A Universal Protocol to Benchmark Camera Calibration for Sports

Floriane Magera[1,2]   Thomas Hoyoux[1]   Olivier Barnich[1]   Marc Van Droogenbroeck[2]

[1] EVS Broadcast Equipment   [2] University of Liège, Belgium

Abstract

Camera calibration is a crucial component in the realm of sports analytics, as it serves as the foundation to extract 3D information out of the broadcast images. Despite the significance of camera calibration research in sports analytics, progress is impeded by outdated benchmarking criteria. Indeed, the annotation data and evaluation metrics provided by most currently available benchmarks strongly favor and incite the development of sports field registration methods, i.e. methods estimating homographies that map the sports field plane to the image plane. However, such homography-based methods are doomed to overlook the broader capabilities of camera calibration in bridging the 3D world to the image. In particular, real-world non-planar sports field elements (such as goals, corner flags, baskets, ...) and image distortion caused by broadcast camera lenses are out of the scope of sports field registration methods. To overcome these limitations, we designed a new benchmarking protocol, named ProCC, based on two principles: (1) the protocol should be agnostic to the camera model chosen for a camera calibration method, and (2) the protocol should fairly evaluate camera calibration methods using the reprojection of arbitrary yet accurately known 3D objects. Indirectly, we also provide insights into the metric used in SoccerNet-calibration, which solely relies on image annotation data of viewed 3D objects as ground truth, thus implementing our protocol. With experiments on the World Cup 2014, CARWC, and SoccerNet datasets, we show that our benchmarking protocol provides fairer evaluations of camera calibration methods. By defining our requirements for proper benchmarking, we hope to pave the way for a new stage in camera calibration for sports applications with high accuracy standards.

1. Introduction

Camera calibration, also known as camera resectioning, is a necessity in many computer vision tasks. It involves estimating the parameters of a camera model, usually the pinhole camera model, that approximates the physical camera that produces a given image. Applications that
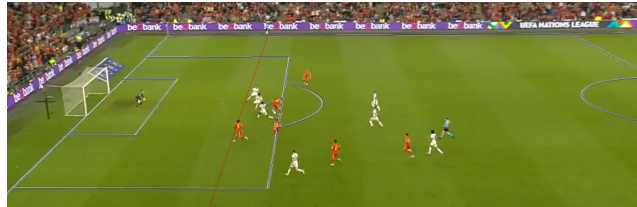

Figure 1. Illustration of a successful camera calibration. Lines superimposed in blue are obtained by projecting the markings of a soccer field. A perfect alignment is mandatory to enable high-precision applications such as offside position assessment (in red, a parallel to the goal line to decide on an offside situation). This paper proposes a new protocol, named ProCC, that has two advantages: (1) it is applicable to any sport, and (2) its evaluation is based on a new metric which is agnostic to the chosen camera type and model.

require accurate calibration are varied and include virtual reality [34], underwater measurement [35, 44], traffic analysis and surveillance [30, 38, 47], vehicle localization [56] or speed estimation [40], person re-identification [54], 3D reconstruction [29], *etc*. In this paper however, we focus on calibration of cameras used in sports event broadcasting as shown in Figure 1, and present a new protocol for the evaluation of the camera calibration quality that is usable for any camera type (static/moving, PTZ, fish-eye, wide angle) and model (pinhole, with/without radial distortion, *etc*.).

Camera calibration for sports. In the context of sport, computer vision systems are used for multiple purposes, such as augmented reality graphics [39], 3D ball trajectory reconstruction [3] or tracking [53], refereeing [19,20,31], or the computation of statistics regarding position, speed, and acceleration of balls, players, bats, sticks, pucks, *etc*. [27, 36, 50, 52]. The expectations in terms of precision and robustness of these systems is constantly rising. Both the fact that it is considered as a valid alternative to GPS trackers for player tracking systems and the fact that the FIFA supported the use of the SAOT (Semi-Automated Offside Technology) for the 2022 Football World Cup are good indicators

of the trust put in computer vision systems. Sports analysis is an active subject of research, with several recent datasets gathering broadcast images with a wide range of annotations serving for game analysis, such as DeepSportRadar dataset for Basketball [51], or the SoccerNet datasets for soccer [8, 12, 17, 32]. The SoccerNet dataset which was initially created for action spotting tasks [2, 17] covers a wide range of tasks for soccer analysis. In particular, the game state recognition task, one of the latest challenges proposed by the SoccerNet team, heavily relies on the ability to calibrate the camera to produce a strategic minimap of the players' localization [28]. This challenge paves the way to a 3D reconstruction that will require to be able to calibrate any camera along the field, including close-up cameras in a multi-view setup.

Camera calibration in the context of sports benefits from the presence of the sports field whose dimensions are specified by the rules of the game (see for example [23, page 32] for soccer, [14] for basketball, [16] for volleyball, and [24] for ice hockey). Their well-known shape and their presence in most sports images make them convenient calibration patterns. Using the field model as a calibration pattern, camera calibration techniques can thus express the parameters by a projection matrix, and sports field registration techniques express the mapping between the field plane and the image by a perspective transform, better known as a homography [46]. Note, however, that the rules of games allow some tolerance on field dimensions, so calibration systems have to cope with some dimension uncertainties. In this regard, goal posts are supposed to be perfectly sized, which makes them adequate to serve as calibration landmarks.

As the terms of "camera" calibration and "sports field" registration can be sometimes used interchangeably in the literature, we deem important to emphasize the difference between them. A sports field registration technique aims to estimate a homography, *i.e.* the transformation between the 3D sports field plane and the image. This mapping is not defined outside these two planes, which is insufficient for all 3D applications that involve elements outside the field plane. This is a blind spot in all papers that mention 3D applications such as player and ball tracking, tactics analysis, augmented reality, *etc*. On the other hand, a camera calibration technique provides a mapping between the 3D world (not only the field plane) and the image, which makes it suited for the aforementioned applications. Sports field registration is an approximate attempt or a first step towards camera calibration, and this is why, in this paper, we consider the homography as a camera model, despite its practical limitations. However, we show that our protocol is applicable to any model, broadening the possibilities in terms of camera models.

Outline. The rest of the paper is organized as follows. We present the related work on camera calibration for sports, in-

cluding a review of existing datasets, in Section 2. Section 3 describes the key elements of our benchmarking protocol, named ProCC, the new metric JaC[1] that we introduced in SoccerNet-calibration, as well as a more relevant ground-truth type, to express the camera calibration quality. Our benchmarking protocol, including the annotation procedure and its metric, are the main contributions of this paper. Section 4 applies our protocol to the evaluation of camera calibration for sports broadcast events. This section also includes an experimental analysis of the protocol for different camera models, illustrating the universality of our protocol and showing that previous protocols have limitations that our new protocol can overcome. Section 5 discusses the results and explains why current calibration techniques developed in the literature overlook the broader capabilities of camera calibration in bridging the 3D world to the image. Finally, Section 6 presents a brief conclusion.

