

BroadTrack: Broadcast Camera Tracking for Soccer

Floriane Magera^{1,2}

Thomas Hoyoux¹

Olivier Barnich¹

Marc Van Droogenbroeck²

¹ EVS Broadcast Equipment

² University of Liège, Belgium

Abstract

Camera calibration and localization, sometimes simply named camera calibration, enables many applications in the context of soccer broadcasting, for instance regarding the interpretation and analysis of the game, or the insertion of augmented reality graphics for storytelling or refereeing purposes. To contribute to such applications, the research community has typically focused on single-view calibration methods, leveraging the near-omnipresence of soccer field markings in wide-angle broadcast views, but leaving all temporal aspects, if considered at all, to general-purpose tracking or filtering techniques. Only a few contributions have been made to leverage any domain-specific knowledge for this tracking task, and, as a result, there lacks a truly performant and off-the-shelf camera tracking system tailored for soccer broadcasting, specifically for elevated tripod-mounted cameras around the stadium. In this work, we present such a system capable of addressing the task of soccer broadcast camera tracking efficiently, robustly, and accurately, outperforming by far the most precise methods of the state-of-the-art. By combining the available open-source soccer field detectors with carefully designed camera and tripod models, our tracking system, BroadTrack, halves the mean reprojection error rate and gains more than 15% in terms of Jaccard index for camera calibration on the SoccerNet dataset. Furthermore, as the SoccerNet dataset videos are relatively short (30 seconds), we also present qualitative results on a 20-minute broadcast clip to showcase the robustness and the soundness of our system.

1. Introduction

Sports content has the advantage of displaying sports fields, which have strictly regulated shapes and dimensions. This particularity makes sports camera calibration and localization possible in the wild, without the need for any other prior knowledge but the ability to correctly detect these field markings. This advantage is well-exploited, as all camera calibration techniques detect field markings, and single-frame camera calibration methods flourish, with ever-increasing accuracy. However, two facts that tend to

be overlooked are (1) sports field markings can be really sparse in some areas, and (2) wide-angle broadcast cameras have typically $25\times$ zoom lenses, which can sometimes lead to quite narrow views. Thus, while the open-source datasets keep the task conveniently framed, there is an unmet need to address the case where few or even none of the field markings are visible for real-world applications. By carefully modeling broadcast cameras, we first include lens distortion effects —something that is systematically overlooked in sports field registration techniques that estimate homographies—, and which we show has a great impact on the accuracy of the results. Then, by including constraints on the camera movements that a tripod actually allows, we demonstrate that with more than accurate reprojections in the image, our models recover consistent position and rotation values, leading to effectively smooth tracking. Besides its novel state-of-the-art performance, our tracking system is robust and includes a reinitialization strategy that is thoroughly validated on the SoccerNet-calibration dataset. Finally, our system is usable in practice, as it works from the first frame and performs at a speed of 16 frames per second (fps) for HD (1920×1080 pixel resolution) images on a server equipped with two RTX 4090 GPUs, without any optimization; by experience, we know that a proper optimization of our code will make it real-time.

To validate our technique, we use the SoccerNet datasets [11, 19, 28]. The SoccerNet initiative has allowed new tracks of research for many problems related to soccer video understanding, including camera calibration and localization, by providing task definitions, corresponding public datasets, and metrics. Among others, the SoccerNet-calibration dataset (denoted as sn-calibration hereafter) is the first open-source dataset that does not use homographies as annotations, but rather the soccer field markings and goal posts. Besides the calibration task, which is about single-frame camera calibration, the novel game state reconstruction task [54] aims to continuously estimate the bird’s-eye view of the soccer field with the players position. The dataset of the game state reconstruction task, denoted by sn-gamestate in the following, consists of soccer videos. Its annotations follow the same conventions defined for sn-calibration. Solving the game state task requires pre-

cise camera calibration and localization, consistently over time, which is a property also required for most applications regarding player performances and game analysis.

Contributions. By leveraging specific knowledge about broadcast cameras, we produce an efficient and accurate tracking system for soccer. We summarize our contributions as follows: **(i)** We propose a tracking system, named BroadTrack for the calibration of moving cameras that is both efficient and state-of-the-art, **(ii)** We define a camera model that is tailored for broadcast applications, and **(iii)** We release the source code of our tracking system at <https://github.com/evs-broadcast/BroadTrack>.

2. Related work

Common approaches to solving camera calibration and localization problems include structure from motion (SfM) [50] and simultaneous localization and mapping (SLAM) [47], both relying on the camera parallax to derive properties of the 3D world in which the camera moves. A typical pipeline could be described as follows: features are detected in the video frames, matched together across time, and finally triangulated. The resulting 2D-3D correspondences are fed in a PnP solver to derive the camera parameters. Compared to the usual problems that both SfM and SLAM solve, broadcast applications require custom approaches. Indeed, real-world applications often require a metric reconstruction of the scene, but for sports applications, the world reference system should not be placed regarding a camera; rather, the cameras should all be expressed relatively to the sports field.

Detectors. Unlike common camera localization or calibration schemes, sports camera calibration algorithms do not usually rely on 2D matching of features between frames, but rather on the detection of sports field elements that provide direct metric correspondences, which is an advantage for tracking, in effect reducing the risk of drifting. Most methods detect field markings in the image, either by binary segmentation or Hough line transforms [1, 21, 25], and sometimes even deriving vanishing points associated to horizontal and vertical lines [34, 35]. Lately, thanks to the performance of deep neural networks, higher levels of semantic interpretation were attained with sports markings being used as zone delimiters [51, 55], dividing the sports field into segmentation zones, or with some methods identifying specific markings as unique classes. Finally, some methods derive directly sports field lines intersections because most of the methods relying on solvers for camera parameters solely use point correspondences.

