Sport video analysis

Mohammadmostafa Talebi-Hamid Mohammadi

Computer Vision Course

Professor Stefano Ghidoni

Date: September 24, 2023

Contents

1 Introduction

In sports analysis, the fusion of technology and athletics has given rise to innovative solutions that delve deep into gameplay dynamics. One such technological marvel is computer vision, a field within artificial intelligence that has revolutionized how we understand and dissect sports events. By harnessing the power of computer vision systems, we can identify and track players on sports teams, extracting a wealth of semantic information about the game's nuances, from player movements to the precise boundaries of the playing field.

Sport video analytics, driven by computer vision systems, seeks to achieve several fundamental objectives. Firstly, it aims to localize and track individual players on the field of play, employing bounding boxes to precisely demarcate their positions. This not only aids in player tracking but also lays the foundation for understanding their movement patterns and interactions within the game. Secondly, the system endeavors to segment the playing field and players from the extraneous visual clutter in the image. This segmentation is crucial for isolating the critical elements of the game. Lastly, the system aspires to classify each detected player according to their respective team, providing valuable insights into team dynamics and strategies.

The crux of this project lies in developing a sophisticated computer vision system tailored explicitly for sport video analysis. This system is engineered to analyze image footage from various sports events, ranging from soccer to basketball and beyond. Its primary objectives are to localize and delineate players within the frame, to segment the playing field and players from extraneous elements, and to classify each detected player according to their respective team.

The system must demonstrate adaptability and robustness across various sports scenarios to meet these goals. It must effortlessly discern players within the playing field, even when a player's attire differs from their teammates, such as in the case of a goalkeeper. Moreover, it should intelligently disregard non-player entities like referees and spectators outside the playing field. Additionally, the system must be adept at segmenting different types of playing fields and accommodating variations in geometries and colors intrinsic to diverse sports.

The core objectives of the system can be summarized as follows:

1. Player Localization: Accurately localizing all players within the image, encapsulating them within bounding boxes, and ensuring they are within the playing field.

2. Segmentation: Segmentation of both the playing field and players from the surrounding elements in the image, allowing for precise and distinct recognition.

3. Team Classification: Classifying each detected player according to their respective team, enabling the differentiation between Team A and Team B players.

2 Overview

This project aims to deliver a versatile and robust system that provides critical insights into various sports scenarios. Since we wanted to challenge our computer vision competences, we implemented this project combining our innovative solutions with existing computer vision methods. It could be feassible to implement whole or part of the project using ML/DL techniques, but as you suggested, we decided to use pure computer vision techniques. We analyze visual information by detecting players and the field and distinguishing between teams. To achieve this, the software follows these steps.

As the first step in our algorithmic approach, we commence image processing by employing the HSV (Hue, Saturation, Value) color space, strategically employed to discern the region of interest within each image. To achieve this, we define distinct HSV ranges tailored to accurately differentiate among the various elements comprising the scene, including the players, the playing field, and the background. This segmentation process creates individual masks, each thoughtfully assigned a specific color designation: the color red signifies team A, the color blue represents team B, and the color green is reserved to symbolize the playing field. This color coding prepares the image for subsequent analysis.

Our subsequent step is determining the precise location of the players within the image. Achieving this goal demands a more sophisticated approach, primarily due to the nature of the images we have. When we employ the Canny Edge detector, it generates multiple contours within the image, introducing complexity into the selection process. To address this challenge, we employ a discerning strategy: specifically, we identify the player by isolating the largest contour among the many present. This choice effectively designates the player, and subsequently, we proceed to establish their precise location by drawing a bounding box around them. This method ensures the accurate localization of the player within the image.

Finally, we turn our attention to the bounding box and checking the predominant color contained within it. In cases where the dominant color is red, we designate the largest contour as representing team A; if it appears as blue, the most significant contour assumes the role of team B; and when the color is identified as green, we classify it as the field. Any elements not falling into these specified categories are uniformly assigned the color black. This color-coding process prepares the image for evaluation and interpretation.

Finally, the segmentation of the given images and the bounding boxes are compared to the ground truth for evaluation purposes. Through this systematic approach, it is possible to develop a computer vision system capable of recognizing different players of the teams and generating semantic information regarding each scenario. This technology can significantly improve sports analysis using computer vision systems.

