Comparative Analysis of Camera Calibration Algorithms for Football Applications

Oleksandr Sorokivskyi[1,*,†], Volodymyr Hotovych[1,†], Oleg Nazarevych[1,†] and Grygorii Shymchuk[1,†]

[1]*Department of Computer Science, Faculty of Computer Information Systems and Software Engineering, Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine*

Abstract

In solving the problem of automated analysis of football match video recordings, special video cameras are currently used. This work presents a comparative characterization of known algorithms and methods for video camera calibration, including those utilizing machine learning and neural networks, with the aim of identifying their shortcomings and forming a theoretical foundation for developing modern, more effective methods and algorithms. Specifically, it examines both algorithms that require more input data but operate quickly [1, 2] and more accurate algorithms using machine learning [3, 4, 5, 6, 5].

It is demonstrated that their main drawback is either accuracy or speed. More accurate algorithms using machine learning often do not specify the algorithm's operational speed, which precludes their use in real-time applications. The examined works that emphasize speed frequently lack the accuracy necessary for practical use in real-life scenarios.

Keywords

computer vision, camera calibration, homography estimation, football application, perspective projection, semantic segmentation, clustering, machine learning,

1. Introduction

Football match analysis uses statistical data, tactics, and player performance metrics to help coaches, scouts, and media professionals understand games better and make data-driven decisions. .In football analytics, determining players' positions on the field plays a crucial role. Based on such information, it is possible not only to analyze [7] but also to predict [8] the game outcome.

One of the most popular solutions is the use of location sensors attached to players' bodies. However, this solution is not always optimal. Body-attached sensors often cause discomfort to players, and moreover, this solution is costly, making it inaccessible for football clubs with limited budgets.

Currently, computer vision technologies are gaining increasing popularity for solving the player localization problem on the field, particularly through automatic analysis of match video recordings. The determination of players' positions on the field occurs in two stages: camera calibration and parameter determination, followed by player localization in the camera image.

This work presents a comparative analysis of known computer vision and machine learning algorithms for camera calibration and parameter determination.

ITTAP'2024: 4th International Workshop on Information Technologies: Theoretical and Applied Problems, October 23-25, 2024, Ternopil, Ukraine, Opole, Poland
*Corresponding author.

†These authors contributed equally.

$ sasha.sorokivski@gmail.com (O. Sorokivskyi); Gotovych@gmail.com (V. Hotovych); taltek.te@gmail.com (O. Nazarevych); gorych@gmail.com (G. Shymchuk)

0009-0006-6477-5878 (O. Sorokivskyi); 0000-0003-2143-6818 (V. Hotovych); 0000-0002-8883-6157 (O. Nazarevych); 0000-0003-2362-7386 (G. Shymchuk)

Copyright © 2024 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-3896/paper18.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

2. Related works

2.1. Camera Model

For the classic pinhole model, the basic formula of perspective projection is given by:

$$\lambda_m m = K[RT]M, \tag{1}$$

where:

M denotes a 3D point and m denotes the corresponding 2D point on image. They are both expressed in homogeneous coordinate and λ_m is an arbitrary scale factor.

R is 3 x 3 rotation matrix that describes the rotational mapping from the world coordinate system into the camera coordinate system.

T is a 3 x 1 vector that describes the translational mapping from the world coordinate system into the camera coordinate system.

K is a 3 x 3 matrix describing the internal camera parameters:

$$K = \begin{bmatrix} f & s & u_0 \\ 0 & \beta f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where:

Scale factor f applies to both the u and v axes of an image, while s describes the skew between these two axes.

Beta accounts for non-isotropic scaling, and the coordinates (u_o, v_o) denote the principal point.

When the observed 3D points lie on a plane, this projection can be simplified to a homography, which is a 3x3 matrix mapping between two planar surfaces. When considering perspective projection, an interesting phenomenon occurs with parallel lines in 3D space. If these lines are not parallel to the image plane, their 2D projections converge to a single point in the image. This point of convergence is termed the vanishing point. Notably, the line that connects this vanishing point to the optical center of the camera runs parallel to the corresponding 3D lines in space. Consequently, all sets of parallel lines in 3D space that share the same direction will correspond to the same vanishing point in the 2D image. This principle is fundamental to understanding how 3D scenes are projected onto 2D images in perspective projection.

2.2. Problem statement

Camera calibration involves determining its internal parameters (focal length, pixel ratio, projection center) and external parameters (rotation and translation expressing the camera's position and orientation relative to the world coordinate system).

Early approaches rely on matching local features in combination with direct linear transformation (DLT) to estimate homography. One of the first algorithms for determining these parameters is vanishingpoint based calibration (VPBC).

The study [9] presents a two-stage camera calibration method. In the first stage, the focal length and location of the principal point (intersection of the optical axis with the image plane) are determined using a single image of a calibration cube, an example of which is shown in Figure 1.