Virtual augmentation of correspondences. A global problem for all single-frame camera calibration and local-

ization methods is the sparse nature of sports field markings, which are rarely uniformly distributed, leading to a lack of visual support in the image as there are few visible markings. Several methods alleviate this by detecting virtual correspondences on the field, or by deriving additional geometric cues from existing elements. For example, some works extend line segments to get new intersections [17, 23, 31, 57], derive tangents to circles or circle keypoints [2, 17, 23, 31], detect grass mowing patterns [16], or even learn to detect grids of keypoints on the sports field surface [10, 14, 45, 46, 48]. These methods leveraging grid-like keypoints on sports fields have the drawback of first relying on homographies to derive the grid projection in the image, which, as shown in previous works [44], are not the best fit to model the projective transformation.

Dictionary methods. Other works leverage directly prior knowledge about the broadcast camera, such that the number of correspondences in the image does not matter. These methods construct a dictionary of plausible broadcast camera views, and then retrieve camera parameters based on the similarity between the dictionary view and the camera view [8, 18, 51, 52, 55, 59]. These methods tend to be slow due to the search time in the dictionary, or if the dictionary is sparsely populated, relies on the ability of a Spatial Transformer Network to regress the parameters of the homography that maps the dictionary view to the current view [51, 55].

Homography regression. Inspired by PoseNet [40] and learned homography estimation [20], previous works directly regress homographies [24, 39, 53]. While the work of DeTone *et al.* [20] estimates the homography between a pair of images, its transposition to sports field registration requires the regressed homography to capture the transformation between a broadcast image and the synthetic bird's-eye view of the sports field. The huge perspective difference between a bird's-eye view of the sports field and one broadcast image probably explains why these methods rarely work in a single forward pass, limiting their use in actual broadcast scenarios.

Tracking. Due to the limited number of datasets with broadcast videos, few methods took interest in the temporal consistency of the camera calibration and localization. Among these methods, we define two main categories: (a) the ones based on homographies between successive frames, and (b) the ones using traditional tracking filters. These categories are further detailed hereafter.

(a) Homography between successive frames: A common strategy to perform tracking is to extract features to obtain point correspondences, or even line ones [33], and ro-

bustly estimate a homography between pairs of frames with RANSAC [30, 45]. The underlying assumption made when computing homographies between successive frames is that the camera is purely in rotation, an assumption that other works refute for soccer [7]. Note that a homography can be computed between frames if the correspondences are taken only on the sports field plane [14, 33]. In the case of rather small sports fields, like tennis, Farin *et al.* [25] assume that the camera velocity is constant, and thus assume that the homography between the future image pair will be equivalent to the homography mapping the present image pair.

(b) Common temporal filters: A range of methods also apply common tracking algorithms such as the Kalman Filter, starting with Claasen and de Villiers [14], who use a Kalman filter to track point correspondences on the sports field, which are later used to track the homography using an extended Kalman filter. Beetz *et al.* [3] use an iterative extended Kalman filter to track their camera parameters, while Citraro *et al.* [13] use a particle filter to model the motion of the camera. Lu *et al.* [41] propose a SLAM algorithm tailored for a PTZ broadcast camera by building a map of 3D rays rather than 3D points, but consider the camera focal point fixed and neglect lens distortion.

Besides NBJW [31] which is the SOTA on the sn-calibration dataset, the methods that inspire us in terms of broadcast modeling are the ones that do model radial distortion [1, 3, 6, 56] and the tripod of the camera [7, 9, 41].

Available data. For completeness, we addressed both camera calibration and sports field registration techniques. However, in terms of datasets, given the recent concerns about the ability of homographies to properly model broadcast cameras [44, 56], we do not use any of the datasets that provide homographies as pseudo ground truth [10, 14, 35]. In this work, by combining high-level semantic analysis of the sports field marking detection and broadcast camera knowledge, we derive a new tracking system that achieves SOTA results on the sn-gamestate dataset.

3. Method

First, we define our models for the camera and for the tripod (Section 3.1), then we outline our tracking system (Section 3.2), which comes with a reinitialization algorithm (Section 3.3).

3.1. Broadcast camera model

To model our camera, we start from the pinhole model with the *calibration matrix* K , defined as follows [32]

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

As stated in the standards of telecommunications for HD [37] and UHD content [38], broadcast pixels are squares, such that the focal length is the same for both axes, and the skew parameter can be ignored. We further assume that the principal point is located at the image center, so that the set of intrinsic parameters reduces to one parameter, f .

The camera pose is modeled by its focal point position $C = (C_x, C_y, C_z)$ and a rotation matrix R parameterized by the *pan* ϕ , *tilt* θ , and *roll* γ angles, defined accordingly in the Euler angles convention: $R = R_z(\gamma)R_x(\theta)R_z(\phi)$. In the absence of lens aberrations, the projection matrix of our camera would be formulated by the following equation [32]: $P = KR \begin{bmatrix} I & -C \end{bmatrix}$. However, it is not uncommon for broadcast cameras to display some radial distortion, a deformation effect that occurs due to the curved nature of the camera lens. While it may be modeled in the pixel space [6, 56], we follow the standard way and apply this deformation in normalized coordinates [32, 60].

Given a point in the normalized image plane $\bar{x} = (\bar{x}, \bar{y})$, the distortion function $\mathcal{L}(r) = 1 + k_1 r^2$ transforms \bar{x} according to its distance to the origin of the image plane $r = \|\bar{x}\|_2$ which finally gives the image point expressed in pixel coordinates:

$$x = f \mathcal{L}(r) \bar{x} + p. \quad (2)$$

We use the model proposed by Brown-Conrady [5], and discard all higher orders of radial distortion and tangential distortion as we experimentally find them superfluous. The set of unknown parameters for our camera is $\kappa : \{f, k_1, \phi, \theta, \gamma, C_x, C_y, C_z\}$, and we define the function $\pi_\kappa : \mathbf{X} \rightarrow \mathbf{x}$ that projects a 3D point to its respective image point. In the next section, we further analyze the specificities of broadcast cameras, and extend our model to consider the tripod on which a broadcast camera is installed, and the constraints that it creates on the camera parameters.