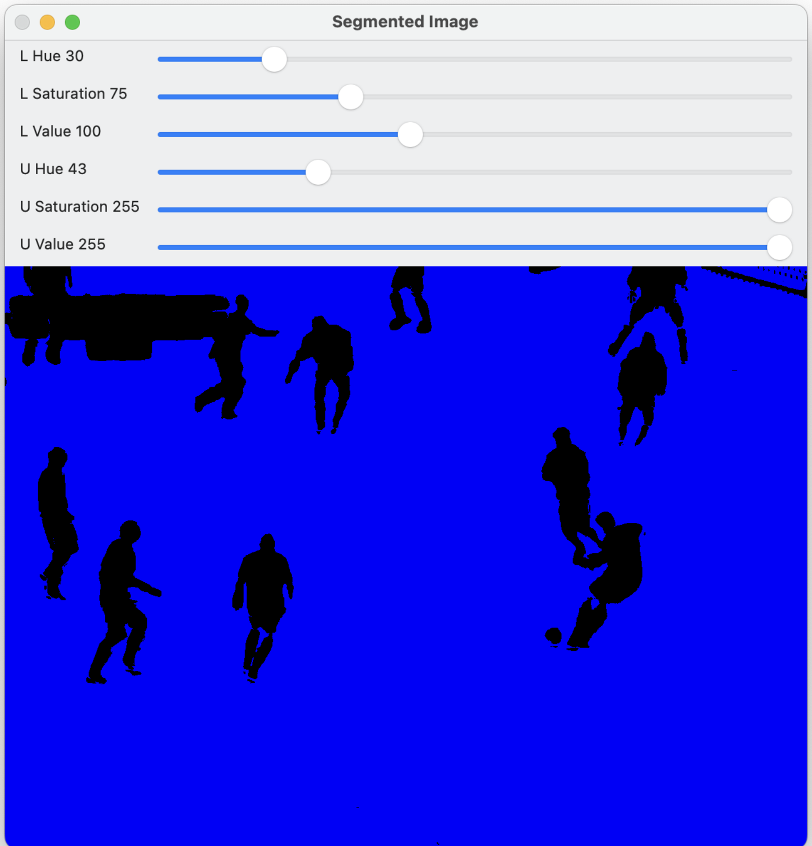


Figure 1: Determining HSV values

3 Methodology

This section outlines the methodology we employed to develop our computer vision system for sport video analysis. As mentioned before, we developed this software purely in computer vision techniques and we did not use any ML/DL solutions. Our approach was divided into three main steps, each explained in detail in the following subsections.

3.1 Segmentation

The segmentation phase in our project has a significant but challenging role. This step entailed the detection and isolation of players in the images, a prerequisite for the subsequent analysis. Initially, we explored various segmentation methods and algorithms. We looked into the watershed algorithm, a widely used technique for image segmentation. However, a major limitation of this algorithm is its dependence on markers for optimal performance, and identifying these markers without human involvement proved impractical. This is not the ideal choice for our project since one of our primary objectives was to reduce or eliminate the need for human intervention.

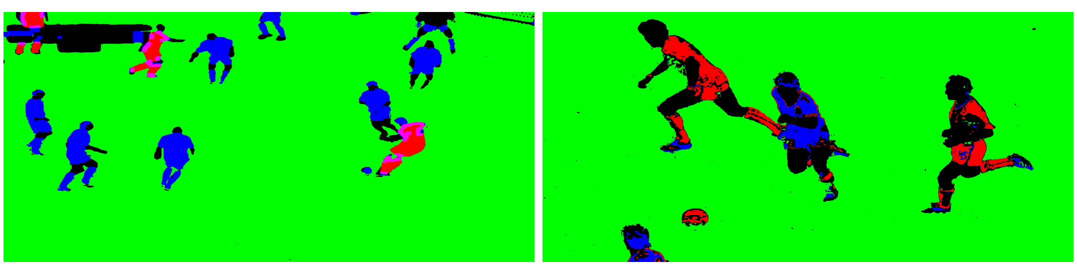
Following thorough deliberation, we decided to use HSV-based segmentation as the method for our project. While determining the suitable HSV color ranges for each element is challenging, this approach demonstrated notable effectiveness. The primary benefit of HSV-based segmentation lies in its independence from human intervention, aligning with our objective of creating an automated system. To implement HSV-based segmentation, we initially needed to specify the HSV color ranges for elements in the image(background and teams). This involved the utilization of a distinct approach provided at test_hsv.cpp, as illustrated in figure 1.