The second stage is dedicated to estimating the rotation matrix and translation vector between two cameras, using a stereo pair of images of a flat calibration pattern. This stage involves finding three corresponding vanishing points on both images, computing the rotation matrix based on these points, and estimating the translation vector through triangulation.

In [10] an improvement proposed to this method. Their approach is based on using only one image with two vanishing points, eliminating the need for a special calibration pattern. The method uses two lines to determine the vanishing points and information about the length of one of these lines

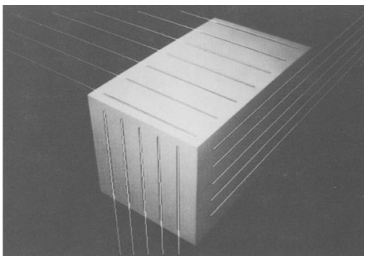

Figure 1: The aluminium block used to calibrate intinsic parameters of each camera

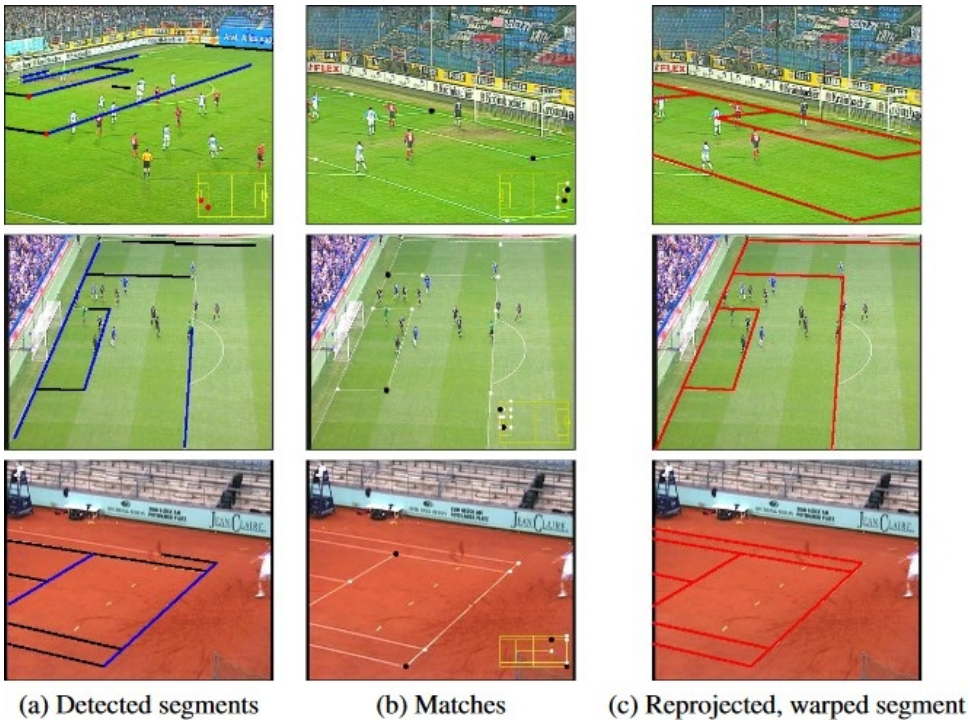

Figure 2: Results examples from [2]: each row corresponds to different video sequence.

to determine the transformation and subsequent calculations. This enhanced method simplifies the calibration process, making it more practical for various applications.

Both studies - [9] and [10] - made significant contributions to the development of camera calibration methods, improving the accuracy and convenience of this process. The first study laid the foundation for using vanishing points in camera calibration, while the second proposed a more efficient approach requiring less input data.

One of the first applications of the aforementioned algorithms in football is described in [2]. This method consists of two main stages.

The first stage involves detecting straight lines or their segments. For this purpose, Hough methods [11] or edge segmentation methods [12] are used. The detected lines are grouped into vertical and horizontal sets.

The second stage involves matching these two sets of image segments with segments in the football field model. Matching occurs by identifying segments that intersect with each other. Having two vanishing points, the algorithm selects segments that best correspond to the field model constructed using these vanishing points. After this, rotation (R) and translation (T) matrices are calculated to obtain the final camera model. An example of the algorithm's sequential operation on real images is shown in Figure 2.

However, all the aforementioned algorithms have limitations and work effectively only under certain

Table 1
Comparison table for static calibration methods

Method	Input data	Accuracy	Applicability
Using Vanishing Points for Camera Calibration [9]	Specific calibration pattern; Calibration Images; 2 Cameras required	Translation errors were about ±3 mm for distances ranging from 13 to 45 cm	Base method for further usage
Using Vanishing Points for Camera Calibration and Coarse 3D Reconstruction from A Single Image [10]	A single image containing at least two vanishing points; Two sets of parallel lines selected by the user to determine the vanishing points; The length of one line segment in 3D space (to determine the translation vector); The principal point is assumed to be the center of the image The aspect ratio is fixed by the user	Not provided	Base method for further usage
Fast 2D model-toimage registration using vanishing points for sports video analysis [2]	A sufficient number of segments of reasonable quality must be extractable from the images for the registration system to work.	Not provided	Applicable only for frames where sufficient number of segments are observable

conditions and with specific input data. In the modern world, this is insufficient for a fully functional analytics system, the foundation of which is determining the homography matrix in each frame of the video stream.

