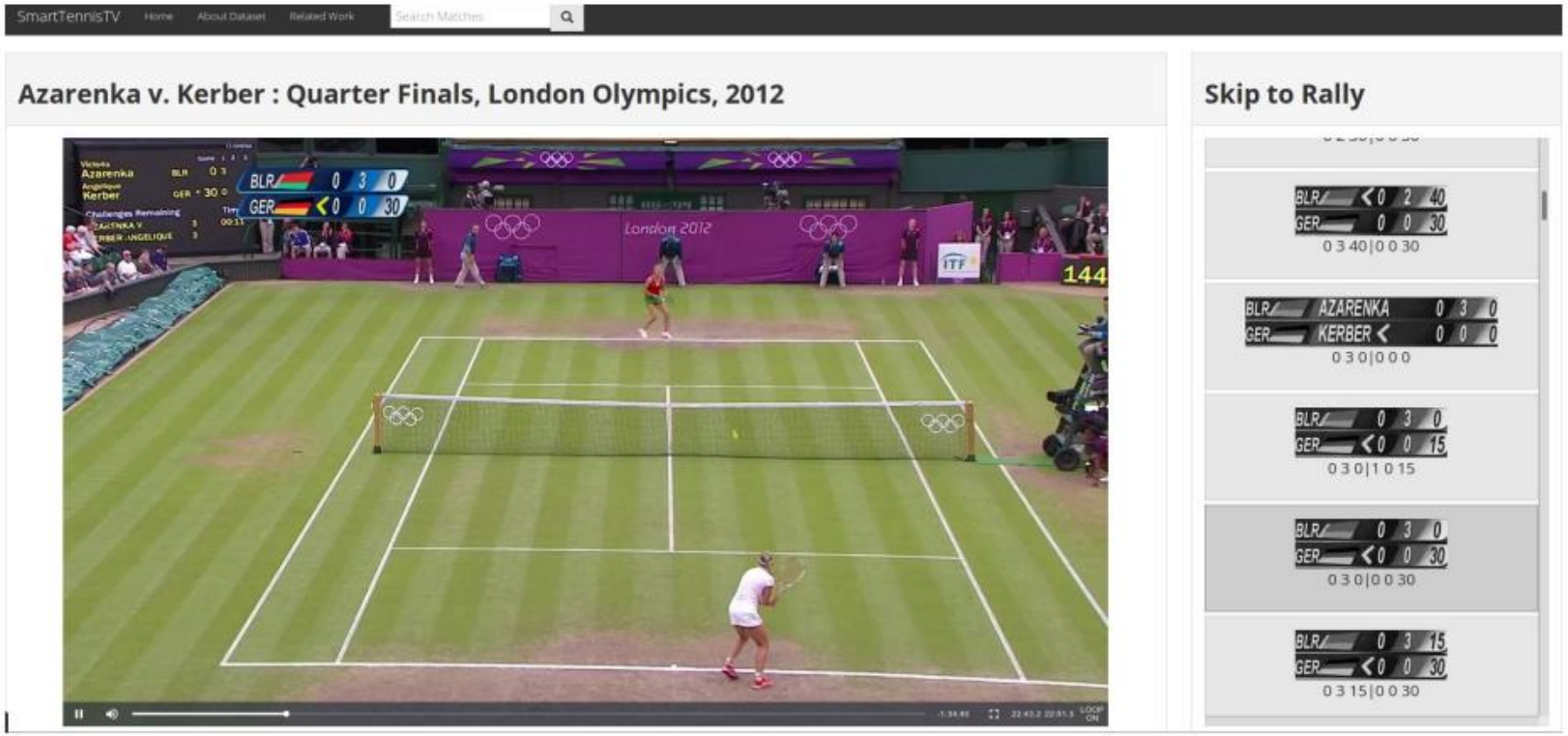
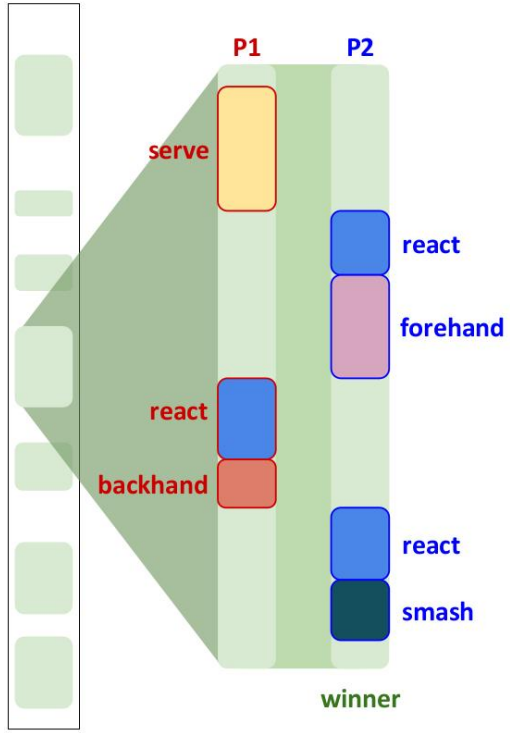


Table 3. Averaged Accuracy score of automatic event tagging (in percentage)

	Match 1		Match 2		Match 3		Match 4	
	Textspot	Ours	Textspot	Ours	Textspot	Ours	Textspot	Ours
Fault	66.66	70.83	52.24	56.71	87.87	90.90	84.44	84.44
Deuce	100.0	100.0	73.68	78.94	100.0	100.0	94.73	94.73
Advantage	100.0	100.0	77.77	77.77	100.0	100.0	95.65	95.65
Overall	75.00	79.41	60.58	64.43	92.45	94.33	89.65	89.65

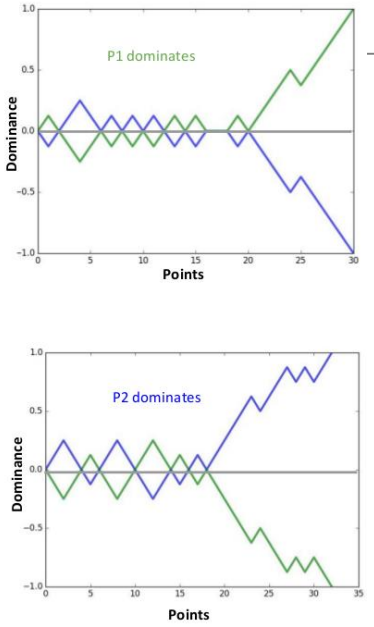
Match Video

Player bbox & tracks



Point Segments

Action Segments



Match and Point Statistics

What's next?

- Described modules are generic, can be used in other racket sports
  - Badminton, Table Tennis etc.
- But still, macro-understanding,
  - Enthusiasts happy!
  - Still need experts to do the analysis themselves.
- Moving toward automated analysis? Look out for our work in this direction. :)

Image Source: (Under Submission) Towards Structured analysis of broadcast badminton videos, Anurag Ghosh, Suriya Singh, C.V.Jawahar

What's next?

- Many interesting problems in this space :)
- Generating highlights automatically
- Coaching recommendation systems and tools
- Predicting player strategies?

Thank you!

Questions?

- If you have any other questions/clarifications please write to
  - anurag.ghosh@research.iiit.ac.in
- I'd love to hear new ideas and possible extensions! :)

Thanks!