After finding the color ranges, we stored these values in a scaler along with a unique color for each team. It is essential to say that for each person we have a separate vector for storing the HSV values for shirts, shorts, and, socks. Finally, we add them to the black image that we created to initialize the processing.

Here are some results after the segmentation.

```
Scalar lowerField(v0: 30, v1: 60, v2: 60);
Scalar upperField(v0: 90, v1: 255, v2: 255);
Scalar lowerTeamAShirt(v0: 100, v1: 100, v2: 100);
Scalar upperTeamAShirt(v0: 130, v1: 255, v2: 255);
Scalar lowerTeamAShort(v0: 100, v1: 100, v2: 100);
Scalar upperTeamAShort(v0: 130, v1: 255, v2: 255);
Scalar lowerTeamASocks(v0: 100, v1: 100, v2: 100);
Scalar upperTeamASocks(v0: 130, v1: 255, v2: 255);
Scalar lowerTeamBShirt(v0: 0, v1: 0, v2: 200);
Scalar upperTeamBShirt(v0: 180, v1: 50, v2: 255);
Scalar lowerTeamBShort(v0: 0, v1: 0, v2: 0);
Scalar upperTeamBShort(v0: 180, v1: 255, v2: 100);
Scalar lowerTeamBSocks(v0: 0, v1: 0, v2: 200);
Scalar upperTeamBSocks(v0: 180, v1: 50, v2: 255);
```

Figure 2: Storing HSV values



(a) Image 01 (b) Image 04

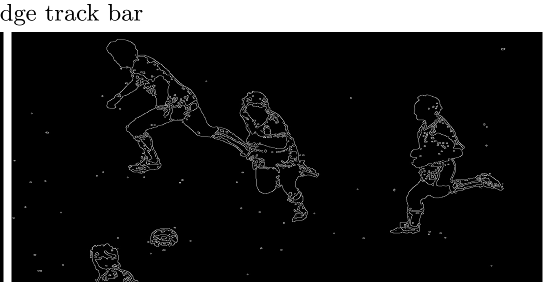
Figure 3: Results of segmentation

3.2 Localization

Now, we turn our attention to the localization of the players in our images. This process is a critical initial step in our methodology. It involves identifying and determining the location of players and isolating them from the background. Practically from this step forward, our region of interest is the players, and our analysis focuses on it. We twice applied the canny edge detector to the images in our algorithm and compared the results—one time before segmentation and one time after it. Comparing the results, we decided to continue with the second approach. Code is provided in text_canny.cpp.

(a) Canny edge track bar

(b) Canny Edge applied before segmentation



(c) Canny Edge applied after segmentation

Figure 4: Image 04

Using the Canny edge detector, numerous contours are produced, and we choose the largest contour, which is most likely to represent the players. Subsequently, the algorithm creates a bounding box around the largest contour, which is most likely to correspond to the player.

3.3 Identification

Recognizing the team in each localized area is a challenging yet crucial part of our project. For this task, we employed an innovative approach which is based on counting the pixels with the same color. We determine the player's teams based on whether the red pixels count is greater than the blue pixels count or vice versa. It assigns the player's team as "TeamA" or "TeamB" accordingly and sets the bounding box color to red or blue, respectively.

3.4 Comprehensive process overview

Our code for this project performs various image processing tasks, including color segmentation, edge detection, and contour analysis, to detect and label players on a playing field. It also overlays bounding

```cpp
// Determine the team label and bounding box color
std::string playerTeam;
if (redPixelCount > bluePixelCount) {
    playerTeam = "TeamA";
    boundingBoxColor = Scalar(0, 0, 255); // Red
} else if (bluePixelCount > redPixelCount) {
    playerTeam = "TeamB";
    boundingBoxColor = Scalar(255, 0, 0); // Blue
}
```

Figure 5: Player localization

boxes and labels on the original image to indicate each player's team. Here's a description of the code's main steps:

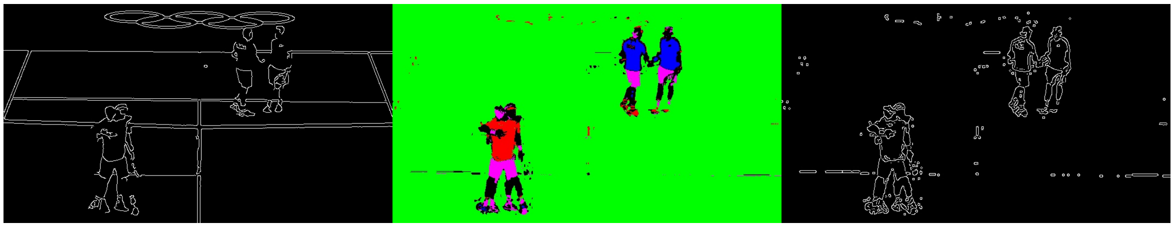
After loading the input image, the algorithm proceeds to convert it into the HSV color space. To facilitate further analysis, masks are generated based on determined color ranges, enabling the segmentation of distinct elements within the image, including the playing field and the players. Subsequently, the algorithm embarks on a segmentation and color-coding process. Leveraging the HSV color ranges, it systematically segments the image into discrete components, discerning between the players affiliated with Team A and Team B. This segmentation culminates in the creation of output images, effectively color-coding player regions based on their respective teams.

The subsequent phase entails edge detection, a computational technique applied diligently to both the original image and the segmented player regions. Employing the Canny edge detection algorithm, the system extracts edge information, facilitating the identification of key boundaries.

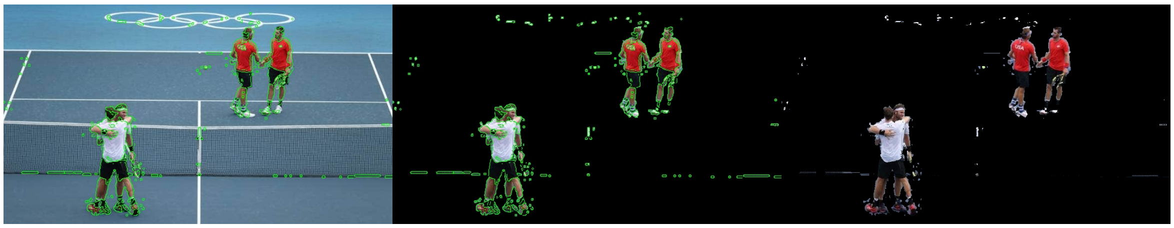
To enhance visual clarity and precision, the algorithm merges the edge maps derived from the original image and the segmented player regions. This fusion operation accentuates the delineation of player edges and field boundaries, providing a comprehensive view of the scene. Contour analysis forms an integral part of the algorithm's computational framework. After extracting contours from the combined edge map, the algorithm engages in data filtering, systematically eliminating small or irregular contours based on predetermined criteria, such as contour area and point count.

In the subsequent phase, player detection and labeling take center stage. The algorithm employs mathematical computations to draw bounding rectangles around detected player regions, facilitating their identification. Moreover, it conducts an in-depth analysis of the dominant color within each player's region to ascertain their team affiliation, subsequently labeling players as members of Team A or Team B. To enhance visual clarity, filled rectangles with semi-transparent color overlays are strategically applied to signify team affiliation.

The algorithm's findings and progress are systematically presented to the user, leveraging OpenCV's image display functions. Furthermore, the results are methodically preserved, encompassing segmented images featuring player detection, the original image adorned with bounding boxes and labels, and the merged Canny edge representations. In "Figure 6" there are images depicting all steps.



(a) Canny Before Segmentation (b) Masks for Segmentation (c) Canny After Segmentation



(d) Overlay with Original Image (e) Overlay with Segmentation (f) Players Localized

(g) [Players Detected] (h) Teams Detected (i) Overlay with Image

Figure 6: Whole Algorithm in a Glance

4 Results and Analysis

In this section, we present the results of our sport video analysis system.

4.1 Players Localization

In table 7 we report the results of player localization using "the mean Average Precision (mAP) calculated at IoU threshold 0.5" as suggested in the project definition. It's important to note that when we refer to a threshold of 0.5, it signifies that we are assessing the model's ability to accurately generate bounding boxes by requiring a minimum of 50 percent overlap (IoU) between the predicted and actual ground truth bounding boxes. The mAp score is calculated by averaging over all of the players in an image. we also considered non-detected bounding boxes as zero.