Further discussion will be devoted to methods that not only find key points but also continue calculating the homography matrix in subsequent frames. These methods allow for the creation of more reliable and flexible systems for analyzing football matches, capable of working in various conditions and with different types of input data. A comparison of the aforementioned algorithms in terms of their application in real conditions of modern football is presented in Table 1.

2.3. Dynamic camera calibration

The study [13] is one of the pioneering works in the field of not only determining the homography matrix for individual frames but also tracking its changes in a sequence of frames. The authors first proposed a combined approach for automatic computation of homography during camera motion. This approach includes using the KLT system [14, 15, 16] for automatic detection of correspondences between frames by extracting characteristic features. Since these correspondences are not perfect and contain outliers, the RANSAC algorithm [17] is applied to filter out incorrect matches. After this additional selection, the DLT algorithm is used to compute a new homography matrix for the current frame. An example of the algorithm's output is shown in Figure 3.

However, this method has a significant drawback: with each new frame, the reprojection error accumulates, leading to inaccuracies in the homography matrix. To address this issue, the authors

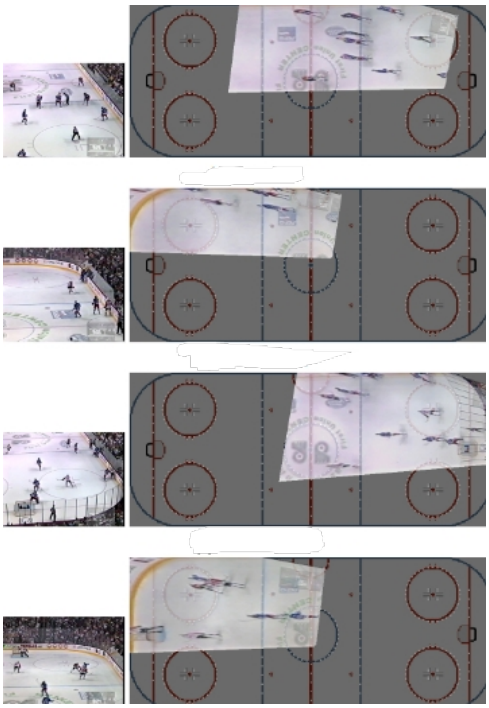

Figure 3: Results examples from [13]: each row corresponds to different frame

propose periodically adjusting the homography matrix using key points on the field lines.

It is important to note that the study does not specify the accuracy of the proposed algorithm. This approach laid the foundation for further research in the field of dynamic homography determination in video sequences.

The study [18] proposes one of the first approaches to determining not only the homography matrix but also the camera rotation angles. The authors developed a method that uses prior information about key points in the goal area to calculate these parameters with a fixed focal length. Experimental verification of the algorithm was conducted on a sample of 500 frames, and the researchers claim to have achieved reprojection accuracy within 2 pixels. It is important to note that the work focuses on theoretical aspects of determining rotation angles without considering the practical application of this method to frames that do not contain the goal area. A visual representation of the reprojection results obtained using this algorithm can be seen in Figure 4.

Despite using lines and camera parameters, [1] proposes using multiple key frames and lines along with ellipses to find the homography matrix. The process begins with system initialization, where key frames, examples of which are shown in Figure 5, are selected from the video sequence to cover the range of camera motion. Point correspondences between these key frames and the geometric model are manually selected to estimate homographies for all key frames. When processing new frames, the algorithm first identifies the closest key frame using local feature matching, applying SFOP key point detection [19] with SIFT descriptors [20]. This provides an initial estimate of the homography between the current frame and the geometric model. Feature finding is then performed by projecting the geometric model onto the current frame using the initial homography estimate. A model-driven approach is used to detect lines and ellipses in the frame, while point correspondences are obtained by back-projecting matches from the nearest key frame. The algorithm then proceeds to estimate homographies using two methods. First, it combines feature matches (lines, points, and ellipses) to obtain a linear estimate of the homography (H_{lin}). Second, it computes a frame-to-frame homography using local feature matches and combines this with the previous frame's homography to obtain an alternative estimate (H_{tr}). For refinement, the algorithm chooses between H_{lin} and H_{tr} based on the residual error area. The selected estimate serves as the initial value for further geometric minimization. The algorithm requires a lot of input data, which may be impossible or time-consuming to obtain.