2. Related work

In the current state of the art, most of the evaluation procedures are based on homography annotations (to the best of our knowledge, basketball [22, 25] and athletics [1] are the only exceptions) which restricts the evaluation to the sports field plane, even if some techniques compute camera parameters. Indeed, all the following camera calibration techniques in the literature [4, 10, 41, 48] use the pinhole model, whose camera parameters can be converted to the homography that maps the image to the sports field plane.

2.1. Datasets to benchmark camera calibration

The latest techniques in camera calibration for sports all leverage the presence of a sports field to understand the mapping between the image and the world.

By browsing the scientific literature, we found a dozen datasets on which researchers evaluated their methods. Table 1 shows that most datasets are only relevant for sports field registration, as their annotations consist of homographies. For some datasets, it seems that the ground truth has not been annotated once and for all, such that several researchers have created their own annotations, which does not ensure the comparability of their results [43].

The nature of the annotations, the relatively small sizes of the datasets for some sports, as well as the difficulty of getting access to the data, make the creation of reliable camera calibration algorithms difficult.

2.2. Evaluation of camera calibration in sports

Today, most authors evaluate their performance using an intersection over union metric, more specifically the

---

[1]We have renamed the metric from AC, used in SoccerNet-calibration, to JaC to avoid any confusion with the notion of accuracy as used in binary classification.

Name	Sport	Open data	Size	Annotations	Techniques
WorldCup 14	Soccer	Yes	395	Homographies	[4, 5, 7, 10, 21, 26, 33, 36, 37, 41–43, 48, 49, 55]
TS-WorldCup	Soccer	Yes	3,812	Homographies	[5]
SoccerNet-calibration	Soccer	Yes	21,132	Field markings	[48]
CARWC	Soccer	Yes	4,207	Homographies	[11]
SportLogiq	Ice Hockey	No	1.67M	Not specified	[21, 26, 43]
SportsFields by Amazon	Multi-sports	No	2,967	Homographies	[37]
Volley ball	Volley ball	Yes	470	Homographies	[4, 43]
College Basketball	Basketball	No	640	Homographies	[41]
DeepSportRadar	Basketball	Yes	728	Pinhole model	[51]
3DMPB	Basketball	Yes	10k	Pinhole model	[22, 45]
Athletics	Athletics	Yes	10k	Pinhole model	[1]

Table 1. Main current datasets dedicated to camera calibration for different sports. In this table, we also mention if the annotations are available, the number of images (Size) for which annotations are provided, if the annotations are specific to homographies or the pinhole model, and list some camera calibration techniques using a given dataset. From this table, one can see that the WorldCup 14 dataset is the most popular dataset for benchmarking, despite its limitations, as explained in this paper.
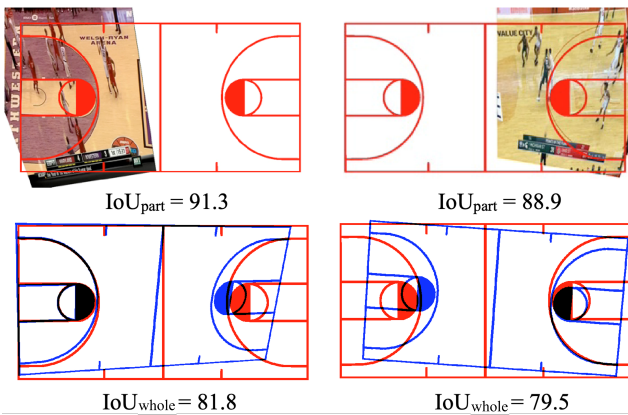

IoU$_{part}$ = 91.3   IoU$_{part}$ = 88.9

IoU$_{whole}$ = 81.8   IoU$_{whole}$ = 79.5

Figure 2. Visualization of the IoU$_{whole}$ and IoU$_{part}$ metrics. Illustration taken from [41] (©IEEE, 2020).

Reference	mIoU$_{whole}$	medIoU$_{whole}$	mIoU$_{part}$	medIoU$_{part}$
Homayounfar et al. [21]	83	-	-	-
Sharma et al. [42]	-	-	91.4	92.7
Chen and Little [4]	89.2	91.0	94.7	96.2
Jiang et al. [26]	89.8	92.9	95.1	96.7
Sha et al. [41]	88.3	92.1	93.2	96.1
Citraro et al. [10]	90.5	91.8	-	-
Cioppa et al. [7]	79.8	81.7	88.5	92.3
Li et al. [33]	92.1	94.3	95.1	96.7
Tsurusaki et al. [49]	-	-	**97**	-
Nie et al. [37]	91.6	93.4	95.9	97.1
Shi et al. [43]	**93.2**	**94.9**	96.6	**97.8**
Chu et al. [5]	91.2	93.1	96.0	97.0
Zhang and Izquierdo [55]	90.0	92.8	95.3	96.9
Theiner and Ewerth [48]	-	-	95.3	96.6
Maglo et al. [36]	92	94.1	96.3	97.4

Table 2. Consolidated leaderboard by collecting published values on the World Cup 14 benchmarking dataset. The values are the mean or median intersections over union (denoted respectively by mIoU and medIoU), on the whole field or on the visible part of it, as mentioned in the references. In this table, references are organized by their year of publication, and the best values are given in bold.

IoU$_{whole}$ metric, that was first introduced by Homayounfar et al. [21] for the performance evaluation of their algorithm on soccer and ice hockey content. This metric measures the intersection over union between the real-world rectangle defined by the field template and its successive projection then deprojection using both the annotated homography and the camera parameters in the 3D world. This metric was extended by Sharma et al. [42] into another metric, denoted by IoU$_{part}$, to only account for the parts of the field that are visible in the image. These two metrics, which are illustrated in Figure 2, were applied to other sports such as volleyball [4, 43], basketball [41], tennis and American football [37].

In 2022, we launched the SoccerNet-calibration challenge, which is an attempt to further improve calibration techniques, but despite the organization of two SoccerNet-calibration challenges [9, 18], the benchmark of the World Cup 14 (WC14) dataset remains popular and the most commonly used benchmark. Therefore, we have collected the results of the current state-of-the-art methods, organized by their year of publication, regarding the WC14 benchmark in a consolidated leaderboard given in Table 2. However, the evaluation of one method stands out: Theiner et al. [48] evaluated their method both on the WC14 and SoccerNet-calibration datasets and, even further, they already showcased some discrepancies between our protocol ProCC and the WC14 evaluation protocol when applied to the same dataset.

Furthermore, Nie et al. [37] and Chu et al. [5] provide an exhaustive evaluation by adding reprojection and projection errors. The projection error measures the average

distance in meters between the projection of pixels sampled in the field image, using the inversion of both the estimated homography and the ground-truth homography. The reprojection error measures the average distance between reprojected points in the image using the ground-truth homography and the predicted homography, this distance is then normalized by the image height.

