doi:10.20944/preprints201706.0013.v1

Line Extraction and Player Tracking in Tennis Videos

Volkan Erol

Computer Engineering Department, Okan University, Tuzla Campus, 34959, Istanbul, Turkey
volkan.erol@okan.edu.tr; Tel.: +90-533-3621947

Abstract: Sports video analysis is an interesting and emerging field in the domain of computer vision. It has become a very popular subject among scholars all over the world in the last decade due to the need of automated classification of events such as offsides, in/out situations, goals and so on. The most important problem in sports video analysis is detection and tracking of ball and players. Extraction of court lines is also another important problem since it is necessary in the classification of aforementioned events. In this study, considering some of the previous work on the subject, we develop a method for tracking the ball and players in tennis videos where the camera position and orientation is supposed fixed.

Keywords: line extraction; morphological image processing; player tracking; sport videos

1. Introduction

Assuming a fixed camera (intrinsic and extrinsic parameters are constant in all frames), our method presented here consists of four major stages:

- The *background extraction* stage is the process of finding the background frame of a given video. The background frame should include all moving objects and no other. So in an ideal background frame we should have the ball and the players. In this stage we use the *median* operator to compute the average of each pixel in the video. Since the players and the ball are moving in most of the frames, the median operator should give us a satisfying result.

- The *line extraction* stage deals with the background image obtained in the previous stage and tries to extract the lines out of it. Although previous work on the subject relies heavily on *Hough transform* in line extraction [4], morphological operations and thresholding are used because of their speed advantage over Hough transform [5].

- The third step is *player segmentation*. For each frame, the background image is extracted from the frame, then a series of morphological operations are applied to the resulting image. This yields a picture with only the player, and some noise. The noise is eliminated by simple thresholding and the resulting image is binarized [5].

- The last step performs *player tracking* on the binarized image. We use a *bounding box* to locate the player and the information from the previous frames is used to compute the current position of the player. Some constraints on the size and position are applied to the bounding box.

Although this subject is getting more popular over time, there hasn't been any attempts in building a test database of tennis videos. So, to test our method we have come up with 2 videos which contain difficult scenes: different illumination in different parts of the field, different intensity in line regions, player shadows, moving people (like ball boys), etc.

2. Background Extraction

Background extraction is the first logical step in tracking of moving objects. The aim of background extraction is to come up with the static objects in the scene so that after substracting the background image from each frame, one can find the moving objects in the frame.

[1] discusses the efficiency of different background extraction techniques. To extract a background from a series of frames, the following is done: for each pixel in the background image, a *median* filter is formed from every *n*-th frame in the sequence from the pixels corresponding to the pixel in the background image. This can be thought as a median filter over the time dimension. This technique yields good results if the moving objects in the scene don't stay at the same position for a long time. Since, in tennis videos, during a service, the players don't change their positions too much, this results in

noise in the background image, but is still acceptable. The results of background extraction with this technique are demonstrated in fig. 1. Notice the noise near serving positions where players spend most of their initial time during a point.

