Moving Sport Player Segmentation

Feng-Ming Chang, Ming Ma
Fall 2015, Cornell University, Electrical and Computer Engineering

ECE 5470

Abstract

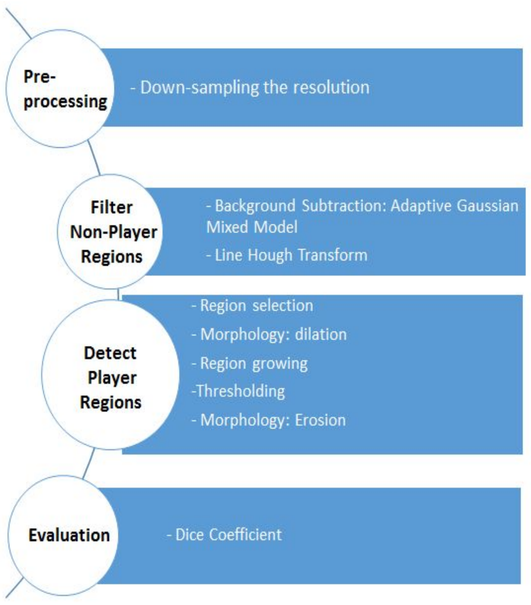
In this project, we propose a graph-based approach for detecting and segmenting the tennis player in temporal sequence of images. In the first stage, the potential position of the players in each frame is determined by filtering out the non-player regions. Based on the properties of the tennis court, straight line Hough transform is introduced in order to eliminate the background court line. The remaining pixels are then grouped using region size filter to identify probable player candidates. Experimented result from several sequence images of tennis matches indicates that the proposed approach is able to segment the player reasonably well even under slight change of camera direction and various illuminations.

Experimental Design

Data Set: The data set comes from the Amsterdam Library of Ordinary Videos for Tracking, ALOV300, which contains 89364 frames totally. It categorizes the data in several subsets based on tracking aspects such as motion smoothness, confusion, occlusion and shaking camera. Our selected set is motion smoothness version, which has slightly non-stationary camera scenes and various illuminations. The total number of frames we have tested is 300. We down-sampled the colorful images to a fixed resolution of 256x256 in that our system requires a constant input dimensionality. Since the given data set does not provide the ground truth of moving player segmentation, we manually marked the moving player of interest as our ground truth.

Image	Frame Size	Image Resolution	Characteristics
Set 1	5	256x256	Non-stationary camera
Set 2	5	256x256	Non-stationary camera
		⋮	
Set 7	5	256x256	Non-stationary camera

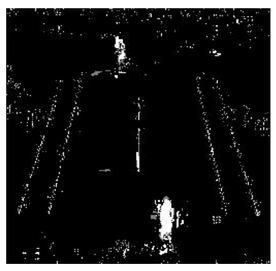
Experiment: First, we down-sampled the data set to 256x256 scale images. We first filter out the non-player region by two background subtraction programs. The naïve method is to compute the differences between consecutive image frames, as our baseline approach. The advanced approach is the combination of Gaussian Mixed Model and Line Hough Transform. After obtaining the binary image containing a few isolated regions, we applied the region filter, which contains based morphology, region-growing algorithm, and sort algorithm. For evaluation section, we mainly focused the indicator which supported by VisionX, e.g., Similarity, Sensitivity, Dice Similarity Coefficient.



Algorithm Verification

Gaussian Mixed Model Background Subtraction :

