Automating NFL Film Study: Using Computer Vision to Analyze All-22 NFL Film

Timothy E. Lee
leete@stanford.edu

Abstract

A series of algorithms were developed in order to automatically extract information from a football play using computer vision techniques. Extraction of play information required two tasks: field reconstruction and player recognition. The field markers of yardlines and hashmarks were extracted using Hough transforms. The player tracking algorithms used blob tracking techniques that focused on the unique color features of the teams that were involved in the play. Despite challenges in reconstructing field coordinates as the play evolved, a team formation was extractable from a sample still image, proving that these algorithms can indeed work in tandem to extract relevant football information from NFL film, showing high promise for future capability of this technology.

1. Introduction

In recent years, professional sports have experienced an inrush of technology. An example of the recent confluence of sport and technology is the 1st & Ten line that appears on television broadcasts for NFL (National Football League) games. Developed by Sportvision, the "yellow line" indicates the location of the first down marker to television audiences, leveraging computer vision techniques to provide an augmented reality to NFL viewers. Furthermore, Sportvision provides real-time object tracking technologies for NFL broadcasts, which are used by television broadcasters to supplement game viewing with overlaid statistics [1].

The current investigation utilizes computer vision techniques to extract player information from video of a football play. This investigation is not meant to augment reality for the NFL viewer; it is instead meant to augment, and eventually automate, a significant aspect of NFL personnel training: study of game film. Utilization of such technology would enable more efficient tactical scouting of opposing NFL teams. If an entire NFL game can be quickly analyzed and summarized, then the time required for NFL team personnel (coaching staff and players) to study film is reduced, allowing greater time for aspects of game preparation that cannot be readily automated, such as conditioning or individual coach-player training.

The capability of extracting meaningful player data from film of a football play leads to the formation of datasets for entire games that can be quickly analyzed. Such a dataset can be leveraged by other branches of artificial intelligence, such as using machine learning principles to characterize the various strategic tendencies of NFL teams, such as overreliance of particular formations or personnel mismatches, and, in so doing, expose potential tactics for defeating opposing teams.

2. Literature Review

2.1. Review of Previous Work

The optical tracking technology provided by the STATS LLC SportVU produce is used by the National Basketball Association (NBA) and in various soccer leagues. Beginning with the 2013-2014 season, this technology is installed in every NBA arena and provides *X*, *Y*, and *Z* coordinates for all objects on the basketball court: players, referees, and the basketball. For NBA games, the SportVU technology utilizes

six high-definition cameras, three per court [2]. This spatial data are later used for various analyses, such as predicting the value of player decisions and automatic detection of a certain kind of basketball play [3, 4].

In contrast to the publically available NBA SportVU data, the data from optical tracking technology used in agreement with the National Football League and Sportvision is not publically available.

2.2. Contributions of the Current Work

The present investigation is unique among published literature in that it proposes methodologies for extraction of relevant information from recorded game film. Unlike the NBA, the NFL has no publically available service to obtain the depth of information available from a tracking technology such as SportVU. However, similar data can be extracted from recorded videos of NFL plays, noting that the richness of such data is limited to the quality and number of cameras used to record the NFL plays. For the current investigation, All-22 film is used, which was recorded from one camera using a 480 x 640 pixel resolution.

3. Technical Methodology

3.1. Methodology Summary

Analyzing a football play from video collected from one camera can be decomposed into two distinct tasks:

1. **Field reconstruction.** This task is intended to capture the stationary objects on the football field that describe the orientation of the field. These objects include the vertical yardlines and horizontal hashmarks of the football field. These objects can be used to estimate the field coordinates of any object on the field from that object's pixel value in the video.
2. **Player recognition.** This task is designed to identify and track moving objects in the football play, which are the 11 players on each team.

Because the methodology tracks the entire football play, the objects of concern are tracked as the play evolves. After tracking is complete, the data are processed offline to extract relevant information about the play. Thus, the algorithm itself requires two steps: a tracking step, where the data are processed frame-by-frame, and a post-processing step.