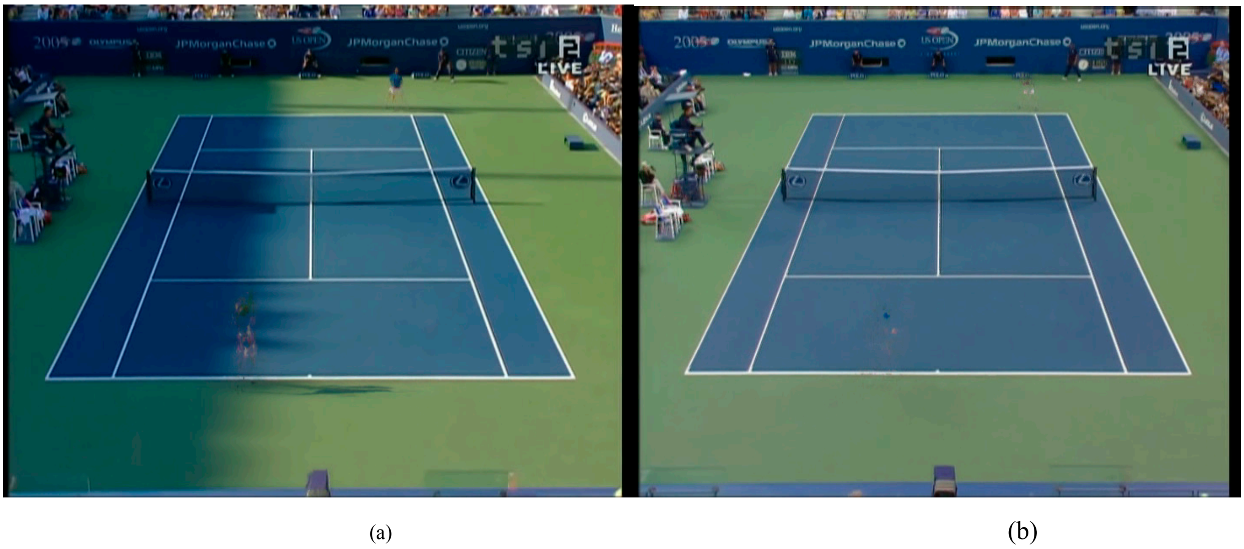


fig 1. Background extraction from: (a) video1 (190 frames), (b) video2 (330 frames)

3. Line Extraction

Before explaining how the line extraction phase is accomplished, we need to explain some of the basic morphological tools that we used in the process.

3.1 – Mathematical Morphology and Basic Morphological Operations

The *mathematical morphology theory* (MM), is a powerful image analysis framework based on geometry, nowadays fully developed for both binary and greyscale images. Its popularity in the image processing community is due mainly to its rigorous mathematical foundation as well as inherent ability to exploit the spatial relationships of pixels. The morphological framework provides a rich set of tools able to perform from the simplest to the most demanding tasks: *noise reduction*, *edge detection*, *segmentation*, *texture and shape analysis*, etc. As a methodology, it has been applied to almost all domains dealing with digital image processing. Consequently, it was only a matter of time before attempting to extend the same concepts to colour and more generally multi-spectral images [6-10].

doi:10.20944/preprints201706.0013.v1

Let *f: E* → *T* be a greyscale image, with *E* the discrete coordinate grid while *T* represents the set of possible grey values. In the present case, a multi-spectral image of n channels is considered as **f**: $\mathbb{Z}^2 \rightarrow \mathbb{Z}^n$ with $f_i : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ denoting its i$^{\text{th}}$ channel.

In shortly, MM studies the transformations of an image, when it interacts through operators with a *matching pattern* B, called *structuring element* (SE). In defining the two basic operations, *dilation* and *erosion* we will consider B as a subset of E. Let's define these two operations respectively:

$$\delta_B(f)(x,y) = \bigvee_{(r,s)\in B} f(x-r, y-s), \quad (x,y)\in E \quad (1)$$

$$\varepsilon_B(f)(x,y) = \bigwedge_{(r,s)\in B} f(x+r, y+s), \quad (x,y)\in E \quad (2)$$

A multitude of operators is then derived from dilation and erosion; such as *opening* and *closing*, used extensively with smoothing ends and defined respectively as:

$$\gamma_B(f) = \delta_B(\varepsilon_B(f)) \quad (3)$$

$$\phi_B(f) = \varepsilon_B(\delta_B(f)) \quad (4)$$

In general we use the opening and closing transforms in order to isolate bright (opening) and dark (closing) structures in images, where bright/dark means brighter/darker than the surrounding features in the images.

3.2– The Top-Hat operator

The "*Top hat*" operator results from the theory of mathematical morphology and allows peaks of intensity to be extracted from an image.

