# VIBOT MSc. Thesis
# Acquisition of 3D tennis ball trajectory and velocity from monocular vision systems


Vujović Dragutin

Heriot-Watt University
Universitat de Girona
Université de Bourgogne

Media Technology
Aalborg University

Supervisor: Rikke Gade, PhD.
Co-supervisor: Thomas B. Moeslund, PhD.



A Thesis Submitted for the Degree of
MSc Erasmus Mundus in Vision and Robotics (VIBOT)

· 2016 ·

**Abstract**

Video annotations are successfully used in many sport related areas, including training, the media coverage and sport science. The key advantages of automatic annotation over conventional manual are speed, and robustness of gathered data. However, current automatic annotations systems are not always practical due installation cost and media rights. The proposed solution provides this kind of data from broadcast videos and single camera shots.

The main contribution of this work is acquisition of 3D trajectory and velocity of tennis ball, including players' positions. The 3D trajectory acquisition is based on ball detection, field lines recognition and prior information about field dimension. The accuracy is tested on all four Grand Slam tournament matches, by available ground-truth data.

*Research is what I'm doing when I don't know what I'm doing. . . .*

Werner von Braun

# Contents

# List of Figures

# List of Tables

# Acknowledgments

I would like to express deep gratitude to many people who provide support, time, and effort to make this thesis possible. First of all, to VIBOT staff for unselfish and exceptional support and understanding during whole master course. Secondly, to Aalborg university for perfect conditions, expertise and support which transformed simple idea into this master thesis. Last but not least, I would like to express gratitude to friends and family for their unique support.

# Chapter 1

# Introduction

Consider ball tracking in tennis, which is still manually done or skipped in cases of insufficient budget. The main advantage of ball tracking is that it provides key information necessary for tactics analysis, training improvement, personal progress tracking. The goal of every tennis player is to advance and at the beginning to compare personal performances with top players. The system that can support this needs to have 'tennis' knowledge, and one of the best way to gain this knowledge is through video analysis.

Let's consider, tennis trainer who wants to improve his player game, and compare it with more advanced players. Significant technique that video analysis can provide is to compare positions and shots of novice player with more advanced players. The first issue in this case is lack of available advanced video statistics related to professional players. Second significant problem is that to provide this statistics for own player, its is necessary to setup expensive environment with multiple high speed cameras.

The lack of advanced video statistics can be solved using the following method: instead of using multi camera environment, it is possible to use single camera to extract key information: players position and 3D ball trajectory. The solution will be explained further in this document. The same solution can be applied on archived tennis video matches, in order to extract advanced statistics about professional players.

This paper will present solution for automatic annotation of broadcast videos. The main contribution of this paper are:

- Automatic tennis players recognition and tracking based on initial labeling (Section ?)

using monocular camera.

- Automatic generation of tennis ball 3D trajectory and velocity (Section ?) using monocular camera. We prove that extracted 3D ball trajectory has more than 90% accuracy comparing with manually extracted ground truth for Grand Slam matches.

# Chapter 2

# Problem definition

# Chapter 3

# State of the art

In [7] the focus is on players tracking. The blobs, after background subtraction, are grouped in a connected graph. Graph nodes contain spatial information (size, color, shape) about blobs, and edges contain distance between blobs, movement of the player in consecutive frames. The player trajectory is found using minimal path searching.

The [9] proposed a new method for generation of background scenes, which enables to remove foreground objects without human assistance. The background scene is created using shots of completely visible playing field. Difference between generated background scene and current frame enables to segment foreign objects. The segmentation of players is enhances using automatic alpha matting technique.

The [10] uses Kalman filter and Minimum Spanning Tree based clustering to improve performances of field lines detection. The tracking objects are represented in real-world frame, after computing homography between image and world coordinate system.

In [5] Kalman filter parameters are adapted to improve tracking of the player further from the camera, where the tracking is corrupted by noise. The uncertainties of the state model and the measurement model are depended on the area of the detected player. When the size of the player is smaller then the predicted size, Kalman filter trusts more prediction then the measurement.

The [6] presented automatic calibration algorithm for court sports. The court lines are found by Hough transform, and ordered on horizontal and vertical. The crossing between two horizontal and two vertical lines are used to find the homography between input court model and reference model. The homography is measured by finding pixel matching between input court model, and projected reference model.

The [8] proposed few improvements for homography calculation and players tracking in basketball videos. The preprocessing step eliminates non-court shots, and generate field mask based on dominant color. The correct homography is selected based on fitting score from different candidates. The players are tracking using adaptive mean shift algorithm (CamShift).

In [2] three major steps are performed. The first is to determine homography, using the field white lines and additionally eliminating outliers that don't fit to the court model. The second step is to construct the background model of the court, using dominant colors and Gaussian mixture model. The last step is to classify foreground objects in two teams and to perform their tracking using Kalman filter.

The [12] did ball data association by generating tracklets. The tracklet is the structure that can originate from the ball or the clutter. The tracklets are connected in directed graph, where edge weight between two nodees are calculted as a compatibility of two tracklets. The ball trajectory is obtained by calculating shortest path between nodes, and then smoothed using Kalman filter.

The [11] used Circular Hough transform to detect ball. Additionally all shots are classified as close, medium or long based on dominant color. Depending on the shot type, circular objects within certain radius are kept as ball candidates. The ball candidates from consecutive frames are connected in acyclic graph, and longest path in graph is considered as the ball trajectory.

The [14] improved precision of camera calibration by grouping frames with the same lookats and the focal length. This method decrease the effect of image noises and neglect frames that are not in any group.

