

Mapping football players to a bidimensional plane

Extended abstract for a Master dissertation in Information and Software Engineering

Francisco Besteiro

franc.best@hotmail.com

Instituto Superior Tecnico, Universidade de Lisboa
Lisbon, Portugal



Figure 1: Frame with the players positions mapped to a 2D plane

ABSTRACT

Mapping football players to a bidimensional plane will allow clubs to have better control of the performance of their players and to better scout other players from rival teams. The purpose of this paper is to describe in the best way possible a project named Field Detector, that was designed to solve the problem of mapping the position of the players from a movie frame to a bidimensional plane. The Field Detector was developed within a bigger system that already identified the players in an image, so together with that identification

and through several image treatment processes and mathematical algorithms, it solves the overall problem of mapping the players.

KEYWORDS

Transformation Matrix, Homography, First Transformation, Morphology, Hough transform, Least Square Fit

Reference Format:

Francisco Besteiro. 2019. Mapping football players to a bidimensional plane: Extended abstract for a Master dissertation in Information and Software Engineering.

1 INTRODUCTION

Mapping players from a football game frame to a bidimensional plane is a complex problem whose solution requires several image processing algorithms and the use of some mathematical algorithms.

This project had several objectives that needed to be achieved so that the solution could be considered a success. These objectives are:

- Develop an application capable of processing, in real time, at least 30fps ;
- Have an error of less than 10% in the precision of the positioning of the players in the field;
- Have an error less than 5% in the field mapping for the 2D plane;
- Integrate with the bigger project, Player Detection, without compromising the system at run time and without interfering with the user interaction.

This extended abstract was structured in such a way that it was easy to understand and follow the theme, being composed of six sections. Section 1 is this introduction. Section 2 explains three other possible solutions and why they are not as good as the one presented by this project. Section 3 presents the problem in its full complexity. Section 4 is where the proposed solution for this project is presented. Section 5 describes the evaluation done to the solution and discusses the results. The final section presents conclusions about the project.

2 RELATED SOLUTIONS

There are three relevant solutions that should be studied to fully understand the complexity of the problem and the advantages of the proposed solution.

2.1 Top 2D view

The first approach that should be analyzed is the one presented by Tomikawa and Saga in [10]. Their solution consists of four important steps explained by:

- (1) Detecting the field and its lines through a black and white (binary) matrix image, in order to calculate the position of the players; The field lines will be used as reference to calculate the real position of the players in the field, since they are the only points that are static/unchangeable in a football field. They use the *Hough Transform* to detect the lines.
- (2) Understanding what the movement of the camera is so that it is possible to adjust the position of the players; This is done by applying a technique named projection transformation that will find a transformation coefficient with eight variables capable of resolving a transformation equation. After finding the coefficient

they apply a technique named "optical-flow", to follow the points movement through the frames, based on their color.

- (3) Doing a selective research in order to find the players; The position of the players found through the selective search is done by joining the results of the previous two steps. That is, with the supposed position of the players and the movement that the camera, it is possible, according to the authors, to find the exact position of the players in the field, discarding in the process all the suppose players that were first detected and that were nothing but noise.
- (4) Creating a Top view of a football field, where the calculated player positions should be drawn. This final step is done by identifying the color of the rectangle of each player and distinguish them by teams. Once this is done the central position of each rectangle is translated to the bidimensional plane and is created a dot in that plane with the color found for that player team.

This solution fails in player detection, which is due to not being able to distinguish when two players are on top of each other on the image. Another failure is the attempt to count players that are missing. And the last fail is allusive to the movement of the camera, if the speed increases, the distortion occurs in the points of appeal and in the colors. [10] [7] [11]

2.2 Wide-angle lens camera

The second approach that needs to be analyzed is the one described in [12]. This solution is distinct from the others by the use of four cameras to capture the images that have to be processed, one GoPro and three cameras with conventional lenses.

