

Context-Aware 3D Object Localization from Single Calibrated Images: A Study of Basketballs

Marcello Davide Caio

Sportradar AG
m.caio.external@sportradar.com

Gabriel Van Zandycke

UCLouvain
gabriel.vanzandycke@uclouvain.be
Sportradar AG
g.vanzandycke@sportradar.com

Christophe De Vleeschouwer

UCLouvain
christophe.devleeschouwer@uclouvain.be



Figure 1: Can you guess the 3D location of the ball in the image? Localizing the ball in 3D from an image of intense gameplay demands insightful cues ranging from players' gaze, to the size and position of players and surrounding objects.

ABSTRACT

Accurately localizing objects in three dimensions (3D) is crucial for various computer vision applications, such as robotics, autonomous driving, and augmented reality. This task finds another important application in sports analytics and, in this work, we present a novel method for 3D basketball localization from a single calibrated image. Our approach predicts the object's height in pixels in image space by estimating its projection onto the ground plane within the image, leveraging the image itself and the object's location as inputs. The 3D coordinates of the ball are then reconstructed by exploiting the known projection matrix. Extensive experiments on the public DeepSport dataset, which provides ground truth annotations for 3D ball location alongside camera calibration information for each image, demonstrate the effectiveness of our method, offering substantial accuracy improvements compared to recent work. Our work opens up new possibilities for enhanced ball tracking and understanding, advancing computer vision in diverse domains. The source code of this work is made publicly available at <https://github.com/gabriel-vanzandycke/deepsport>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MMSports '23, October 29, 2023, Ottawa, ON, Canada

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0269-3/23/10...\$15.00

<https://doi.org/10.1145/3606038.3616173>

CCS CONCEPTS

• **Computing methodologies** → **Computer vision tasks**; *Scene understanding*; *Object detection*; *Interest point and salient region detections*; *Reconstruction*; *Matching*.

KEYWORDS

deep learning, computer vision, image understanding, sports, basketball, ball 3D localization, scene understanding, 3D

ACM Reference Format:

Marcello D. Caio, Gabriel Van Zandycke, & Christophe De Vleeschouwer. 2023. Context-Aware 3D Object Localization from Single Calibrated Images: A Study of Basketballs. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports (MMSports '23)*, October 29, 2023, Ottawa, ON, Canada. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3606038.3616173>

1 INTRODUCTION

Three-dimensional (3D) localization plays a crucial role in computer vision, enabling a wide range of applications across various domains. While 3D localization is relatively straightforward when certain assumptions can be made, such as objects being located on the ground plane, or when triangulation is possible, the task becomes significantly more challenging when objects are positioned off the ground and only a single image is available. Under these circumstances, this challenge extends to human perception, emphasizing the importance of developing robust computational methods for 3D localization.

In sports analytics, understanding the precise 3D localization of objects, such as balls [11] and players [19, 35], can offer valuable insights for player performance analysis [5, 26], referee assistance [1], and audience engagement. While existing commercial solutions in sports analytics often rely on multi-camera setups [4, 14, 16, 18, 20–23], these often come with logistical and financial constraints. In scenarios where access to premium multi-camera setups is limited, the ability to leverage single calibrated camera systems for accurate ball localization becomes invaluable. In this work, we focus on the open problem of estimating the 3D location of balls in basketball games from single calibrated images.

While 3D ball localization has been extensively studied for parabolic trajectories in calibrated videos [2, 3, 12, 15, 21, 24, 25, 27], little attention has been given to the general challenge of 3D ball localization from a single calibrated image, leaving a significant gap in research. Recent works addressing this task involved the estimation of the ball’s diameter in pixels and the reconstruction of its 3D location based on this estimation [17, 32]. However, this approach suffers from various limitations, including sensitivity to variations in the ball’s size within the image, issues with motion blur and image resolution, and the requirement for prior knowledge of the ball’s real-world size.

The observation that motivates this work is that in basketball, much like in other sports, human observers rely on contextual cues such as the position, occlusion, and size of nearby players, crowd, and court objects like basketball boards to accurately locate the ball during a game. If tasked with estimating the 3D position of the ball in a paused video, a person would use the position of the players’ feet on the court as a reference for X and Y coordinates, while making an informed guess about the height at which the ball is held. During passes or shots, the task becomes even more challenging and cues like players gaze orientation are required to improve the height perception. The idea is to give enough context to the model to guess the ball height in image space, and use geometry to retrieve the 3D ball coordinates.