The current study utilizes various MathWorks products, such as MATLAB, the Image Processing Toolbox, and the Computer Vision System Toolbox.

3.2. Detailed Presentation of Methodology

3.2.1. Dataset Description

The choice of NFL film for this technology is key. While most television viewers of NFL games are familiar with the broadcast feed, which is captured from cameras on the side of the football field, this feed is not well suited for this technology. The broadcast feed is designed to focus on the line of scrimmage – the initial center of action of any football play. While the broadcast feed usually captures all 11 players on offense, not all defensive personnel may be visible in the feed. Thus, to ensure that all 22 players are visible, this technology utilizes "All-22" film. This film is not a part of the television broadcast feed, but is available by scouts, analysts, players, and coaches, and is specifically used to study game footage.

The dataset specifically used in the study is the All-22 film from Super Bowl XLVIII, played on February 3, 2013 between the Baltimore Ravens and the San Francisco 49ers. The All-22 video was recorded with an

image size of 640 pixels (width) by 480 pixels (height) and with an approximate frame rate of 24 frames per second. The video for the study was acquired from the National Football League through NFL Game Rewind. NFL Game Rewind contains replays of the television broadcast feed as well as a condensed feed and All-22 film for all games from the 2009—2013 NFL seasons.

Because Super Bowl XLVIII contained over a hundred football plays, it is critical to first assess the capability of augmenting film study with computer vision by processing just one play. Once the algorithms are demonstrated successfully for one play as a proof-of-concept, the algorithms can be extended naturally to the remainder of the plays. The play that was highlighted for the study was Baltimore's first touchdown. It was selected because it was a passing play. Therefore, the players were spread out in the field of play. Such a play would be less difficult to track players as the play evolved, instead of a running play, where most of the players would be grouped closely together and more difficult to distinguish individually.

The play was approximately 10 seconds in duration and consisted of approximately 250 frames. A selection of frames is shown in Figure 1 to convey the nature of the play. Note that the football was not snapped to the quarterback until about 4 seconds of the play had elapsed.

Figure 1: *Progression of the football play of interest in four frames (order is row-wise).*

3.2.2. Field Recognition Methodology

Prior to discussions of field recognition, it is necessary to first define the coordinate systems of the world correspondences and the image correspondences. For image correspondences, the origin is located at the upper-left corner of the image, with increasing *X* values to the right (column-wise) and increasing *Y* values downward (row-wise). For world coordinates, the coordinate system of the football field is established using the regulated dimensions of a football field. The field is oriented horizontally, with the origin located at the intersection of the lower boundary line and the 0-yard line. Increasing *X* values move towards the right end zone, and increasing *Y* values move towards the upper part of the playing field. Positive *Z* values represent heights above the football playing surface, which has a *Z* value of 0. The following regulated dimensions are utilized to build the world coordinate system [5]:

- End zones are 10 yards wide.
- The entire width of the field, including both end zones, is 120 yards.
- The field is horizontally separated by 10 yard intervals by vertical yardlines.
- The entire height of the field is 160 ft.
- The lower hashmarks are located 70.75 ft from the lower boundary line.
- The upper hashmarks are located 70.75 ft below the upper boundary line.
- Therefore, the distance between the hashmarks is 18.5 ft.
- The crossbar of each goal post is 10 ft above the field, and the projection of the crossbar onto the playing field lies on the back line of the end zone.
- The width of the goal post crossbar is the same as the distance between the hashmarks.

The vertical yardlines and horizontal hashmarks are critical components in field reconstruction. The yardlines run vertically, spanning the entire horizontal range of the field. The hashmarks are, literally, hashmarks that run horizontally, indicating the vertical position on the football field. These are pronounced objects on the field and, because they are linear in nature, the Hough transform is well suited to extract their positions [6].