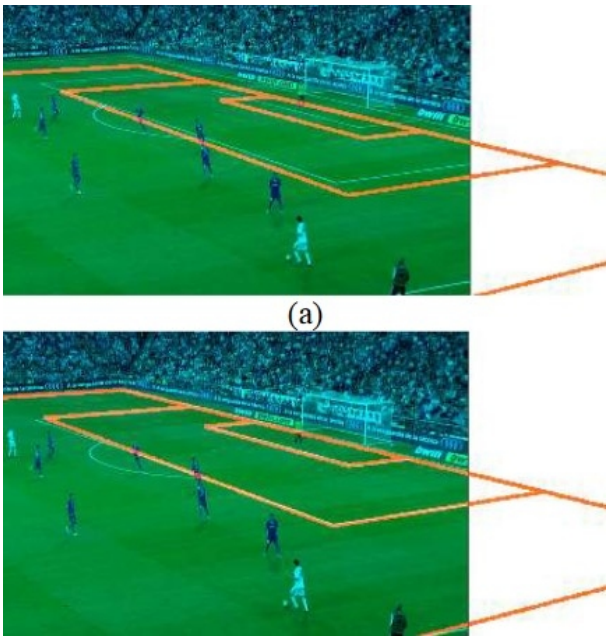

Figure 4: Results examples from [18].

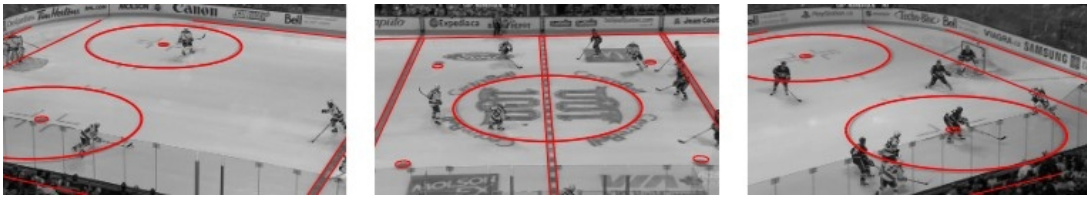

Figure 5: Key frames used in [1].

The accuracy of this approach is also not specified.

The study by Zhang et al. [21] describes a method for simplifying camera calibration and finding the transformation matrix by leveraging the specifics of a particular task. Traditionally, the DLT algorithm required four non-collinear key points to obtain the homography matrix. Instead, the authors propose the PCC (Pan-tilt camera calibration) algorithm, which takes into account the specifics of game filming where the camera remains stationary and only pan-tilt parameters change. This approach allows reducing the number of required key points for calibration to two. The PCC calibration process consists of two stages: first, initial camera calibration is performed using four points to determine fixed parameters, and then the Levenberg-Marquardt algorithm [22] is applied to find the homography matrix using only two key points. An important innovation in this work is the use of the offset line as a source of key points for determining the transformation matrix. The authors conducted a comparative analysis of the accuracy of both algorithms using computer simulation, which showed that the accuracy of the PCC algorithm surpasses that of the DLT algorithm.

A comparison of the aforementioned algorithms in terms of their application in real conditions of modern football is presented in Table 2.

Table 2
Comparison table for dynamic calibration methods

Method	Input data	Accuracy	Applicability

AUTOMATIC RECTI-FICATION OF LONG IMAGE SEQUENCES [13]	For the initial frame only, manually selected point correspondences between the image and the rink model	Not provided	Ice rink contains more unique elements and key points. Football field is larger and it will lead to an accuracy drop. Authors haven't provided accuracy metrics, thus it is hard to estimate applicability
Camera pose estimation in football scenes based on vanishing points [18]	Known key points for goal zone	Tested on simulated data with added noise. Reports 2 pixel accuracy for projections. Error in pan/tilt estimation correlates with roll error	Algorithm is based on goal zone detections that are not visible all the time in real match recording.
Using Line and Ellipse Features for Rectification of Broadcast Hockey Video [1]	A set of manually annotated key-frames with point correspondences to the geometric mode	Not provided	Ice rink contains more unique elements and key points. Football field is larger and it will lead to an accuracy drop. Also it requires manually annotated frames for every new camera.
Research on Camera Calibration in Football Broadcast Videos [21]	Details are not provided	Better than DLT algorithm, but overall accuracy is not provided	Base method for further usage

2.4. Machine learning based camera calibration

Machine learning is a powerful approach in the field of artificial intelligence that uses statistical methods to analyze large volumes of data. This technology allows algorithms to detect complex patterns and make accurate predictions, finding applications in many areas - from speech recognition to autonomous vehicle control.

In the context of camera calibration and finding the homography matrix, [3] proposes an innovative approach. They use a branch and bound method in a Markov random field, where the energy function is based on semantic features such as field surface, lines, and circles. These features are obtained through semantic segmentation - one of the tasks of machine learning. The process involves minimizing an energy function that takes into account that the field should predominantly consist of field surface pixels, and the projections of field primitives should correspond to detected primitives in the image. To optimize this function, the authors applied a Structured SVM algorithm trained on data from 9 unique stadiums. The accuracy of the algorithm, measured on 186 labeled images, reached an IOU score of 0.86.