This paper introduces a novel approach to 3D ball localization in basketball from single calibrated images. Building upon previous work [8, 31], we assume that the ball has already been detected in the image, allowing us to crop the image region around the ball. By providing sufficient context within the crop, our model predicts the height of the ball in image space, measured in pixels. Furthermore, by placing the ball at the center of the crop, we implicitly convey its location to the model.

Besides lifting the limitations of the previous method, this approach is more versatile and easily transferable to other sports such as soccer, volleyball, handball, and potentially extensible to domains beyond sports.

2 METHOD

We propose to estimate the 3D location of a ball by training a neural network model to predict the distance in pixels between the ball and its ground projection in image space. Retrieving the 3D ball coordinates is then easily achieved by leveraging the projection matrix from the camera calibration data.

2.1 Ball height estimation in pixels

Our model has a Convolutional Neural Network (CNN) backbone to extract image features, which are subsequently processed by a regression head supervised to predict the distance in pixels between the ball and its ground projection in the image space. The model operates on square image crops centered on the ball, ensuring sufficient contextual information is provided.

2.2 Pixels height to real-world coordinates

Let us define a 3D coordinates system (X, Y, Z) with the horizontal ground plane defined by $Z = 0$, for the real world, and a 2D coordinate system (x, y) , for image space. Details of the camera calibration allow us to project 3D points from the real world to the image space, and to convert 2D points in image space into light rays in 3D. To transform the estimated height of the ball in image space into its corresponding 3D coordinates in the real world, we apply the following procedure; see Fig. 2.

First, the input image is undistorted, ensuring that straight lines in the real world correspond to straight lines in image space. Then, the image is rotated, aligning the Z -axis in the real world with the vertical direction y in the image space. After rotation, the predicted height of the ball in pixels $\hat{h}$ effectively represents the vertical displacement between the ball’s position in the image $(\tilde{x}, \tilde{y})$, located at the center of the crop, and its projection onto the ground in image space $(\tilde{x}, \tilde{y} + \hat{h})$. As the camera is calibrated, we can project $(\tilde{x}, \tilde{y} + \hat{h})$ into 3D space on the $Z = 0$ ground plane, determining the position of the ball’s projection on the ground $(\tilde{X}, \tilde{Y}, 0)$. Finally, to obtain the 3D world coordinates of the ball, we perform a reprojection process. We trace the path of the light ray passing through the ball in the image space, determining the two 3D points at which it intersects the $Y = \tilde{Y}$ and $X = \tilde{X}$ vertical planes. By taking the average of these two points, we retrieve the 3D ball coordinates.

3 DATASETS

In this study, we aim to evaluate the effectiveness of our proposed method for ball 3D localization in basketball. To accomplish this, we require a suitable dataset that provides high-resolution images with accurate ball annotations. While datasets such as Basket-APIIDIS [34] and Soccer-ISSIA [7] have been previously utilized for related tasks, we exclude them from consideration as the concerns raised in reference [32] also apply to this work.

DeepSport dataset. Following Ref. [33], we utilize for training of our models the public DeepSport dataset [29, 30], introduced in Ref. [31] and expanded in Ref. [33], which offers 364 high-resolution panoramic images captured during professional basketball matches across 15 different arenas. The images in the dataset vary in resolution, typically reaching approximately 4500×1500 pixels. The ball’s 3D annotation is provided through the annotation of its center and vertical projection on the court within the image space.

Ballistic dataset. Beside evaluation of our model on the DeepSport dataset, in Section 4.6, we benchmark it also on the evaluation dataset introduced in Ref. [32] that features highly reliable 3D ball annotations from ballistic trajectories. This dataset is composed of 233 images of 2336×1756 pixels resolution, coming from 2 different basketball arenas.