3.1.1 Pan-tilt head and tripod

Compared to the usual cameras used in the computer vision literature for robotics or augmented reality with mobile phone captures, broadcast cameras are another kind of beast. A typical broadcast camera for soccer has a lens that can zoom up to 25 times, which consists of an arrangement of optics that can weigh up to a few kilograms [26]. When we use the pinhole model to represent the arrangement of optics inside those lenses, we approximate it with a single optic. While Chen *et al.* [7] assume that the camera focal point is at a fixed location inside the camera lens, we argue that this virtual projection center may not be located inside the camera and that it should evolve along the optical axis as the camera zooms. Therefore, our only assumption about the position of the focal point regarding the physical camera is that the focal point is located on the optical axis of

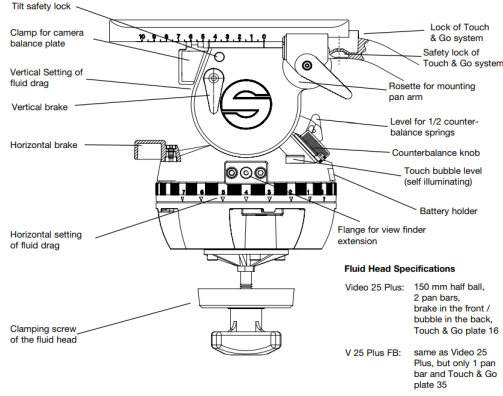


Figure 1. Usual pan-tilt head for $25\times$ lens broadcast camera. (taken from [49], others pan-tilt heads can be found online [58]).

the camera. We choose to rather constrain the optical axis position regarding the tripod.

A camera comes in several parts: the tripod, the pan-tilt head, and the camera itself. The challenging part for our extrinsic parameters model is the pan-tilt head. This is the part that is mounted on the tripod on which the camera is rigged. The pan-tilt head allows the camera to rotate smoothly while allowing the camera operators to make quick sweeps if they lose track of the action. By construction, it allows the camera to rotate along two axes, *i.e.* to pan and tilt the camera. To better visualize the rotation axes, a common professional pan-tilt head is depicted in Figure 1.

According to this blueprint, we make the assumption that the pan and tilt rotation axes intersect in a point T , which remains fixed during the whole game. Our model states that there is a point O in the camera which belongs to the optical axis of the camera, and which remains at a fixed distance δ of the rotation center T as the camera moves. Let r_i be the i th column of the orientation matrix of the camera R^T , the vector $-r_2$ is the upvector of the camera, while the vector r_3 defines its optical axis direction. If we further assume that the camera is centered on the pan-tilt head, this point O is then determined by the upvector $-r_2$ and its distance to the tripod rotation center $O = T - \delta r_2$. As the focal point of the camera changes with the zoom level of the camera, the position of the focal point is finally given by $C = T - \delta r_2 + \lambda r_3$. The parameters $\{T, \delta\}$ are fixed over time, while λ varies with the camera view. A visualization of the model is shown in Figure 2.

3.2. Complete description of our tracking system

Our system comprises different steps, described hereafter: the detection of the field markings (Section 3.2.1), which is augmented using optical flow (Section 3.2.2), a parameter update procedure (Section 3.2.3), and an evaluation of the tracking confidence (Section 3.2.4).

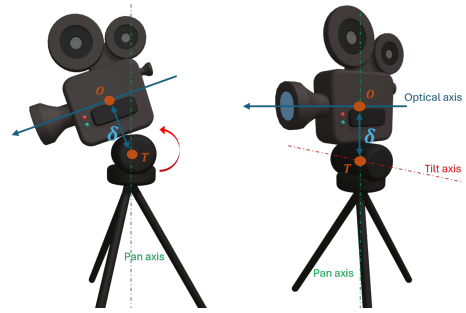


Figure 2. Tripod model. The center of rotation T remains fixed as the camera rotates, and the point O which belongs to the optical axis of the camera remains at a fixed δ distance of T .

3.2.1 Sports field detection

Let \mathcal{S} be the set of semantic classes constituent of soccer field markings, such as “left goal line”, “center circle”, *etc.* Each one can be viewed as a simple geometric element. The laws of the game [36] further specify their position and dimensions, such that we can define the soccer field template $\mathcal{F} : \mathcal{S} \rightarrow \mathbb{E}$, which maps each semantic class $c \in \mathcal{S}$ to its corresponding 3D geometric element $e \in \mathbb{E}$. For convenience, the projection of the soccer field markings in the camera view κ will be denoted by $\pi_\kappa(\mathcal{F})$.

To obtain 2D-3D correspondences, we only need to detect and identify soccer field elements in the videos. We leverage existing open-source sports field detectors such as the keypoint detector of Gutiérrez-Pérez and Agudo [31] that detects the intersection points of the soccer field markings. While our reinitialization algorithm uses the detected points, the tracking leverages the denser information retrieved from semantic segmentation of field markings. We rely on the semantic field markings detector of Theiner and Ewerth [56]. Semantic segmentation of field markings provides blobs of pixels towards 3D line or circle equation correspondences. As segmentation blobs provide a robust, but maybe too rich source of information, we synthesize it with the mean shift algorithm [15] to fit a set of points that approximate the segmented blob, thus generating a set of image points x_c^1, \dots, x_c^n per visible soccer field marking class c of \mathcal{S} .