All the aforementioned metrics rely on the annotated homographies of the dataset, which are a limited and simplified interpretation of the observable field markings in the image. Moreover, with the small size of the World Cup 14 dataset and the small improvements in performance obtained in the last few years, concerns have been raised about the relevance of this dataset. Indeed, Chu *et al*. [5] proposed a new dataset called TS-WorldCup to increase the dataset size and allow for tracking evaluation, Claasen *et al*. [11] further corrected both the World Cup 14 and the TS-WorldCup annotations in a revised version named CARWC, which proposes homographies that are more precise. While the latter two improvements address some limitations of the current benchmarking protocol, we wish to go one step further by proposing to get rid of homographies as ground-truth data for camera calibration in sports. In our opinion, one major problem with the current benchmarks is the imperative of a restricted and limited camera model for fulfilling the requirements of professional use. Indeed, around the field for a top-tier sports game, there can be up to 50 cameras, of a wide variety of quality: from wide-angle to super slow-motion cameras or fish-eye cameras, one model does not fit all of them. To circumvent this issue, we propose a new benchmarking protocol that allows for the use of any camera model. Furthermore, if we were to actually choose only one camera model for a set of broadcast cameras, our protocol would allow deducing which model would be the best fit according to our model-agnostic metric —or usual metrics such as the reprojection error— instead of choosing an axiomatic, arbitrary camera model from the outset.

3. A benchmarking protocol for model-agnostic camera calibration

In this section, we describe our new benchmarking protocol that is based on two main pillars: annotations and a metric. Both are designed by assuming that a good camera calibration algorithm will be able to produce results that allow a minimal reprojection error, which is, in fact, the only reasonable assumption when there is no actual ground-truth knowledge about the broadcast cameras that captured the images of the datasets.

3.1. Annotations: beyond homographies

As an essential requirement for evaluating camera calibration is to handle different camera models, we must first
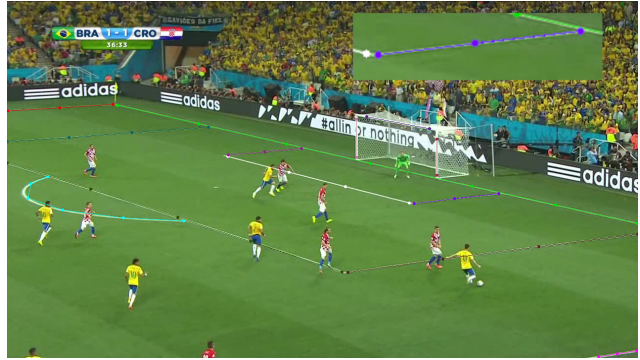

Figure 3. Illustration of annotations on a SoccerNet-v3 [6] image as used for the camera calibration challenge. As shown in the zoomed snapshot, annotations consist of points and the labels of the objects they belong to.

change the type of the annotations and, subsequently, provide a revised evaluation metric.

In contrast to camera calibration techniques that rely on a specific pattern that will only be seen once and self-calibration techniques that do not require any calibration target and instead usually rely on reference frames, sports images have the unique advantage of always displaying at least partially a specific pattern, which is the sports field. We suggest leveraging this advantage, which is already done in most benchmarking and calibration techniques, but also pushing forward, by using as ground truth a lower-level interpretation of the sports field image: semantic point annotations of the sports field markings, which are annotations that are valid for any type of camera.

A sports field can be decomposed into a set of simple geometrical elements that are points, lines, circles, or ellipses. For each of these simple constituent elements of the field, we propose to assign a unique semantic label. In general, the field is only partially visible in the image, so for a given image, only a few elements need to be annotated. Practically, we annotate points along each element and thus, obtain a set of points for each semantic element of the field. To illustrate this, in SoccerNet, the annotations consist of polylines, *i.e.* sequential lists of 2D points along each soccer field element. Ideally, annotations should be regularly spaced, and the density of annotated points might be increased depending on the curvature of the field element as it appears in the image. Each element of the soccer field corresponds to a class, and in total, we count 26 semantic classes. An example of annotation is shown in Figure 3; in this figure, the points are provided by human annotators, while polylines are superimposed by an annotation tool.

By employing a semantic annotation of soccer field elements represented by a polyline, we meet the first requirement of a universal evaluation protocol, in the sense that the ground truth is independent of the used camera model.

The main advantages of this annotation type in the image domain are its independence from the task to be solved and the ease to validate it by the human eye. As the annotations do not necessarily cover all field markings pixels, we will show that our metric addresses the issue that while the annotated points are part of field elements, it is possible that neighboring pixels also belong to the same element, despite being unannotated.

Finally, the use of semantic annotations opens up possibilities for more complex calibration scenarios. For example, if a system can triangulate a 3D point and knows its projection in the image, it may use it for calibration, even if the point belongs to a moving object such as a player. This opens the way to close-up camera calibration in a multi-view setup.

We now show how to evaluate a camera calibration based on this ground truth.

3.2. Evaluation metric

Given a field model and the corresponding annotations in the image, we can evaluate the quality of camera parameters by assessing how well the camera projection of the 3D field superimposes on the annotations. A camera model definition includes the mathematical imaging function that allows for the computation of the 2D image of a 3D object. This function is also called the *projection* function. In our protocol, since the evaluation must be independent of the model, the only requirement for a camera model is that this function must be defined and, when it is given a 3D point X, it must be able to provide its corresponding 2D location x in the image plane. We denote the projection function of a camera model by π and, therefore, $x = \pi(X)$.

The metric that we propose is inspired by the reprojection error, while aiming to address its shortcomings. The reprojection error usually measures the Euclidean distance between the observation (*e.g.* the annotation) of an object and the projection of that 3D object model using the estimated camera model. The mechanism of projection is the source of several difficulties. First, depending on the chosen camera model, the reprojection error may be undefined when the projection of the 3D object does not land in the image and, more generally, is unsuited to handle hallucinated or missing objects. Second, a projection may not exactly match an annotation, but may still be superimposed on visible field marks, since the field markings have a thickness that can make them visible for more than one pixel in width. Finally, it is inconvenient to grasp the meaning of an unbounded metric that, in our case, can vary between zero and infinity. For these reasons, by thresholding the reprojection errors, we obtain another metric that can be intuitively understood as the proportion of field elements that are correctly imaged, bridging the camera calibration evaluation with an object detection problem. Indeed, the underlying

intuition behind our metric is that the quality of the camera parameters reflects how well the parameters can reproject objects close to their image.

Since we defined that each field marking element is annotated with as many points as necessary to constitute a fitting polyline, we mark a field element as correctly detected if its reprojection in the image is close enough to each of the annotated points of the polyline. The closeness criterion is defined based on a threshold value τ in pixels. More formally, a field element is defined by a polyline L, which is an ordered list of 3D points. Its projection $\pi(L)$ gives a 2D polyline (*i.e.* an ordered list of 2D points) which also defines a list of line segments S by considering pairs of the consecutive 2D points. The distance between an annotated point x_i and the projection of the field element $\pi(L)$ is given by the minimal Euclidean distance between the point x_i and the segments S constituting $\pi(L)$.