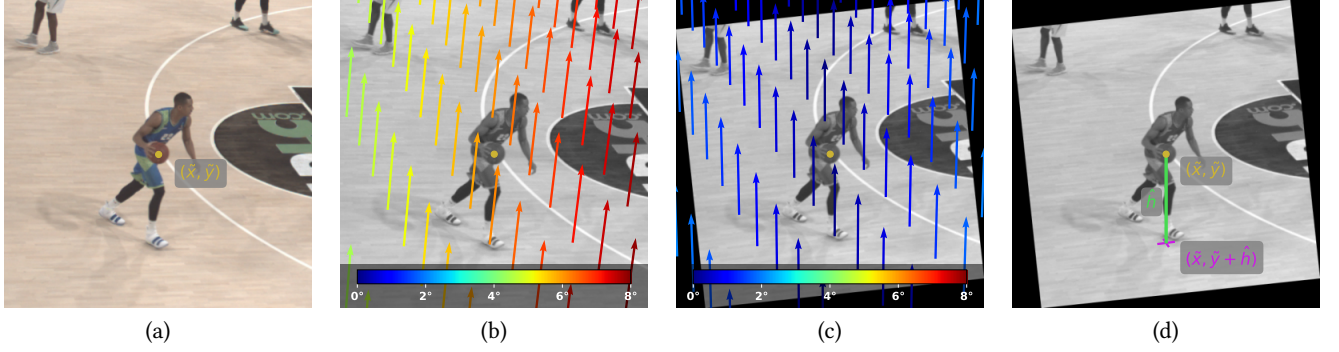


Figure 2: The predicted height and the given 2D position are converted to 3D coordinates using the camera calibration data. (a) Original image with ball position $(\tilde{x}, \tilde{y})$. (b), (c) Colored vectors perpendicular to the court floor show how verticals are oriented relative to image y -axis, with a scale given in degrees. (c) Image is rectified aligning the world Z -axis with the image y -axis. (d) Predicted height $\hat{h}$ is used to find the ball vertical projection on the ground floor $(\tilde{x}, \tilde{y} + \hat{h})$ which can be projected to $(\tilde{X}, \tilde{Y}, 0)$ using the calibration data. The Z coordinate is obtained by re-projecting $(\tilde{x}, \tilde{y})$ to the planes $X = \tilde{X}$ or $Y = \tilde{Y}$.

4 EXPERIMENTS

4.1 Experiment configurations

Implementation details and metrics used to conduct this research are presented hereafter.

4.1.1 Dataset. As mentioned in Section 3, we use the DeepSport dataset for training, and have adopted the split defined in [30], where each fold contains images from basketball arenas exclusive to that fold. The fold A remained unseen and was used for testing.

4.1.2 Baseline. As a baseline for our approach, we use a model introduced in Ref. [32] whereby the ball 3D position is obtained by estimating its diameter in the image space and combining knowledge of real ball size with calibration data.

4.1.3 Architecture. We study the impact of the feature extractor by experimenting with several popular architectures, including VGG, ResNet, MobileNet and EfficientNet; all pre-trained on ImageNet [6] from the TensorFlow library [28]. The regression head, borrowed from Ref. [32], contains 3 layers, and outputs a scalar supervised to predict the ball height in pixels.

4.1.4 Context size. We thoroughly investigate the impact of the square crop size on our model’s performance. It is evident that a small crop around the basketball would lack sufficient context for the model to accurately estimate its projection onto the ground. By enlarging the crop size and incorporating elements from the surrounding scene, we anticipate a significant enhancement in model performance. However, it is important to note that as the crop size increases, the improvements in performance may eventually plateau, leading to diminishing returns.

4.1.5 Image scale. In our research, we investigate the influence of the input image scale on our model’s performance, as it directly affects the representation of the basketball and all relevant objects and people within the scene. Moreover, the combined effect of image scale and context size also influences the inference speed of the model, a critical consideration for real-time analysis in production systems.

4.1.6 Training procedure. The backbone is pre-trained on ImageNet [6] from the TensorFlow library [28], and we initialize the regression head with the Glorot initializer [9]. The neural network is then trained on image patches centered around the oracle ball positions for 100 epochs using Adam optimizer [13] with a learning rate of 10^{-4} . The final height prediction is supervised with a Huber loss [10] with $\delta = 1.0$. Each experiment is repeated with 8 different initialization of the regression head and every reported metric is composed of mean and standard deviation computed over the 8 repetitions.

4.1.7 Metrics. We evaluate the performance of our 3D ball localization method using the following metrics:

- Mean Absolute Error (MAE [px]): This metric measures the average error in pixels for the predicted height of the ball in image space.
- Mean and Median Absolute Projection Error (MAPE [m] and MdnAPE [m]): These metrics quantify the average and median error in meters for the predicted projection of the ball on the court floor, with respect to the ground truth basketball projection.
- Mean and Median Absolute 3D Error (MA3DE [m] and MdnA3DE [m]): These metrics assess the average and median error in meters for the reconstructed 3D ball position, relative to the ground truth 3D ball position.

By utilizing these metrics, we can comprehensively assess the accuracy and performance of our proposed 3D ball localization method. As the metrics are computed on oracle ball positions in image space, they reveal the true errors from our proposed approach, decoupled from any error that a detector upstream could introduce.