This filter can be used by supposing that the required lineaments are brighter than the environment. The main advantage of this filter is to be able to detect an over brightness even if the environment is not uniform. Moreover it is possible to regulate the size or the width of the over brightness very easily. The principle is based on the subtraction of an image from its ``opening". The opening consists of an erosion followed by a dilation, the size of the structuring element being conditioned by the width of the lineament to be detected. So, the extraction of lines is actually reduced to the careful choice of this structuring element.

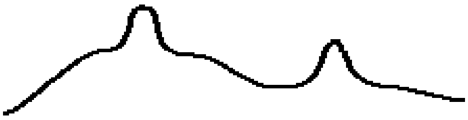
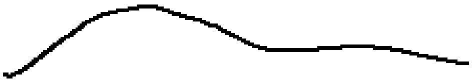
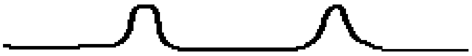
Original signal	
Opening applied to signal	
Subtraction (Original image - Opening)	

fig 2. Top hat operatör

3.3– Line Extraction Method

Given the background image, we begin by transforming the image from *RBG* color space to *HSV* color space. The aim here is to separate the *luminance* of the *colors* present in the frame from their color. Since the correlation between pixels in the RGB space is too high we need another space to work with. The *V* channel of HSV space is such a space. We then apply a *top-hat operator* with a *structuring element* chosen as disk on the background image. This brightens up the lines in the image and can be thought as an enhancing stage. Denote this image as *E*. *Weighted mean thresholding* is applied to E, we obtain a binary image *B*, which will be used later as a mask.

The next step is to extract *horizontal lines*. With 20x20 horizontal structuring element we apply *gray-opening-by-reconstruction* (*GOBR*) to E. Then we do an *opening* with a structuring element of 150 horizontal pixels and one with a structuring element of

3 vertical pixels. What results is the horizontal lines longer than 150 pixels and 3 pixels wide.

Vertical lines are detected similarly: we apply GOBR to E, then we do an opening with a structuring element of 75 vertical pixels and one with a structuring element of 3 horizontal pixels.

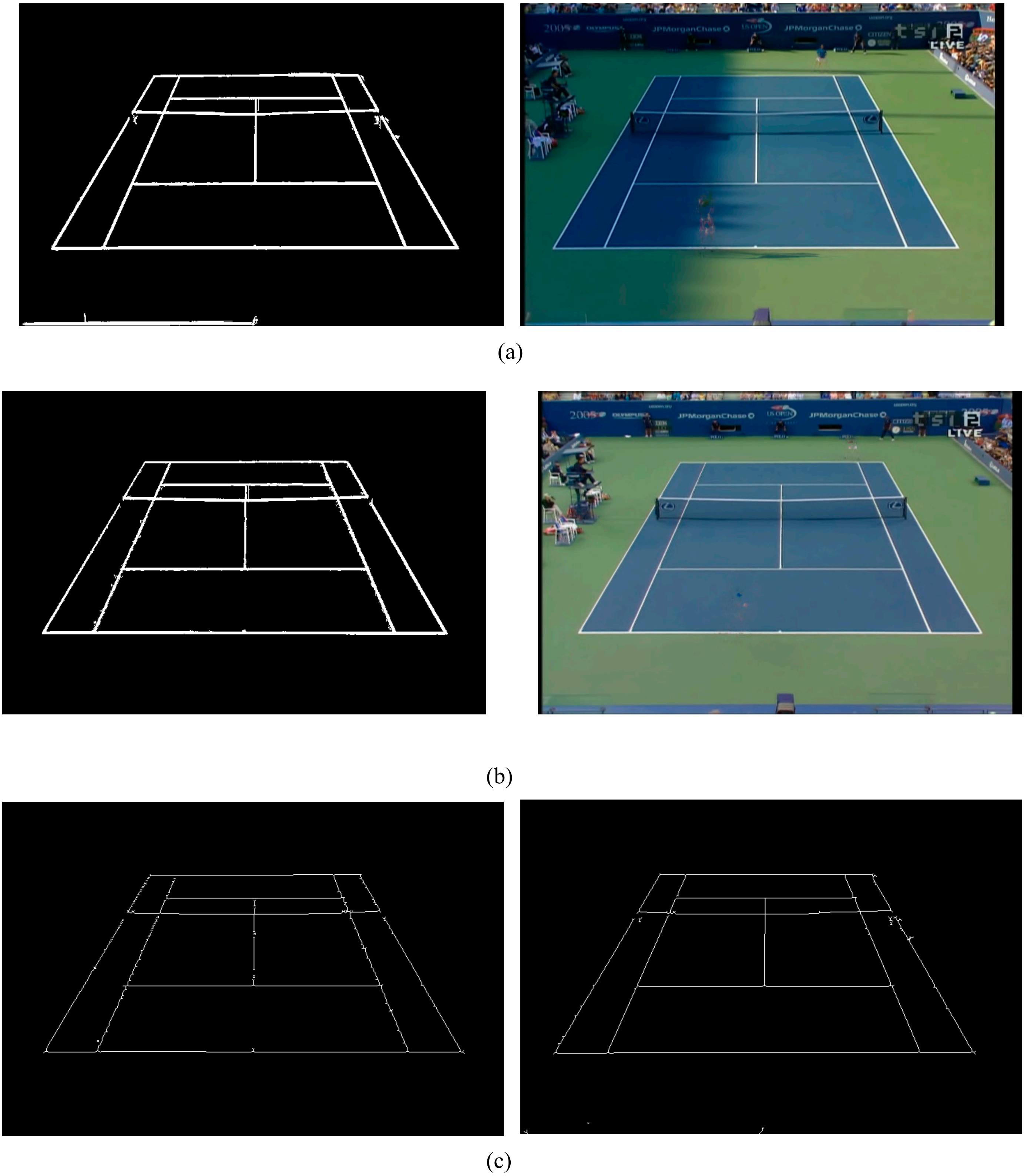
Vertical and horizontal lines are summed in an image called *L*. To L we do the following:

To get rid of unwanted lines, we apply another weighted thresholding, then by using closing operator we connect the unconnected line segments. By using B we apply a *binary reconstruction by dilation* and then do 2 openings as a 'cleaning' step: One with 3x3 vertical se and one with 3x3 horizontal se.

So to sum up these steps:

1. Transform background image to HSV and take the V channel. Call it V. Apply top hat with 7x7 disk se to V. Call it E
2. E thresholded. Call it B.
3. Extract horizontal lines: GOBRH (20x20) to E. Opening (150 horizontal 3 vertical)
4. Extract vertical lines GOBRV (20x20) to E. Opening (75 vertical 3 horizontal)
5. Compute L from 3 and 4.
6. Apply binary reconstruction by dilation to L (use B as a mask).
7. Clean the image with 3x3 horizontal and 3x3 vertical openings.

The resulting binary image is the lines extracted from the image, just in case unwanted lines are present we apply connected component labeling and keep only the most connected component. Then we apply a thinning to get a smoother result. The results are shown in fig.3.



(a)

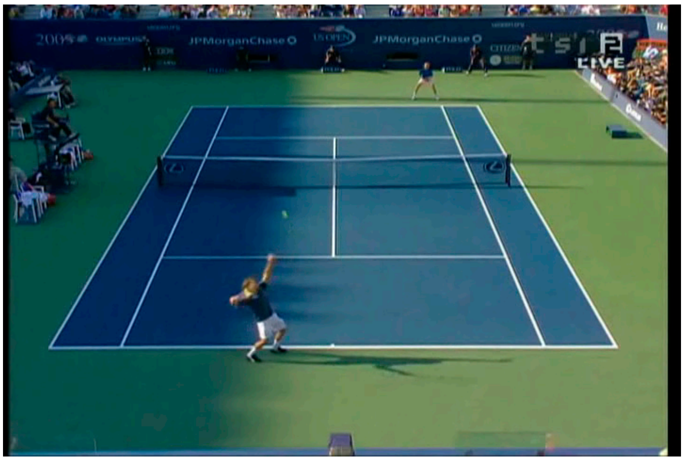
(b)

(c)

Fig 3 . Line detection results from background1 (a) and background2 (b) and edge thinning applied to these images

4. Player Segmentation and Tracking

Player segmentation stage starts with substracting the *background frame* from the *current frame* and obtain a *foreground frame* (Fg). Then an opening operator with a 3x9 structuring element is applied to Fg. The result is thresholded (mean of image is calculated and pixels brighter than the mean are eliminated). The result for a frame is given in fig. 4(a,b,c,d).