In the following, we define the function $d(.)$ computing the distance between a point and a line segment. More specifically, given the projection c of the point x on the line passing through the segment extremities a and b, we define the distance between the point x and the line segment S_{ab} defined by a and b with the traditional Euclidean distance function as follows:

$$d(x, S_{ab}) = \begin{cases} \|x - a\|_2 & z \leq 0, \\ \|x - c\|_2 & 0 < z < 1, z = \dfrac{c-a}{b-a}, \\ \|x - b\|_2 & z \geq 1, \end{cases} \quad (1)$$

where the variable z is introduced to account for the fact that the distance to the line segment is equal to the distance to one of its extremities if the point projection c lands outside the segment. Finally, we say that a field element L is correct if all the points x_i annotated for this element are less than τ pixels away from the projection of said field element L. In mathematical terms, we then have that:

$$\min_{S \in \pi(L)} d(x_i, S) < \tau, \forall x_i. \quad (2)$$

The correctly detected field markings are counted as true positives TP_τ. Both the hallucinated field markings and wrongly detected field elements whose reprojection lands further than τ pixels from the annotations are false positives FP. Missing field elements are counted as false negatives FN. Finally, we define our JaC_τ metric, the Jaccard index for camera calibration or JaC, to evaluate the calibration "accuracy" as follows:

$$\mathsf{JaC}_\tau = \frac{\mathsf{TP}_\tau}{\mathsf{TP}_\tau + \mathsf{FN} + \mathsf{FP}}. \quad (3)$$

We propose to use this metric, parametrized by the reprojection error τ, for the best comprehension of the camera parameters quality.

4. Results

For practical reasons, our experiments focus on soccer because this sport has the most available techniques and data. Furthermore, we are restricted to soccer due to a lack of semantic annotations for other sports, despite that ProCC is applicable to any sport for which there are measurable 3D references. Hereafter, we describe experiments on the WC14, CARWC, and the SoccerNet datasets to demonstrate our benchmarking protocol.

4.1. Description of the datasets

For all the datasets, the experiments employ soccer field markings and goal posts that are decomposed into distinct classes.

SoccerNet. The SoccerNet dataset contains 21,132 images that were recently annotated with soccer field markings elements, as well as goal posts elements, which provides 3D correspondences, thus outside the field plane [6]. In total, the annotations consist of 167,589 field marking lines or circles and 53,577 goal posts elements that are grouped into 26 classes for soccer fields. Each element is annotated with its two extremities for rectilinear elements and by as many points as needed to fit curves for non-rectilinear elements.

World Cup 14. For the sake of comparison, the World Cup 14 test set has been manually annotated following our convention. Indeed, we annotated the 186 images with points along soccer field elements, resulting in a total of 1,681 soccer field elements annotated with several points.

4.2. Establishing a better camera model for broadcast cameras

The goal of our experiments is twofold: validate our evaluation protocol, and demonstrate that there is a better camera model for broadcast cameras. This is why we first compare both qualitatively and quantitatively ground-truth homographies of the WC14 and CARWC datasets with estimated camera parameters following a richer camera model.

In our experiments, we selected a threshold τ of 5 pixels for our JaC_5 metric, which is a value that is suited for distinguishing between methods given the current quality standards of the different camera calibration methods that exist in the literature. When we want to differentiate quite precise methods, we tune the pixel threshold to 2 pixels.

As illustrated in Figure 4 and explained in the above, the reprojection of the soccer field model using *ground-truth homographies* (see columns 1 and 2) fails to match the images of the field markings. This is an indication that the use of these annotations is suboptimal if we agree that a good camera calibration algorithm should be able to match the image of known 3D objects such as soccer fields, which is a reasonable and tractable solution in the absence of actual ground truth for the camera parameters.

Camera model	JaC_5 ($\uparrow$)	Reprojection error ($\downarrow$)
Homography (WC14 annotations [21])	67.4	3.07
Homography (CARWC annotations [11])	79.1	1.79
Pinhole camera parameters with one radial distortion coefficient ([13])	92.5	1.44

Table 3. Comparison of camera models on the World Cup 14 dataset.

In Table 3, we compare different camera models and highlight the limitations of using homographies as ground truth. The manual annotations and JaC_5 metric allow us to quantify the quality of the annotations of the WC14 and CARWC datasets compared to the camera parameters estimated with the Xeebra product. Xeebra [13] is a professional product for Video Assistant Referee, whose Offside Technology features are certified by the FIFA Quality Program [15]. With these experiments, our evaluation protocol confirms the concerns raised [5, 11, 48] about the quality of the WC14 ground truth. We obtain different results for the WC14 homographies evaluation than Theiner *et al.* [48] which is explained by the fact that we used our annotations. We also establish that the consolidated annotations provided by Claasen *et al.* [11] in CARWC are a welcome enhancement in terms of precision.

However, despite visible enhancements provided by the CARWC annotations to the WC14 dataset, the homography models fail to reproduce the images of the field markings, while we obtain better results by estimating camera parameters with Xeebra. This experiment demonstrates that the qualitative insights shown in Figure 4 are further supported by both the JaC_5 metric and the reprojection error, comforting the idea that our evaluation protocol is relevant and fulfills a need for correct evaluation. Moreover, the middling results of the careful CARWC annotations suggest that the problem does not lie with the quality of the provided annotations, but rather with the nature of the annotation. Indeed, we attribute the better results of the camera parameters of Xeebra to its inclusion of radial distortion parameters. In an attempt to demonstrate both our protocol's ability to assess the quality of different camera models and to further prove our last hypothesis, we evaluated Xeebra's results on SoccerNet as well.

In the next experiment, as can be seen in Table 4, we establish that on the SoccerNet test set, which is a much larger and thus more diverse and challenging dataset, it is a necessity to consider the radial distortion. Indeed, when radial distortion is considered in the optimization of the camera