4.2 Players and Playing field segmentation

Table 8 is representing the results for segmentation based on The mean Intersection over Union (mIoU)

The results of our system are generally satisfactory. However, there are some segmentation and detection errors. Specifically, using HSV color ranges for segmentation sometimes made segmenting each player with their corresponding teams difficult. In some images, the players were inside each other; it is easy to find their team and categorize them as a team, but it is challenging to separate each team member because the Canny edge detector considers them as a whole. We selected the most

Image	mAp
Image 01	0.8
Image 02	0.3
Image 03	0.25
Image 04	0.75
Image 05	0.42
Image 06	0.9
Image 07	0.5
Image 08	0.5
Image 09	0.25
Image 10	0.33
Image 11	0.2
Image 12	0.2
Image 13	0.33
Image 14	0.25
Image 15	0.5

Image	Team A	Team B	Field
Image 01	0.291835	0.504264	0.0.935188
Image 02	0.134724	0.196879	0.678339
Image 03	0.0714812	0.101949	0.801511
Image 04	0.280562	0.282034	0.951973
Image 05	0.172718	0.205275	0.72032
Image 06	0.258454	0.377884	0.728439
Image 07	0.326249	0.45295	0.904268
Image 08	0.292983	0.297893	0.522614
Image 09	0.557575	0.432455	0.836387
Image 10	0.418689	0.345147	0.722139
Image 11	0.156876	0.251356	0.759184
Image 12	0.340244	0.336487	0.64701
Image 13	0.174324	0.251942	0.923778
Image 14	0.252234	0.527446	0.882008
Image 15	0.570238	0.38681	0.860067

reliable results by trying different HSV ranges and ranges for the Canny edge detector. Related to this issue, in some cases, the colors of the background match the color of the teams, and results are heavily affected. Despite these challenges, our system demonstrated its potential in automating the task of sports video analysis. With further improvements, particularly in the segmentation step, we believe that our system could become a valuable tool for sports analysis in any setting.

5 Challenges

We assessed our system's performance using the dataset provided in the project description, which comprised fifteen distinct images, each representing a unique scenario. Overall, our system produced generally satisfactory results. However, we encountered some challenges in the segmentation phase, primarily stemming from the complexities inherent in this step.

Specifically, our use of HSV color ranges for segmentation occasionally posed difficulties in accurately distinguishing between various elements. For instance, in image 3, distinguishing the player from the audience proved nearly impossible due to the striking similarity in attire color. Consequently, both the audience and the players were grouped together.

In certain situations, such as in image 8, where players overlapped, our algorithm erroneously treated them as a single player. This issue likely resulted from the application of the Canny edge detector, which identified a single contour.

Despite these obstacles, our system exhibited promise in the field of sports video analysis. With further refinements, particularly in the segmentation phase, we believe that our system has the potential to become a valuable tool for sports analysis.

6 Conclusion

In conclusion, the developed computer vision system represents a significant result in the realm of sports video analysis. It has shown promising capabilities in the critical tasks of player localization and the precise segmentation of various elements within images, thereby offering invaluable insights into sports analysis.

One of the system's notable strengths is the use of well-established techniques. The employment of the Canny edge detector, for instance, enhances the system's ability to identify and outline the contours of objects, while the utilization of the Hue, Saturation, and Value (HSV) color space provides a robust foundation for segmentation, localization, and player identification.

The core functionality of the system is its capacity to accurately localize players and discern their respective teams. This achievement is vital in enabling sports analysts and coaches for providing valuable tactical insights. However, it's important to acknowledge that the accuracy of the system is intrinsically tied to the effectiveness of the segmentation step.

The segmentation process hinges on the precise definition of HSV color ranges for each element of interest, such as players and teams. Any inconsistencies or inaccuracies in these color definitions can impact the system's performance, potentially leading to misclassifications or incomplete segmentations. As such, meticulous attention to detail in defining these color ranges is paramount for optimal results.

Furthermore, it's essential to recognize that real-world scenarios may introduce complexities, such as image noise and color similarities between the background, players, and teams. These challenges can pose hurdles to achieving perfect segmentation and player identification. Thus, ongoing refinement and adaptation of the system to handle these real-world nuances will be critical for its continued success.

In summary, the developed computer vision system represents a significant step forward in the field of sports video analysis. Its ability to localize players, segment playing fields, and classify teams opens up exciting possibilities for sports enthusiasts, analysts, and coaches alike. As the system evolves and adapts to real-world challenges, it holds the potential to revolutionize the way we analyze and understand sports events, ultimately benefiting athletes and teams in their quest for excellence.

Name	Student ID	Hours	Parts
MohammadMostafa Talebi	2043994	80	Segmentation-Identification-Detection
Hamid Mohammadi	2005552	80	Segmentation-Identification-Detection

Table 1: Contributions