Examples of the algorithm's results are shown in Figure 6.

An alternative approach using machine learning was proposed [4]. The authors developed their own camera simulator to create 75 labeled images that imitate field edges. These images and corresponding transformation matrices are stored in a separate database. When processing a real match image, the KNN algorithm searches for the most similar image in the database using one of three strategies: Chamfer matching, HOG, or CNN-based. To extract field edges from real images, the

stroke width transform (SWT) algorithm is applied, which demonstrates better noise resistance compared to traditional methods

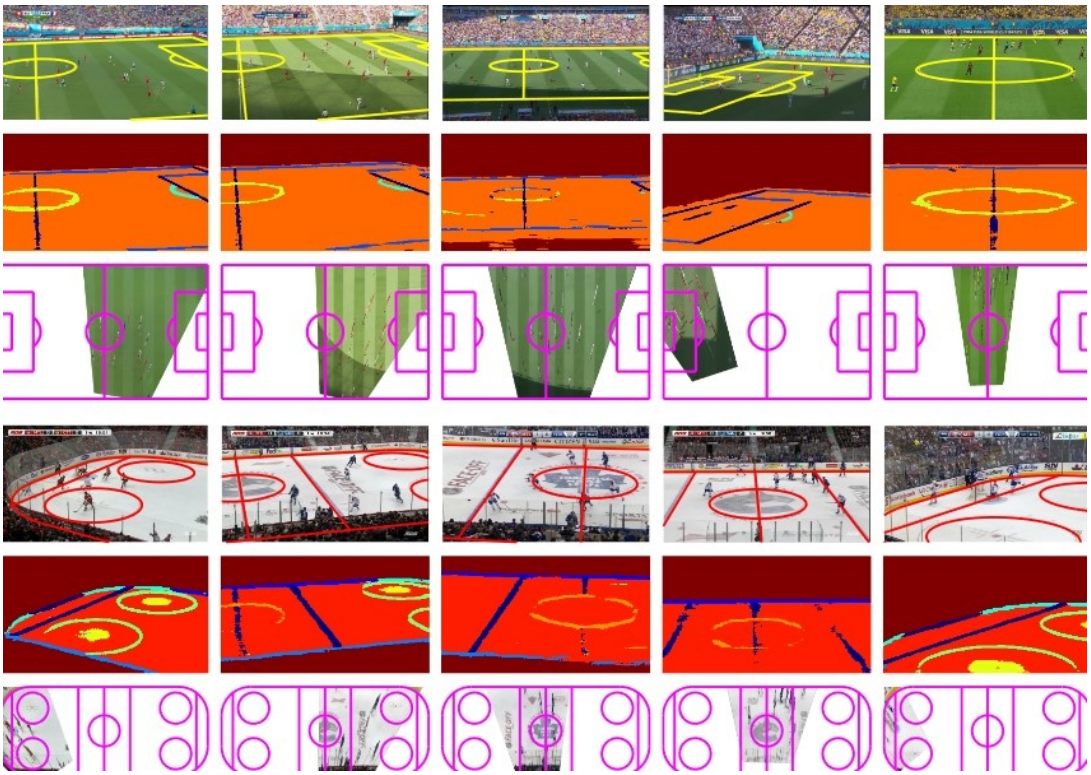

Figure 6: Examples of the obtained homographies and semantic segmentations in [3].

such as the Canny edge detector. Additionally, the authors remove the crowd from the image (using color-based field segmentation) and players (applying the Faster-RCNN human detector) to obtain an edge map that predominantly contains only field lines with minimal noise.

In [23], a method is proposed that uses two neural networks: the first determines the initial homography matrix, while the second estimates the registration error. The process involves transforming the sports field template to the current perspective, combining the transformed image with the current one, and iteratively updating the homography parameters to minimize the error. The authors evaluated the accuracy of their method on WorldCup and hockey match data [4], achieving an IOU score of 89.8,
which surpasses previous results.

A similar method is presented in [6], but with an important distinction. They first perform semantic segmentation of the field lines using DeepLabV3 ResNet [24], and then determine camera parameters through iterative optimization. This approach considers the reprojection error calculated from the found segments and their counterparts in the 2D image. The method was tested on the SoccerNetV3Calibration [25] and WorldCup datasets, achieving scores of 76.9 Compound Score and 96.1 IOU_{part}, respectively.