To extract the yardlines, first, a Sobel edge detector is utilized on a greyscale version of the image. This exposes the edges on the field and increases the prominence of the yardlines. Then, a Hough transform is utilized to detect the locations of these yardlines. An example of identification of yardlines using the Hough transform is seen in Figure 2. The result of this operation is obtaining the pixel coordinates of these yardlines.

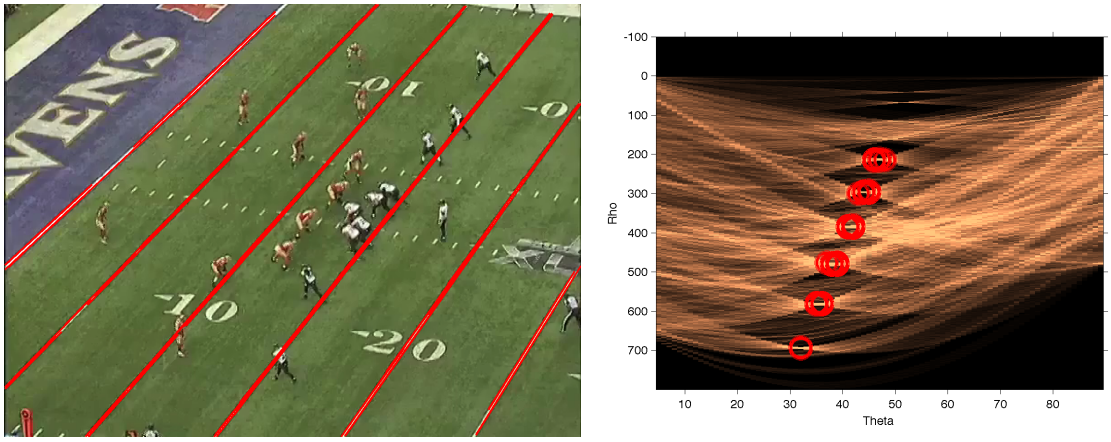

Figure 2: *Automatic detection of yardlines on a football field (left) using the Hough transform (right).*

Extraction of hashmarks is done is a similar manner, but prior to applying the Hough transform, the image is converted from the RGB colorspace to the L*a*b* (referred to as CIELAB) colorspace. The advantage of the CIELAB colorspace is that it separates colors into luminance (L*) and chrominance (a* and b*) components, and does so in a way that is perceptually motivated by the human visual system [7]. The luminance portion of the image is filtered and a blob detector is used to ultimately extract the locations of bright white "spots" on the field. These locations are passed through a Hough transform which extracts the line that connects these hashmarks together. The identification of hashmarks can be seen in Figure 3.

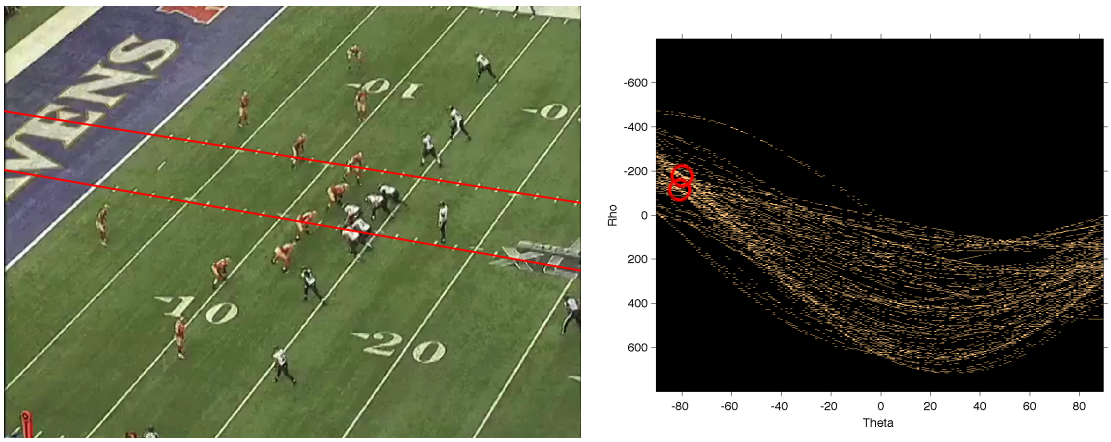

Figure 3: *Automatic detection of hashmarks on a football field (left) using the Hough transform (right).*

