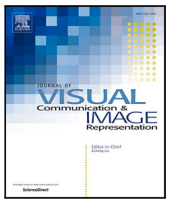
J. Vis. Commun. Image R. 117 (2026) 104781

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci



Full Length Article

Viewpoint-invariant soccer pitch registration using geometric and learned features☆

Carlos Cuevas *, Daniel Berjón, Narciso García

Grupo de Tratamiento de Imágenes (GTI), Information Processing and Telecommunications Center (IPTC), and ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Dataset link: https://www.gti.ssr.upm.es/data

Keywords:
Soccer field registration
Homography estimation
Projective invariants
Line and ellipse detection
Grass-band analysis
Sports video analytics

ABSTRACT

Automatic registration of broadcast soccer images to a standardized field model enables advanced analytics, augmented reality overlays, and precise player tracking. We propose a fully automatic, viewpoint-independent homography estimation pipeline fusing three complementary geometric cues: white field markings (lines and elliptical arcs), grass-band delimitations, and a binary playing-field mask. Detected primitives are first richly labeled — classifying lines as longitudinal or transversal, characterizing grass-tone transitions, and encoding four-quadrant intersection patterns — to reduce correspondence ambiguity. We then generate and prune candidate subsets of primitives, establish plausible matches to model elements via intersection-pattern rules and projective cross-ratio invariants, and systematically evaluate homography hypotheses using bidirectional mask-projection accuracies and mean reprojection error. An experimental evaluation on the LaSoDa benchmark demonstrates that the proposed method achieves highly accurate registrations with ground-truth primitives and robust performance in the fully automatic end-to-end pipeline. Furthermore, comparative experiments with recent state-of-the-art approaches confirm improved precision and robustness across diverse broadcast scenarios.

1. Introduction

Automatic analysis of soccer broadcast videos has recently attracted significant attention, as reflected in several recent surveys on player detection and tracking [1], soccer video analysis [2], and camera calibration [3]. In parallel, numerous recent works have addressed specific tasks such as multi-object tracking [4], tactical analysis [5], and event recognition [6]. Within this context, sports field registration, particularly in soccer broadcasts, has become an essential component in advanced sports analytics, augmented reality applications, and automated game understanding systems. This process establishes a geometric correspondence between the image captured by broadcast cameras and a standardized model of the playing field [7,8]. By computing this mapping, typically represented as a homography matrix, systems can translate image coordinates to real-world positions, enabling precise player tracking, tactical analysis, and enhanced viewing experiences [9,10].

The challenge of automatic sports field registration has attracted significant research attention due to its complexity and practical importance. Early approaches relied on detecting specific landmarks or geometric features such as field lines, circles, and corner points [9, 11]. While foundational, these methods often struggled with varying camera perspectives, partial field visibility, and occlusions. More recent work has explored both traditional computer vision techniques and deep learning-based approaches to address these limitations [7,8, 12]. In this context, existing methods can be broadly categorized into learning-based approaches, which rely on deep networks trained on large annotated datasets, and geometry-driven strategies that exploit the known structure of the playing field. Learning-based methods have shown strong performance in many scenarios, but often require extensive training data, may generalize poorly across leagues or broadcast styles, and offer limited interpretability [8]. Conversely, purely geometric approaches typically depend on a restricted subset of visual cues and may struggle under challenging viewpoints, partial field visibility, or heavy occlusions [9].

In contrast to existing approaches, this paper presents a novel method for automatic soccer field registration that explicitly combines multiple complementary geometric cues — including straight and curved white field markings, grass-band delimitation lines, and a playing-field mask — within a unified and fully automatic framework. By introducing a rich data labeling stage and leveraging local

☆ This paper has been recommended for acceptance by Junsong Yuan.

* Corresponding author.
E-mail addresses: carlos.cuevas@upm.es (C. Cuevas), daniel.berjon@upm.es (D. Berjón), narciso.garcia@upm.es (N. García).

https://doi.org/10.1016/j.jvcir.2026.104781
Received 6 October 2025; Received in revised form 6 February 2026; Accepted 14 March 2026
Available online 23 March 2026

intersection patterns together with projective invariants, the proposed approach substantially reduces correspondence ambiguity prior to homography estimation. This structured preprocessing, coupled with robust hypothesis testing, enables accurate and viewpoint-independent registration without requiring task-specific training. As a result, the method preserves interpretability, offers fine-grained control over failure modes, and maintains reliable performance even when certain field elements are occluded or poorly visible, making it particularly suitable for real-world broadcast scenarios.

The proposed approach consists of two main stages. First, a preprocessing phase characterizes the detected field elements and determines their most likely correspondences in the standard field model. This stage leverages both the geometric properties of the detected lines and their spatial relationships to establish initial matching hypotheses. Second, a systematic hypothesis testing framework evaluates different potential homographies between the image and field model. This evaluation incorporates both the projection of the field mask onto the model and the analysis of cross-ratios between intersection points of field lines, providing a geometrically robust criterion for selecting the optimal mapping.

Our contributions can be summarized as follows:

- A fully automatic field registration system that works from arbitrary camera perspectives without requiring specific assumptions about field visibility or camera position;
- A comprehensive feature extraction approach that leverages multiple visual cues including white field markings, grass strip delimitation lines, and the overall field mask;
- A novel preprocessing stage that characterizes detected elements and establishes their correspondence with the field model;
- A robust homography hypothesis testing framework based on field mask projection and cross-ratio analysis.

The remainder of this paper is organized as follows. Section 2 discusses related work in automatic sports field registration, covering both classical geometric methods and recent deep learning-based approaches. Section 3 presents our system overview and details the four main stages: data labeling (Section 4), data combination and selection (Section 5), hypothesis obtention (Section 6), and hypothesis evaluation and selection (Section 7). Section 8 reports experimental results, including performance with ground-truth inputs, end-to-end evaluation, and comparison to state-of-the-art. Section 9 concludes with a summary of our contributions and outlines future work.

2. Related work

Automatic soccer field registration has evolved from classical geometric pipelines to modern deep learning frameworks. We organize prior work into three categories — classical, deep learning, and hybrid methods — and highlight remaining challenges that motivate our approach.

2.1. Classical methods

Classical computer vision techniques have historically formed the foundation for soccer field of play registration. These methods typically involve the identification and subsequent matching of salient features present in the image to a predefined model of the soccer field of play [3]. While the field has seen a significant surge in the application of deep learning methodologies, classical approaches continue to provide a fundamental understanding of the problem and may still offer advantages in specific contexts, such as scenarios with limited computational resources or when interpretability of the registration process is paramount.

A significant portion of classical soccer field registration techniques relies on the detection and utilization of the white lines that delineate the playing area [13]. These lines, including goal lines, sidelines, the center circle, and penalty areas, serve as prominent and geometrically well-defined features [14]. Methods like the Hough transform are frequently employed to detect straight line segments and elliptical arcs that correspond to these field markings [15,16]. The RANdom SAmple Consensus (RANSAC) algorithm [17] is often applied to robustly estimate the homography, which represents the projective transformation between the image plane and the abstract model of the field of play. This is particularly useful in the presence of noise or when the detected lines are incomplete [14,18].

Many approaches further enhance registration accuracy and robustness by leveraging the inherent geometric constraints of a soccer field, such as the parallelism and orthogonality of certain lines and the known dimensions of the field and its markings. Some techniques focus on identifying key points at the intersections of these detected field lines [13], using the correspondences between these image points and their known locations on the field model to estimate the required transformation. Probabilistic decision trees have also been utilized to classify the detected lines and associate them correctly with the field model [19]. Furthermore, the concept of vanishing points, which are points in the image plane where parallel lines in the 3D world appear to converge, has been explored as a geometric cue for both camera calibration and subsequent field registration [20]. Despite the effectiveness of these methods in many scenarios, a significant limitation arises when there are insufficient visible keypoints on the field in the image, which can hinder robust registration [19]. This often occurs due to factors such as zoomed-in camera views or significant occlusions by players. The accuracy of these classical methods is intrinsically linked to the quality of the initial feature extraction, and they can be particularly challenged by poor lighting conditions or unusual camera angles.

Classical computer vision methods for soccer field registration possess several key strengths. They often rely on well-established geometric principles, and in certain scenarios, they can achieve real-time performance [21]. Furthermore, the registration process in classical methods is more interpretable compared to data-driven approaches. However, these methods also have inherent limitations. They can be sensitive to the quality of the input image, particularly regarding noise and blur, and they often struggle with occlusions of the field markings and significant changes in the camera's viewpoint [22]. Many classical techniques may require manual tuning of parameters to achieve optimal performance across different scenarios. Additionally, handling non-planar distortions in the image can be challenging for these methods. Achieving robust registration from arbitrary viewpoints without making prior assumptions about the camera's intrinsic or extrinsic parameters remains a significant hurdle [23]. For methods that process video sequences, the accumulation of reprojection errors from one frame to the next can lead to inaccuracies over time [24].

2.2. Deep learning methods

Recent years have witnessed a significant shift towards the application of machine learning, and particularly deep learning, techniques to address the problem of soccer image registration [8]. These data-driven methods possess the capability to learn intricate patterns and features directly from image data, which can potentially lead to improved robustness and accuracy in comparison to traditional methods that rely on handcrafted features [25]. Indeed, studies have shown promising results from deep learning-based approaches in soccer field registration when compared against traditional baselines [26].

A critical aspect of effectively training machine learning models for soccer image registration is the availability of large and diverse training datasets [25]. Creating such datasets, especially those encompassing the wide range of camera viewpoints encountered in real-world soccer matches, presents significant challenges [25]. Obtaining a sufficient quantity of accurately labeled data that covers various viewing angles,

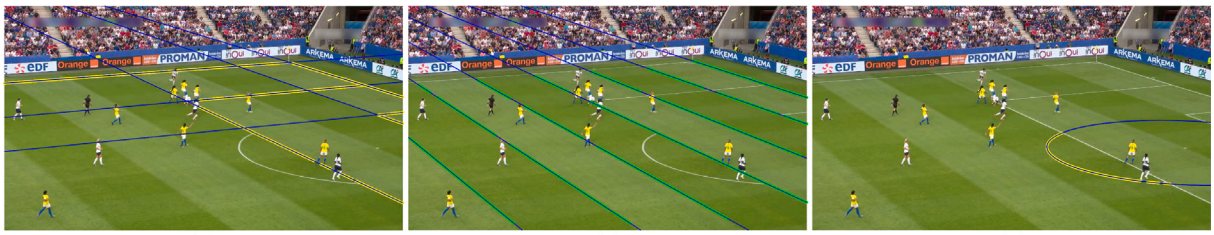


Fig. 1. Example of the data sets used as a starting point for carrying out the proposed image registration approach. (a) Set of straight white lines in W: six lines depicted in blue and points detected along the lines depicted in yellow. (b) Set of straight lines delimiting the grass bands, G: seven lines depicted in blue and points detected along the lines depicted in green. (c) Ellipses in E: one ellipse depicted in blue and points detected along the ellipse depicted in yellow.

lighting conditions, and field variations can be a laborious and expensive process. The use of synthetic data generation is a potential solution to mitigate the scarcity of real-world labeled data and to enhance the generalization capabilities of the models [27].

The machine learning literature also addresses the potential bias and limitations of models that are trained predominantly on data captured from a fixed "master camera" perspective, which is a common setup in broadcast scenarios [19]. Models trained on such a limited distribution of viewpoints may struggle to generalize effectively to images taken from significantly different angles, such as those from cameras held by fans or professional cameras that offer different coverage of the field (e.g., the aerial camera) [8]. Achieving true viewpoint invariance is therefore of the utmost importance for robust soccer image registration in diverse real-world scenarios. A key differentiator for a proposed method would be its ability to achieve viewpoint invariance, especially when compared to methods that might be biased towards a dominant camera perspective.

Various deep learning architectures have been employed for the task of soccer image registration. Convolutional Neural Networks (CNNs) are frequently used for feature extraction directly from the image pixels and for estimating the homography [26]. Encoder–decoder architectures have gained popularity for predicting keypoints on the field and generating heatmaps that represent the likelihood of these keypoints being present at specific image locations [25]. Vision Transformers (ViTs), which utilize attention mechanisms, have also been adopted to capture global features within the image, leading to improvements in registration accuracy [24]. Multi-task learning networks are employed to simultaneously perform tasks such as detecting the marker lines on the field and calculating the homography matrix [26]. Some approaches involve training deep networks to directly regress the registration error and then iteratively optimize the registration parameters based on this learned error [28].

2.3. Hybrid and projective-invariant approaches

Recent work seeks to combine learned detection with hard geometric constraints. Gutiérrez-Pérez and Agudo enforce cross-ratio and parallelism invariants on CNN-detected keypoints to prune false matches and achieve viewpoint invariance without manual tuning [8]. Magera et al. integrate lens-distortion calibration into a deep tracking framework, enabling robust registration on real broadcast cameras [29]. Such hybrids retain interpretability and enforce projective laws, yet most address only a subset of available cues (e.g. lines or keypoints) and lack a unified hypothesis selection mechanism.