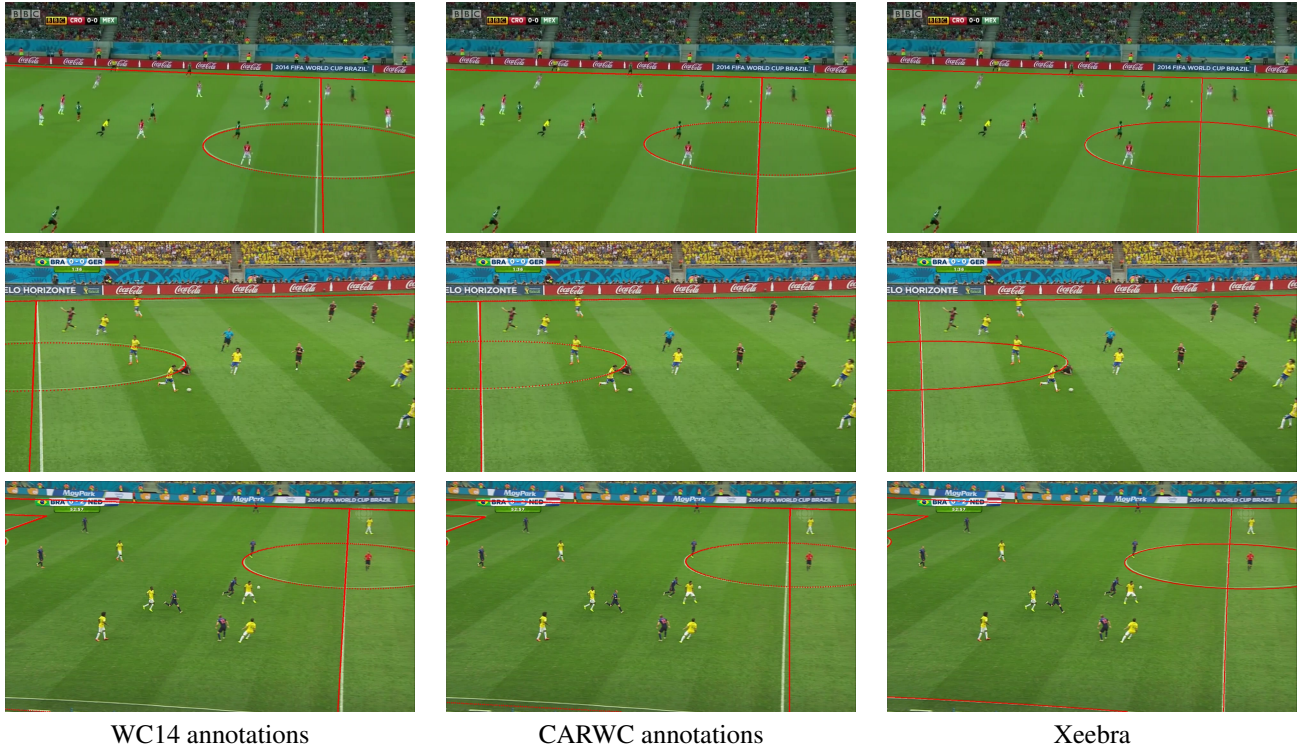| | WC14 annotations | CARWC annotations | Xeebra |

Figure 4. Comparison of reprojected field elements based on different types of annotations: in the first column, the WC14 homographies are used to project the soccer field model. The second column corresponds to the CARWC annotations. Both these annotations fail to provide correct reprojections. Finally, the third column displays results obtained with the Xeebra product.

	JaC$_5$ ($\uparrow$)	JaC$_2$ ($\uparrow$)	Reprojection error ($\downarrow$)
P	78.7	40.2	4.51
R	83.1	54.3	4.01

Table 4. Comparison of camera models on the SoccerNet-calibration dataset. P corresponds to the simplified pinhole camera model, and R corresponds to the simplified pinhole model extended with one radial distortion coefficient.

parameters, better results are obtained according to our protocol.

In these experiments, we have shown that our evaluation protocol obtains results in agreement with qualitative and quantitative evaluations, and that it allows to evaluate different camera models, which enables us to support our second contribution, stating that soccer broadcast cameras are better modeled with radial distortion.

5. Discussion

With the previous results, we have shown that better camera modeling allows for better precision in the image. In this section, we want to stress the impact of such improvements on the real-world applications of camera calibration algorithms.

Considering that many sports field registration methods emphasize the relevance of their research in applications like player tracking, it is essential to highlight the pivotal role of selecting an appropriate camera model for efficient player tracking on a sports field. In Figure 5, we show that, depending on the chosen camera model to link the image and the physical world, the differences in the estimated 3D positions can often exceed one meter. Such variations will inevitably impact the accuracy of player tracking throughout an entire sports game. Hence, we propose initiating a discussion on optimal modeling strategies before venturing into applications that may not be practically feasible. With our newly introduced benchmarking protocol, we enable a systematic evaluation that facilitates the exploration and analysis of the best models in this context.

Still, our protocol presents some shortcomings. The inherent limitations of our protocol lies in the quantity of field elements present in the images. Yet, the quality of the evaluation gets better when the number of elements increases and when the field markings are well distributed in the image. Indeed, camera models may become overparameterized (and be overfitted) in the case of images that show only few field elements. Another limitation is the fact
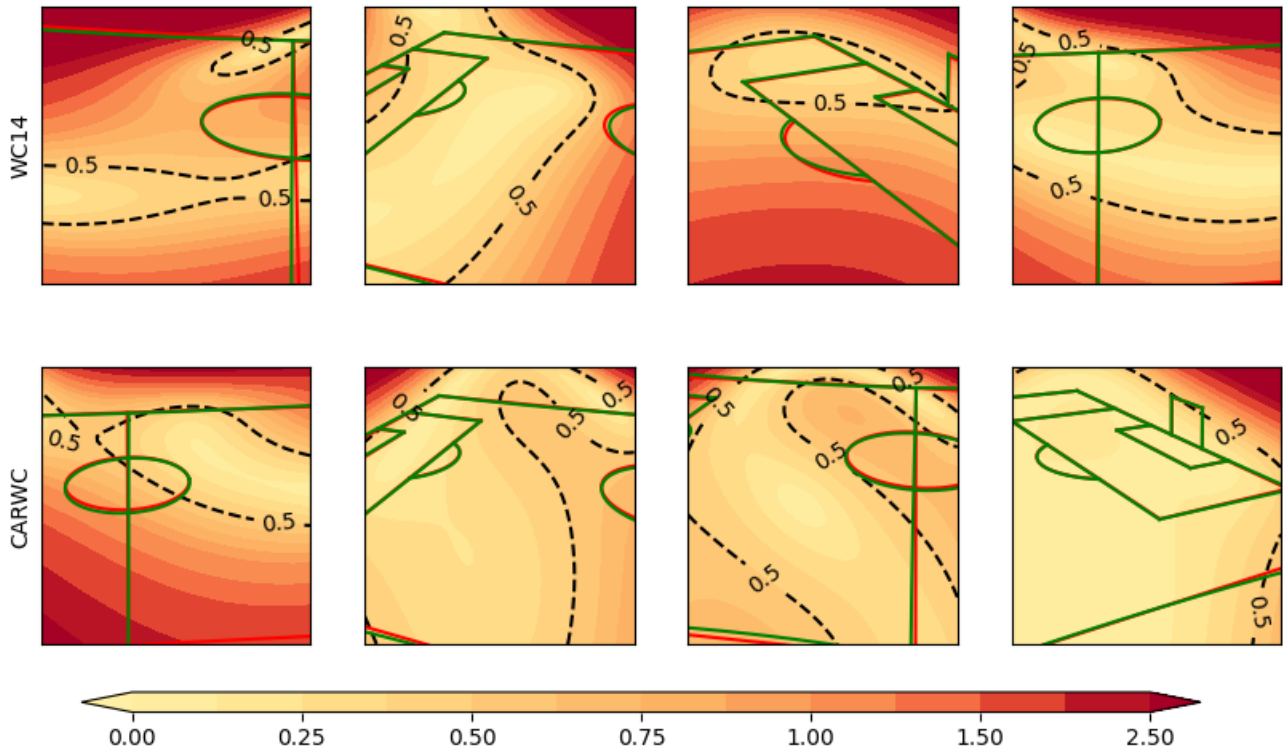
Figure 5. Illustration of the disagreement between estimations from two camera models: the ground-truth homography (see the red wireframes) and the richer model combining the pinhole and radial distortion used in Xeebra (see the green wireframes). The small difference between the two reprojected wireframes is misleading, as the superimposition of contour plots on the field shows that some parts of the sports field are seen over 2.5 meters apart by these two camera models. We have also plotted the contour line for a difference of 50 cm (see the dashed lines), which is the minimal accuracy standard demanded by the FIFA for Offside Technologies, to highlight why methods developed with older calibration protocols such as the WC14 or CARWC would fail to meet professional requirements.