In this algorithm is proposed a way to correct the image distortion, which can be summarized as follows:

- (1) For the first frame of the video, the set of lines is automatically extracted using a modified *Hough Transform* together with a division model to a parameter. This results in a set of distorted lines and an estimated initial value of the first distortion parameter;
- (2) Next, and still for the first frame of the video, the lens distortion model is improved using non-linear optimization techniques, such as the use of a said energy function, which calculates the distance from a point to its ideal position, and the calculation of the center of distortion as a function of the Taylor expansion, the center of the image and the energy function. Resulting in an optimized 2-parameter division model and obtaining the ideal center of distortion;

- (3) Finally, the radial distortion of each frame of the video is corrected using the lens distortion model, found in the previous step.

After correcting the frames and marking them, the calibration points and the trajectories of the players were reconstructed by means of the direct linear transformation method, which provided the 2D coordinates of the player in relation to the coordinates of the frame.

The results of using this approach show that the use of corrected radial distortion image sequences from a single wide-angle lens camera using the *Hough Transform*, is a highly applicable and reliable way to conduct a cinematic investigation of trajectories and distances covered by futsal players.[12][3] [4] [2]

2.3 Player Detection System

The last approach that should be studied, to fully understand both the problem and the solution, is the one done by Mackowiak, and his team, in [9]. Where a system for detecting football players was developed having for main operations: the field detection, the field 3D model, the object region recognition and the object tracking model.

The main operations can be described as:

- (1) **Field Detection** - This operation is the most important one and is done by the assumption that the playing area is a homogeneous region with a relatively uniform, green color. So to detect the field the algorithm looks for areas that are mostly green. After detecting the field zone the search for the field lines begins with the execution of a modified *Hough Transform* followed up with morphological treatment. The result of the execution of these two treatments is a set of lines corresponding to the ones of the field. These procedures are done for every frame, and for each one of those frames, the detected lines are readjusted with the values found within the frames prior of the one that is being analyzed, so that the lines are as precise as possible. [9] [13] [8]
- (2) **Field 3D Model** - A predefined area model is used as a reference to find the area lines with well-defined dimension ratios, and game field lines are detected based on these reference parameters, recreating a kind of 3D model for the field. [9]
- (3) **Object Region Recognition** - To detect objects, a multiple-scale approach has been used to ensure proper detection of players regardless of their distance from the camera. Having has only limits the minimum and maximum dimensions of the window. In this approach, a histogram of oriented gradients (HOG) descriptors is used, which are classified by a linear SVM algorithm and are stored in a database of player models. Later

the objects detected are compared with the ones of the database and if they are similar they are set has players. [9]

- (4) **Object Tracking Model** - a single player can cause multiple detections through the frames, and therefore the bounding boxes resulting from the single-player detector overlap many times. To overcome this problem and to efficiently do the tracking, an additional merger operation was proposed, which consists of the following operations:
 - (a) All detection boxes are discarded where their size does not respect the imposed limits or if they contain many pixels of playing field;
 - (b) An overlay test is performed for each pair of boxes detected by pixel comparison of each of the boxes not corresponding to the field, where it is tested whether those areas and formats of such pixels are similar and overlap;
 - (c) Aggregation of boxes where two overlapping boxes are joined together in a box.

For a conclusion this approach deserves more research, especially in the parts of it referring to the database of images used to train the SVM detectors and the number of detectors used to find the players. [9] [5]

3 PROBLEM AND GOALS

The problem trying to be solved by this dissertation is the following:

Create an application that by receiving a video with a football game is able, at all times of the gameplay, to map the position of the players visible to a 2D plan of a football field.

The context for this problem goes around the necessity that exists in the world of sports, more specifically in the world of football, where some processes are still manual and need to become more automated and precise.

One of these processes is the performance analyses of the football players, that for smaller clubs is still done manually without the help of GPS chips or another kind of similar technologies. So one problem rouse: **Is there a way of developing a low-cost application that can do this?** This problem eventually gave birth to several others, being one of them the one described earlier and that is the target of this project.

Despite being a definition that does not fully demonstrate the complexity of the problem, when put together with the objectives described below, it can give a good notion of what it is that needs to be solved.

The main objectives of this project are the following ones:

- Create an application capable of processing in real time, at least 30fps;

- Have an error less than 5% in the field mapping for a two-dimensional plane;
- Have an error of less than 10% in the precision of the positioning of the players in the 2D football field;
- Integrate with the Player Detection system without compromising the run time and without interfering with user interaction.