The [13] explored audio clues to detect key tennis events: serve, player hitting the ball, bounce, ball hitting the net. The audio Mel-frequency cepstral coefficients are classified using maximum likelihood. The audio clues are combined with the ball tracking ( [12]) in order to perform automatic annotation.

| Ref | Task | Image Stream | Feature extracted | Method |
|---|---|---|---|---|
| [2] | Volleyball ball tracking | 352x240 MPEG-1 | Nearest neighbor to the estimation in the next frame | |
| [7] | Football players tracking | 720x480 MPEG-2 | Blobs from background subtraction | Graph of components |
| [9] | Automatic annotation | 720x480 Broadcast videos | Score box, Ball hit sound | Kalman filter |
| [10] | Football's tactic pattern recognition | 704x576 MPEG-2 | Field lines, Ball | Global Motion Estimation, Kalman Filter, MST Clustering |
| [5] | Player tracking | 720x480 MPEG-2 | Blobs from background subtraction | Adaptive Kalman Filter |
| [6] | Automatic camera calibration | 720x480 Broadcast videos | Field lines | Hough transform |
| [8] | Camera calibration and basketball players tracking | 720x480 MPEG-2 | Field lines, Dominant field color | RANSAC, CamShift tracking |
| [3] | Basketball players tracking | 640x352 Broadcast videos | Field lines, Blobs from background subtraction | Hough transform, Kalman filter, K-Means segmentation |
| [12] | Tennis ball tracking | Broadcast videos | Blobs from background subtraction | Kalman filter |
| [1] | Camera calibration | 1920x1080 Images | Field lines | Hough transform, |
| [11] | Ball trajectory | Broadcast videos | Dominant color, Blobs from background subtraction | Sobel edge detector, Circular Hough transform |
| [14] | Camera calibration, ball detection and tracking | 704x576 MPEG-2 | Field lines | Hough like search for homography transform, |
| [13] | Automatic annotation | Broadcast videos | Audio (Mel-frequency cepstral coefficients) | Maximum likelihood |
| [4] | Valleyball 3D ball trajectory | 352x240 MPEG-1 | Field lines, Blobs from background subtraction | Hough transform |

Table 3.1: A review of works related to camera calibration, ball and player tracking from single camera

# Chapter 4

# Methodology

## 4.1 Methods

## 4.2 Data Collection Methods

## 4.3 Analysis

# Chapter 5

# Results

# Chapter 6

# Conclusions

# Appendix A

# The first appendix

If you need to add any appendix, do it here... Etc.

# Bibliography


[1] M Alemán-Flores, L Alvarez, L Gomez, P Henriquez, and L Mazorra. Camera calibration in sport event scenarios. *Pattern Recognition*, 47(1):89–95, January 2014.

[2] Hua-Tsung Chen, Hsuan-Shen Chen, and Suh-Yin Lee. Physics-Based Ball Tracking in Volleyball Videos with its Applications to Set Type Recognition and Action Detection. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 1:I–1097–I–1100, 2007.

[3] Hua-Tsung Chen, Chien-Li Chou, Tsung-Sheng Fu, Suh-Yin Lee, and Bao-Shuh P Lin. Recognizing tactic patterns in broadcast basketball video using player trajectory. *Journal of Visual Communication and Image Representation*, 23(6):932–947, August 2012.

[4] Hua-Tsung Chen, Chien-Li Chou, Wen-Jin Tsai, and Suh-Yin Lee. 3D ball trajectory reconstruction from single-camera sports video for free viewpoint virtual replay. In *Visual Communications and Image Processing (VCIP), 2011 IEEE*, pages 1–4. IEEE, 2011.

[5] M Y Fang, C K Chang, N C Yang, and C M Kuo. Robust player tracking for broadcast tennis videos with adaptive Kalman filtering. *source: Journal of . . .*, 2014.

[6] Dirk Farin, Susanne Krabbe, Peter H N de With, and Wolfgang Effelsberg. Robust camera calibration for sport videos using court models. *Storage and Retrieval Methods and Applications for Multimedia*, 5307:80–91, 2004.

[7] P J Figueroa, N J Leite, and RML Barros. Tracking soccer players aiming their kinematical motion analysis. *Computer Vision and Image Understanding*, 101(2):122–135, 2006.

[8] Min-Chun Hu, Ming-Hsiu Chang, Ja-Ling Wu, and Lin Chi. Robust Camera Calibration and Player Tracking in Broadcast Basketball Video. *IEEE Transactions on Multimedia*, 13(2):266–279, March 2011.

[9] Jui-Hsin Lai, Chieh-Li Chen, Chieh-Chi Kao, and Shao-Yi Chien. Tennis Video 2.0: A new presentation of sports videos with content separation and rendering. *Journal of Visual Communication and Image Representation*, 22(3):271–283, April 2011.

[10] Zhenxing Niu, Xinbo Gao, and Qi Tian. Tactic analysis based on real-world ball trajectory in soccer video. *Pattern Recognition*, 45(5):1937–1947, May 2012.

[11] V Pallavi, J Mukherjee, A K Majumdar, and S Sural. Ball detection from broadcast soccer videos using static and dynamic features. *Journal of Visual Communication and Image Representation*, 19(7):426–436, 2008.

[12] Fei Yan, William Christmas, and Josef Kittler. Ball Tracking for Tennis Video Annotation. In *Computer Vision in Sports*, pages 25–45. Springer International Publishing, Cham, January 2015.

[13] Fei Yan, Josef Kittler, David Windridge, William Christmas, Krystian Mikolajczyk, Stephen Cox, and Qiang Huang. Automatic annotation of tennis games: An integration of audio, vision, and learning. *IMAVIS*, 32(11):896–903, November 2014.

[14] Xinguo Yu, Nianjuan Jiang, Loong-Fah Cheong, Hon Wai Leong, and Xin Yan. Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking. *Computer Vision and Image Understanding*, 113(5):643–652, May 2009.