that it might be possible that very different camera parameters obtain the same results due to the symmetry of the sports field template, for instance; the uniqueness of the results is not guaranteed. However, introducing a means of managing all kinds of ambiguities, which in practice are rather rare, would make the evaluation protocol considerably more complex. A possible yet costly solution to this issue is to acquire real ground truth for the camera parameters, which requires new datasets and much greater means to equip the production cameras with sensors.

6. Conclusion

In conclusion, this study demonstrates the inadequacy of current sport field registration benchmarks due to suboptimal ground truth. Considering the availability of superior camera models, we advocate abandoning metrics reliant on homographies, and propose a new type of annotation, combined with proper metrics, leading to the definition of a camera model-agnostic evaluation protocol. Furthermore, our protocol proves that richer camera models, such as the pinhole model augmented with radial distortion instead of

homographies obtain better results on broadcast cameras, which emphasizes the need to evaluate camera calibration algorithms independently of the camera model. Finally, it should be mentioned that our evaluation methodology applies to multi-view camera systems, as one can imagine that 3D elements visible in the calibrated cameras can be used, like any annotated data, as key points to calibrate other cameras. By doing so, not only have we have proven that our protocol improves camera calibration in sports, but also that there are new opportunities that are beyond the reach of methods solely based on field registration.

References

[1] Tobias Baumgartner and Stefanie Klatt. Monocular 3D human pose estimation for sports broadcasts using partial sports field registration. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*,

pages 5109–5118, Vancouver, Can., Jun. 2023. Inst. Electr. Electron. Eng. (IEEE). 2, 3

[2] Bruno Cabado, Anthony Cioppa, Silvio Giancola, Andrés Villa, Bertha Guijarro-Berdiñas, Emilio Padrón, Bernard Ghanem, and Marc Van Droogenbroeck. Beyond the Premier: Assessing action spotting transfer capability across diverse domains. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, Seattle, WA, USA, Jun. 2024. 2

[3] Vanyi Chao, Ankhzaya Jamsrandorj, Yin May Oo, Kyung-Ryoul Mun, and Jinwook Kim. 3D ball trajectory reconstruction of a ballistic shot from a monocular basketball video. In *Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, pages 1–6, Singapore, Singapore, Oct. 2023. Inst. Electr. Electron. Eng. (IEEE). 1

[4] Jianhui Chen and James J. Little. Sports camera calibration via synthetic data. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 2497–2504, Long Beach, CA, USA, Jun. 2019. Inst. Electr. Electron. Eng. (IEEE). 2, 3

[5] Yen-Jui Chu, Jheng-Wei Su, Kai-Wen Hsiao, Chi-Yu Lien, Shu-Ho Fan, Min-Chun Hu, Ruen-Rone Lee, Chih-Yuan Yao, and Hung-Kuo Chu. Sports field registration via keypoints-aware label condition. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3522–3529, New Orleans, LA, USA, Jun. 2022. Inst. Electr. Electron. Eng. (IEEE). 3, 4, 6

[6] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. Scaling up SoccerNet with multi-view spatial localization and re-identification. *Sci. Data*, 9(1):1–9, Jun. 2022. 4, 6

[7] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Floriane Magera, Olivier Barnich, Bernard Ghanem, and Marc Van Droogenbroeck. Camera calibration and player localization in SoccerNet-v2 and investigation of their representations for action spotting. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 4532–4541, Nashville, TN, USA, Jun. 2021. 3

[8] Anthony Cioppa, Silvio Giancola, Adrien Deliege, Le Kang, Xin Zhou, Zhiyu Cheng, Bernard Ghanem, and Marc Van Droogenbroeck. SoccerNet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 3490–3501, New Orleans, LA, USA, Jun. 2022. Inst. Electr. Electron. Eng. (IEEE). 2

[9] Anthony Cioppa, Silvio Giancola, Vladimir Somers, Floriane Magera, Xin Zhou, Hassan Mkhallati, Adrien Deliège, Jan Held, Carlos Hinojosa, Amir M. Mansourian, Pierre Miralles, Olivier Barnich, Christophe De Vleeschouwer, Alexandre Alahi, Bernard Ghanem, Marc Van Droogenbroeck, Abdullah Kamal, Adrien Maglo, Albert Clapés, Amr Abdelaziz, Artur Xarles, Astrid Orcesi, Atom Scott, Bin Liu, Byoungkwon Lim, Chen Chen, Fabian Deuser, Feng Yan, Fufu Yu, Gal Shitrit, Guanshuo Wang, Gyusik Choi, Hankyul Kim, Hao Guo, Hasby Fahrudin, Hidenari Koguchi, Håkan Ardö, Ibrahim Salah, Ido Yerushalmy, Iftikar Muhammad, Ikuma Uchida, Ishay Be'ery, Jaonary Rabarisoa, Jeongae Lee, Jiajun Fu, Jianqin Yin, Jinghang Xu, Jongho Nang, Julien Denize, Junjie Li, Junpei Zhang, Juntae Kim, Kamil Synowiec, Kenji Kobayashi, Kexin Zhang, Konrad Habel, Kota Nakajima, Licheng Jiao, Lin Ma, Lizhi Wang, Luping Wang, Menglong Li, Mengying Zhou, Mohamed Nasr, Mohamed Abdelwahed, Mykola Liashuha, Nikolay Falaleev, Norbert Oswald, Qiong Jia, Quoc-Cuong Pham, Ran Song, Romain Hérault, Rui Peng, Ruilong Chen, Ruixuan Liu, Ruslan Baikulov, Ryuto Fukushima, Sergio Escalera, Seungcheon Lee, Shimin Chen, Shouhong Ding, Taiga Someya, Thomas B. Moeslund, Tianjiao Li, Wei Shen, Wei Zhang, Wei Li, Wei Dai, Weixin Luo, Wending Zhao, Wenjie Zhang, Xinquan Yang, Yanbiao Ma, Yeeun Joo, Yingsen Zeng, Yiyang Gan, Yongqiang Zhu, Yujie Zhong, Zheng Ruan, Zhiheng Li, Zhijian Huangi, and Ziyu Meng. SoccerNet 2023 challenges results. *CoRR*, abs/2309.06006, 2023. 3