4 PROPOSED SOLUTION

The proposed solution to the problem, described in the previous section, 3, was developed with the purpose of perceiving, in real time and frame by frame, the position of the players in a football field, and is denominated by Field Detector

The development of the project was divided into five main phases: (1) Image Treatment; (2) Line Identification; (3) 2D Plane Creation; (4) Image mapping to a 2D map; and (5) Mapping players.

Through the sequence diagram, visible in the figure 2, is possible to visualize where each phase enters in the project and what it intends to accomplish.

In the next subsections, it will be explained each of the phases, so that it is easier to understand how each one was developed and what it intends to achieve.

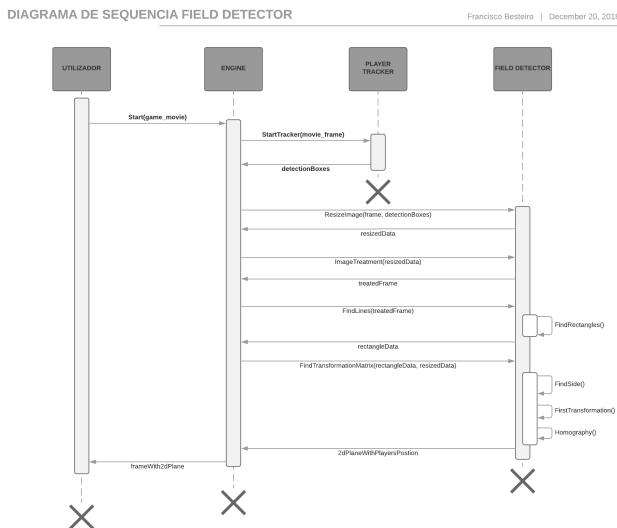


Figure 2: Sequence diagram of the proposed solution

4.1 Image Treatment

This phase of the solution is responsible for cleaning the image of garbage that can interfere with the area detection and consequently with the mapping of the field to the bidimensional plane.

The operations performed by this phase start with a resize of the frame that decreases it for an easier size to analyze.

The **resize of the image** corresponds to the number of times 500 pixels fits within the width of the frame.

After the resize of the frame is done, a search for the area corresponding to the playable part of the field begins with the help of the **HSL color pattern**.

By setting a search for pixels between the **Hue values of 45 and 150** (values corresponding to green pixels), making them white and the other ones black, it is obtained one first version of a mask that will correspond to the field area.

In order to polish this mask and to clean the residue, two morphological techniques are used: **Erode** and **Close**. The result of the application of these two techniques is an almost perfect mask that can clearly represent the field zone.

After having the mask it is applied on the original image, in a way that a search for white pixels that are within the original frame and within the field area of the mask (pixels corresponding to lines) is started, putting all these pixels in white and all the rest in black. As result, a **black and white picture of the field lines** is given, almost without residue, like can be seen in the figure 3.

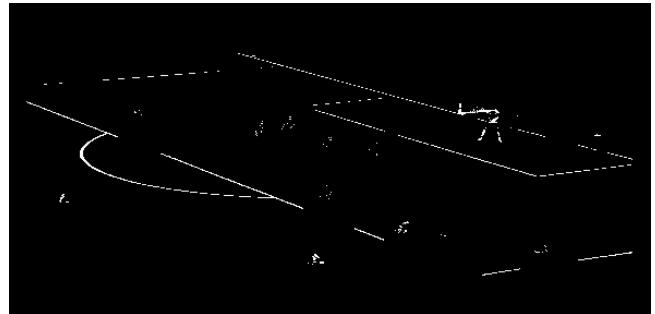


Figure 3: Result of the Image Treatment

4.2 Line Identification

The Line Identification phase is responsible for finding the lines that correspond to the great areas and their corresponding interception points so that the mapping can be done by taking them as a reference.

Firstly the application needs to identify that the white pixels in the picture, resulting from the previous phase (figure 3), so that they may represent the lines of the field. To do this the app was developed to use an algorithm named **Hough Transform**, that by searching white lines in a black and white image **returns all the possible segments of lines that can represent a field line**.