2.4. Challenges and motivation

A comparative analysis reveals distinct strengths and weaknesses in both classical and machine learning approaches for achieving viewpoint-independent automatic registration. Classical methods, grounded in geometric principles, offer interpretability and can be computationally efficient. However, their reliance on accurate feature extraction makes them susceptible to image quality issues, occlusions, and viewpoint variations. Machine learning, particularly deep learning, has demonstrated the potential to learn robust features directly from data, leading to improved accuracy and the ability to handle complex image conditions. However, deep learning methods often require large, diverse, and accurately labeled datasets, which can be challenging to create, especially for a wide range of viewpoints. Furthermore, models trained on limited viewpoint data may exhibit biases and struggle to generalize to unseen perspectives.

Current trends in soccer image registration show an increasing adoption of deep learning techniques to overcome the limitations of classical methods. There is a growing focus on achieving viewpoint invariance to handle the diverse camera setups encountered in both professional broadcasts and user-generated content.

A notable gap in the current state of the art appears to be the explicit utilization of both white field lines (including elliptical markings) and grass stripes for viewpoint-independent automatic registration. While classical methods have explored both features to some extent, their combined use for robust viewpoint-independent registration is not prominently featured in the literature. Similarly, deep learning approaches, while powerful in learning features directly from data, might not explicitly model the geometric properties of grass stripes in conjunction with field lines for registration. The proposed method focuses on this specific combination of features to achieve viewpoint independence and therefore represents a significant contribution in this field.

Our proposed pipeline bridges this gap by jointly leveraging white lines, ellipses, and grass bands as complementary geometric cues. We perform rich data labeling (line types, intersection patterns), generate and prune correspondence hypotheses using projective invariants, and select the optimal homography via bidirectional mask-projection accuracy and reprojection error. This unified approach combines the interpretability and minimal supervision of classical cues with the robustness of deep detection, delivering reliable, viewpoint-independent registration even under severe player occlusions and challenging camera viewpoints.

3. System overview

Let I be an original RGB image of a soccer match and let M_{PF} be its binary playing-field mask. From I we also extract three complementary sets of geometric primitives:

1. $W = \{w_i\}_{i=1}^{N_w}$, the set of N_w straight white field-line segments;
2. $G = \{g_i\}_{i=1}^{N_g}$, the set of N_g straight lines delimiting the grass bands;
3. $E = \{e_i\}_{i=1}^{N_e}$, the set of N_e elliptical arcs corresponding to the penalty arcs and center circle.

Each element in W, E and G carries both its analytic equation and the set of image points supporting it (see Fig. 1). From these inputs, our overall registration pipeline proceeds through four main stages (see block-diagram in Fig. 2):

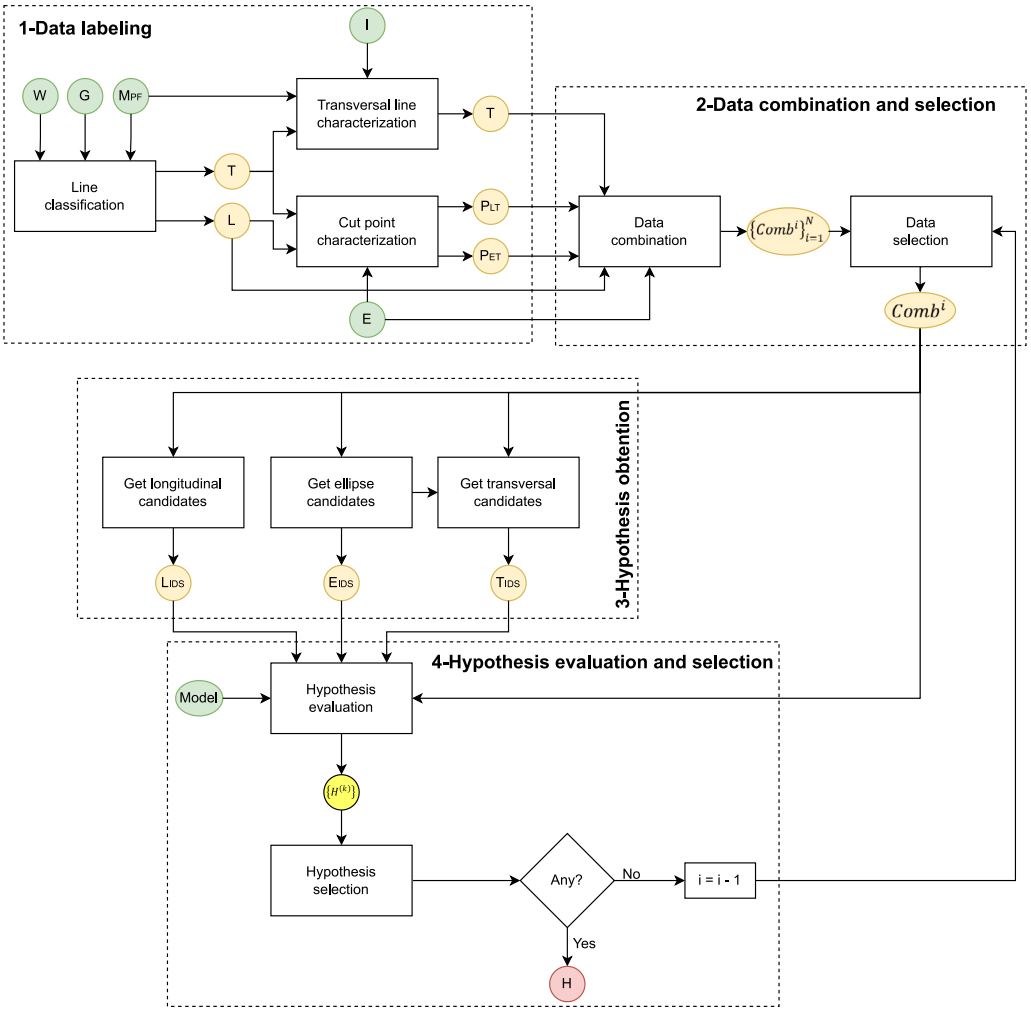


Fig. 2. Block diagram of the proposed strategy. Rectangular blocks denote processing blocks, diamond blocks denote decision making, and circular blocks denote data (inputs in green, intermediate in yellow, and outputs in red).

1. Data labeling (Section 4): We first classify each detected straight line in $W \cup G$ as either longitudinal or transversal, and we further characterize each transversal according to the direction of grass-tone transition (lighter → darker, darker → lighter, or indeterminate) relative to a reference point. We then compute all intersection ("cut") points between longitudinal–transversal lines and between ellipses–transversal lines, encoding for each the local pattern of white-line presence in four quadrants. This rich labeling dramatically reduces later combinatorial ambiguity.
2. Data combination and selection (Section 5): Denoting by N the total number of labeled elements, we form all subsets of size $i = N, N-1, \ldots$. We will attempt registration first using all N elements; if the resulting homography fails our quality criteria (Sections 6–7), we progressively consider subsets of size $N-1$, $N-2$, etc., thereby discarding false or poorly localized detections.
3. Hypothesis obtention (Section 6): For each candidate subset Comb^i, we generate plausible correspondences to the standard field model by: (a) matching each detected ellipse to one of the three arcs present in the model (left penalty, center circle, right penalty) via intersection-pattern rules and ordering according to a transversal reference point; (b) matching each longitudinal line to one of the six model sidelines using intersection patterns and projective-invariant cross-ratio tests; (c) matching each transversal line to one of the 21 model cross-lines by combining intersection patterns, cross-ratios, grass-transition types and relative position constraints with respect to the ellipses.
4. Hypothesis evaluation and selection (Section 7): Each full assignment (ellipse + longitudinal + transversal correspondences) yields a candidate homography. We evaluate each homography using three quantitative metrics in parallel: (i) the mask-projection accuracy from the image to the model; (ii) the mask-projection accuracy from the model back to the image; (iii) the mean reprojection error of the original line/ellipse points. Hypotheses passing the prescribed thresholds are then disambiguated by favoring the most complete sequence of transversal lines and, in case of ties, the lowest reprojection error. The surviving homography is output as the final image-to-model registration H.

By leveraging the complementary strengths of white-line geometry, grass-band structure and robust projective invariants, this four-stage pipeline achieves reliable, viewpoint-independent soccer-field registration without any manual initialization.

4. Data labeling

This section details the process of labeling the input data extracted from the image. The primary objective of the stages described herein is to accurately characterize the detected elements (straight lines and

ellipses) in order to significantly reduce the number of potential correspondences when attempting to associate them with the geometric elements of the standard soccer field model. This characterization process first involves classifying the detected straight lines as either longitudinal or transversal. Following this, a detailed characterization of the identified transversal lines is performed, considering the adjacent grass tones and establishing a reference point for ordering. Finally, the intersection points between the different straight lines and ellipses are identified and characterized by analyzing the presence of white line points in their immediate vicinity. The information obtained in this section is fundamental for the subsequent generation and evaluation of homography hypotheses.

4.1. Line classification

In this stage, the detected straight-line sets corresponding to grass-band delimitation lines G and white field markings W are partitioned into transversal lines T and longitudinal lines L. This classification step constitutes the first data labeling stage of the pipeline (see Fig. 2) and plays a key role in reducing the ambiguity of later correspondence and homography hypotheses.

Initialization using grass-band lines. We initially assume that all detected lines in G correspond to transversal lines and compute their associated vanishing point as

$$V_T = \text{vanishing\_point}(G). \tag{1}$$

This vanishing point provides an initial estimate of the dominant transversal direction in the image.

Compatibility-based expansion with white lines. Next, we analyze the set of detected white field markings W to identify those lines that are compatible with the transversal vanishing point V_T. Let $\text{Comp}(w, V_T)$ denote the compatibility test between a white line $w \in W$ and the vanishing point V_T, defined as in [30]. This test evaluates whether the orientation of w is consistent with convergence towards V_T under perspective projection.

The set of transversal lines is then defined as

$$T = G \cup \{\, w \in W \mid \text{Comp}(w, V_T) = \text{true} \,\}. \tag{2}$$

After updating T, the transversal vanishing point V_T is recomputed using all lines currently assigned to T. This process is iterated until no additional white lines from W satisfy the compatibility criterion. This refinement strategy ensures that V_T converges to a vanishing point supported jointly by grass-band lines and geometrically consistent white field markings.

Analysis of remaining white lines. Once the transversal set T has been finalized, the remaining white lines are given by

$$W' = W \setminus T. \tag{3}$$

These lines are candidates for the longitudinal set or correspond to spurious detections that must be discarded.

Among the lines in W', we first select the largest subset whose vanishing point lies outside the playing-field mask M_{PF}, as expected for valid longitudinal field lines. Previously, in order to remove false positives, we discard those lines in W' that intersect any transversal line outside the playing-field mask but within the image domain.

Let $p(w, t)$ denote the intersection point between a candidate white line $w \in W'$ and a transversal line $t \in T$. We define the validity condition for w as

$$\text{valid}(w) = \begin{cases} \text{true}, & \text{if } p(w,t) \in M_{PF},\ \forall t \in T, \\ \text{true}, & \text{if } p(w,t) \notin I,\ \forall t \in T, \\ \text{false}, & \text{otherwise}, \end{cases} \tag{4}$$

where I denotes the image domain.

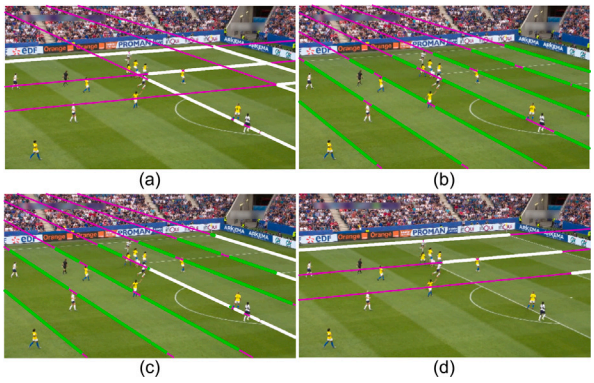


Fig. 3. Illustration of the line classification stage. (a) Original set of white field-line segments, W: lines depicted in magenta and points detected along the lines depicted in white. (b) Original set of straight lines delimiting the grass bands, G: lines depicted in magenta and points detected along the lines depicted in green. (c) Set of final transversal lines, T: lines depicted in magenta, points along the lines corresponding to white lines in white, and points along the lines corresponding to grass bands in green. (d) Set of final longitudinal lines, L: lines depicted in magenta and points along the lines in white.