3.2.2 Optical flow

As broadcast cameras can zoom in on areas of the soccer field that are sparse in terms of markings, optical flow correspondences are necessary to prevent drifting. Furthermore, these correspondences can be sampled uniformly in the image, which helps to distribute the visual support in the whole image, unlike field markings which can be condensed into small areas. We use the pyramidal version of Lucas-Kanade’s algorithm [4] to retrieve N_o point matches

$\{\mathbf{x}_{t-1}^i \leftrightarrow \mathbf{x}_t^i\}_{i=0}^{N_o}$ between the previous image \mathcal{I}_{t-1} and the current one \mathcal{I}_t .

3.2.3 Update through non-linear optimization

Starting from the previous camera estimate κ_{t-1} , we update the camera parameters $\kappa_t : \{f, k_1, \phi, \theta, \gamma, C_x, C_y, C_z\}$ by minimizing the sum of three error functions.

First, given soccer field markings correspondences, if we denote by \mathcal{S}^* the subset of \mathcal{S} that is detected in the current frame \mathcal{I}_t , and by $\{\mathbf{x}_c^i\}_{i=0}^{N_c}$ the set of points extracted by mean shift from the segmentation maps for the soccer field marking class c , we minimize the reprojection error between each extracted point and the closest point \mathbf{p} belonging to the projected soccer field element $\pi_\kappa(\mathcal{F}(c))$:

$$\mathcal{L}_{\mathcal{F}} = \sum_{c \in \mathcal{S}^*} \sum_{i=0}^{N_c} \rho \left(\min_{\mathbf{p} \in \pi_\kappa(\mathcal{F}(c))} \|\mathbf{p} - \mathbf{x}_c^i\|_2 \right), \quad (3)$$

where $\rho(\cdot)$ is the Cauchy loss, used to filter outliers.

Secondly, given optical flow correspondences, we can filter out the point correspondences landing outside the polygon obtained by the projection of the soccer field side lines. To minimize the noise coming from the players motion, the players are detected with RTMDet [42], and the correspondences are discarded around their bounding boxes. A point correspondence \mathbf{x}_{t-1}^i in frame \mathcal{I}_{t-1} can be mapped to a ray by inverting the projection function: $\mathbf{L}_{t-1}^i = \pi_{\kappa_{t-1}}^{-1}(\mathbf{x}_{t-1}^i)$. As we only keep correspondences on the surface of the sports field, its intersection with the plane $Z = 0$ allows us to retrieve a 3D point $\mathbf{X}_{t-1}^i = \mathbf{L}_{t-1}^i (0 \ 0 \ 1 \ 0)^\top$. As a result, for all detected optical flow correspondences, we minimize the Cauchy loss of the reprojection error:

$$\mathcal{L}_{OF} = \sum_{i=0}^{N_o} \rho \left(\|\pi_\kappa(\mathbf{X}_{t-1}^i) - \mathbf{x}_t^i\|_2 \right). \quad (4)$$

The third term of our objective function is a constraint that ensures that the camera rotation and focal point position satisfy our tripod model. We derive the position of the point \mathbf{O}^* as the closest point to \mathbf{T} belonging to the camera optical axis:

$$\mathbf{O}^* = \mathbf{C} + \frac{\langle (\mathbf{T} - \mathbf{C}), \mathbf{r}_3 \rangle \mathbf{r}_3}{\|\mathbf{r}_3\|_2^2}, \quad (5)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product. Since the distance between this point and the tripod rotation center should be δ meters, we define the following loss term:

$$\mathcal{L}_T = \delta - \|\mathbf{O}^* - \mathbf{T}\|_2. \quad (6)$$

The previous losses are scaled and summed into a final objective function $\mathcal{L} = \mathcal{L}_{\mathcal{F}} + \mathcal{L}_{OF} + \omega \mathcal{L}_T$, with ω being a scalar hyperparameter, to optimize the camera parameters κ using the Levenberg-Marquardt algorithm.

3.2.4 Online confidence evaluation

Given the semantic segmentation of the soccer field markings \mathcal{S} , we produce a binary map B of all the soccer field markings in the image: $B(x, y) = 1, S(x, y) = c, \forall c \in \mathcal{S}$. Considering the projection of the soccer field template \mathcal{F} in the image $\pi_\kappa(\mathcal{F})$, our confidence score is the Jaccard index between the two generated masks:

$$s = \frac{|B \cap \pi_\kappa(\mathcal{F})|}{|B \cup \pi_\kappa(\mathcal{F})|}. \quad (7)$$

This score is used to detect when a reinitialization of the tracker is needed, and when the optical flow correspondences are to be discarded to prevent them from damaging the optimization, as it relies on the previous camera parameters κ_{t-1} .

3.3. Reinitialization algorithm

For initialization, or when the confidence score becomes low, we propose a reinitialization algorithm. With the capabilities of deep neural networks today, soccer field markings detection is not a very difficult task. Usually, the tracking starts drifting when the sports field markings become very sparse in the image, which is why we include a reinitialization algorithm that only needs two point correspondences.

Given two point correspondences between the image and the top-view sports field model, our reinitialization algorithm estimates the focal length, the pan, and the tilt of the camera. The focal point is set to the tripod rotation center, and the roll value is set to zero.