The researchers in [26] introduce a novel approach using evenly spaced keypoints as field-specific features, framing the task as an instance segmentation problem with dynamic filter learning. To validate their method, they created the TS-WorldCup dataset, which comprises 3,812 sequential images from 43 videos of the 2014 and 2018 FIFA World Cup tournaments, featuring precise field markings. The method employs a standard encoder-decoder architecture similar to U-Net, with a ResNet-34 backbone for the encoder. It introduces a keypoints-aware label condition, using 91 pre-defined keypoints and dynamically generated convolution kernels. The approach utilizes a keypoints-specific controller and dynamic head to predict keypoint heatmaps, which are then merged to estimate the final homography using DLT and RANSAC. The proposed method demonstrated strong performance

when evaluated on the WC and TSWC datasets, achieving IOU_{par} and IOU_{whole} scores of 0.96 and 0.91 for WorldCup, and 0.97 and 0.93 for TSWC, respectively.
In their study, [27] introduce speed metrics as a measure of performance in keypoint detection. The authors propose a dual-model approach, employing separate deep learning models for point and line detection, both utilizing heatmap-based techniques. For keypoint extraction, they adopt the widely-used HRNetV2-w48 model as their backbone architecture. The researchers report a processing time of 33.6 ms per image using a single Nvidia GeForce RTX 3090 GPU. While this performance approaches real-time capabilities, further optimization is necessary for widespread practical application.

In contrast, [5] do not explicitly report processing times for their approach. They also use the HRNetV2-w48 backbone, but employ a single network for both keypoint and line detection, potentially offering computational advantages. Their model is trained on a less powerful NVIDIA GeForce RTX 2080 Ti GPU, which may impact processing speed. While both papers use heatmap-based techniques, the "No Bells, Just Whistles" approach integrates an additional boundary channel to enhance global information capture. This difference in architecture and the use of a single network for multiple tasks may lead to different performance characteristics, though direct speed comparisons are not possible without explicit timing data from the second paper. The effectiveness of this algorithm was evaluated on the SN23, WC14, and TSWC datasets, with the most significant improvement (98.6) achieved on the TSWC dataset.

3.Comparative analysis

3.1. Calibration methods comparison

Static calibration methods paved the way for more accurate and powerful approaches. The methods of [9] and [10] are utilized in nearly every dynamic and machine learning-based algorithm. However, without additional refinement, they cannot be employed to determine the homography matrix for frame coordinate transformation. In contrast, [2] can be applied to certain frames containing a sufficient number of visible segments, with the accuracy of this approach primarily dependent on the quality and quantity of identified segments. In typical football match video recordings, conditions are often suboptimal, with insufficient visible lines for this algorithm to function effectively.

The importance lies not only in the accuracy and speed of algorithms but also in their field coverage. A football field does not always contain enough lines to determine key points based on their intersections. The existing field coverage issues of previous approaches are partially addressed in [13]. However, the authors do not specify the algorithm's accuracy, only its speed - 1900 frames per hour, or 2 frames per second on a 2.8 GHz Pentium IV processor. Modern cameras record at a minimum of 24 frames per second, which would result in a very long wait time to process a 90-minute football match.

Camera parameter determination in [18] does not resolve accuracy or speed issues of systems, focusing more on adding content to existing broadcasts or videos, which does not require high precision. Accuracy assessment is mentioned in [1] - visual evaluation by experts is conducted. However, neither accuracy nor speed metrics are provided, though the approaches in this work improve upon the results of [13]. Field coverage is also increased by using not only lines but also their combination with ellipses on the field, found in the central and goal areas.

However, improvements in accuracy and coverage depend on well-annotated key frames required for the algorithm's operation. This condition precludes application to real football match videos, as new annotated key frames would need to be created for each new camera or stadium.

With the advancement of machine learning, camera calibration methods have also evolved. In [3], the reprojection accuracy after homography determination is 0.88 *IOU*, significantly surpassing all previous approaches. The algorithm also does not focus on field parts where many key elements are visible but works on all field areas. The authors also indicate the speed of the homography matrix determination algorithm but do not specify the speed of the segmentation model. Typically, segmentation models are resource-intensive, precluding their use in real-time and significantly increasing the resources required for processing pre-recorded video.

Table 3
Comparison table for machine learning based calibration methods

Method	Input data	Accuracy	Applicability
Sports Field Localization via Deep Structured Models [3]	No input data required	Authors collected 186 images from 10 games as a test set. Based on that test set accuracy 0.88 *IOU* with the manually labeled data.	For offline usage
Automated Top View Registration of Broadcast Football Videos [4]	Database with edge images and corresponding homography matrices	Authors manually annotated 500 images from 16 different matches. *IOU* measure is 0.86 for the best approach.	For offline usage
Optimizing Through Learned Errors for Accurate Sports Field Registration [23]	Models weights	WorldCup *IOU*: 0.88 Hokey dataset *IOU*: 0.967	For offline usage
TVCalib: Camera Calibration for Sports Field Registration in Football [6]	Model weights	WorldCUP IOU_{par}: 0.96 SoccerNetV3 CR: 76.9	For offline usage
Sports Field Registration via Keypointsaware Label Condition [26]	Model weights	WorldCup IOU_{par}: 0.96 WorldCup IOU_{whole}: 0.91 TSWC IOU_{par}: 0.97 TSWC IOU_{whole}: 0.93	For offline usage
Enhancing Football Camera Calibration Through Keypoint Exploitation [27]	Model weights	Soccernet Camera Calibration Challenge 2023 [28] *Acc*@5: 0.73	For offline usage
No Bells, Just Whistles: Sports Field Registration by Leveraging Geometric Properties [5]	Models weights	WorldCup IOU_{par}: 0.96 WorldCup IOU_{whole}: 0.92 TSWC IOU_{par}: 0.98 TSWC IOU_{whole}: 0.96	For offline usage