After having the list of the segment of lines, that represent the field lines, there is a need to put together all the ones that correspond to the same lanes, to execute this a mathematical approach was used. The mathematical algorithm that helped with this problem was the **Least Square Fit**, that traces the

best fit line between a collection of points. The result of the recursive usage of this algorithm, together with some conditions to identify the correct segments that needed to be put together, usually is **good lines that fully represent some part of the field**.

Once a list of well-defined lines is found is important, for the application, to recognize what side of the field is being analyzed that is the mapping can be done correctly.

To identify the part of the field that is being processed is important to detect the rectangles that the lines, found in the previous step, create by intercepting each other. Once the rectangles are calculated the one with the biggest area is identified and the signal of the **slope of its biggest line** is saved, because it **will represent the side of the field that is being analysed**, a positive slope means that the right side of the field is being analyzed and a negative slope means that is the left side.

The last step within this part of the phase is one that is almost transverse to the other steps, however, is very important for the mapping of the field to the bidimensional plane. Because of this, the step was adapted to do two distinct things:

- Intercept all lines detected and **check and save the points that within a limit defined by $2xWidth$ and $1,4xHeight$** of the resized frame;
- **Check the points of the rectangles** that can be printed on the frame image (so that the tests can be done in order to correct and test the transformation matrices).

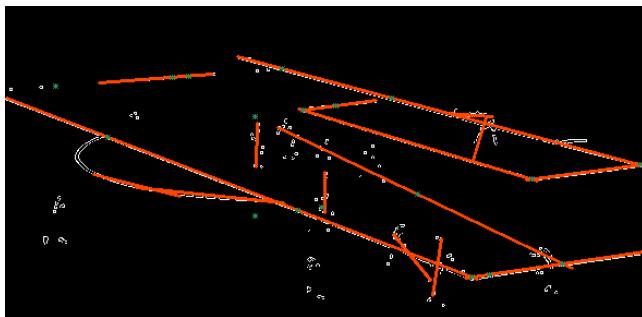


Figure 4: Points of interception and clean lines, from the rectangles found

The result of this phase is the one visible in the figure 4, where is possible to see both the clean lines, from step two, and the interception points of those lines, that are within the acceptable range.

4.3 2D Plane Creation

The bidimensional plane creation phase is the fastest one to be executed since its operations don't depend on the results from the previous phases.

What this phase was developed to do is create a small 2D plane of a football field. To be able to do this initially is important to find the maximum scale of the plane, the way that the value of the scale is found is by the number of times that the width of a field, in pixels, can fit in the width of the resized frame.

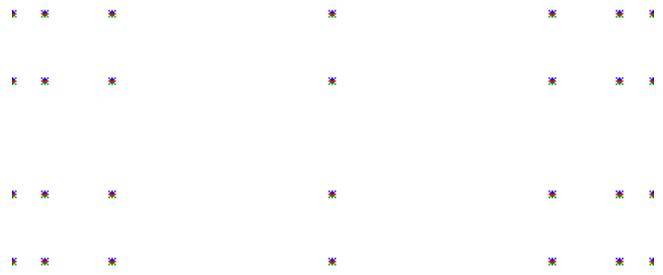


Figure 5: Bidimensional plane of a football field

Next, the interception points given by the lines of the field are found and scaled with the value given by the previous step. Once these points are found and properly rescaled the 2D plane is created, resulting in the figure 5.

4.4 Image mapping to a 2D map

The transformation matrix is a great challenge due to being responsible for translating the positions recorded from a single camera to a 2D plane. To be able to do this is important to identify the transformations suffered by an object when a picture is taken. After some investigation is possible to find that the transformations are four (rotation, translation, scale, and perspective) and they all need to be dealt with accordingly.

To deal with the rotation, scale, and translation a new algorithm was developed, the *First Transformation*, besides dealing with this transformations it also has the purpose of decreasing the error that would come from trying to just solve the perspective transformation.