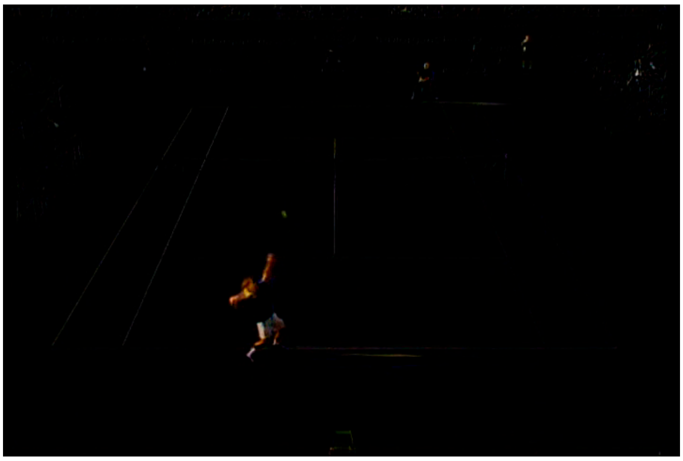
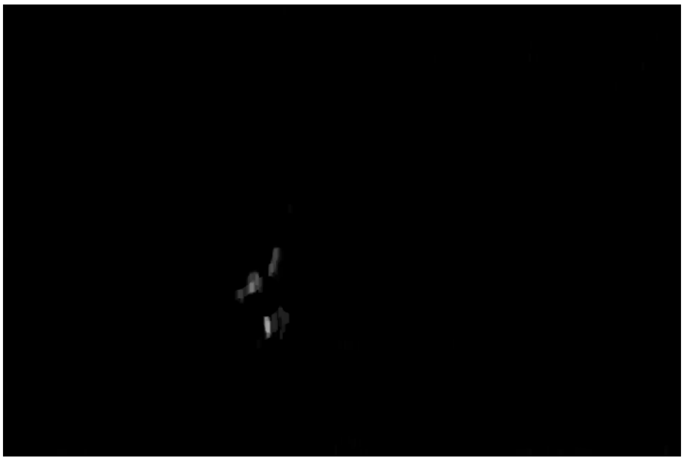
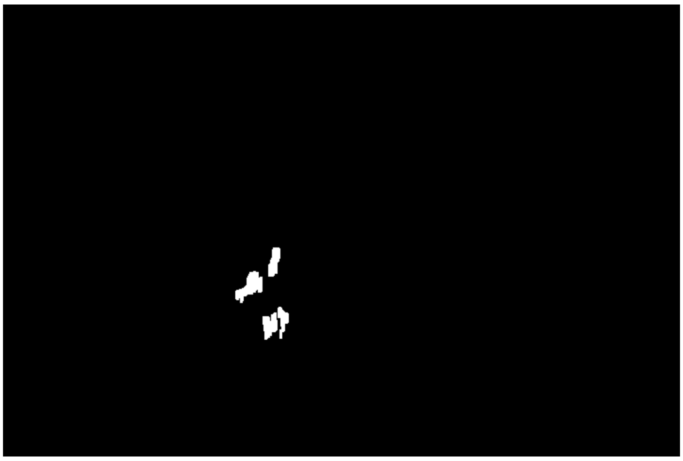


Fig 4(a) ***Fig 4(b)***

Fig 4(c) ***Fig 4(d)***





doi:10.20944/preprints201706.0013.v1

Fig 4(a) is the original frame. Fig 4(b) is the foreground frame (Fig 4(a) – background). Fig 4(c) is the result of opening applied to Fig 4(b). Fig 4(d) is computed by applying weighted mean thresholding to Fig 4(c).