4.2 Evaluation against the baseline

As a first experiment, we validate our proposed method for 3D ball localization against the model presented in Ref. [32], which relied on ball size estimation. For our model, we select a reasonable set of hyperparameters: the crop size is fixed to 512 pixels (about 1/3rd of the average dataset image height), the backbone is VGG16,



Figure 3: Various crop sizes for a given image in the dataset. Larger crop sizes offer more contextual information to the model.

and the image resolution is the one of the dataset. These choice of hyperparameters will be taken as default in the following sections. To ensure a fair comparison, we retrained the baseline model using the same hyperparameters as described in the previous paper and ran the experiments under similar conditions.

Table 1: Comparison of performance metrics between the proposed method and baseline model [32].

Metrics	Our Method	Baseline
MAE [px]	34 ± 3	N/A
MAPE [m]	1.25 ± 0.11	2.88 ± 0.22
MdnAPE [m]	0.92 ± 0.15	2.10 ± 0.38
MA3DE [m]	1.29 ± 0.10	2.97 ± 0.23
MdnA3DE [m]	0.95 ± 0.16	2.18 ± 0.40

The results in Table 1 demonstrate the superiority of our proposed method over the baseline model. The significant improvement in MAPE and MA3DE validates the effectiveness of our approach for accurate 3D ball localization, setting a solid foundation for further experimentation and analysis.

The notable discrepancy between median and mean values observed across all experiments, including both our proposed method and the baseline, indicates that current solutions for ball 3D localization are significantly affected by outliers. Consequently, the performance of these models in production may be better than what is suggested by mean value of metrics.

4.3 Impact of backbone

We considered several popular backbone architectures including VGG, ResNet, MobileNet and EfficientNet. Our findings show that they all have similar performance when trained on the DeepSport

dataset. We anticipate that when exposed to a larger dataset, larger models would perform better, at the cost of longer inference time.

4.4 Impact of context size

Here, we investigate the influence of context size on the predictive power of our model for 3D ball localization; see Fig. 3. Smaller crop sizes provide faster inference and require fewer resources, but may lack sufficient context for accurate height estimation. Conversely, larger crop sizes offer more contextual information, potentially improving model performance. Our goal is to identify the ideal crop size for a production environment where both performance and speed are important. We explore a range of crop sizes: 64, 96, 128, 256, 320, 480, 512, 640, and 800 pixels. For each crop size, we report in Table 2 the resulting MAPE.

Table 2: Impact of context size on MAPE. The crop size is varied whilst keeping the image scale fixed.

Crop Size (pixels)	MAPE [m]	MA3DE [m]
64	1.71 ± 0.17	1.75 ± 0.17
96	1.52 ± 0.13	1.56 ± 0.13
128	1.46 ± 0.06	1.50 ± 0.06
256	1.04 ± 0.11	1.07 ± 0.12
320	1.17 ± 0.09	1.20 ± 0.10
480	1.18 ± 0.07	1.21 ± 0.07
512	1.25 ± 0.11	1.22 ± 0.11
640	1.16 ± 0.07	1.20 ± 0.08
800	1.08 ± 0.06	1.11 ± 0.06

As expected, a small crop size such as 64×64 yields a higher MAPE, indicating limited context for accurate height estimation. As the crop size increases, MAPE decreases, indicating improved

performance. However, MAPE already plateaus beyond a crop size of 256 pixels. We thus identify the crop size of 256 pixels as the optimal choice, striking a balance between model performance and efficient inference, with promising implications for real-time applications.

4.5 Impact of image scale

In this section, we present a comprehensive study to assess the robustness of our model with respect to input image scale. The dataset comprises panoramic images of the basketball court. To evaluate the impact of image scale, effectively we shrink the images at various ratios (1, 1/2, 1/4, and 1/8). The square crop used as input to the model is also scaled accordingly, such that the model has access to the same contextual information independently of the scale.

Table 3: Performance of the model with respect to input image size.

Scale Ratio	MAPE [m]
1	1.25 ± 0.11
1/2	1.33 ± 0.11
1/4	1.32 ± 0.08
1/8	1.22 ± 0.07

From the results in Table 3, we observe that our proposed method exhibits remarkable robustness concerning image resolution. Despite shrinking the images up to 1/8th of the original size, the model’s performance remains consistent. This robustness stands in stark contrast to the approach presented in Ref. [32], where high-resolution images were a prerequisite for achieving accurate ball localization.