The longitudinal set is then defined as

$$L = \{\, w \in W' \mid \text{valid}(w) = \text{true} \,\}. \tag{5}$$

Output of the classification stage. By construction, the set T contains all transversal lines, including grass-band delimitation lines and compatible white field markings, while L contains the validated longitudinal lines. The remaining non-valid elements of W' are discarded. The vanishing points associated with these sets are denoted as $V_T = \text{vanishing\_point}(T)$ and $V_L = \text{vanishing\_point}(L)$, and will be used in the subsequent characterization of transversal lines.

Fig. 3 illustrates the main steps of the line classification process described above. Panels (a) and (b) show the two initial sets of detected straight-line primitives: the white field markings W and the grass-band delimitation lines G, respectively. Starting from these inputs, the set of transversal lines T is constructed by combining the grass-band lines with those white lines that are compatible with the transversal vanishing point, as shown in Fig. 3.c. Finally, the remaining white lines are analyzed to identify valid longitudinal lines, yielding the final set L in Fig. 3.d.

4.2. Transversal line characterization

Once the straight lines have been classified into transversal lines T and longitudinal lines L, and their associated vanishing points V_T and V_L have been estimated, we characterize each transversal line by the direction of the grass–tone transition it induces. This information will later be used to constrain the correspondence between image primitives and the elements of the field model.

4.2.1. Dominant vanishing point and line ordering

We first determine the *dominant vanishing point*, denoted V_{dom}, as the vanishing point — either V_T or V_L — that lies closest to the image center $C = (c_x, c_y)$:

$$V_{\text{dom}} = \arg \min_{V \in \{V_L, V_T\}} \|V - C\|. \tag{6}$$

This point provides a stable geometric reference to consistently order both transversal and longitudinal lines.

If $V_{\text{dom}} = V_T$, transversal lines are ordered in decreasing angular order with respect to V_T, while longitudinal lines are ordered by

increasing Euclidean distance to V_T (from nearest to farthest). Conversely, if $V_{\text{dom}} = V_L$, longitudinal lines are ordered in increasing angular order with respect to V_L, while transversal lines are ordered by increasing distance to V_L.

In both cases, angles are measured clockwise with respect to the positive vertical image axis. This ordering induces a consistent notion of direction along each set of lines, which is essential for defining local neighborhoods and transition sides in a coherent manner.

4.2.2. Local neighborhood of a transversal line

For each transversal line $t \in T$, we define a local neighborhood around the line as

$$N(t) = \{\, x \mid \text{dist}(x, t) \le t_{\text{LN}} \,\}, \tag{7}$$

where t_{LN} is a small distance (e.g., 10 pixels for FullHD images), chosen to ensure that the neighborhood does not cross into adjacent grass bands.

This neighborhood is used to analyze the grass–tone variation across the line.

4.2.3. Incoming and outgoing half-planes

The neighborhood $N(t)$ is partitioned into two half-planes, corresponding to the *incoming* side $H_{\text{in}}(t)$ and the *outgoing* side $H_{\text{out}}(t)$. The definition of these two regions depends on the dominant vanishing point.

Case 1: $V_{\text{dom}} = V_L$. In this case, the separation between H_{in} and H_{out} is determined measuring the signed distance to the transversal line t. Let t be represented in homogeneous coordinates by (a, b, c). Points whose signed distance has the same sign as the distance from V_L are assigned to the incoming side, while points with opposite sign are assigned to the outgoing side:

$$\begin{aligned} H_{\text{in}}(t) = \{\, x \in N(t) \mid \text{sign}(ax_1 + bx_2 + c) \\ = \text{sign}(aV_{L,1} + bV_{L,2} + c) \,\}, \end{aligned} \tag{8}$$

$$H_{\text{out}}(t) = N(t) \setminus H_{\text{in}}(t). \tag{9}$$

Case 2: $V_{\text{dom}} = V_T$. When the dominant vanishing point is transversal, the notion of incoming and outgoing sides is defined angularly. Let $\theta(x, V_T)$ denote the angle of the ray connecting V_T to a point x, measured clockwise with respect to the positive vertical image axis, and let $\theta(t, V_T)$ denote the angle of the line t itself, defined by its direction with respect to V_T. The incoming side is then defined as the set of points whose angular position exceeds that of the line, while the outgoing side contains the remaining points:

$$H_{\text{in}}(t) = \{\, x \in N(t) \mid \theta(x, V_T) > \theta(t, V_T) \,\}, \tag{10}$$

$$H_{\text{out}}(t) = \{\, x \in N(t) \mid \theta(x, V_T) < \theta(t, V_T) \,\}. \tag{11}$$

4.2.4. Grass–tone transition characterization

Once the incoming and outgoing half-planes $H_{\text{in}}(t)$ and $H_{\text{out}}(t)$ have been defined for each transversal line $t \in T$, we characterize the grass–tone transition across the line.

Following the analysis reported in [30], we compute this transition using the blue channel of the RGB color space, which has been shown to provide better discrimination between different shades of green than alternative color components. The analysis is restricted to pixels belonging to the playing-field mask M_{PF} in order to avoid contamination from non-field regions. Let $\overline{I}_b(\cdot)$ denote the mean intensity of the blue channel within a given region. We compute the average grass intensity on each side of the line as

$$\mu_{\text{in}}(t) = \overline{I}_b\big(H_{\text{in}}(t) \cap M_{PF}\big), \tag{12}$$

$$\mu_{\text{out}}(t) = \overline{I}_b\big(H_{\text{out}}(t) \cap M_{PF}\big), \tag{13}$$

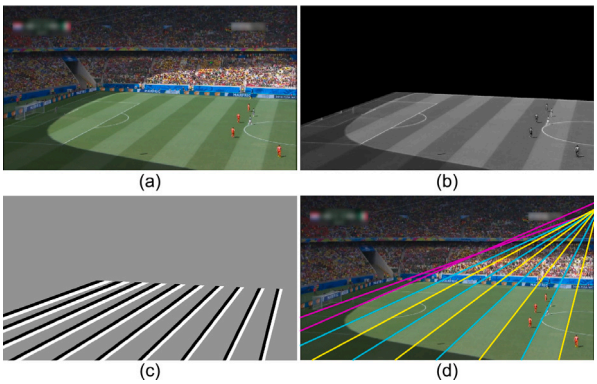


Fig. 4. Illustration of the transversal line characterization stage. (a) Original RGB image. (b) Blue channel of the original RGB image. (c) Incoming (in black) and outgoing (in white) half-planes corresponding to each transversal line in the image. (d) Result of the stage. The yellow lines are type 1 transversals, the cyan ones are type 2 transversals, and the magenta ones are type 3 transversals.

and define the grass–tone transition magnitude as

$$\Delta_b(t) = \mu_{\text{out}}(t) - \mu_{\text{in}}(t). \tag{14}$$

Based on this value, each transversal line is assigned a discrete transition type

$$\text{type}(t) = \begin{cases} 1, & \Delta_b(t) > \tau, \\ 2, & \Delta_b(t) < -\tau, \\ 3, & |\Delta_b(t)| \le \tau, \end{cases} \tag{15}$$

where τ is a small threshold introduced to suppress noise and insignificant intensity variations.

This classification distinguishes three types of transversal lines:

- Type 1: lines for which the grass on the incoming side (H_{in}) is lighter than on the outgoing side (H_{out});
- Type 2: lines for which the grass on the incoming side (H_{in}) is darker than on the outgoing side (H_{out});
- Type 3: lines for which no significant grass–tone difference can be reliably determined.

An illustrative example of the transversal line characterization process is shown in Fig. 4. Fig. 4.a presents the original RGB input image, while Fig. 4.b shows its corresponding blue channel, which is used for grass-tone analysis. Fig. 4.c depicts, for each detected transversal line, the associated incoming and outgoing half-planes (H_{in} and H_{out}), restricted to the playing-field mask. Finally, Fig. 4.d shows the output of this stage, where transversal lines are color-coded according to the assigned grass-tone transition type: type 1 (yellow), type 2 (cyan), and type 3 (magenta).

4.3. Cut point characterization

Once the longitudinal and transversal lines have been identified and ordered, we characterize the intersection points (cut points) arising from the interaction between different geometric primitives. Two types of cut points are considered: intersections between longitudinal and transversal straight lines, and intersections between transversal lines and elliptical arcs corresponding to the center circle and the penalty arcs.[1]

[1] It is worth noting that intersections between ellipses and longitudinal lines are not considered, since such configurations do not occur in the standardized soccer field model defined by the FIFA Laws of the Game [31].

Line–line cut points. Let L and T denote the sets of longitudinal and transversal lines, respectively. The set of line–line cut points is defined as

$$P_{LT} = \{\, p_{i,j} = \ell_i \cap t_j \mid \ell_i \in L,\ t_j \in T \,\}. \tag{16}$$

Each cut point $p_{i,j}$ represents a potential semantic intersection in the field layout (e.g., touchline–yard line, box corner).

For each cut point $p \in P_{LT}$, we define a neighborhood

$$N(p) = \{\, x \in I \mid \|x - p\| \le \delta_p \,\}, \tag{17}$$

where δ_p is chosen small enough to capture local line evidence while avoiding interference from adjacent markings (e.g., $\delta_p = t_{LN}$).

This neighborhood is partitioned into four regions (quadrants) induced by the two intersecting lines ℓ_i and t_j. The orientation of these quadrants is defined according to the ordering of the lines in the sets L and T, yielding a consistent notion of relative directions across the image.

For each quadrant Q_k, we determine whether white-line evidence is present:

$$b_k(p) = \begin{cases} 1, & \text{if white-line pixels are detected in } Q_k, \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

This results in a 4-bit binary descriptor

$$\mathbf{b}(p) = \big(b_1(p),\, b_2(p),\, b_3(p),\, b_4(p)\big), \tag{19}$$

which encodes the local intersection pattern around the cut point.

The ordering of the four components in the descriptor $\mathbf{b}(p)$ is defined consistently with the orientation induced by the vanishing points of the line sets. Specifically,

- $b_1(p)$ and $b_3(p)$ correspond, respectively, to the quadrants lying in the positive and negative directions of the axis defined by the transversal vanishing point V_T;
- $b_2(p)$ and $b_4(p)$ correspond, respectively, to the quadrants lying in the positive and negative directions of the axis defined by the longitudinal vanishing point V_L.

This convention ensures a consistent and viewpoint-invariant ordering of the binary patterns across images.

Line–ellipse cut points. We also consider cut points arising from the intersection between transversal lines and elliptical arcs corresponding to the center circle and the penalty arcs. Let E denote the set of detected ellipse primitives. The set of line–ellipse cut points is defined as

$$P_{LE} = \{\, p_{j,k} = t_j \cap e_k \mid t_j \in T,\ e_k \in E \,\}. \tag{20}$$

For each cut point $p \in P_{LE}$, a neighborhood $N(p)$ is defined analogously. In this case, the neighborhood is partitioned by the transversal line t_j and by the tangent direction of the ellipse at the intersection point. This induces four regions around p, for which the presence or absence of white-line evidence is evaluated in the same manner as for line–line cut points, yielding a 4-bit binary pattern.

These binary patterns provide a compact and viewpoint-invariant description of the local geometric configuration at each cut point. They are later matched against the corresponding patterns in the field model to establish plausible associations between detected image primitives and model elements, significantly reducing the ambiguity in the subsequent data combination stage.

Fig. 5 illustrates one example of the cut point characterization process described above. The top part of the figure shows line–line cut points obtained from the intersections between a longitudinal line and multiple transversal lines, together with their associated four-region patterns. The bottom part depicts line–ellipse cut points corresponding to intersections between transversal lines and the center circle. In all cases, cut points are highlighted in red, while the local neighborhoods are partitioned into four regions whose binary configurations are encoded in the descriptor $\mathbf{b}(p)$.

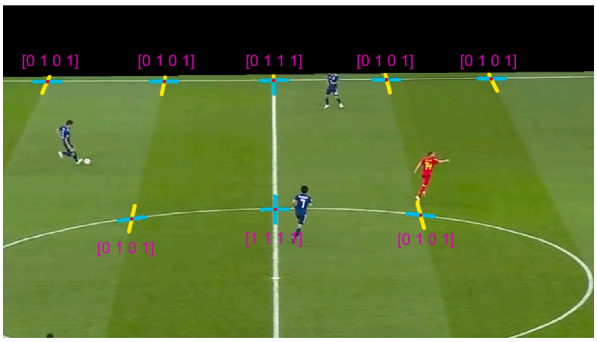


Fig. 5. Illustration of cut point characterization and associated intersection patterns. Top: line–line cut points arising from the intersections between one longitudinal line and five transversal lines. Bottom: line–ellipse cut points corresponding to intersections between transversal lines and the center circle. Cut points are marked in red. For each cut point, the four-region neighborhood is shown, with regions containing white-line evidence highlighted in cyan and regions without evidence shown in yellow. The corresponding binary descriptor $\mathbf{b}(p)$ is displayed in magenta next to each pattern.