The *First Transformation* can be explained as the execution in cycle of a set of small operations, which are done on the detected points of interception, in order to obtain or get closer to the ones from the bidimensional plane (created in the previous phase); and reduce the difference of scales between the planes. The operations within the cycle of execution are:

- Calculate the **Least Square Line** of the two sets of points, the one from the interception plane and the one from the 2D plane;

- Execute the **Rotation** on the points of intersection, based on the difference of slopes, from the two lines found in the previous operation. This is the only operation that will be done in two different ways: in the first iteration, the rotation will be done on the point of origin, and in the rest, it will be done on the centroid of interception points;
- Calculate the **centroid** of each list of points;
- Once centroids are found, a **translation** must be performed, on the one that corresponds to the points of interception and on the respective points, so that both centroids overlap;
- Calculate the **variance** for each set of the points, in relation to their centroid;
- Finally the **scale** needs to be found, so to do this the percentage of variance of the points of the plane in relation to the variance of the points of interception, must be calculated and then applied to the vector that transforms the centroid into each point in the set of points.

The cycle, with all where all the operations described earlier, is executed, runs until the rotation of the points is equivalent to a 360-degree rotation around the centroid of the set of interception points. Through the iterations of the cycle the distance between the corresponding points of the two planes is calculated with the use of a clustering technique named KNN, and for every time that the distance arrives at a new minimum the number of the corresponding iteration is saved, to be possible to know, when the cycle finishes, the best number of iterations to reduce the error between the two planes.

This phase finishes its operations by dealing with the perspective transformation, after dealing with the rotation, translation, and scale between the two planes. The way that this is done is by using an algorithm named Homography, that works as follows: [1] [6]

- (1) It is necessary to **map the corresponding points in the two planes**, having to have at least four points, for the transformation to be possible. For the case of this dissertation the vertices of the rectangle of the great area were selected and were mapped to the respective points in the 2D plane;
- (2) These points must be placed with the same center/centroid in both planes, making necessary adjustments, in one of the planes, **translation**. What has already been done with the execution of the *first transformation*.
- (3) Temporary lines are created that connect the previously mapped points, and these are extended until all intersected in a single point. When this point is found a 3x3 matrix is created, which contains the transformations necessary to deal with the **perspective**.

- (4) Finally, the matrix must be applied to the points that are intended to be mapped, so that they are placed in the correct position.

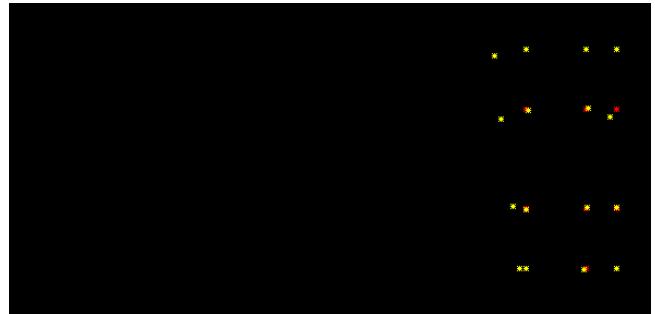


Figure 6: Result of the application of the first transformation and the Homography

The result of this phase is the one visible in the figure 6, where is possible to see that the interception point that came from the frame (yellow points) are almost aligned with the ones from the bidimensional plane (red points).

4.5 Mapping players

For the mapping of players is necessary to get the position of the players in the image. In the case this dissertation these positions come from the integration with the rest of *Player Detection* system, more specifically with other application named *Player Tracker*, that gives the exact positions of the players visible on the frame.

After getting the position of the players these positions should suffer the same transformations as the from the interception points of the field, this is, the ones calculated through the first transformation and the homography from the previous phase.

By applying the transformations for the ideal number of iterations, found in the first transformation, and the 3x3 matrix, found by the homography, to the player positions the result turns out to be the one from the figure 1, which turns out to be a very good mapping from a picture/frame to a bidimensional plane.

5 EVALUATION

The tests were performed after the first version of the project solution was developed, which affected the results of the tests in a negative way. However, the results should still be considered very positive with the exception of three of the four cases described here, which are rare occurrences in a football movie record.

To better understand the results it is important to know that the carried out tests contained two types of video: one

with static images of a frame; and another with videos of games.