Once the pixel coordinates of the yardlines and hashmarks are known, it is possible to utilize an algorithm to transform pixel coordinates to field coordinates. One such way of doing this is to derive the camera matrix. Initial investigation of the data suggested that it is possible to calculate the camera matrix uniquely from every frame of the NFL play. Unfortunately, this was only possible when the goal posts are visible, as only the goal posts provide datapoints in a separate plane than the playing surface. Because the goal posts are usually not visible, except only for plays near the end zones, it became necessary to develop an algorithm to reconstruct the field from only yardlines and hashmarks, which are always visible. The method utilized in the current study is to assume an affine transformation from the location of the yardlines and hashmarks in the image and on the football field. This introduces some error into the transformation between the image and the football field, but an advantage of this algorithm is that it only requires yardlines and hashmarks, which are readily extractable from any video frame.

3.2.3. Player Recognition Methodology

Unlike the field markers of yardlines and hashmarks, which are extractable using Hough transforms, tracking players requires a more complicated algorithm. Previous work had shown that tracking players using a point cloud consisting of Harris features for each player was successful only until the players began moving. Then, the points become confounded, stuck on lines, and migrate between players. Success was achieved by the determination that football players are uniquely identified by their uniform colors. Therefore, an algorithm was developed the tracked the color features of the players.

The Baltimore football player uniform consists of black pants and white jerseys. Because white is a prevalent color in the image, black is used as the discriminating feature. To track the player, the image Is first preprocessed by calculating the distance of each pixel's RGB color from the color black (RGB: [0 0 0]) and running a 2-dimensional smoothing filter. Then, a heuristic threshold is used to identify the local minima of the images that correspond to objects with a localized concentration of black, such as the pants of the Baltimore players. A blob detection algorithm is then applied to this processed image that identifies the location and size of these minima, which is reality are the Baltimore players. Sample results of this recognition algorithm are shown in Figure 4.

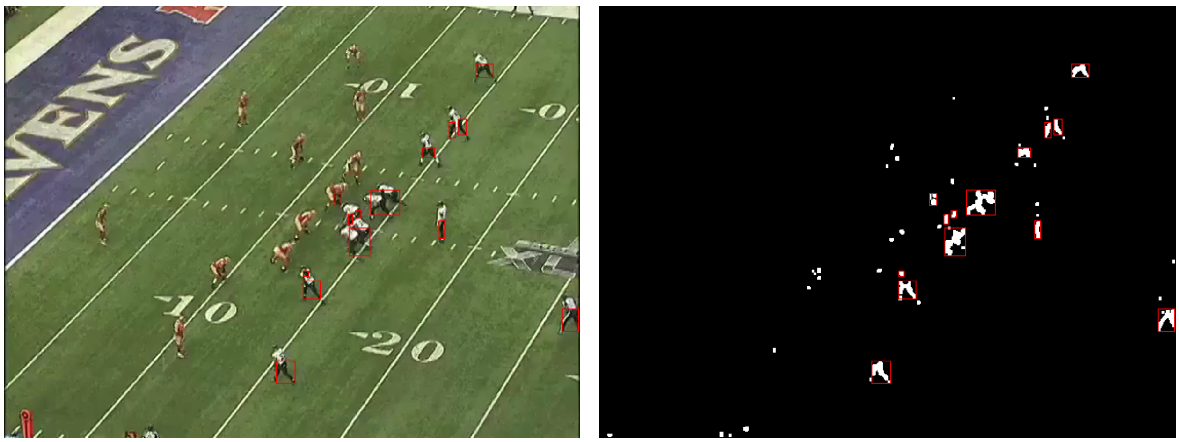

Figure 4: *Automatic recognition of Baltimore players, with bounding box overlaid on play image (left) and on the underlying processed image (right) that exposes the unique color features of the Baltimore players: black pants.*