In [4, 23, 6, 26], the focus is mainly on improving accuracy and increasing field coverage. Moreover, with the achieved maximum accuracy of full field homography IOU_{part} of 0.92, using the algorithm on offline video is quite feasible. Speed is mentioned only in [23], taking 9.58 seconds per frame.

Recent advancements in real-time keypoint detection algorithms have demonstrated significant progress in processing speed and efficiency. Falaleev et al. [27] and Gutierrez et al. [5] have reported state-of-the-art performance using the HRNet keypoint detection model, achieving a processing time of 33ms per frame on an Nvidia GeForce RTX 3090 GPU. These studies utilized the HRNetV2-w48 model, which belongs to the second-largest category within the HRNetV2 model family and comprises 67.1 million parameters. While the aforementioned research focused on larger models, it is worth noting that smaller variants within the HRNet family exist. These include the HRNet-W40-C with 57.6 million parameters and the most compact version, HRNet-W18-C-Small-v1, containing 13.2 million parameters.

However, the performance characteristics of these smaller models in real-time keypoint detection tasks remain unexplored in the current literature. Furthermore, the studies by Falaleev et al. [27] and Gutierrez et al. [5] did not investigate additional techniques for model size reduction or processing speed enhancement. This presents an opportunity for future research to explore optimization strategies that could potentially improve the efficiency and applicability of keypoint detection models across various computational resources and real-time scenarios.

3.2. Future directions of research

An analysis of current literature in the field of machine learning for camera calibration and homography matrix determination reveals a significant shortcoming in algorithm description and evaluation: a considerable portion of the presented methods lacks comprehensive information regarding quality metrics or performance speed. This limitation is particularly noticeable in machine learning algorithms,
where information about the algorithm's operational speed is often absent. Such a situation creates serious obstacles for objective assessment of algorithm efficiency and their comparison.

The lack of performance data significantly limits the application of these algorithms in real-time systems, where data processing speed is a critical factor. Many algorithms that demonstrate high accuracy on test datasets may prove unsuitable for practical use due to low operational speed.

To overcome these limitations, it is necessary to focus on developing algorithms in the direction of improving their performance. This includes optimizing existing algorithms, developing new approaches with an emphasis on computational efficiency, as well as applying parallel computing and specialized hardware. It is important for researchers and developers to pay more attention to comprehensive algorithm evaluation, including both quality and performance metrics, which will expand their scope of application and increase efficiency in real conditions.

It is worth noting that the authors of the aforementioned works do not consider a number of important optimization techniques for computer vision models. In particular, methods such as Pruning, Quantization, Knowledge Distillation, and Sparsity remain overlooked. These techniques have significant potential for substantially accelerating model performance while maintaining their effectiveness.

The use of these methods allows for the optimization of large and powerful models, reducing their computational requirements without significant loss of quality. For example, Pruning allows for the removal of the least important weights in a neural network, Quantization reduces the precision of parameter representation, Knowledge Distillation transfers knowledge from a large model to a smaller one, and Sparsity introduces sparseness into the network architecture.

The absence of consideration of these techniques in the analyzed works indicates a potential direction for further research and improvements in the field of computer vision model optimization.

Conclusions

In this work, studies related to camera calibration for determining the homography matrix for subsequent transformation of 2D coordinates into 3D coordinates were analyzed. The analysis encompassed both foundational works associated with general camera calibration algorithms and more contemporary developments utilizing machine learning. It was ascertained that works without using of machine learning do not specify algorithm accuracy. Machine learning approach demonstrate sufficient accuracy but lack the necessary speed for comfortable use on offline recordings, as well as for potential future real-time application during broadcasts. The study analyzed and presented the main shortcomings of these works, conducted a comparative analysis of solutions, and identified directions and ideas for future improvements in this field.

The research covered a range of approaches, from basic camera calibration methods to advanced machine learning techniques. It revealed that traditional methods often lack precise accuracy metrics,

while machine learning approach, though accurate, fall short in terms of processing speed. This speed limitation hinders their practical application in both offline video analysis and real-time broadcast scenarios.