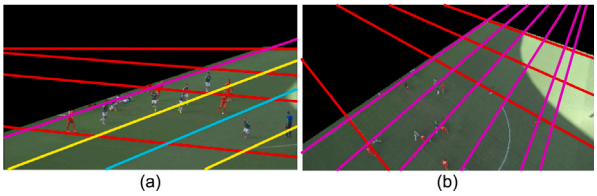


Fig. 6. Examples of images in which some of the initial elements are not suitable for calculating the homography that relates the images to the playing field model. The color notation is the same as that used in Fig. 4.d. (a) The upper longitudinal line is inaccurate. (b) The right transversal line is a false detection.

5. Data combination and selection

Once the detected primitives have been labeled and characterized as described in Section 4, the next stage consists in combining this information to generate a reduced set of plausible image–to–model associations. The goal of this stage is not to directly estimate the final homography, but to construct a small number of geometrically consistent hypotheses that will later be evaluated and ranked.

Let N_L, N_T, and N_E denote the number of longitudinal lines, transversal lines, and ellipses resulting from the data labeling stage, respectively. The total number of detected elements is then $N = N_L + N_T + N_E$.

5.1. Generation of data combinations

We define a hierarchy of combination sets $\{Comb^i\}_{i=1}^{N}$, where each $Comb^i$ contains all possible combinations formed by selecting exactly i elements from the initial set of N detected primitives. Formally,

$$Comb^i = \left\{C^{i,k}\right\}_{k=1}^{K^i}, \qquad K^i = \frac{N!}{i!\,(N-i)!}, \tag{21}$$

where K^i is the number of distinct combinations of size i.

Each combination $C^{i,k}$ groups a subset of longitudinal lines, transversal lines, and ellipses, together with the cut points induced by their intersections, and is defined as

$$C^{i,k} = \left(\{L_l^{i,k}\}_{l=1}^{N_L^{i,k}}, \{T_t^{i,k}\}_{t=1}^{N_T^{i,k}}, \{E_e^{i,k}\}_{e=1}^{N_E^{i,k}}, P_{LT}^{i,k}, P_{ET}^{i,k}\right), \tag{22}$$

where

$$N_L^{i,k} + N_T^{i,k} + N_E^{i,k} = i.$$

Here, $\{L_l^{i,k}\}$, $\{T_t^{i,k}\}$, and $\{E_e^{i,k}\}$ denote the subsets of longitudinal lines, transversal lines, and ellipses included in the combination, respectively, while $P_{LT}^{i,k}$ and $P_{ET}^{i,k}$ represent the sets of line–line and line–ellipse cut points associated with that combination.

5.2. Progressive reduction strategy

At the beginning of the process, it is assumed that all N detected elements correspond to correct and sufficiently accurate observations. Under this assumption, only the combination set $Comb^N$ (i.e., $i = N$) needs to be analyzed, which contains a single combination involving all detected primitives.

If this combination allows obtaining a homography that registers the image onto the field model with sufficient quality (as described in Sections 6 and 7), the process terminates and no further combinations are considered.

In practice, however, some detected elements may correspond to false positives or to detections that are not accurate enough to be jointly explained by a single homography (see Fig. 6). In such cases, it is not possible to obtain a valid registration using all N elements simultaneously.

To handle this situation, the value of i is progressively decreased, starting from $i = N - 1$, and the combinations in $Comb^i$ are evaluated. If any combination in $Comb^i$ yields a homography of sufficient quality, the process stops and that solution is retained. Otherwise, the value of i is further reduced, allowing the method to discard multiple incorrect or unreliable elements if necessary.

This progressive reduction strategy ensures robustness to spurious or inaccurate detections while keeping the number of evaluated combinations manageable in practice because the number of false detections is typically very low.

6. Hypothesis obtention

The data combination and selection stage (Section 5) yields candidate sets $Comb^i$ containing geometric primitives that potentially correspond to actual field elements. However, establishing the correct mapping between these detected elements and their counterparts in the standardized field model remains a challenging correspondence problem. A naive approach that considers all possible associations would generate an exponential number of hypotheses, which would be computationally prohibitive to evaluate and also wasteful, since most of the hypotheses would be geometrically implausible.

To address this challenge, we leverage the rich characterization performed during the data labeling stage (Section 4) to drastically prune the correspondence search space. Our approach exploits three key insights: (1) the intersection patterns between lines and ellipses provide strong geometric signatures that eliminate many impossible matches; (2) projective invariants such as cross-ratios remain constant under homographic transformations, allowing us to verify correspondence consistency; and (3) the grass-band transition types provide additional constraints that further reduce ambiguity. For each combination $C^{i,k}$ within $Comb^i$, we systematically determine plausible correspondences between the detected image elements and the standard field model through three specialized modules. Each module exploits different aspects of the geometric structure to establish reliable element-to-model associations while maintaining computational efficiency.

6.1. Ellipse correspondence

The first step in our correspondence framework focuses on associating detected elliptical arcs with the three canonical ellipses in the FIFA field model: the left penalty arc, center circle, and right penalty arc. This initial correspondence is particularly valuable because ellipses provide strong geometric anchors that constrain the subsequent line matching process. Our approach exploits the spatial relationships between ellipses and transversal lines, specifically analyzing the intersection patterns encoded during the cut point characterization stage (Section 4.3). When an ellipse intersects a transversal line, the resulting intersection points carry distinctive four-quadrant patterns that reveal which sides of the intersection contain white field markings. The correspondence rules are derived from the known geometry of soccer fields. If an ellipse intersects a transversal line such that white markings appear on both sides of the transversal (as indicated by the intersection pattern $\pi(p)$), this configuration can only occur at the center circle, where the circle boundary extends across multiple field regions. Conversely, if white markings appear only on one side of the intersection, the ellipse must correspond to one of the penalty arcs, with the specific side indicating whether it belongs to the left or right penalty area. Once we establish a confident identification for one ellipse (particularly the center circle), we can exploit the known spatial ordering to constrain the remaining associations. If additional ellipses are detected, their positions relative to the reference transversal directions (incoming side H_{in} versus outgoing side H_{out}) directly indicate whether they correspond to the left or right penalty arcs.

6.2. Longitudinal correspondence

Once ellipse correspondences have been established, we proceed to associate detected longitudinal lines with the six primary longitudinal elements in the standard field model (designated L_1 through L_6). This matching process combines two complementary geometric constraints: local intersection patterns and global cross-ratio invariants.

The intersection pattern analysis compares the four-quadrant encodings at longitudinal-transversal intersection points against the known patterns derivable from the field model geometry. Each intersection pattern $\pi(p) = [\pi_u(p), \pi_r(p), \pi_d(p), \pi_\ell(p)]^T$ acts as a geometric fingerprint that significantly reduces the number of plausible model line candidates for each detected longitudinal element.

For scenarios with multiple detected longitudinal lines ($N_L^{i,k} > 3$), we employ cross-ratio analysis to further validate correspondences. The cross-ratio represents a fundamental projective invariant—a quantity that remains constant under homographic transformations. For any four collinear intersection points P_1, P_2, P_3, P_4, their cross-ratio is defined as:

$$CR = Cross\left(P_1, P_2, P_3, P_4\right) = \frac{\left(P_1P_2\right) \cdot \left(P_3P_4\right)}{\left(P_1P_3\right) \cdot \left(P_2P_4\right)}, \tag{23}$$

where $\left(P_iP_j\right)$ represents the signed distance between points P_i and P_j. Since the mapping between image and field model is precisely a homographic transformation, the cross-ratio computed from four intersection points along any detected longitudinal line must match the cross-ratio of the corresponding points on the true model line. This constraint provides a powerful verification mechanism: we compute cross-ratios for all possible four-point combinations along each detected line and compare them against pre-computed cross-ratios for the candidate model lines. Only correspondences with compatible cross-ratios (within a specified tolerance) are retained as plausible associations.

6.3. Transversal correspondence

The final correspondence stage addresses the most complex matching problem: associating detected transversal lines with the 21 distinct transversal elements in the standard field model (T_1 through T_{21}). This complexity arises from the larger number of model candidates and the varying geometric contexts in which transversal lines appear.

Our approach integrates four complementary constraint types to achieve reliable correspondences. First, we employ the two analyses used for longitudinal lines (intersection pattern and cross-ratio), comparing detected intersection signatures against those expected from the model geometry.

Second, we exploit the grass-band transition characterization performed in Section 4.2. The classification of transversal lines into three types — Type 1 (light-to-dark grass transition), Type 2 (dark-to-light transition), and Type 3 (indeterminate transition) — provides a powerful filtering mechanism. Based on the known grass-band structure of FIFA-compliant fields, Type 1 lines can only correspond to even-numbered model transversals (T_2, T_4, …, T_{20}), while Type 2 lines are restricted to odd-numbered elements (T_1, T_3, …, T_{21}). Type 3 lines, lacking clear transitions, remain compatible with any model transversal.

Finally, we impose spatial ordering constraints derived from the established ellipse correspondences. If a detected ellipse has been associated with a specific model ellipse (e.g., the center circle), then transversal lines positioned before this ellipse cannot correspond to model elements located beyond the center line, and vice versa. These relational constraints, based on the consistent ordering of field elements, eliminate geometrically impossible associations.

By systematically applying these four constraint types — intersection patterns, cross-ratios, grass-band transitions, and spatial relationships — we generate a pruned set of plausible correspondences for each detected transversal line. This comprehensive filtering dramatically reduces the hypothesis space for the subsequent evaluation stage.

7. Hypothesis evaluation and selection

The correspondence generation process described in the previous section produces candidate lists of plausible matches between detected elements and model components. However, multiple correspondence combinations may satisfy the geometric constraints, requiring a systematic evaluation framework to identify the optimal registration. The challenge lies in distinguishing between geometrically consistent hypotheses and selecting the most accurate among potentially several viable options. Our evaluation strategy addresses this challenge through a two-stage process. First, we assess the quality of each complete correspondence hypothesis evaluating both the local accuracy of feature alignments and the global coherence of the overall field mapping. Second, when multiple hypotheses pass our quality thresholds, we apply a principled selection mechanism that prioritizes structural completeness and geometric precision. By combining multiple independent quality measures, we ensure that the selected homography maintains high accuracy across diverse viewing conditions and field configurations.

7.1. Hypothesis quality assessment

Each complete correspondence hypothesis — comprising ellipse, longitudinal, and transversal line associations — yields a candidate homography matrix that maps between image coordinates and the standard field model. To assess the quality of this mapping, we employ three complementary evaluation metrics that together provide a comprehensive measure of geometric consistency. The evaluation approach recognizes that a correct homography should satisfy multiple geometric properties simultaneously. Local feature alignments should be precise, meaning that detected points should project accurately to their corresponding model locations. Additionally, the global field structure should be preserved, ensuring that the overall playing area maintains proper spatial relationships under the transformation. Finally, the mapping should be bidirectionally consistent, meaning that projection from image to model and back should yield minimal distortion.

7.1.1. Bidirectional mask projection analysis

The first component of our evaluation framework assesses how well the homography preserves the overall spatial structure of the playing field through bidirectional mask projection analysis. This evaluation exploits the fact that a correct homography should map field regions to field regions and non-field regions to non-field regions, both in the forward and reverse directions.

For the forward direction (image to model), we generate a uniform grid of evaluation points covering both the detected playing field mask M_{PF} and its immediate surroundings in the image. These grid points are classified into three categories: points clearly inside M_{PF}, points clearly outside M_{PF}, and boundary points where the classification is uncertain due to potential segmentation inaccuracies (see Fig. 7.a). To ensure robust evaluation, we exclude the uncertain boundary points from our analysis.

The remaining grid points are projected onto the standard field model coordinate system using the candidate homography. We then verify whether points originally inside the image mask correctly project inside the model field boundaries, and whether points originally outside the mask correctly project outside the model boundaries. The forward projection accuracy is computed as:

$$Acc_{m\to M} = 100\frac{N_{m\_in\to M\_in} + N_{m\_out\to M\_out}}{N_{total\_valid}}, \tag{24}$$

where $N_{m\_in\to M\_in}$ represents points mapped from inside the image mask to inside the model field, $N_{m\_out\to M\_out}$ represents points correctly mapped from outside the image mask to outside the model field, and N_{total_valid} is the total number of valid evaluation points.

The reverse direction (model to image) follows a similar procedure but inverts the projection direction. Grid points are generated over the standard field model (see Fig. 7.b) and projected back to the image plane using the inverse homography. The reverse projection accuracy is calculated as:

$$Acc_{M\to m} = 100\frac{N_{M\_in\to m\_in} + N_{M\_out\to m\_out}}{N'_{total\_valid}}, \tag{25}$$

where $N_{M\_in\to m\_in}$ represents points correctly mapped from inside the model field to inside the image mask, $N_{M\_out\to m\_out}$ represents points correctly mapped from outside the model field to outside the image mask, and N'_{total_valid} is the total number of valid evaluation points.