In the case of San Francisco, their players have uniforms containing scarlet (jersey) and gold (pants), so the player recognition algorithm will be different than the one used for Baltimore. Because these colors could be influenced by the relative luminance of the image, the image preprocessing algorithm converted the image to CIELAB to separate the influence of luminance from chrominance. Then, two color distance calculations were computed for the equivalent of the color scarlet and the color gold in the CIELAB colorspace. These two distance calculations were then combined into one metric by calculating the resultant of these two distances. The resulting image has local minima that are found using a heuristic threshold and then sent to a blob detector to extract the location and shape of the minima, which most often corresponded to San Francisco players. Sample results are shown in Figure 5.

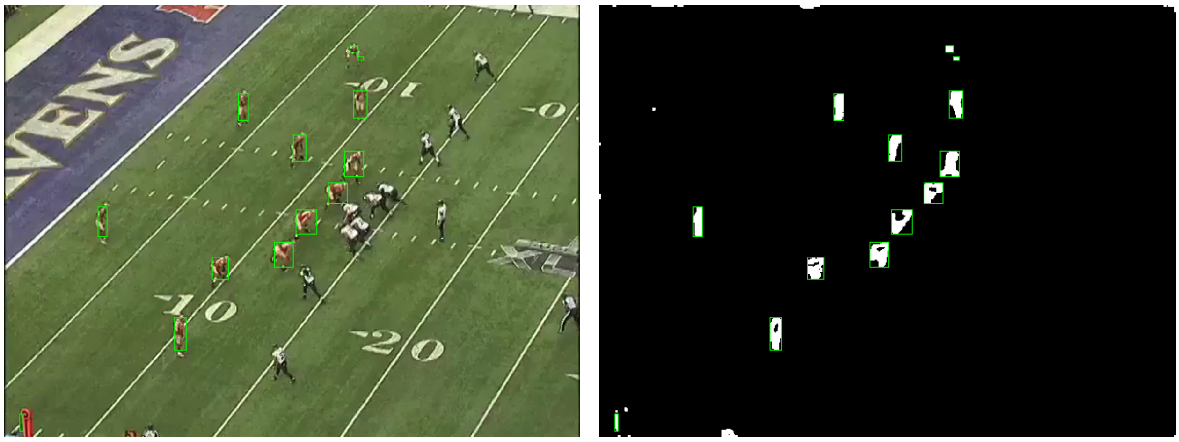

Figure 5: *Automatic recognition of San Francisco players, with bounding box overlaid on play image (left) and on the underlying processed image (right) that exposes the unique color features of the San Francisco players: scarlet jerseys and gold pants.*

In both cases, there were some false positives with the algorithms. Most commonly, the algorithm would occasionally identify objects that weren't players, such as the yardage markers on the side of the football playing field. In other cases, due to occlusions, the algorithm would combine two players into one blob. With work, these issues can be addressed in post processing and are seen as future areas of improvement of the methodology.

It should be noted that this color tracking methodology only works for certain uniform combinations of their respective teams. Each team may have multiple uniform sets (e.g., home and away), so for a comprehensive automated NFL film algorithm, the algorithm would need to know which teams are playing and what uniform combination they are wearing. However, for the current study, it suffices to consider the uniform combination for each team as distinct for the entirety of one football game.