Then, given the image points $\mathbf{x}_i = (x_i, y_i)$ and their corresponding world points $\mathbf{X}_i = (X_i, Y_i, Z_i)$, we derive the camera parameters $\{f, \phi, \theta\}$. The focal length f is estimated with the equations originally derived by Gedikli [27], and summarized in Appendix 6.1 of [9]. The pan and tilt angles, respectively ϕ and θ , are estimated by an iterative algorithm assuming that ϕ only affects the projection along the x -axis, and that θ only affects the projection along the y -axis. We initialize ϕ and θ by steering the optical axis towards the barycenter of the detected world points \mathbf{X}_i . Then we estimate updates depending on the correspondences:

$$d\phi = \sum_i \frac{1}{2} \tan^{-1}(x_i - \pi_\kappa(\mathbf{X}_i)_x, f), \quad (8)$$

$$d\theta = \sum_i \frac{1}{2} \tan^{-1}(y_i - \pi_\kappa(\mathbf{X}_i)_y, f), \quad (9)$$

with $\pi_\kappa(\cdot)_x$ denoting the x coordinate of the point projection. The pan and tilt are thus iteratively refined as $\phi_{k+1} = \phi_k + d\phi_k$ and $\theta_{k+1} = \theta_k + d\theta_k$, with k being less or equal to 5.

If there are more than two correspondences in the image, we use RANSAC to filter out potential outliers. As a

last step, the optimization described in Section 3.2.3 is performed without the \mathcal{L}_{OF} error function.

4. Results

To validate our tracking system, we conducted several experiments. In Section 4.1, we first report the performance of BroadTrack on sn-gamestate according to the metrics proposed with the sn-calibration dataset. Then, in Section 4.2, we validate our reinitialization algorithm on the 2023 version of the sn-calibration challenge. To further validate the effects of optical flow, tripod constraint, and lens distortion, we provide an ablation study in Section 4.3. Finally, since the metrics are meant for single-frame camera calibration and only evaluate the quality of the reprojection in the images, in Section 4.4, we qualitatively illustrate the suitability of the camera parameters derived by BroadTrack.

4.1. Tracking system on SoccerNet-gamestate

We evaluate our system on the sn-gamestate dataset, which consists of sequences of 30 seconds, where each frame is annotated with point correspondences along the soccer field markings and goal posts. The test set contains 49 sequences of 750 images each, extracted from 3 games of the Swiss Football League.

For our experiments, we use the JaC metric [44] (previously denoted by AC in the sn-calibration challenges [12, 29]), which expresses the percentage of soccer field elements that are correctly reprojected in the image, the correctness being tuned by a tolerance parameter τ in pixels. We report this metric for both 5 and 10 pixels, which are quite challenging threshold values for the HD frames of the dataset. For comparison, note that the SoccerNet camera calibration challenge evaluates this metric for 5 pixels, but for 960×540 -sized images, a quarter of this dataset resolution. We also give the mean reprojection error and the completeness rate, the latter indicating the proportion of the dataset for which the technique produced camera parameters.

To estimate the tripod rotation center T , we run our tracking system without the tripod constraint, and then optimize the tripod position and distance to the optical axis δ given the estimated camera parameters. From the keypoints detected by the neural network of Gutiérrez-Pérez and Agudo [31], we only keep the ones that are actual marks, line intersections, or line and circle (arc) intersections, *i.e.* we discard all keypoints that are derived from other geometric cues. The reinitialization algorithm is used to get the parameters of the first frame, and during the tracking, the reinitialization is performed when the score confidence score s falls under 0.5, which leads to frequent reinitialization.

During our evaluation, out of the 49 sequences of 750 frames in the test set, the reinitialization had to be used 33

	JaC ₅ (↑)	JaC ₁₀ (↑)	MRE(↓)	CR(↑)
TVCalib [56]	19.88	50.42	12.4	99.93
NBJW [31]	37.14	68.24	10.28	93.67
PTZ-SLAM [41]	25.87	45.28	27.64	26.67*
Ours (fixed C)	50.97	74.93	5.39	100
BroadTrack	56.88	79.79	5.02	100

Table 1. Comparison metrics on the sn-gamestate dataset. MRE stands for Mean Reprojection Error and is measured in pixels; CR stands for completeness rate in percent. All results are reported for HD, 1920×1080 frames. For the PTZ-SLAM method, the code provided by the authors crashes after processing about 200 frames; hence the low completeness rate (see *).

times, and the loss of tracking lasted for 15 frames on average. From the result reported in Table 1, BroadTrack outperforms all available open-source methods. We also experimentally confirm that the modeling of a broadcast camera as a PTZ camera with a fixed focal point deters the performance of the tracking. This emphasizes the specificity of broadcast cameras, which are not well modeled by PTZ cameras, even if they are rigged on a tripod.

4.2. Reinitialization on SoccerNet-calibration

We evaluate our reinitialization algorithm on the sn-calibration dataset. The test set comprises 3,141 images from a wide range of cameras used during soccer broadcast, including wide-angle cameras, and fish-eye cameras. The JaC evaluation of our method is computed only for views that obtain a confidence score $s < 0.2$, which lowers our completeness rate. We also expect this score condition to filter out views for which our tracking system is not especially designed, *e.g.* views from fish-eye cameras. The default focal point localizations are defined for common wide-angle cameras, that is for main, 16 meters, and high behind the goal (*HBG*) cameras respectively as $C_{main} = (0, 55, -12)$, $C_{16m} = (\pm 36, 55, -12)$, $C_{HBG} = (-65, 0, -15)$, all expressed in meters in the sn-calibration world reference system [43].

From the results reported in Table 2, we establish that our algorithm can reach state-of-the-art performance on a smaller part of the dataset. Since the camera diversity of the dataset is higher than the one we design our system for, the high performance of the reinitialization part comforts our strategy of frequent reinitialization.

4.3. Ablation study

We conduct our ablation study on the sn-gamestate test set. To run our algorithm without any prior knowledge of the tripod rotation center, we choose a default position as input to our initialization algorithm. We arbitrarily set the camera focal point to a default main location in $C = (0, 55, -12)$ meters. As shown in Table 3, the biggest

	JaC ₅ (↑)	JaC ₁₀ (↑)	CR(↑)
TVCalib [56]	52.9	73.4	66.5
NBJW [31]	73.7	86.7	77.5
Ours	75.25	86.8	69.8

Table 2. Two-point reinitialization algorithm evaluation on the sn-calibration dataset of 2023. JaC metrics are reported in percent at 5 and 10 pixels for 960×540 frames. Our reinitialization algorithm equals or achieves the SOTA performance of the NBJW method on a slightly smaller subset of the dataset. To enable fairer comparison with TVCalib, we filter out their calibrations with $\text{JaC}_5 = 0$, because, by design, their method never signals failure.