[10] Leonardo Citraro, Pablo Márquez-Neila, Stefano Savarè, Vivek Jayaram, Charles Dubout, Félix Renaut, Andrés Hasfura, Horesh Ben Shitrit, and Pascal Fua. Real-time camera pose estimation for sports fields. *Mach. Vis. Appl.*, 31(3), Mar. 2020. 2, 3

[11] Paul J. Claasen and Jill P. de Villiers. Video-based sequential Bayesian homography estimation for soccer field registration. *CoRR*, abs/2311.10361, 2023. 3, 4, 6

[12] Adrien Deliège, Anthony Cioppa, Silvio Giancola, Meisam J. Seikavandi, Jacob V. Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B. Moeslund, and Marc Van Droogenbroeck. SoccerNet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, pages 4508–4519, Nashville, TN, USA, Jun. 2021. 2

[13] EVS Broadcast Equipment. Multi-camera review system - Xeebra. https://evs.com/products/video-assistance/xeebra, Jun. 2022. 6

[14] FIBA. 2022 official basketball rules: Basketball rules & basketball equipment. Technical report, Fédération

Internationale de Basketball, Mies, Switzerland, 2022. 2

[15] FIFA. Handbook of test methods for virtual offside line assessment. Technical report, Fédération Internationale de Football Association, Zurich, Switzerland, Jun. 2019. 6

[16] FIVB. Official volleyball rules. Technical report, Fédération Internationale de Volleyball, Lausanne, Switzerland, 2021. 2

[17] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. SoccerNet: A scalable dataset for action spotting in soccer videos. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 1792–179210, Salt Lake City, UT, USA, Jun. 2018. Inst. Electr. Electron. Eng. (IEEE). 2

[18] Silvio Giancola, Anthony Cioppa, Adrien Deliège, Floriane Magera, Vladimir Somers, Le Kang, Xin Zhou, Olivier Barnich, Christophe De Vleeschouwer, Alexandre Alahi, Bernard Ghanem, Marc Van Droogenbroeck, Abdulrahman Darwish, Adrien Maglo, Albert Clapés, Andreas Luyts, Andrei Boiarov, Artur Xarles, Astrid Orcesi, Avijit Shah, Baoyu Fan, Bharath Comandur, Chen Chen, Chen Zhang, Chen Zhao, Chengzhi Lin, Cheuk-Yiu Chan, Chun Chuen Hui, Dengjie Li, Fan Yang, Fan Liang, Fang Da, Feng Yan, Fufu Yu, Guanshuo Wang, H. Anthony Chan, He Zhu, Hongwei Kan, Jiaming Chu, Jianming Hu, Jianyang Gu, Jin Chen, João V. B. Soares, Jonas Theiner, Jorge De Corte, José Henrique Brito, Jun Zhang, Junjie Li, Junwei Liang, Leqi Shen, Lin Ma, Lingchi Chen, Miguel Santos Marques, Mike Azatov, Nikita Kasatkin, Ning Wang, Qiong Jia, Quoc Cuong Pham, Ralph Ewerth, Ran Song, Rengang Li, Rikke Gade, Ruben Debien, Runze Zhang, Sangrok Lee, Sergio Escalera, Shan Jiang, Shigeyuki Odashima, Shimin Chen, Shoichi Masui, Shouhong Ding, Sin-wai Chan, Siyu Chen, Tallal El-Shabrawy, Tao He, Thomas B. Moeslund, Wan-Chi Siu, Wei Zhang, Wei Li, Xiangwei Wang, Xiao Tan, Xiaochuan Li, Xiaolin Wei, Xiaoqing Ye, Xing Liu, Xinying Wang, Yandong Guo, Yaqian Zhao, Yi Yu, Yingying Li, Yue He, Yujie Zhong, Zhenhua Guo, and Zhiheng Li. SoccerNet 2022 challenges results. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, pages 75–86, Lisbon, Port., Oct. 2022. ACM. 3

[19] Jan Held, Anthony Cioppa, Silvio Giancola, Abdullah Hamdi, Bernard Ghanem, and Marc Van Droogenbroeck. VARS: Video assistant referee system for automated soccer decision making from multiple views. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*

Work. (CVPRW), pages 5086–5097, Vancouver, Can., Jun. 2023. Inst. Electr. Electron. Eng. (IEEE). 1

[20] Jan Held, Hani Itani, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. X-vars: Introducing explainability in football refereeing with multi-modal large language models. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, Seattle, WA, USA, Jun. 2024. 1

[21] Namdar Homayounfar, Sanja Fidler, and Raquel Urtasun. Sports field localization via deep structured models. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 4012–4020, Honolulu, HI, USA, Jul. 2017. Inst. Electr. Electron. Eng. (IEEE). 3, 6

[22] Buzhen Huang, Tianshu Zhang, and Yangang Wang. Pose2UV: Single-shot multiperson mesh recovery with deep UV prior. *IEEE Trans. Image Process.*, 31:4679–4692, 2022. 2, 3

[23] IFAB. Laws of the game. Technical report, The International Football Association Board, Zurich, Switzerland, 2022. 2

[24] IIHF. IIHF Official rule book 2022/23. Technical report, International Ice Hockey Federation, Zurich, Switzerland, Jul. 2022. 2

[25] Maxime Istasse, Vladimir Somers, Pratheeban Elancheliyan, Jaydeep De, and Davide Zambrano. DeepSportradar-v2: A multi-sport computer vision dataset for sport understandings. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, pages 23–29, Ottawa, Ontario, Can., Oct. 2023. ACM. 2

[26] Wei Jiang, Juan Camilo Gamboa Higuera, Baptiste Angles, Weiwei Sun, Mehrsan Javan, and Kwang Moo Yi. Optimizing through learned errors for accurate sports field registration. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 201–210, Snowmass, CO, USA, Mar. 2020. Inst. Electr. Electron. Eng. (IEEE). 3

[27] Zhongyu Jiang, Haorui Ji, Samuel Menaker, and Jenq-Neng Hwang. GolfPose: Golf swing analyses with a monocular camera based human pose estimation. In *IEEE Int. Conf. Multimedia Expo Work. (ICMEW)*, pages 1–6, Taipei City, Taiwan, Jul. 2022. Inst. Electr. Electron. Eng. (IEEE). 1

[28] Victor Joos, Silvio Giancola, Anthony Cioppa, Seyed Abolfazl Ghasemzadeh, Floriane Magera, Baptiste Standaert, Amir Mohammad Mansourian, Xin Zhou, Shohreh Kasaei, Bernard Ghanem, Alexandre Alahi, Marc Van Droogenbroeck, and Christophe De Vleeschouwer. SoccerNet game state reconstruction: End-to-end athlete tracking and identification on a minimap. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, Seattle, WA, USA, Jun. 2024. 2

[29] Masanori Kano, Hidehiko Okubo, Masaki Takahashi, Kensuke Ikeya, Kensuke Hisatomi, and Tomoyuki Mishina. Accurate and practical calibration of multiple pan-tilt-zoom cameras for live broadcasts. *IEEE Access*, 8:153993–154006, 2020. 1