A critical evaluation of existing methodologies highlighted their respective strengths and weaknesses. The comparative analysis provided insights into the effectiveness of various solutions, considering factors such as accuracy, field coverage, and computational efficiency. This comprehensive review served to pinpoint areas where current approaches fall short and where future research efforts should be concentrated.

Based on the findings, several directions for future research and development were identified. These include the need for improved algorithm speed optimization, especially for machine learning-based methods, without compromising accuracy. Additionally, the potential for incorporating advanced optimization techniques such as pruning, quantization, knowledge distillation, and sparsity in model architectures was emphasized as a promising avenue for enhancing both accuracy and computational efficiency.

The work underscores the importance of developing algorithms that not only achieve high accuracy but also demonstrate practical applicability in real-world scenarios, particularly in the context of sports analytics and broadcast technologies. By highlighting these areas for improvement, this analysis provides a valuable foundation for future research aimed at advancing the task of camera calibration and homography matrix determination.

References

[1] A. Gupta, J. J. Little, R. J. Woodham, Using line and ellipse features for rectification of broadcast hockey video, in: 2011 Canadian conference on computer and robot vision, IEEE, 2011, pp. 32–39.

[2] J.-B. Hayet, J. H. Piater, J. G. Verly, Fast 2d model-to-image registration using vanishing points for sports video analysis, in: IEEE International Conference on Image Processing 2005, volume 3, IEEE, 2005, pp. III–417.

[3] 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.

[4] R. A. Sharma, B. Bhat, V. Gandhi, C. Jawahar, Automated top view registration of broadcast football videos, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 305–313.

[5] 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.

[6] 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.

[7] K. Kim, M. Grundmann, A. Shamir, I. Matthews, J. K. Hodgins, I. Essa, Motion fields to predict play evolution in dynamic sport scenes, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2010) 840–847. URL: https://api.semanticscholar.org/CorpusID:8455859.

[8] F. Li, R. J. Woodham, Video analysis of hockey play in selected game situations, Image and Vision Computing 27 (2009) 45–58.

[9] B. Caprile, V. Torre, Using vanishing points for camera calibration, International journal of computer vision 4 (1990) 127–139.

[10] E. Guillou, D. Meneveaux, E. Maisel, K. Bouatouch, Using vanishing points for camera calibration and coarse 3d reconstruction from a single image, The Visual Computer 16 (2000) 396–410.

[11] D. Farin, S. Krabbe, W. Effelsberg, et al., Robust camera calibration for sport videos using court models, in: Storage and Retrieval Methods and Applications for Multimedia 2004, volume 5307, SPIE, 2003, pp. 80–91.

[12] J.-B. Hayet, J. Piater, J. Verly, Incremental rectification of sports fields in video streams with application to soccer, in: Advanced Concepts for Intelligent Vision Systems (ACIVS 2004), 2004.

[13] K. Okuma, J. J. Little, D. G. Lowe, Automatic rectification of long image sequences, in: Asian conference on computer vision, volume 9, 2004.

[14] S. T. Birchfield, Depth and motion discontinuities, stanford university, 1999.

[15] J. Shi, C. Tomasi, Good features to track, Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 600 (2000). doi:10.1109/CVPR.1994.323794.

[16] C. Tomasi, T. Kanade, Detection and tracking of point, Int J Comput Vis 9 (1991) 3.

[17] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.

[18] V. Babaee-Kashany, H. R. Pourreza, Camera pose estimation in soccer scenes based on vanishing points, in: 2010 IEEE International Symposium on Haptic Audio Visual Environments and Games, IEEE, 2010, pp. 1–6.

[19] W. Förstner, T. Dickscheid, F. Schindler, Detecting interpretable and accurate scale-invariant keypoints, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 2256–2263.

[20] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2004) 91–110.

[21] S. Zhang, Research on camera calibration in football broadcast videos, International Journal of u- and e-Service, Science and Technology 8 (2015) 89–98.

[22] Z. Niu, X. Gao, D. Tao, X. Li, Semantic video shot segmentation based on color ratio feature and svm, in: 2008 International Conference on Cyberworlds, IEEE, 2008, pp. 157–162.

[23] 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.

[24] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587 (2017).

[25] A. Cioppa, A. Deliege, S. Giancola, B. Ghanem, M. Van Droogenbroeck, Scaling up soccernet with multi-view spatial localization and re-identification, Scientific data 9 (2022) 355.

[26] 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.

[27] N. S. Falaleev, R. Chen, Enhancing soccer camera calibration through keypoint exploitation, arXiv preprint arXiv:2410.07401 (2024).

[28] A. Cioppa, S. Giancola, V. Somers, F. Magera, X. Zhou, H. Mkhallati, A. Deliège, J. Held, C. Hinojosa, A. M. Mansourian, et al., Soccernet 2023 challenges results, Sports Engineering 27 (2024) 24.