OF	T	k_1	JaC ₅ (↑)	JaC ₁₀ (↑)	MRE(↓)	MedRE(↓)
✗	✗	✗	42.09	74.09	5.74	3.13
✗	✗	✓	54.1	78.35	5.04	2.47
✓	✗	✓	55.96	79.07	4.85	2.43
✗	✓	✓	54.9	78.99	4.95	2.44
✓	✓	✓	56.88	79.79	5.02	2.37

Table 3. Ablation study on sn-gamestate. JaC metrics are reported in percent at 5 and 10 pixels for HD frames. Mean (MRE) and Median (MedRE) Reprojection Errors are reported in pixels. OF stands for optical flow, T is for our tripod constraint, and finally k_1 represents the inclusion of radial distortion in our model.

improvement in performance comes from the inclusion of radial distortion in the camera model, which legitimates our concerns about the previous datasets based on homographies and confirms our choice of not using their annotations as it would lead to an unfair evaluation. It is also worth noticing that the optimization procedure starting from the previous camera estimate does not lead to much of a performance boost when compared with the NBJW method. This demonstrates that their strategy of deriving virtual keypoints pays off, even if it is at the expense of physical modeling of the camera parameters, as discussed in Section 4.4. Both optical flow and tripod constraints demonstrate a smaller contribution to the performance of our algorithm according to the reported metrics of Table 3. We argue and show in the next section that their contribution is only partially reflected by the single-view metrics used until now. In the next section, we show their benefits through qualitative evaluation and comparisons.

4.4. Physical soundness

As explained in Section 3.1.1, professional pan-tilt heads are made to ensure the smooth motion of the camera. This means that we expect pan and tilt values to be particularly smooth over time. We also expect our tripod constraint to make the camera focal points more localized, even if, per se, there is no restriction in terms of distance to the tripod. These intuitions are confirmed by our visualization of Fig-

ure 3. We notice that, while the focal lengths estimated by BroadTrack are smoother or display lower variations than the other methods, it still shows some high-frequency variations. We explain that because of a vertigo effect, as the scene is far away, some uncertainty in terms of position can be compensated by a focal length adjustment, and conversely, without displaying perspective distortions.

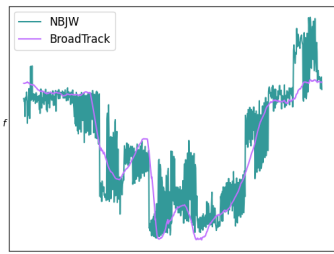
4.5. Long-term tracking

To further demonstrate the performance of BroadTrack, we show results on 20 minutes (60,000 frames) of the main camera footage taken from a game of the German Bundesliga. As the stadium is much bigger than the ones of SoccerNet, we set the default focal point position $C_{default} = (0, 90, -18)$, and we use a commercial tool for keypoints and markings detections [22]. To derive the tripod position, we run our system without the tripod constraint on the first 5,000 frames, and perform the optimization procedure as described in Section 4.1 to derive T and δ . This sequence is not annotated; hence, only the overlay of the soccer field projection provides a qualitative assessment. As displayed in Figure 4, the projection of the soccer field overlays almost perfectly on top of the actual field markings. BroadTrack maintains the same quality of overlay for the complete video; this illustrative video is given in the supplementary material. Moreover, the reinitialization step is only performed 60 times, and lasts for 4 frames on average. We explain this improvement by the quality of the commercial keypoints and markings detection.

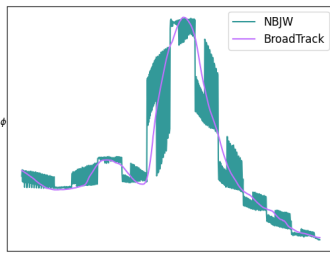
5. Conclusions

In this paper, we have presented BroadTrack, a new tracking system specially designed for wide-angle broadcast cameras. Through diverse qualitative and quantitative evaluation, we show that our system is both accurate and robust, outperforming available solutions in both aspects. Our results also corroborate the suitability of our broadcast camera lens and tripod models, motivating further exploration and refinement. Future works entail the dynamic incorporation of the tripod constraint, even if BroadTrack obtains convincing results with a roughly appropriate focal point. Another exciting piece of research lies in the vertigo effect noticed with the high-frequency variation of the focal length, as we believe modeling this effect might lead to even better focal point position.

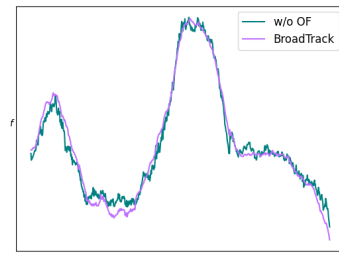
Acknowledgments. This work was supported by the Service Public de Wallonie (SPW) Recherche, Belgium, under Grant N°8573.



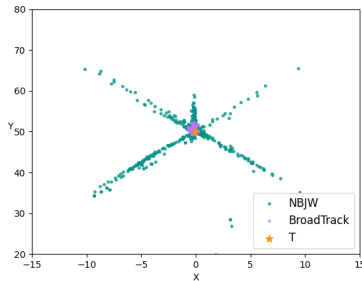
(a) Focal length value variations.