4. Experimental Results

While the affine transformation between image pixel coordinates and field coordinates is successful by hand for an individual frame (with some introduction of error), unfortunately, a completely automated method of determining the field position as the play evolved was not successful despite best efforts otherwise. However, it is possible to show the possibilities of using the affine transformation data with the player tracking algorithms. Shown in Figure 6 are the field position coordinates of the Baltimore players, before the snap. This formation is determined by applying the affine transformation to each of the tracked object, the Baltimore players. The affine transformation is calculated by hand from the points of the hashmarks and yardlines, which were identified by the Hough transformation. Such results scratch the surface of the potential of these algorithms and indicate the usefulness of technology when fully mature.

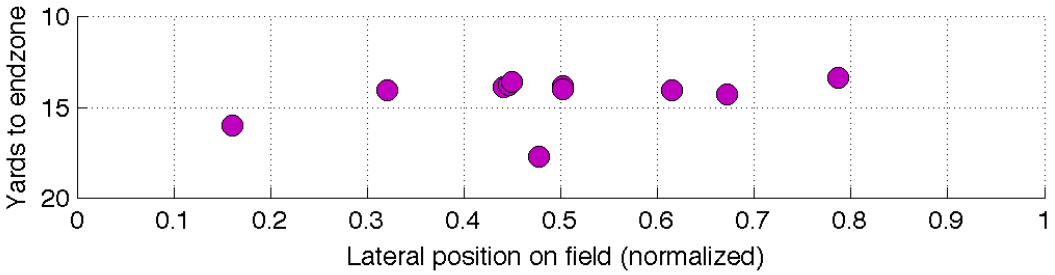

Figure 6: *Extraction of field position for the Baltimore players, before the snap.*

5. Conclusions

As evidenced, the capability of using computer vision techniques to extract meaningful information from NFL film is promising. The current work was successful in providing a proof-of-concept for extraction of relevant information about a football play using only recorded video from one camera. While difficulties were experienced with field reconstruction for the entire play, it is shown that the concept of linking together automatically identified field markers and players to calculate the positions of the players on the field is feasible.

5.1. Challenges

It is worth noting that the methodology proposed in the current investigation should not be compared on the same basis as the professional sports tracking technology. Several challenges of the dataset that impacted the current investigation were the limitation of using one camera and the relatively lower resolution available of the data recorded from this camera. In contrast, the professional sports tracking technology SportVU utilizes six cameras for one NBA basketball court, and it is possible that the camera matrices are available directly, without need for the challenges of automatic world reconstruction. The relatively lower resolution also caused difficulty in tracking players when they were not isolated. As such, obtaining quality data of the movements of the linemen (offense and defensive line), who were usually closely confined, was difficult as compared to the skill positions (wide receiver and defensive back).

5.2. Future Work

Despite the challenges faced by the existing data, there is great promise for future extensions of the work. Future work would mostly be focused on reconstruction of the field for any number of yardlines and hashmarks, which was a source of great difficulty in the current work. A Kalman filter model would potentially provide greater accuracy of detected objects moving in time by providing a predictive component towards how objects will move between frames.

6. References

[1] Sportvision, Description of Football Technologies, www.sportvision.com/football.

[2] STATS LLC, "SportVU Player Tracking Technology," www.stats.com/sportvu/sportvu.asp.

[3] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, "POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data," 8th Annual MIT Sloan Sports Analytics Conference, February 28 – March 1, 2014.

[4] A. McQueen, J. Wiens, and J. Guttag, "Automatically Recognizing On-Ball Screens," 8th Annual MIT Sloan Sports Analytics Conference, February 28 – March 1, 2014.

[5] National Football League, "NFL Rules Digest: Field," www.nfl.com/rulebook/field.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004.
[7] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011.

7. Disclaimer

The All-22 film utilized in this investigation is property of the National Football League. Utilization of this data is strictly done in an academic and non-commercial manner, and such use is congruent with the Terms & Conditions of the National Football League. Out of respect for the copyright of the National Football League, the investigation dataset is not available for public distribution.