There are four types of situations that may occur through the execution of the application:

- (1) **The good case/General case** - Where the frame to be analyzed has the correct image processing and the line identifications occur without any error, that could compromise the mapping of the field and the players. This is the most common case in system tests, corresponding to the vast majority of analyzed frames (happening for every 90 to -95 frames in 100);
- (2) **Light Trails** - This case happens when the games are held in fields whose lawns are part in the shade and part in the sun. Causing problems, because at least one rectangle referring to the area where the light is will be detected as a good rectangle, causing the detection of lines that despite being bad are classified as good, which can originate bad detection of the rectangle that corresponds to the great area. The result of this case is an image with similar to one in the figure 7;

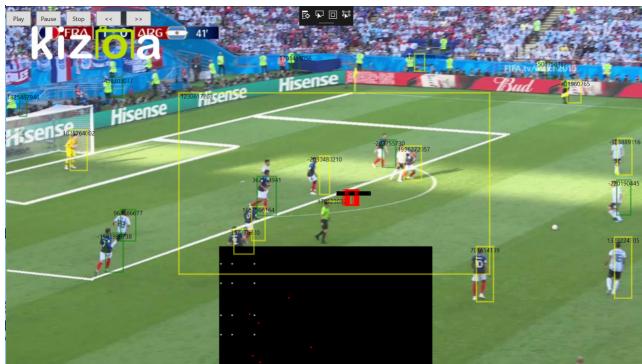


Figure 7: The result of a poor mapping, coming from a video with traces of light

- (3) **Failure to identify the great area** - In this second case there are two errors, which cause the application to malfunction, the first is the poor recognition of the lines in the great area and the second is the poor recognition of the player's size in the frame. This particular case occurs very rarely in the execution of the video by the application and its impact on the overall results is almost nil, because it is not too damaging to the statistics of each player, since it causes great discrepancies with the previous values and therefore their positions are ignored, as can be seen in the figure 8;
- (4) **Excess of people detected** - This is a case that does not come directly from the Field Detector application, which is the solution that intends to solve the problem

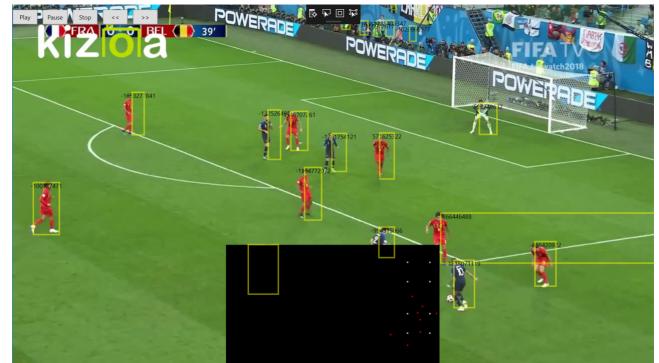


Figure 8: The result of a bad mapping, coming from a deficient recognition of the players and the lines of the great area

addressed by this dissertation, but that has some impact on it, because creates problems in the detection of field lines, due to generating more white data that is supposed to.

With this larger cluster of data, several lines are detected near the lines of large area, when applying the algorithm of line detection, which means that when applying the algorithm to join all segments of lines detected the great area lines are negatively influenced by the residue.

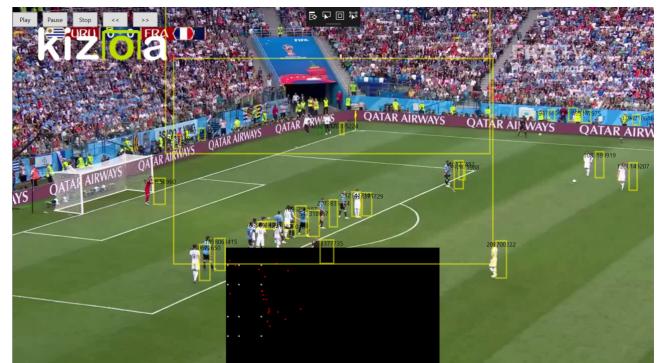


Figure 9: The result of a bad mapping, coming from a deficient recognition of the players and of a deficient agglomeration of lines

This can create problems in the detection of the area, or because the lines formed are very close to the lines of the great area and form rectangles very similar to the one intended, confusing the application, or due to these lines being so similar to the ones that are intended to be detected that the application cluster them together, forming a line that does not correspond correctly to

the ones from the great area and thus deforming the mapping, as can be seen in the figure 9.