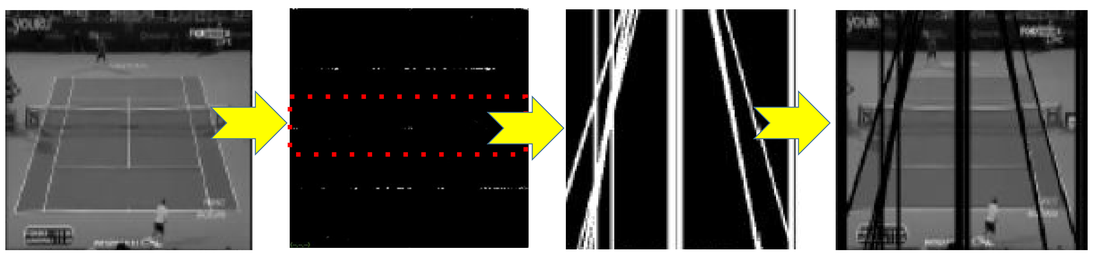
The background subtraction method that we employ essentially is to model each pixel as a mixture of Gaussians and using an on-line approximation to update the model. Then we evaluate the Gaussian distributions of the adaptive mixture model by setting threshold to determine which are most likely to result from a background process. We can update the training set by adding new samples and discarding the old ones to adapt to changes. Thus, a new object could belong to foreground or be removed from foreground. The mixed weights are updated by recursive equations each time and an on-line clustering algorithm is employed for background process.



background removal image

Straight Line Hough Transform :

The playground usually contains boundary lines to limit the tennis player moving area. Due to non-stationary camera, in some frames the boundary lines were regarded as part of foreground. In order to remove these lines, we employed Line Hough Transform. Here, we only filter the near vertical lines region due to the concern of player region separated by horizontal lines.



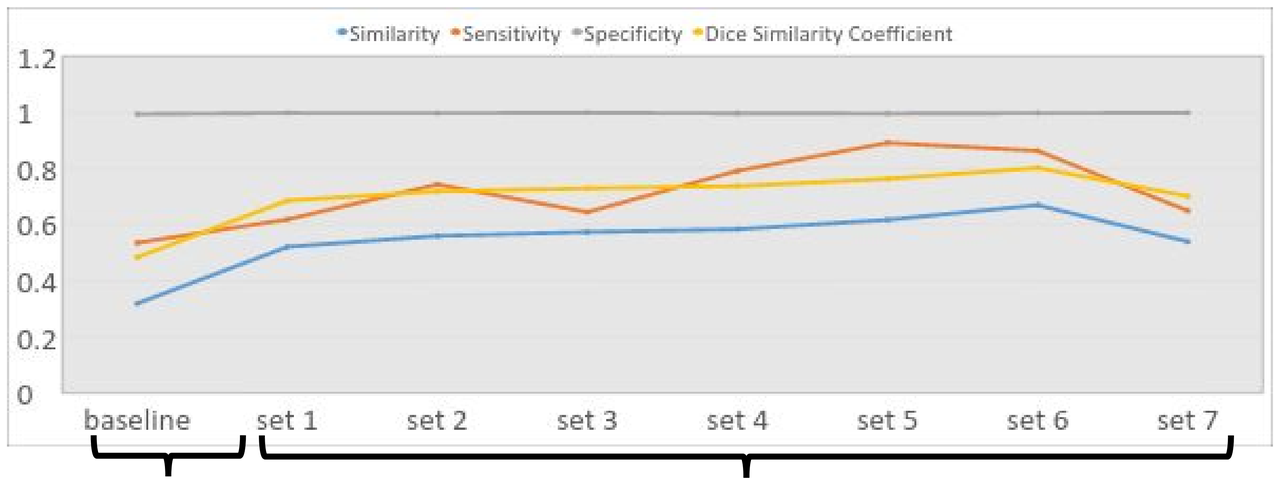
Original Image **Hough space** **reverse lines** **image without lines**