High values for both $Acc_{m\to M}$ and $Acc_{M\to m}$ indicate strong geometric consistency, confirming that the homography correctly preserves the overall field structure in both projection directions (see examples in Fig. 7).

7.1.2. Feature reprojection accuracy

The third evaluation metric directly measures the geometric precision of the homography by computing reprojection errors for the original feature points used in the correspondence generation. This metric provides a local assessment of alignment quality, complementing the global structure evaluation provided by mask projection analysis.

For each feature point p_{img} detected on lines or ellipses within the current combination $C^{i,k}$, we compute its corresponding location on the standard field model using the candidate homography: $p_{model} = H \cdot p_{img}$. This model point is then projected back to the image plane using the inverse transformation: $p'_{img} = H^{-1} \cdot p_{model}$.

The reprojection error for each point is computed as the Euclidean distance between the original image point and its reprojected position. The overall quality metric is the mean reprojection error across all relevant feature points:

$$E_{reproj} = \frac{1}{N_{points}} \sum_{i=1}^{N_{points}} d(p_{img}, p'_{img}). \tag{26}$$

Low values of E_{reproj} indicate that the homography accurately aligns detected features with their corresponding model elements, suggesting strong local geometric consistency. Conversely, high mean reprojection errors signify substantial misalignments, indicating that the correspondence hypothesis is likely incorrect.

Fig. 7 provides illustrative examples of these evaluation metrics for both correct and incorrect hypotheses. As shown in the first example (Case 1), a correct hypothesis typically yields high accuracy values for both projection analyses ($Acc_{m\to M}$ and $Acc_{M\to m}$), often approaching 100%, indicating strong consistency between the image mask and the

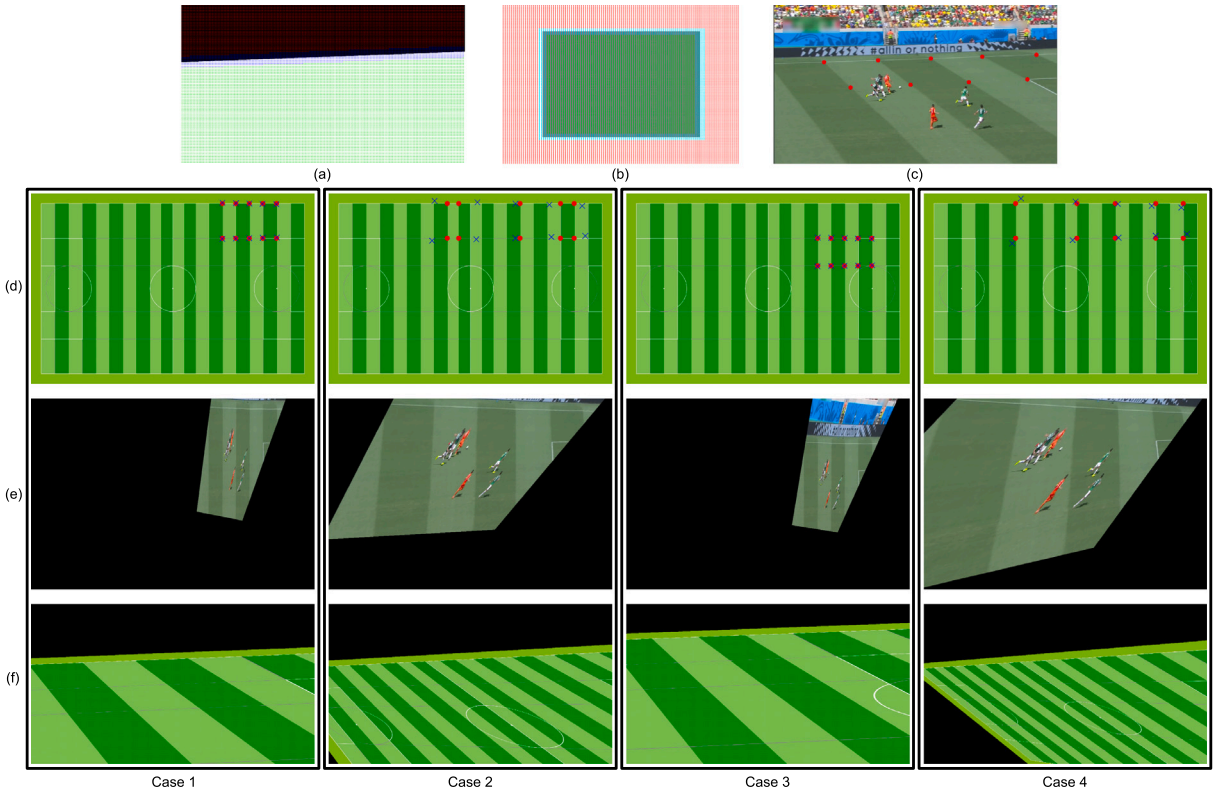


Fig. 7. (a) Grid of points covering both the playing field mask (M_{PF}) and its immediate exterior in the image. (b) Grid of points covering both the playing field model and its immediate exterior. (c) Original image, I, and the intersection points used in the hypotheses under evaluation. (d) Field of play model with the points considered in four different hypotheses (red dots) and their reprojections (blue crosses). (e) Result of projecting M_{PF} on the field of play model canvas. (f) Result of projecting the field of play model on the image canvas. Case 1: Correct hypothesis (5 pixels of reprojection error, $Acc_{m\to M} = 100$, $Acc_{M\to m} = 99.7$). Case 2: The projection of the masks is good (close to $Acc_{m\to M} = 99.7$ and $Acc_{M\to m} = 99.5$), but the reprojection error of the points is not (114 pixels). Case 3: The reprojection error of the points is good (4 pixels), but the projection of the masks is not ($Acc_{m\to M} = 85.6$, $Acc_{M\to m} = 78.2$). Case 4: Both the reprojection error (44 pixels) and the projection of the masks ($Acc_{m\to M} = 89.7$, $Acc_{M\to m} = 85.5$) are wrong.

projected model. Furthermore, the mean reprojection error (E_{reproj}) for a correct hypothesis is generally very low (e.g., around 5 pixels in the example), signifying excellent geometric alignment of the detected features. The figure also showcases examples corresponding to poor hypotheses. In these cases, either the mean reprojection error is substantially high (Case 2), or the projection accuracy metrics yield significantly lower values (Case 3), or potentially both conditions occur simultaneously (Case 4), clearly distinguishing incorrect registrations from valid ones.

7.2. Optimal hypothesis selection

Once a set of candidate homographies $\{H^{(k)}\}$ has been generated and evaluated, a final decision must be made to select the most reliable one. This step is critical, since even if several hypotheses satisfy the basic evaluation thresholds, only one can be retained as the image-to-model registration.

First, we measure the density of detected transversal lines. Let $T^{(k)}$ be the set of transversal lines in the model under $H^{(k)}$, and $\hat{T}^{(k)}$ be the set of transversal lines actually detected in the image that match $T^{(k)}$. We then define the transversal density ratio as

$$\rho^{(k)} = \frac{\hat{T}^{(k)}}{T^{(k)}}. \tag{27}$$

This ratio quantifies the completeness of the transversal-line sequence associated with hypothesis $H^{(k)}$. In practice, higher values of $\rho^{(k)}$ indicate more plausible hypotheses, since missing transversal detections are unlikely to occur between correctly detected ones.

The optimal hypothesis is thus selected as

$$H = \arg\max_{H^{(k)}} \rho^{(k)}. \tag{28}$$

In the event of ties, i.e. when multiple hypotheses achieve the same transversal density ratio, we resolve ambiguity by selecting the hypothesis with the smallest mean reprojection error:

$$H = \arg\min_{H^{(k)}\in\mathcal{H}_{\max}} E^{(k)}_{\text{reproj}}, \tag{29}$$

where $\mathcal{H}_{\max} = \{H^{(k)} \mid \rho^{(k)} = \max_j \rho^{(j)}\}$.

This selection strategy ensures that the chosen homography H simultaneously maximizes the structural consistency of transversal lines and minimizes geometric reprojection error, thereby providing a robust and accurate registration outcome.

8. Results

Throughout this section, only the most representative results are presented directly. The complete set of results and supplementary materials is publicly available at https://www.gti.ssr.upm.es/data.

The section is organized as follows:

- Section 8.1 describes the LaSoDa dataset used to assess the quality of the proposed strategy. It also explains the rationale for choosing this dataset over alternative options.
- Section 8.2 presents the quantitative metrics employed to evaluate registration accuracy.
- Section 8.3 reports the results obtained when the system uses ground-truth data for the playing field mask, white lines, and

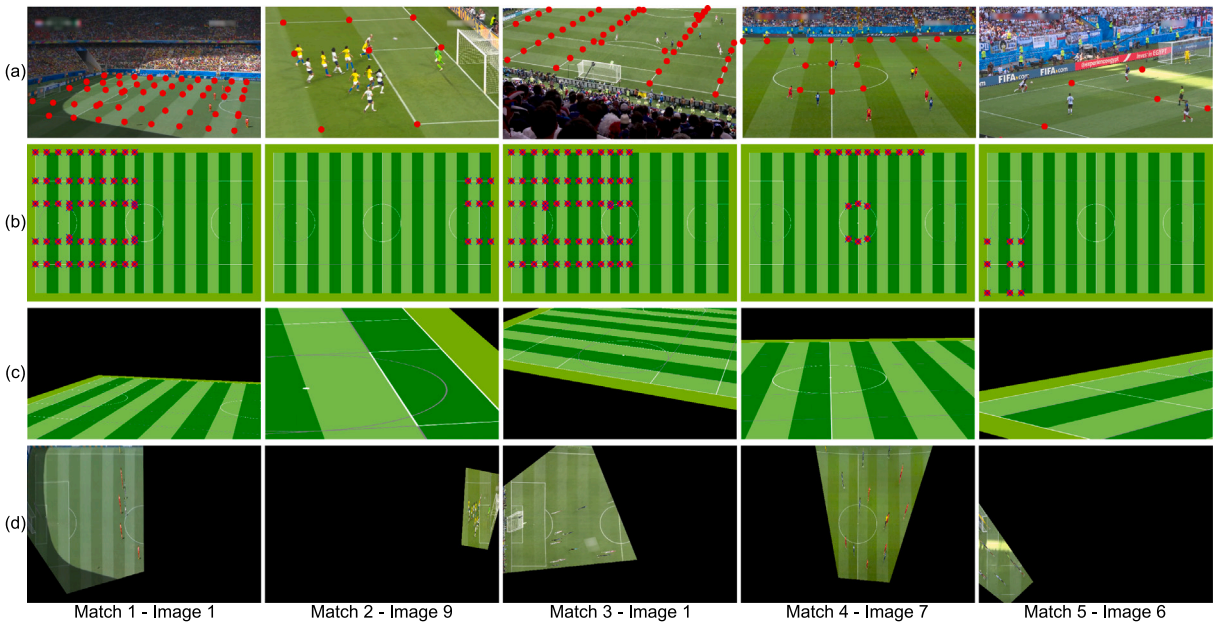


Fig. 8. Representative results corresponding to the performance evaluation with ground truth inputs. (a) Original image, I, and the intersection points used. (b) Field of play model with the points considered (red dots) and their reprojections (blue crosses). (c) Result of projecting M_{PF} on the field of play model canvas. (d) Result of projecting the field of play model on the image canvas.

grass bands (as provided by LaSoDa) as input. This setup allows for an isolated assessment of the core registration algorithm, independent of possible errors introduced by earlier feature detection stages.

- Section 8.4 evaluates the performance of the complete pipeline, using the actual outputs of the feature extraction procedures described in Section 3 as input. This analysis reflects the real-world behavior of the system under imperfect input conditions, including false detections and missing primitives.
- Section 8.5 provides a comparative analysis, benchmarking the proposed strategy against several existing methods for automatic soccer field registration.
- Section 8.6 analyzes the computational cost of the proposed method, reporting runtime statistics for the main processing stages under both ground-truth and end-to-end configurations.

8.1. Data

To comprehensively evaluate the performance of the proposed registration strategy, we utilize the LaSoDa dataset [22]. LaSoDa comprises 60 fully annotated images captured during matches across five different stadiums, exhibiting diverse characteristics such as varying camera positions, view angles, and grass color patterns, under different lighting conditions (both day and night). The images within LaSoDa cover all areas of the playing field and feature five distinct zoom levels, ranging from 1 (closest zoom) to 5 (widest zoom). Furthermore, the data was acquired using four different camera types – master camera (MC), side camera (SC), end camera (EC), and aerial camera (AC) – and includes challenging scenarios like heavily shaded images.