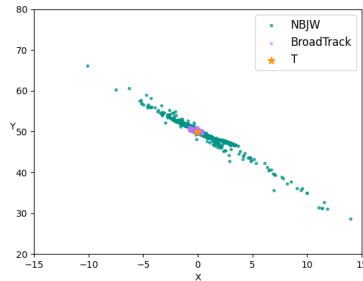
(b) Pan value variations.



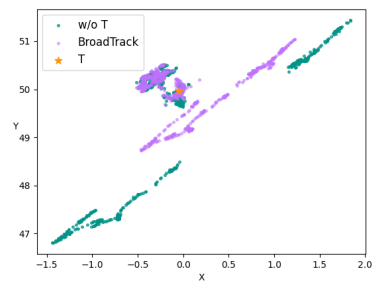
(c) Focal length with or without optical flow.



(d) Focal point position in the XY plane.



(e) Focal point position in the XY plane.



(f) Benefit of including the tripod constraint on the focal point position in the XY plane.

Figure 3. Camera parameters visualizations, best viewed on screen. The first row displays pan and focal length values along test sequences of the sn-gamestate dataset. Figure 3a and 3b show the jitter of the parameters extracted by NBJW compared to BroadTrack. Figure 3c shows that the optical flow helps to smooth focal length values. The second row displays the focal point C position in the XY plane. Figure 3d and 3e show that the focal point estimated by NBJW can travel up to 20 meters along a single sequence, while our focal point remains in a close neighborhood of the estimated tripod position. Finally, Figure 3f shows the benefit of including the tripod constraint on the camera position, which remains closer to the estimated center of rotation T .



Figure 4. Qualitative evaluation of the 20 minutes sequence from the Bundesliga. One out of 10,000 images is displayed. Soccer field markings are reprojected in red using the estimated camera parameters κ .

References

[1] Miguel. Alemán-Flores, Luis Alvarez, Luis Gómez, Pedro Henriquez, and Luis Mazorra. Camera calibration in sport

event scenarios. *Pattern Recognit.*, 47(1):89–95, Jan. 2014. 2, 3

[2] Peter Andrews, Njål Borch, and Morten Fjeld. FootyVi-

- sion: Multi-object tracking, localisation, and augmentation of players and ball in football video. In *Int. Conf. Multimedia Image Process.*, volume 15, pages 15–25, Osaka, Japan, Apr. 2024. [2](#)
- [3] Michael Beetz, Suat Gedikli, Jan Bandouch, Bernhard Kirchlechner, Nico Hoyningen-Huene, and Alexander Perzlyo. Visually tracking football games based on tv broadcasts. In *Int. Jt. Conf. Artif. Intell. (IJCAI)*, pages 2066–2071, Hyderabad, India, Jan. 2007. [3](#)
- [4] Jean-Yves Bouguet. Pyramidal implementation of the Lucas Kanade feature tracker – Description of the algorithm. OpenCV documentation, 2001. [4](#)
- [5] Dean Brown. Decentering distortion of lenses. *Photogramm. Eng. Remote Sens.*, 32(3):444–462, 1966. [3](#)
- [6] Peter Carr, Yaser Sheikh, and Iain Matthews. Point-less calibration: Camera parameters from gradient-based alignment to edge images. In *IEEE Work. Appl. Comput. Vis. (WACV)*, pages 377–384, Breckenridge, CO, USA, Jan. 2012. [3](#)
- [7] Jianhui Chen and Peter Carr. Mimicking human camera operators. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 215–222, Waikoloa, HI, USA, Jan. 2015. [3](#)
- [8] 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. [2](#)
- [9] Jianhui Chen, Fangrui Zhu, and James J. Little. A two-point method for PTZ camera calibration in sports. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pages 287–295, Lake Tahoe, NV, USA, Mar. 2018. [3](#), [5](#)
- [10] 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. [2](#), [3](#)
- [11] Anthony Cioppa, Adrien Delière, 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. [1](#)
- [12] Anthony Cioppa, Silvio Giancola, Vladimir Somers, Floriane Magera, Xin Zhou, Hassan Mkhallati, Adrien Delière, 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, Gysuk 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 Huang, and Ziyu Meng. SoccerNet 2023 challenges results. *Sports Eng.*, 27(2):1–18, July 2024. [6](#)
- [13] Leonardo Citraro, Pablo Márquez-Neila, Stefano Savaré, Vivek Jayaram, Charles Dubout, Félix Renaut, Andrés HSFura, Horesh Ben Shitrit, and Pascal Fua. Real-time camera pose estimation for sports fields. *Mach. Vis. Appl.*, 31(3), Mar. 2020. [3](#)
- [14] Paul J. Claasen and Jill P. de Villiers. Video-based sequential Bayesian homography estimation for soccer field registration. *arXiv*, abs/2311.10361, 2023. [2](#), [3](#)
- [15] Dorin Comaniciu and Peter Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002. [4](#)
- [16] Carlos Cuevas, Daniel Berjón, and Narciso García. Grass band detection in soccer images for improved image registration. *Signal Process.: Image Commun.*, 109:1–16, Nov. 2022. [2](#)
- [17] Carlos Cuevas, Daniel Quilón, and Narciso García. Automatic soccer field of play registration. *Pattern Recognit.*, 103, Jul. 2020. [2](#)
- [18] Giacomo D’Amicantonio, Egor Bondarev, and Peter H. N. De. Automated camera calibration via homography estimation with GNNs. In *IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pages 5864–5471, Waikoloa, HI, USA, Jan. 2024. [2](#)
- [19] Adrien Delière, 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/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 4503–4514, Nashville, TN, USA, Jun. 2021. [1](#)
- [20] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv*, abs/1606.03798, 2016. [2](#)
- [21] Elan Dubrofsky and Robert J. Woodham. Combining line and point correspondences for homography estimation. In *Adv. Vis. Comput.*, volume 5339 of *Lect. Notes Comput. Sci.*, pages 202–213, 2008. [2](#)
- [22] EVS Broadcast Equipment. Multi-camera review system - Xeebra. <https://evs.com/products/video-assistance/xeebra>, Jun. 2022. [7](#)
- [23] Nikolay Falaleev. Top-1 solution of SoccerNet camera calibration challenge 2023. <https://github.com/NikolasEnt/soccernet-calibration-sportlight>, Jun. 2023. [2](#)
- [24] Mehrnaz Fani, Pascale Berunelle Walters, David A. Clausi, John Zelek, and Alexander Wong. Localization of ice-rink