The tracking algorithm adaptively calculates the coordinates of the bounding box encapsulating the player. Shortly, for each frame, the algorithm calculates local *minx, miny, maxx and maxy*. Some assumptions and constraints are applied to the bounding box: The nearer player cannot be in the upper half of the frame, and if the newly calculated coordinate values varies inexpectedly the box is drawn according to the values of the previous frame. The result of applying the tracking algorithm to the frame in Fig 4(a) is shown in Fig 5.

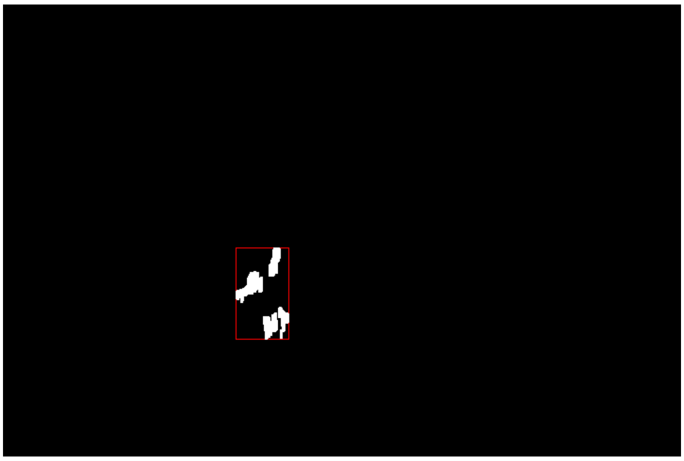


Fig 5. Tracking result for Fig 4(a).

5. Conclusion

We have presented a scheme for detecting players and lines in tennis videos. The method uses common image processing tools such as the median filter, thresholding, connected component labeling, thinning and morphological operations such as opening, closing, top-hat and reconstruction by dilation.

The method has been tested on 2 real life tennis videos and yields satisfying results. The subject is still an open issue in the domain of computer vision and future work can be done by including ball detection and event classification such as stroke recognition.

References

[1] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, "Wallflower: Principles and Practice of Background Maintenance", *Microsoft Research*.

[2] Y. Seo, S. Choi, H. Kim, K. S. Hong, "Where Are the Ball and Players? Soccer Game Analysis with Color-Based Tracking and Image Mosaick", *Int. Conf. on Image Analysis and Processing,* Florence, Italy, pp. 196-203, Sept. 1997.

[3] Y. Ohno, J. Miura, Y. Shirai, "Tracking Players and a Ball in Soccer Games", *Proceeding of the 1999 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Taipei, Taiwan, R.O.C, August 1999.

[4] K. Matsui, M. Iwase, M. Agata, T. Tanaka, N. Ohnishi, "Soccer Image Sequence Computed by a Virtual Camera", Department of Information Engineering, Nagoya University.

[5] T. Bloom, A. P. Bradley, "Player Tracking and Stroke Recognition in Tennis Video", School of Electrical Engineering and Telecommunications, Sydney, Australia, Centre for Sensor Signal and Information, The University of Queensland, St. Lucia, Australia.

[6] E. Aptoula, S. Lefèvre. "A comparative study on multivariate mathematical morphology." Pattern Recognition 40.11 (2007): 2914-2929.

[7] E. Aptoula, "Extending morphological covariance." Pattern Recognition 45.12 (2012): 4524-4535.

[8] E. Aptoula, "Comparative study of moment based parameterization for morphological texture description." Journal of Visual Communication and Image Representation 23.8 (2012): 1213-1224.

[9] S. Lefèvre et al. "Morphological template matching in color images." Advances in Low-Level Color Image Processing. Springer Netherlands, 2014. 241-277.

[10] E. Aptoula, "The impact of multivariate quasi-flat zones on the morphological description of hyperspectral images." International Journal of Remote Sensing 35.10 (2014): 3482-3498.