To the best of our knowledge, only two other datasets — the WorldCup 2014 dataset [20] and the TS-WorldCup dataset [12] — offer full annotations, including homography matrices that map each image to a field model, thereby enabling quantitative evaluation of registration performance. However, both datasets suffer from two key limitations: first, all images were captured from similar wide-angle viewpoints and from essentially the same camera position, limiting the diversity of perspectives represented; second, and more critically, many of the provided homography matrices exhibit insufficient accuracy to be considered reliable ground truth for precise model-to-image registration. These shortcomings make LaSoDa a more appropriate benchmark for evaluating the robustness and accuracy of registration methods under diverse and realistic conditions.

Another relevant dataset is SoccerNetV3-Calibration [32], which has been used by several authors to evaluate soccer field registration and calibration strategies. However, unlike the previously mentioned datasets, it does not provide homographies that map the images to a canonical field model. Instead, it only includes sparse coordinates of some white field lines within each image. Furthermore, the dataset consists of images from professional matches played in competitions where the exact field dimensions (including the width of grass bands) are not standardized or publicly available. In contrast, all images in LaSoDa comply with the field dimensions defined by the FIFA [31,33], enabling accurate and consistent registration to a known field model.

8.2. Evaluation metrics

To evaluate the quality of the registration results, a uniform grid of 1800 control points was generated over the field model. These points were projected onto the original image using the ground-truth homography matrix. Among these, the N_Q points that fall within the playing area in the image were selected for evaluation.

Each of these N_Q points was then reprojected back onto the model field using the homography matrix H estimated by the proposed method. The reprojection error was computed as the average Euclidean distance between the ground-truth projections and the estimated ones, both in pixels, E_{pix}, and in centimeters, E_{cm}. Since the field model was defined such that one pixel corresponds to 4 cm, the two errors are related by a simple scaling factor.

These metrics provide a robust and interpretable measure of geometric alignment accuracy between the estimated and true homographies.

8.3. Performance with ground-truth inputs

To isolate the accuracy of our homography estimation from any errors in line, ellipse or grass-band detection, we first feed the system with perfect ground-truth inputs provided by the LaSoDa dataset: the

Table 1
Results corresponding to the performance evaluation with ground truth inputs. Lines highlighted in red correspond to cases where the estimated homography was completely incorrect.

Match	Image id.	N_P	$Acc_{m\to M}$	$Acc_{M\to m}$	E_{reproj}	N_Q	E_{pix}	E_{cm}
1	1	54	100%	100%	2.4	773	1.8	7.1
	2	13	100%	100%	0.1	343	0.7	2.7
	3	130	100%	100%	16.1	1700	14.6	58.2
	4	27	100%	99%	1.3	342	1.6	6.6
	5	12	100%	100%	0.5	247	0.8	3.1
	6	36	100%	97%	0.9	655	0.5	1.9
	7	9	100%	100%	0.2	166	0.8	3.1
	8	25	100%	99%	2.0	217	0.8	3.2
	9	50	100%	100%	1.6	537	0.9	3.5
	10	42	100%	100%	1.8	521	0.8	3.2
	11	20	100%	100%	5.4	286	10.7	42.8
	12	20	100%	100%	0.7	294	0.8	3.3
2	1	16	100%	100%	0.2	526	0.3	1.2
	2	30	100%	100%	0.8	268	0.6	2.5
	3	23	100%	100%	1.7	322	1.6	6.5
	4	4	100%	100%	0.0	223	0.5	2.2
	5	32	100%	100%	0.5	405	0.5	1.9
	6	42	100%	100%	2.2	424	1.6	6.6
	7	14	100%	100%	0.2	452	0.7	2.7
	8	42	100%	100%	1.8	459	0.6	2.5
	9	9	100%	100%	0.6	123	0.5	2.1
	10	9	100%	100%	0.7	100	0.6	2.3
	11	30	100%	100%	0.4	334	1.5	6.1
	12	38	100%	100%	5.7	757	11.1	44.6
3	1	66	100%	100%	1.2	649	0.3	1.2
	2	56	100%	100%	1.3	580	0.7	2.9
	3	15	100%	100%	0.5	185	0.6	2.3
	4	15	100%	100%	0.5	191	0.4	1.5
	5	92	99%	99%	6.0	1013	3.2	12.8
	6	7	100%	100%	0.1	241	0.6	2.4
	7	27	100%	100%	0.4	288	0.4	1.5
	8	110	100%	100%	6.9	1234	5.7	22.6
	9	54	100%	100%	1.2	715	0.8	3.1
	10	20	100%	100%	0.9	322	0.7	2.9
	11	20	100%	100%	0.5	264	0.6	2.2
	12	61	100%	100%	1.6	719	0.8	3.1
4	1	61	100%	100%	2.0	729	2.2	8.7
	2	47	100%	100%	1.5	524	1.1	4.4
	3	20	100%	100%	0.4	301	0.5	1.9
	4	42	100%	100%	2.6	542	2.3	9.2
	5	20	100%	100%	0.6	312	0.4	1.5
	6	54	100%	100%	5.7	866	3.5	13.9
	7	17	100%	100%	0.2	588	0.7	2.9
	8	11	100%	100%	0.5	166	0.5	2.1
	9	39	100%	100%	0.6	694	0.3	1.2
	10	20	100%	100%	0.6	319	0.3	1.4
	11	47	100%	100%	1.4	552	0.2	0.9
	12	118	100%	100%	2.3	1369	2.2	8.9
5	1	24	100%	100%	1.1	588	0.8	3.3
	2	14	100%	100%	0.1	368	0.4	1.5
	3	23	100%	100%	1.0	346	0.7	2.7
	4	23	100%	100%	0.4	306	0.9	3.4
	5	6	100%	100%	0.3	57	369.2	1477.0
	6	9	100%	100%	0.3	122	0.5	2.1
	7	11	100%	100%	0.5	202	0.6	2.4
	8	9	90%	96%	2.5	739	2557.3	10 229.4
	9	27	100%	100%	0.8	323	0.5	2.2
	10	42	100%	100%	3.3	513	1.6	6.6
	11	76	100%	100%	3.6	791	2.9	11.7
	12	20	100%	100%	0.7	254	0.5	2.0

exact playing field mask, white-line segments, ellipse arcs and grass-band boundaries. In this configuration, the only sources of error are the hypothesis generation, evaluation, and selection stages, as well as slight inaccuracies in the ground-truth annotations and the presence of radial distortion in the images.

Fig. 8 presents representative examples — one for each of the five matches in LaSoDa — illustrating (a) the original image with its annotated intersection points, (b) the corresponding field model with the reprojected points, (c) the projection of the field model onto the image, and (d) the projection of the image mask onto the model canvas. Table 1 reports, for each of the 60 test images, the number of feature points used to compute the homography, N_{points}; the mask-projection accuracies $Acc_{m\to M}$ and $Acc_{M\to m}$; the mean reprojection error E_{reproj}; the number of reprojected points used in the evaluation metric defined in the previous subsection, N_Q; and the corresponding evaluation error in both pixels, E_{pix}, and centimeters, E_{cm}.

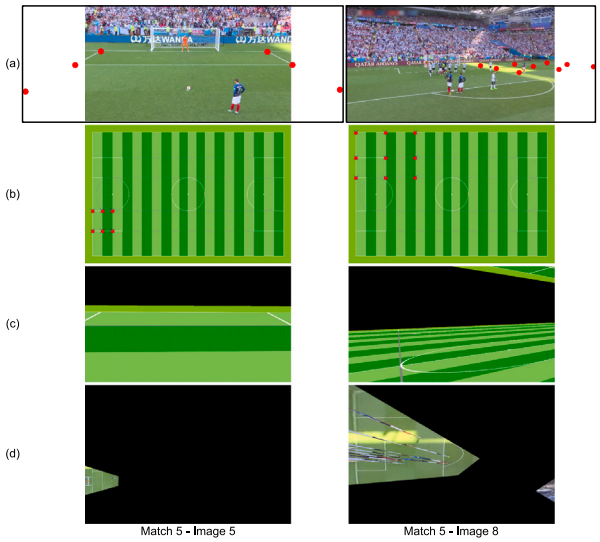


Fig. 9. Registration failures using ground truth inputs. (a) Original image, I, and the intersection points used. (b) Field of play model with the points considered (red dots) and their reprojections (blue crosses). (c) Result of projecting M_{PF} on the field of play model canvas. (d) Result of projecting the field of play model on the image canvas.

Across all 60 cases, only 2 images exhibited distinctly higher reprojection error values, indicating a failure in registration for these specific instances (lines highlighted in red in Table 1). For the remaining images, the registration proved highly precise, achieving a mean reprojection error of $E_{\text{pix}} = 1.6$ pixels (corresponding to $E_{\text{cm}} = 6.4$ cm). These results, obtained even under extreme viewpoints and varying zoom levels, confirm the sustained precision of the calculated homographies. This demonstrates that, when supplied with precise geometric primitives, our hypothesis framework is capable of reliably recovering the true image-to-model mapping with negligible drift for the vast majority of cases.

Regarding the cases where registration has failed, their corresponding results are illustrated in Fig. 9. Match 5 - Image 5 corresponds to a view where the system is unable to detect a sufficient number of reliable intersection points or identify distinctive geometric patterns, which prevents the correct localization of the corresponding field area. In Match 5 - Image 8, the error results from a strongly distorted perspective: the image is captured from a very low camera position, leading to extreme foreshortening.

8.4. End-to-end system performance

In this end-to-end evaluation, all the inputs required by the registration algorithm are obtained automatically through dedicated detection modules, rather than using ground-truth annotations. Specifically, the playing-field mask M_{PF} is generated using the fully-automatic segmentation strategy described in [34]. The set of white field-line segments W and the elliptical arcs E (penalty arcs and center circle) are detected following the procedures in [22], while the grass-band lines G are extracted using the method proposed in [30]. These automatically obtained primitives serve as the input for the registration process, ensuring that the evaluation reflects realistic operating conditions where feature extraction may introduce false positives, false negatives, or slight localization errors.

Table 2 summarizes the performance metrics (the same as in Table 1) for each image. These results reflect the behavior of the system when all elements—from primitive detection to hypothesis selection—are performed automatically.

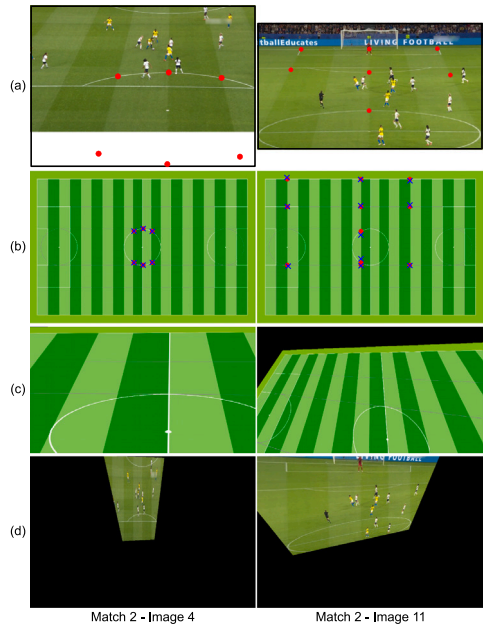


Fig. 10. Registration failures in the end-to-end system. (a) Original image, I, and the intersection points used. (b) Field of play model with the points considered (red dots) and their reprojections (blue crosses). (c) Result of projecting M_{PF} on the field of play model canvas. (d) Result of projecting the field of play model on the image canvas.

Although the results are largely consistent with those obtained using ground-truth annotations, some differences emerge. The number of points used to compute the homographies, N_{points}, is generally lower due to occasional false negatives in the automatic detection of lines and ellipses. Likewise, the reprojection error, E_{reproj}, tends to be slightly higher, as the localization of reference points is less precise than in the ground-truth setup. Consequently, the reprojection error (E_{pix} and E_{cm}) is also somewhat larger. Beyond the two failure cases already identified in the previous experiment, two additional failures are observed in Match 2 - Image 4 and Match 2 - Image 11. In these two cases (illustrated in Fig. 10), the perspective is perpendicular to the goal and makes several longitudinal grass bands visible, leading the system to confuse longitudinal and transversal lines. Moreover, no result is produced for Match 5 - Image 12, as the number of intersections between the automatically detected primitives is insufficient to compute a valid homography. Despite these three new cases, the proposed system achieves accurate and reliable registration in 90% of the dataset, with a mean reprojection error of $E_{\text{pix}} = 7.7$ pixels (corresponding to $E_{\text{cm}} = 30.8$ cm).

8.5. Comparison with state-of-the-art methods

To contextualize the performance of our proposal within the current state of the art, we conduct a comparative evaluation under fair and reproducible experimental conditions. In the context of soccer field registration, meaningful comparisons require that all methods be evaluated on a common dataset using consistent metrics, which significantly restricts the set of approaches that can be directly assessed. Many recent methods rely on proprietary datasets, private annotations, or lack publicly available implementations. For these reasons, we compare our method against the recent approach by Gutiérrez-Pérez and Agudo [8], which represents one of the most competitive geometric strategies to date and provides publicly available code, enabling a fair and reproducible evaluation on the LaSoDa dataset.