- for broadcast hockey videos. *arXiv*, abs/2104.10847, 2021. **2**
- [25] Dirk Farin, Susanne Krabbe, Peter H. N. de With, and Wolfgang Effelsberg. Robust camera calibration for sport videos using court models. In *Storage and Retrieval Methods and Applications for Multimedia*, volume 5307 of *Proc. SPIE*, pages 1–12, San Jose, CA, USA, 2004. **2, 3**
- [26] Fujifilm. Television lenses & cine lenses. Technical note, 2024. **3**
- [27] Suat Gedikli. *Continual and Robust Estimation of Camera Parameters in Broadcasted Sports Games*. PhD thesis, Technische Universität München, Apr. 2008. **5**
- [28] 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. **1**
- [29] Silvio Giancola, Anthony Cioppa, Adrien Delière, 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 Debieen, 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. **6**
- [30] Ankur Gupta, James J. Little, and Robert J. Woodham. Using line and ellipse features for rectification of broadcast hockey video. In *Can. Conf. Comput. Robot. Vis.*, pages 32–39, St. Johns, Canada, May 2011. **3**
- [31] Marc Gutiérrez-Pérez and Antonio Agudo. No bells, just whistles: Sports field registration by leveraging geometric properties. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3325–3334, Seattle, WA, USA, Jun. 2024. **2, 3, 4, 6, 7**
- [32] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, second edition, 2004. **3**
- [33] Jean-Bernard Hayet, Justus Piater, and Jacques Verly. Incremental rectification of sports fields in video streams with application to soccer. In *Adv. Concepts Intell. Vis. Syst. (ACIVS)*, pages 1–8, Brussels, Belg., Aug.-Sept. 2004. **2, 3**
- [34] Jean-Bernard Hayet, Justus Piater, and Jacques Verly. Fast 2D model-to-image registration using vanishing points for sports video analysis. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1–4, Genova, Italy, Sept. 2005. **2**
- [35] 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. **2, 3**
- [36] IFAB. Laws of the game. Technical report, The International Football Association Board, Zurich, Switzerland, 2022. **4**
- [37] ITU. Parameter values for the HDTV standards for production and international programme exchange, 2015. Recommendation ITU-R BT.709-6. **3**
- [38] ITU. Parameter values for ultra-high definition television systems for production and international programme exchange, 2015. Recommendation ITU-R BT.2020-2. **3**
- [39] 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. **2**
- [40] Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In *IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 2938–2946, Santiago, Chile, Dec. 2015. **2**
- [41] Jikai Lu, Jianhui Chen, and James J. Little. Pan-tilt-zoom SLAM for sports videos. In *Br. Mach. Vis. Conf. (BMVC)*, pages 1–14, Cardiff, Wales, Sept. 2019. **3, 6**
- [42] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. RTMDet: An empirical study of designing real-time object detectors. *arXiv*, abs/2212.07784, 2022. **5**
- [43] Floriane Magera. SoccerNet camera calibration challenge. <https://github.com/SoccerNet/sn-calibration>, Jun. 2022. **6**
- [44] Floriane Magera, Thomas Hoyoux, Olivier Barnich, and Marc Van Droogenbroeck. A universal protocol to benchmark camera calibration for sports. In *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3335–3346, Seattle, WA, USA, Jun. 2024. **2, 3, 6**
- [45] 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. **2, 3**
- [46] Adrien Maglo, Astrid Orcesi, and Quoc Cuong Pham. Kali-Calib: A framework for basketball court registration. *arXiv*, abs/2209.07795, 2022. **2**
- [47] Raul Mur-Artal, Jose M. M. Montiel, and Juan D. Tardos. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.*, 31(5):1147–1163, Oct. 2015. **2**
- [48] 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. **2**

- [49] Sachtler. Video 25 Plus and Video 25 Plus FB: Manual. Product description, 2023. [4](#)
- [50] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 4104–4113, Las Vegas, NV, USA, Jun. 2016. [2](#)
- [51] 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. [2](#)
- [52] 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. [2](#)
- [53] 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. [2](#)
- [54] Vladimir Somers, Victor Joos, Anthony Cioppa, Silvio Giancola, Seyed Abolfazl Ghasemzadeh, Floriane Magera, Baptiste Standaert, Amir M. 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/CVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW)*, pages 3293–3305, Seattle, WA, USA, Jun. 2024. [1](#)
- [55] Shuhei Tarashima. Sports field recognition using deep multi-task learning. *Journal of Information Processing*, 29(0):328–335, 2021. [2](#)
- [56] 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. [3](#), [4](#), [6](#), [7](#)
- [57] Hiroki Tsurusaki, Keisuke Nonaka, Ryosuke Watanabe, Tomoaki Konno, and Sei Naito. Sports camera calibration using flexible intersection selection and refinement. *ITE Trans. Media Technol. Appl.*, 9(1):95–104, 2021. [2](#)
- [58] Vinten. Vision 250 pan & tilt head. Specifications, 2023. [4](#)
- [59] Neng Zhang and Ebroul Izquierdo. A high accuracy camera calibration method for sport videos. In *IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, pages 1–5, Munich, Germany, Dec. 2021. [2](#)
- [60] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000. [3](#)