[30] Susumu Kikkawa, Fumio Okura, Daigo Muramatsu, Yasushi Yagi, and Hideo Saito. Accuracy evaluation and prediction of single-image camera calibration. *IEEE Access*, 11:19312–19323, 2023. 1

[31] P. Kurowski, K. Szelag, W. Zaluski, and R. Sitnik. Accurate ball tracking in volleyball actions to support referees. *Opto-Electronics Review*, 26(4):296–306, Dec. 2018. 1

[32] Arnaud Leduc, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. SoccerNet-Depth: a scalable dataset for monocular depth estimation in sports videos. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), CVsports*, Seattle, WA, USA, Jun. 2024. 2

[33] Pengjie Li, Jianwei Li, Shouxin Zong, and Kaiyu Zhang. Soccer field registration based on geometric constraint and deep learning method. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, volume 13020 of *Lect. Notes Comput. Sci.*, pages 287–298, Beijing, China, 2021. Springer Int. Publ. 3

[34] Wang Lu, Suya You, and Ulrich Neumann. Single view camera calibration for augmented virtual environments. In *IEEE Virtual Reality Conference*, pages 255–258, Charlotte, NC, 2007. Inst. Electr. Electron. Eng. (IEEE). 1

[35] Zhenling Ma, Xu Zhong, Hong Xie, Yongjun Zhou, Yuan Chen, and Jiali Wang. A combined physical and mathematical calibration method for low-cost cameras in the air and underwater environment. *Sensors*, 23(4):1–18, Feb. 2023. 1

[36] Adrien Maglo, Astrid Orcesi, Julien Denize, and Quoc Cuong Pham. Individual locating of soccer players from a single moving view. *Sensors*, 23(18):1–28, Sept. 2023. 1, 3

[37] Xiaohan Nie, Shixing Chen, and Raffay Hamid. A robust and efficient framework for sports-field registration. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 1935–1943, Waikoloa, HI, USA, Jan. 2021. Inst. Electr. Electron. Eng. (IEEE). 3

[38] Hanan Quispe, Jorshinno Sumire, Patricia Condori, Edwin Alvarez, and Harley Vera. I see you: A vehicle-pedestrian interaction dataset from traffic surveillance cameras. *CoRR*, abs/2211.09342, 2022. 1

[39] Konstantinos Rematas, Ira Kemelmacher-Shlizerman, Brian Curless, and Steve Seitz. Soccer on your table-

top. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 4738–4747, Salt Lake City, UT, USA, Jun. 2018. 1

[40] T.N. Schoepflin and D.J. Dailey. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. *IEEE Trans. Intell. Transp. Syst.*, 4(2):90–98, Jun. 2003. 1

[41] Long Sha, Jennifer Hobbs, Panna Felsen, Winyu Wei, Patrick Lucey, and Sujoy Ganguly. End-to-end camera calibration for broadcast videos. In *IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 13627–13636, Seattle, WA, USA, Jun. 2020. Inst. Electr. Electron. Eng. (IEEE). 2, 3

[42] Rahul Anand Sharma, Bharath Bhat, Vineet Gandhi, and C. V. Jawahar. Automated top view registration of broadcast football videos. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 305–313, Lake Tahoe, NV, USA, Mar. 2018. Inst. Electr. Electron. Eng. (IEEE). 3

[43] Feng Shi, Paul Marchwica, Juan Camilo Gamboa Higuera, Mike Jamieson, Mehrsan Javan, and Parthipan Siva. Self-supervised shape alignment for sports field registration. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 3768–3777, Waikoloa, HI, USA, Jan. 2022. Inst. Electr. Electron. Eng. (IEEE). 2, 3

[44] Mark Shortis. Camera calibration techniques for accurate measurement underwater. In *3D Recording and Interpretation for Maritime Archaeology*, volume 31 of *Coast. Res. Libr.*, pages 11–27. Springer Int. Publ., London, Engl., 2019. 1

[45] Yuan Shu and Yangang Wang. A camera calibration method based on scene and human semantics. In *Youth Acad. Annu. Conf. Chin. Assoc. Autom. (YAC)*, pages 216–221, Beijing, China, Nov. 2022. Inst. Electr. Electron. Eng. (IEEE). 3

[46] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, second edition, 2022. 2

[47] Xinyao Tang, Wei Wang, Huansheng Song, and Ying Li. Novel optimization approach for camera calibration in traffic scenes. *Transp. Res. Rec.: J. Transp. Res. Board*, 2677(3):1048–1066, Sept. 2022. 1

[48] Jonas Theiner and Ralph Ewerth. TVCalib: Camera calibration for sports field registration in soccer. In *IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pages 1166–1175, Waikoloa, HI, USA, Jan. 2023. Inst. Electr. Electron. Eng. (IEEE). 2, 3, 6

[49] Hiroki Tsurusaki, Keisuke Nonaka, Ryosuke Watanabe, Tomoaki Konno, and Sei Naito. Sports camera calibration using flexible intersection selection and re-

finement. *ITE Trans. Media Technol. Appl.*, 9(1):95–104, 2021. 3

[50] Gabriel Van Zandycke and Christophe De Vleeschouwer. 3D ball localization from a single calibrated image. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3471–3479, New Orleans, LA, USA, Jun. 2022. Inst. Electr. Electron. Eng. (IEEE). 1

[51] Gabriel Van Zandycke, Vladimir Somers, Maxime Istasse, Carlo Del Don, and Davide Zambrano. DeepSportradar-v1: Computer vision dataset for sports understanding with high quality annotations. In *Int. ACM Work. Multimedia Content Anal. Sports (MMSports)*, pages 1–8, Lisbon, Port., Oct. 2022. ACM. 2, 3

[52] Kanav Vats, Mehrnaz Fani, David A Clausi, and John Zelek. Puck localization and multi-task event recognition in broadcast hockey videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4567–4575, 2021. 1

[53] Wanneng Wu, Min Xu, Qiaokang Liang, Li Mei, and Yu Peng. Multi-camera 3D ball tracking framework for sports video. *IET Image Process.*, 14(15):3751–3761, Dec. 2020. 1

[54] Yan Xu, Yu-Jhe Li, Xinshuo Weng, and Kris Kitani. Wide-baseline multi-camera calibration using person re-identification. *CoRR*, abs/2104.08568, 2021. 1

[55] Neng Zhang and Ebroul Izquierdo. A fast and effective framework for camera calibration in sport videos. In *IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, pages 1–5, Suzhou, China, Dec. 2022. Inst. Electr. Electron. Eng. (IEEE). 3

[56] Wentao Zhang, Huansheng Song, Lichen Liu, Congliang Li, Bochen Mu, and Qian Gao. Vehicle localisation and deep model for automatic calibration of monocular camera in expressway scenes. *IET Intell. Transp. Syst.*, 16(4):459–473, Dec. 2021. 1