For the comparison, we evaluate the performance of their system under the four sets of pre-trained weights made available by the authors in their official repository: (i) the single-view (SV) configuration,

Table 2
Results corresponding to the performance evaluation of the complete end-to-end registration pipeline. Lines highlighted in red correspond to cases where the estimated homography was completely incorrect.

Match	Image id.	N_P	$Acc_{m\to M}$	$Acc_{M\to m}$	E_{reproj}	N_Q	E_{pix}	E_{cm}
1	1	34	99%	100%	17.3	773	20.8	83.0
	2	13	99%	100%	2.9	343	1.9	7.6
	3	19	99%	99%	2.1	1700	3.1	12.6
	4	17	100%	98%	3.5	342	3.3	13.0
	5	10	100%	100%	2.2	247	9.6	38.6
	6	36	99%	97%	11.5	655	13.9	55.7
	7	9	100%	97%	1.4	166	3.8	15.3
	8	25	100%	98%	5.2	217	6.2	24.8
	9	48	99%	99%	5.1	537	4.9	19.6
	10	42	100%	99%	9.0	521	10.0	40.1
	11	15	99%	97%	3.2	286	7.8	31.1
	12	20	100%	99%	9.0	294	8.3	33.1
2	1	16	99%	100%	3.7	526	6.9	27.8
	2	30	100%	97%	7.1	268	3.2	12.8
	3	23	100%	100%	4.1	322	3.6	14.2
	4	6	100%	100%	1.4	223	756.4	3025.7
	5	32	100%	96%	5.1	405	4.6	18.6
	6	37	100%	99%	6.9	424	4.7	18.6
	7	14	99%	100%	4.4	452	3.2	12.6
	8	40	100%	100%	8.0	459	2.6	10.6
	9	9	100%	94%	2.1	123	2.1	8.2
	10	6	100%	88%	1.7	100	5.5	22.0
	11	11	95%	98%	19.4	334	886.9	3547.7
	12	47	100%	100%	19.9	757	10.4	41.6
3	1	20	100%	98%	3.8	649	15.9	63.5
	2	42	100%	98%	5.3	580	5.8	23.3
	3	15	100%	97%	1.8	185	4.9	19.6
	4	25	100%	98%	7.1	191	4.6	18.3
	5	47	100%	99%	13.8	1013	10.4	41.5
	6	5	100%	100%	0.0	241	20.3	81.1
	7	27	100%	100%	5.5	288	2.3	9.1
	8	86	99%	100%	18.6	1234	20.8	83.4
	9	34	98%	100%	11.3	715	19.8	79.1
	10	18	100%	96%	2.1	322	5.4	21.7
	11	18	100%	100%	6.0	264	4.9	19.6
	12	39	99%	100%	18.5	719	6.7	26.8
4	1	34	99%	100%	13.7	729	8.0	32.1
	2	47	99%	100%	12.7	524	10.7	42.6
	3	20	100%	99%	3.2	301	6.2	24.9
	4	42	100%	99%	7.4	542	6.0	24.0
	5	20	100%	98%	5.5	312	9.7	38.7
	6	34	99%	99%	13.5	866	25.3	101.3
	7	16	99%	100%	3.7	588	5.5	22.2
	8	9	100%	100%	2.1	166	1.7	6.9
	9	16	100%	95%	3.4	694	8.4	33.5
	10	20	100%	100%	0.9	319	2.4	9.5
	11	42	100%	100%	5.6	552	5.5	22.1
	12	28	100%	100%	11.8	1369	8.8	35.0
5	1	24	100%	100%	6.4	588	7.1	28.2
	2	13	99%	100%	2.2	368	3.1	12.4
	3	23	100%	100%	5.0	346	3.6	14.4
	4	20	100%	100%	1.6	306	2.8	11.0
	5	4	100%	100%	0.0	57	480.1	1920.6
	6	6	100%	97%	1.1	122	6.9	27.4
	7	9	100%	97%	1.8	202	3.1	12.5
	8	10	100%	100%	19.4	739	2598.4	10 393.5
	9	27	100%	98%	5.8	323	7.7	30.7
	10	37	100%	99%	10.8	513	6.9	27.6
	11	57	98%	95%	12.0	791	21.5	85.9
	12	–	–	–	–	–	–	–

trained on the SoccerNet distribution; (ii) the multi-view (MV) configuration, also trained on SoccerNet but optimized for diverse viewpoints; and (iii–iv) the SV configurations fine-tuned on the WorldCup 2014 (WC14) and TS-WorldCup (TSWC) datasets, respectively. According to the authors, these fine-tuned models are designed to enhance performance on broadcast images by adapting the SV configuration to datasets that provide reliable ground-truth homographies.

Table 3 reports the reprojection errors E_{pix} obtained on the LaSoDa dataset for the four configurations of the method in [8], as well as for our proposed strategy. Complementarily, Fig. 11 illustrates representative qualitative examples of the registration results for this comparison.

The results in Table 3 show that the proposed strategy achieves the best performance for the majority of the images: in particular, our method outperforms all alternatives in 41 out of the 60 test images.

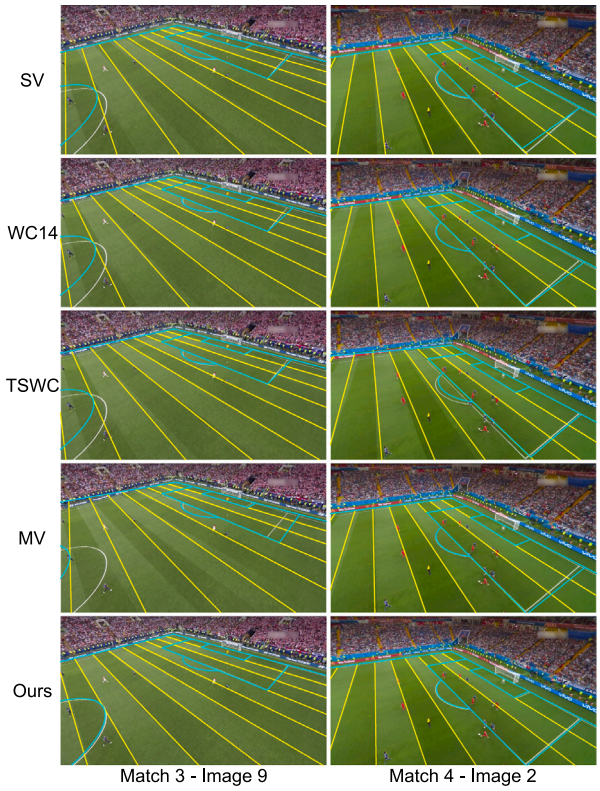


Fig. 11. Representative registration results on LaSoDa for the proposed strategy and the four configurations of [8] (SV, WC14, TSWC, and MV).

Moreover, it is also the most reliable approach, as it produces the smallest number of registration failures—i.e., cases where the field elements are mismatched and the corresponding reprojection errors remain relatively high (highlighted in red in Table 3). In addition, our strategy yields the fewest empty results (shown as dashes in the table), corresponding to cases where no valid homography could be estimated.

In addition, the analysis of the reprojection error distributions reveals that our strategy achieves the lowest median error across the dataset.[2] This indicates that our method not only succeeds in establishing correct correspondences between field elements in the images and those in the model, but also does so with greater geometric precision.

This improvement in accuracy can also be visually appreciated in the examples presented in Fig. 11, where our method consistently produces tighter alignments of the field markings and grass-band structures compared to the competing configurations.

Beyond the quantitative results reported above, it is also important to contextualize the proposed method under challenging broadcast conditions. The LaSoDa dataset used in this work includes a wide variety of demanding scenarios, such as highly oblique camera viewpoints, strong perspective distortions, large zoom variations, and frequent occlusions of field markings caused by players. Although the dataset consists of static images rather than full video sequences, these conditions are representative of real-world soccer broadcast imagery. In practice, we observed that the proposed method remains robust as long as a sufficient subset of field primitives (lines, arcs, or grass-band delimiters) is reliably detected, even when other elements are partially or fully occluded. Failure cases typically arise in situations where several key primitives are simultaneously missing or inaccurately detected, rather than from a systematic sensitivity to camera viewpoint or occlusion. To promote transparency and facilitate qualitative assessment, all registration results obtained on the LaSoDa images are made publicly available.

[2] The median is reported instead of the mean because it is less sensitive to outliers caused by homographies that are not merely imprecise but completely incorrect, in which case the reprojection errors can be arbitrarily large.

8.6. Computational cost analysis

This section analyzes the computational cost of the proposed registration pipeline, focusing exclusively on the stages introduced in this work. The detection of field primitives (lines, ellipses, grass-band delimiters, and playing-field masks) relies on existing methods and is therefore excluded from the timing analysis, as it does not constitute part of the proposed contribution.

All experiments were conducted on a desktop computer equipped with an Intel Core i7 CPU running at 1.10 GHz and 16 GB of RAM. The complete pipeline was implemented in MATLAB, using a straightforward and non-optimized implementation intended to favor clarity and reproducibility over execution speed. No parallelization, hardware acceleration, or low-level code optimization was employed. Consequently, the reported timings should be regarded as conservative estimates of the computational cost of the proposed strategy.

Tables 4 and 5 report the mean processing times and standard deviations for each stage of the pipeline under two different evaluation settings: (i) using ground-truth primitives as input, and (ii) using automatically detected primitives in a fully end-to-end configuration. In both cases, the reported times correspond exclusively to the data labeling, data combination and selection, hypothesis obtaining, and hypothesis evaluation stages.

When ground-truth primitives are used, the overall processing time is dominated by the data labeling stage, which accounts for approximately 70% of the total runtime. This behavior is expected, as ground-truth annotations typically include a larger number of primitives than those obtained through automatic detection, leading to a higher computational cost during the labeling and characterization of straight lines, ellipses, and intersection patterns. The remaining stages — data combination, hypothesis obtaining, and hypothesis evaluation — require comparatively less computation, resulting in a mean total processing time of 6.66 s per image.

In contrast, in the fully automatic end-to-end configuration, the computational burden shifts towards the hypothesis evaluation and selection stage, which represents 57% of the total runtime. This increase is mainly caused by the presence of false or imprecise detections, which may lead to a larger number of candidate combinations that need to be evaluated before identifying a valid homography. As a result, although the data labeling stage becomes less expensive due to the reduced number of detected primitives, the total mean processing time increases to 11.36 s per image. The larger standard deviation observed in this setting reflects the image-dependent nature of the number of hypotheses explored.

Overall, these results indicate that the computational cost of the proposed approach is primarily driven by the number and quality of the input primitives, rather than by intrinsically expensive geometric operations. In particular, the most time-consuming stages consist of evaluating independent homography hypotheses, which makes the pipeline highly amenable to parallel execution. A reimplementation in a compiled language such as C/C++, combined with multi-threading or GPU-based parallelism, is therefore expected to significantly reduce the execution time and enable practical deployment in time-constrained scenarios.

9. Conclusion

In this work, we have presented a novel strategy for automatic soccer field registration that integrates classical geometric cues with robust feature detection mechanisms. By jointly exploiting line and ellipse fitting, grass-band transitions, and projective invariants, the

Table 3
Reprojection errors (E_{pix}) on LaSoDa for the four configurations of [8] and for the proposed strategy. The best result for each image is highlighted in bold. Results highlighted in red correspond to cases where the estimated homography was completely incorrect.