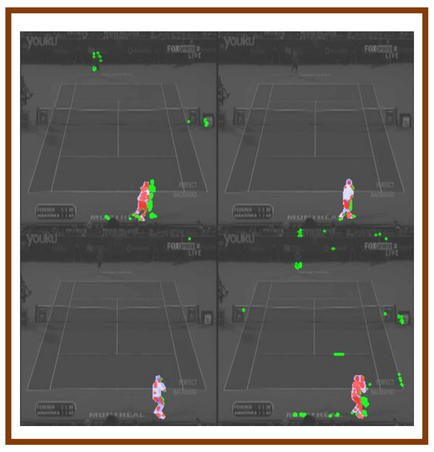
Algorithm Verification

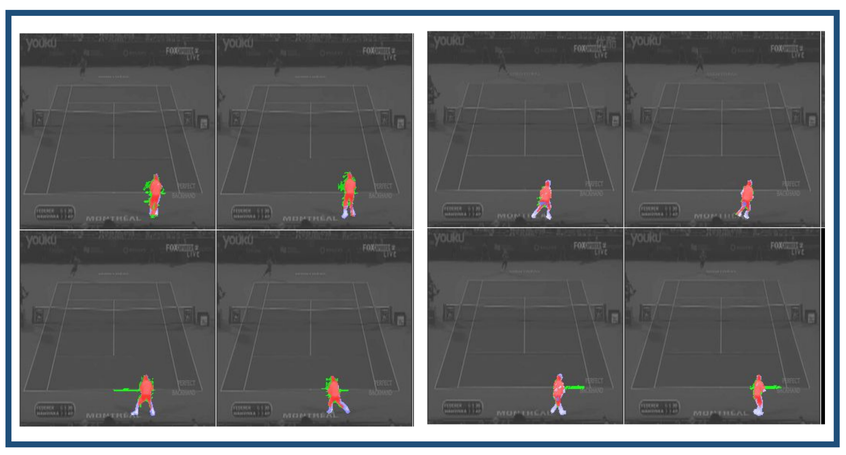
Region Size Filter: The input images of this filter is binary images of the original frame consists of a few isolated regions. Nevertheless, not all the regions can be regarded as player region, such as the region caused by background subtraction due to non-stationary scenes. Our region filter mainly utilizes morphology and region-growing algorithm. First, we only remain the area where the player mostly presents since we realize that such area is limited. Second, we apply region growing algorithm, of which the threshold to be used for segmentation is based on the area distribution of the segmented regions. This type of segmentation also helps remove too small segments, which may arise due to noise or other processing errors. A player region may sometimes be divided into multiple regions because of previous processes. Hence, before selecting the player moving area and after the segmentation, we apply dilation and erosion respectively in order to handle this accidental division of players.

Performance Testing

	Similarity	Sensitivity	Specificity	Dice Similarity Coefficient
baseline	0.319	0.535	0.993	0.483
set 1	0.521	0.617	0.998	0.685
set 2	0.560	0.743	0.997	0.718
set 3	0.573	0.644	0.998	0.729
set 4	0.583	0.791	0.996	0.736
set 5	0.616	0.891	0.995	0.762
set 6	0.669	0.862	0.997	0.801
set 7	0.539	0.648	0.998	0.701
Set Average	0.581	0.743	0.998	0.734
Improvement	**82.0 %**	**38.7 %**	**0.4 %**	**51.7 %**







Discussion: Our baseline approach for background subtraction failed to capture the behaviors when the players moved fast. It might resulted in the detected player with two bodies connected. Gaussian Mixed Model rarely has such problems, while it has other flaws– it failed to remove the whole background in some cases since this method is ideal for the stationary camera. Besides the player pixels, other pixels such as lines, audience in the back were detected as well. The main weakness of Line Hough Transform is that the detected lines might split the player regions. In addition, the region size filter is another powerful tool to remove the remaining noise pixel. Due to the properties of input image, the player has the largest area compare to other moving object. Hence, the region filter was applicable for both naïve method and the advanced method.

Conclusion and Future Work

- In conclusion, we proposed a procedure that can detect and segment the moving tennis player from a video. The experiment result supports that the proposed scheme might also be applicable to other sport videos such as badminton and Ping-Pong players.
- For the region filter part, we notice that some research papers that provide methods of removing the shadows. It is pity that we did not have time to separate the shadow of the player. If we were able to separate the shadow from the player, the overall performance would be better than current accuracy. Besides, in the future, we would like to modify the length of line for inverse Hough image based on the layout of tennis court in order to lower the possibility of player region separated by lines.