6 CONCLUSIONS

It is possible to conclude that the mapping of a frame to a two-dimensional plane is closely related to the image treatment exerted on it, having a direct influence on the obtained results.

Within the scope of the problem, which is discussed in this dissertation, it is crucial to understand that:

- Is important to find which reference points should be taken into account in order to make the references for the mapping as accurate as possible. For the proposed solution the **reference points** were the ones from the **interception of the lines belonging to the great areas**;
- To obtain the lines of the great areas in a well-defined manner and with the least possible residue is necessary to apply and develop an **image treatment**, through the **HLS color model**, and **detection of rectangles** with restricted rules;
- The transformations that the points suffer (**rotation, translation, scale, and perspective**) while they are in a frame need to be treated properly. To be able to deal with this the proposed solution develops and uses specific methods and formulas (**first transformation and homography**).

As for the **proposed solution** is safe to say that it is **good in solving the problem**, since it **fills three of the four objectives** intended for the project (described at section 3) and because it can map correctly the players visible and detected in a frame.

The only objective that the project fails to achieve is the one related to the process-rate of the application where it fails to get the ideal 30fps. Due to the hardware being outdated is possible that it had compromised the process-rate of the solution, however, it was not possible to say that it was only because of the hardware.

REFERENCES

- [1] 2013. The Homography transformation. Retrieved October 24, 2018 from http://www.corrrmap.com/features/homography_transformation.php
- [2] Miguel Aleman-Flores, Luis Alvarez, Luis Gomez, and Daniel Santana-Cedres. 2014. Line detection in images showing significant lens distortion and application to distortion correction. *Pattern Recognition Letters* 36 (2014), 261-271.
- [3] Luis Alvarez, Luis Gomez, and Rafael Sendra. 2009. An Algebraic Approach to Lens Distortion by Line. *Journal of Mathematical Imaging and Vision* 35, 1 (2009), 36-50.
- [4] Luis Alvarez, Luis Gomez, and Rafael Sendra. 2010. Algebraic Lens Distortion Model Estimation. *Image Processing On Line* (2010), 1-10.
- [5] Navneet Dalal and Bill Triggs. 2005. Histograms of Oriented Gradients for Human Detection - ISBN: 0769523722. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (2005), 886-893.
- [6] Elan Dubrofsky. 2009. *Master essay: Homography Estimation*. Master's thesis. THE UNIVERSITY OF BRITISH COLUMBIA.
- [7] James J. Gibson. 1950. *The Perception of the Visual World* - ISBN: 9780837178363. KOUGHTON MIFFLIN COMPANY, Boston.
- [8] Qin Li, Lei Zhang, Jane You, David Zhang, and Prabir Bhattacharya. 2008. Dark line detection with line width extraction. *IEEE International Conference on Image Processing* (2008), 621-624.
- [9] Slawomir Mackowiak, Jacek Konieczny, Maciej Kurc, and Przemyslaw Mackowiak. 2010. Football Player Detection in Video Broadcast - ISBN: 9783642159060. In *Computer Vision and Graphics* (, ed.). Springer, Cham, Berlin, Heidelberg, 118–125.
- [10] K. Tomikawa and R. Saga. 2018. Conversion of Player Locations from Football Goal Scene Videos to a 2D Top View - ISBN: 9783319922690. In *Communications in Computer and Information Science*. Springer, Cham, 211–21.
- [11] J.R.R. Uijling, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders. 2012. Technical Report: Selective Search for Object Recognition. (2012).
- [12] Luiz H. P. Vieira, Emilio A. Pagnoca, Fabio Milioni, Ricardo A. Barbieri, Rafael P. Menezes, Luis Alvarez, Luis G. D'Alniz, Daniel Santana-Cedres, and Paulo R. P. Santiago. 2017. Tracking futsal players with a wide-angle lens camera: accuracy analysis of the radial distortion correction based on an improved Hough transform algorithm. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualization* 5, 3 (2017), 221-231.
- [13] Xinguo Yu, H.C. Lai, Sophie Liu, and Hon Wai Leong. 2005. A gridding Hough transform for detecting the straight lines in sports vÃrdeo - ISBN: 0780393317. *2005 IEEE International Conference on Multimedia and Expo* (2005).