Match	Image id.	SV	WC14	TSWC	MV	Ours
1	1	23.5	**8.0**	45.0	28.8	20.8
	2	12.6	9.5	12.6	10.5	**1.9**
	3	1807.7	1208.3	1784.5	1722.3	**3.1**
	4	507.3	64.8	978.9	39.8	**3.3**
	5	–	–	–	–	**9.6**
	6	16.7	18.6	25.8	**6.5**	13.9
	7	3.4	**3.3**	4.0	3.9	3.8
	8	11.6	8.8	9.9	**5.9**	6.2
	9	9.9	10.5	8.4	9.0	**4.9**
	10	923.2	978.9	14.3	13.6	**10.0**
	11	40 547.9	5220.6	5220.6	2243.0	**7.8**
	12	**7.1**	8.1	7.6	8.7	8.3
2	1	**6.0**	10.8	10.9	7.2	6.9
	2	8.2	3.3	4.8	4.5	**3.2**
	3	8.0	8.7	8.9	8.2	**3.6**
	4	–	–	–	–	756.4
	5	17.9	15.8	12.8	10.4	**4.6**
	6	1527.3	48.8	20.3	5.2	**4.7**
	7	5.0	3.3	8.8	9.1	**3.2**
	8	9.5	8.5	7.7	8.8	**2.6**
	9	7.0	838.2	9.5	8.1	**2.1**
	10	**4.0**	6.1	315.3	4.1	5.5
	11	–	–	–	**6.5**	886.9
	12	313.1	209.4	4297.5	161.2	**10.4**
3	1	–	–	–	1002.2	**15.9**
	2	11.6	11.0	17.9	10.3	**5.8**
	3	12.8	14.6	17.3	11.8	**4.9**
	4	4.5	**1.5**	5.2	4.0	4.6
	5	18.0	23.2	13.9	26.9	**10.4**
	6	–	–	–	–	**20.3**
	7	1284.2	1284.2	1284.2	3.9	**2.3**
	8	37.4	**18.8**	197.0	24.0	20.8
	9	36.2	46.5	51.7	85.7	**19.8**
	10	14.0	12.9	10.8	7.0	**5.4**
	11	33.0	2732.6	3885.8	**4.8**	4.9
	12	71.8	326.0	1140.8	16.9	**6.7**
4	1	3502.2	68.6	5980.6	48.3	**8.0**
	2	11.0	21.9	51.0	22.2	**10.7**
	3	13.9	7.5	32.5	12.4	**6.2**
	4	**5.6**	15.3	8.5	5.0	6.0
	5	26.8	31.0	20.3	27.3	**9.7**
	6	900.7	661.4	886.5	**19.5**	25.3
	7	11.1	8.0	23.5	12.8	**5.5**
	8	2444.4	2444.4	2444.4	10.4	**1.7**
	9	824.9	692.4	692.4	**6.9**	8.4
	10	45.0	31.8	13.0	5.8	**2.4**
	11	7.2	5.8	12.1	7.1	**5.5**
	12	1787.0	1365.7	3949.0	21.6	**8.8**
5	1	8.5	11.6	10.8	14.2	**7.1**
	2	4.8	3.1	7.5	11.9	**3.1**
	3	9.7	9.6	10.0	9.5	**3.6**
	4	8.0	9.6	705.9	4.2	**2.8**
	5	888.3	1180.6	1422.0	**3.5**	480.1
	6	144.7	103.2	90.6	114.4	**6.9**
	7	583.3	1839.0	12.0	4.8	**3.1**
	8	29.7	29.7	517.6	**19.9**	2598.4
	9	23.1	**5.5**	9.2	7.3	7.7
	10	17.3	24.4	9.8	8.0	**6.9**
	11	1708.8	54.8	40.4	56.1	**21.5**
	12	1424.8	1583.3	598.8	**5.1**	–
Overall (median)		17.9	18.8	20.3	9.5	6.2

proposed approach achieves a balance between the interpretability and viewpoint invariance of geometric methods, and the robustness of modern feature extraction pipelines.

Table 4
Mean processing time (in seconds) and standard deviation for each stage of the proposed pipeline when using ground-truth field primitives as input. Percentages indicate the relative contribution of each stage to the total processing time.

Stage	Mean	Std	%
Data labeling	4.65	1.99	70%
Data combination and selection	0.36	0.63	5%
Hypothesis obtention	0.46	1.08	7%
Hypothesis evaluation and selection	1.18	6.10	18%
Total	6.66	6.64	100%

Table 5
Mean processing time (in seconds) and standard deviation for each stage of the proposed pipeline in the fully automatic end-to-end configuration. Percentages indicate the relative contribution of each stage to the total processing time.

Stage	Mean	Std	%
Data labeling	3.76	1.68	33%
Data combination and selection	0.18	0.15	2%
Hypothesis obtention	0.93	3.88	8%
Hypothesis evaluation and selection	6.49	19.47	57%
Total	11.36	20.91	100%

The extensive evaluation carried out on the LaSoDa dataset — comprising 60 annotated images from matches played in five different stadiums and covering a wide range of camera viewpoints, zoom levels, and illumination conditions — demonstrates the effectiveness of the proposed strategy. Results show high registration accuracy when ground-truth inputs are available, and consistently strong performance under realistic end-to-end conditions, with successful and precise registrations achieved in 90% of the test images.

Furthermore, a comparative study with the recent state-of-the-art confirms the advantages of our approach. In particular, our strategy achieves the best results in the majority of test images (41 out of 60), produces the lowest median reprojection error, and exhibits fewer complete registration failures, thus highlighting both its robustness and accuracy across diverse scenarios.

Beyond the quantitative performance reported above, it is also important to clarify the current scope and limitations of the proposed approach. While the proposed method is designed around the standardized geometry of a soccer field, this assumption also defines its current scope. The approach relies on a predefined field model composed of straight lines, circular arcs, and grass-band delimiters, which correspond to the markings specified in the official Laws of the Game. As a result, fields with significantly different layouts or additional non-standard markings are not explicitly handled in the current formulation. Nevertheless, the proposed framework is inherently modular: extending it to more complex field designs or to scenarios with additional field features would mainly require incorporating new geometric primitives and updating the corresponding intersection patterns and model associations, without altering the core hypothesis generation and evaluation strategy. This flexibility makes the approach potentially applicable to other sports or enriched field representations, provided that an appropriate geometric model is available.

Despite the demonstrated performance and the flexibility of the proposed framework, a small number of failure cases remain, mainly under extreme zoom or forced perspectives where few usable field elements are visible. Addressing these situations will be the focus of future work, which may include incorporating temporal consistency across video sequences, learning-based refinement modules, or hybrid pipelines combining geometric reasoning with deep representations.

In summary, the proposed registration system provides a reliable, precise, and interpretable solution for soccer field alignment under real broadcast conditions, offering a solid foundation for downstream applications such as player tracking, tactical analysis, or augmented broadcast graphics.

CRediT authorship contribution statement

Carlos Cuevas: Writing – original draft, Software, Conceptualization. **Daniel Berjón:** Writing – review & editing, Software. **Narciso García:** Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI) in order to improve the writing style, polish the English text, and refine the structure of certain sections. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been partially supported by project PID2023-148922OA-I00 (EEVOCATIONS) funded by MCIU/AEI/10.13039/501100011033 of the Spanish Government, and by project TEC-2024/COM-322 (IDEALCV-CM) funded by Comunidad de Madrid.

Data availability

The complete set of results and supplementary materials is publicly available at https://www.gti.ssr.upm.es/data.

References

[1] C. Yang, M. Yang, H. Li, L. Jiang, X. Suo, L. Mao, W. Meng, Z. Li, A survey on soccer player detection and tracking with videos, Vis. Comput. 41 (2) (2025) 815–829.
[2] C. Cuevas, D. Quilón, N. García, Techniques and applications for soccer video analysis: A survey, Multimedia Tools Appl. 79 (39) (2020) 29685–29721.
[3] M. Manafifard, A review on camera calibration in soccer videos, Multimedia Tools Appl. 83 (6) (2024) 18427–18458.
[4] K. Qi, W. Xu, W. Chen, X. Tao, P. Chen, Multiple object tracking with segmentation and interactive multiple model, J. Vis. Commun. Image Represent. 99 (2024) 104064.
[5] C.S. Ranganathan, P. Pandey, M. Arulprakash, K. Gopalakrishnan, T. Ganesh-Babu, S. Murugan, Artificial neural networks for enhancing soccer team performance through tactical data analysis, in: IEEE International Conference on Machine Learning and Autonomous Systems, ICMLAS, 2025, pp. 642–647.
[6] M. Afzal, J.H. Shah, S. ur Rehman, F.A. Khokhar, M. Yasmin, S. Kadry, Automated soccer event detection and highlight generation for short and long views, Multimedia Tools Appl. 84 (26) (2025) 30971–30991.
[7] N. Jacquelin, R. Vuillemot, S. Duffner, Efficient one-shot sports field image registration with arbitrary keypoint segmentation, in: IEEE International Conference on Image Processing, ICIP, 2022, pp. 1771–1775.
[8] M. Gutiérrez-Pérez, A. Agudo, No bells just whistles: Sports field registration by leveraging geometric properties, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3325–3334.
[9] X. Tong, W. Li, T. Wang, Y. Zhang, Playfield registration in broadcast soccer video, Int. J. Multimed. Intell. Secur. 1 (2) (2010) 120–138.
[10] A. Linnemann, S. Gerke, S. Kriener, P. Ndjiki-Nya, Temporally consistent soccer field registration, in: IEEE International Conference on Image Processing, 2013, pp. 1316–1320.
[11] F. Wang, L. Sun, B. Yang, S. Yang, Fast arc detection algorithm for play field registration in soccer video mining, in: IEEE International Conference on Systems, Man and Cybernetics, Vol. 6, 2006, pp. 4932–4936.
[12] Y.-J. Chu, J.-W. Su, K.-W. Hsiao, C.-Y. Lien, S.-H. Fan, M.-C. Hu, R.-R. Lee, C.-Y. Yao, H.-K. Chu, Sports field registration via keypoints-aware label condition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3523–3530.
[13] N.S. Falaleev, R. Chen, Enhancing soccer camera calibration through keypoint exploitation, in: Proceedings of the 7th ACM International Workshop on Multimedia Content Analysis in Sports, 2024, pp. 65–73.
[14] P.J. Claasen, J.P. de Villiers, Video-based sequential bayesian homography estimation for soccer field registration, Expert Syst. Appl. 252 (2024) 124156.
[15] S. Zhang, Research on effective field lines detection and tracking algorithm in soccer videos, Int. Multimed. Ubiquitous Eng. 10 (7) (2015) 75–84.
[16] T. Rianthong, S. Thewsuwan, T. Charoenpong, K. Pattanaworapan, A method for detecting lines on soccer field by color of grass variation, in: IEEE 12th International Conference on Knowledge and Smart Technology, KST, 2020, pp. 131–134.
[17] Y. Wang, J. Zheng, Q.-Z. Xu, B. Li, H.-M. Hu, An improved ransac based on the scale variation homogeneity, J. Vis. Commun. Image Represent. 40 (2016) 751–764.
[18] A. Maglo, A. Orcesi, J. Denize, Q.C. Pham, Individual locating of soccer players from a single moving view, Sensors 23 (18) (2023) 7938.
[19] C. Cuevas, D. Quilon, N. García, Automatic soccer field of play registration, Pattern Recognit. 103 (2020) 107278.
[20] N. Homayounfar, S. Fidler, R. Urtasun, Sports field localization via deep structured models, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5212–5220.
[21] L. Citraro, P. Márquez-Neila, S. Savare, V. Jayaram, C. Dubout, F. Renaut, A. Hasfura, H. Ben Shitrit, P. Fua, Real-time camera pose estimation for sports fields, Mach. Vis. Appl. 31 (3) (2020) 16.
[22] D. Berjón, C. Cuevas, N. García, Soccer line mark segmentation and classification with stochastic watershed transform, Signal Process., Image Commun. 118 (2023) 117014.
[23] P. Li, J. Li, S. Zong, K. Zhang, Soccer field registration based on geometric constraint and deep learning method, in: Chinese Conference on Pattern Recognition and Computer Vision, PRCV, Springer, 2021, pp. 287–298.
[24] J. Theiner, R. Ewerth, Tvcalib: Camera calibration for sports field registration in soccer, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 1166–1175.
[25] F. Shi, P. Marchwica, J.C.G. Higuera, M. Jamieson, M. Javan, P. Siva, Self-supervised shape alignment for sports field registration, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 287–296.
[26] X. Nie, S. Chen, R. Hamid, A robust and efficient framework for sports-field registration, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1936–1944.
[27] J. Chen, J.J. Little, Sports camera calibration via synthetic data, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, IEEE Computer Society, 2019, pp. 2497–2504.
[28] W. Jiang, J.C.G. Higuera, B. Angles, W. Sun, M. Javan, K.M. Yi, Optimizing through learned errors for accurate sports field registration, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 201–210.
[29] F. Magera, T. Hoyoux, O. Barnich, M. Van Droogenbroeck, Broadtrack: Broadcast camera tracking for soccer, in: IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2025, pp. 6177–6187.
[30] C. Cuevas, D. Berjón, N. García, Grass band detection in soccer images for improved image registration, Signal Process., Image Commun. 109 (2022) 116837.
[31] The IFAB, Laws of the game, 2025, (accessed 08 September 2025). URL https://www.theifab.com/.
[32] A. Cioppa, A. Deliege, S. Giancola, B. Ghanem, M. Van Droogenbroeck, Scaling up soccernet with multi-view spatial localization and re-identification, Sci. Data 9 (1) (2022) 355.
[33] The IFAB, Stadium guidelines, 2025, (accessed 08 September 2025). URL https://inside.fifa.com/innovation/stadium-guidelines.
[34] C. Cuevas, D. Berjón, N. García, A fully automatic method for segmentation of soccer playing fields, Sci. Rep. 13 (1) (2023) 1464.