4.6 Benchmark on ballistic dataset

We evaluate our method on a benchmark dataset of balls moving on ballistic trajectories, as introduced in Ref. [32]. The obtained Mean Absolute Projection Error (MAPE) on this dataset is 2.68 ± 0.12 m, significantly higher than the results obtained on the DeepSport test set; see Table 1. The poor performance on the ballistic dataset is promptly explained by observing the distinct distribution of ball heights in the two datasets; see Fig. 4. The ballistic dataset shows a skew towards higher ball heights, with nearly half of the samples (102/233) having balls above 3 m, while the DeepSport dataset primarily comprises balls below this mark, with only 60 of 801 training samples reaching 3 meters.

To mitigate this limitation, we attempted to balance the number of balls above and below 2 meters in the training set. This approach resulted in an improved performance on the ballistic dataset, reducing the MAPE to 2.21 ± 0.27 m. Nonetheless, the limited number of high-ball samples in the training dataset remains a problem: feeding the same few high-ball samples to the neural network many times during training can not help much the model to generalise to those situations. We anticipate that a larger dataset, featuring more samples with high-ball instances, could narrow the gap between the performance on the ballistic dataset and the DeepSport test set.

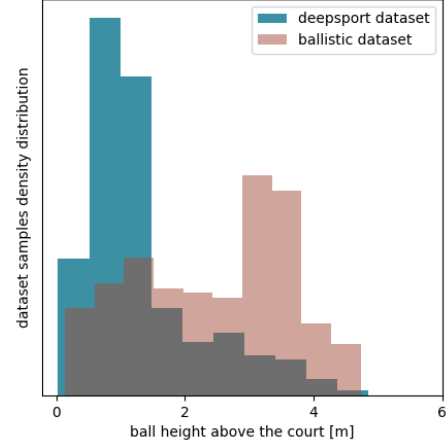


Figure 4: Ball height distribution in meters for the DeepSport dataset (training and evaluation sets combined), and the ballistic dataset introduced in Ref. [32].

5 DISCUSSION

In this paper, we introduced a novel method for 3D ball localization, achieving impressive results with MAPE as low as 1.1 meters. Notably, our proposed approach exhibits remarkable robustness with respect to image resolution, allowing the use of low-resolution images with the same crop size, thereby providing more context while maintaining good performance and inference speed.

Although our method demonstrates significant improvements in various scenarios, it underperforms on the ballistic dataset [32]. However, this outcome is not surprising given its skewed distribution of ball heights, which significantly differs from the one on which our models were trained.

While our method is based on single images, there is potential for even better performance when applied to videos, by means of temporal smoothing or tracking. However, this goes beyond the scopes of this paper and would require an ad-hoc dataset, at least for evaluation.

Lastly, the performance of our approach could be further optimized with a more extensive and diverse training dataset. With additional data, our approach has the potential to achieve even better results, and the use of larger backbones may yield higher accuracy.

6 CONCLUSION

In conclusion, we have presented a novel method for 3D ball localization that offers a significant improvement over existing approaches in terms of accuracy and robustness. Importantly, this approach can be readily extended to other sports, given the creation of similar datasets tailored to each sport.

Moreover, our method’s underlying concept of height detection in the image space and subsequent 3D position reconstruction holds promise for general 3D object localization. While each application may adapt the idea to suit specific tasks, the fundamental framework remains applicable.

ACKNOWLEDGMENTS

Part of this work has been funded by the Walloon Region project DeepSport and by Sportradar AG. C. De Vleeschouwer is a Research Director of the Fonds de la Recherche Scientifique - FNRS. Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region. M. D. Caio and G. Van Zandycke thank Sportradar AG for enabling and sponsoring this work, and for its commitment to scientific research, which made it possible for us to make significant contributions to the sports technology field. We also thank Davide Zambrano and the rest of the Computer Vision and Deep Learning team at Sportradar AG for their overall support of this work and for the fruitful conversations that led to the findings presented in this paper.

REFERENCES

- [1] Baljinder Bal and Gaurav Dureja. 2012. Hawk Eye: A Logical Innovative Technology Use in Sports for Effective Decision Making. *Sport Science Review* 21, 1-2 (may 2012), 107–119. <https://doi.org/10.2478/v10237-012-0006-6>
- [2] Hua Tsung Chen, Ming Chun Tien, Yi Wen Chen, Wen Jiin Tsai, and Suh Yin Lee. 2009. Physics-based ball tracking and 3D trajectory reconstruction with applications to shooting location estimation in basketball video. *Journal of Visual Communication and Image Representation* 20, 3 (2009), 204–216. <https://doi.org/10.1016/j.jvcir.2008.11.008>
- [3] Hua Tsung Chen, Wen Jiin Tsai, Suh Yin Lee, and Jen Yu Yu. 2012. Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimedia Tools and Applications* 60, 3 (2012), 641–667. <https://doi.org/10.1007/S11042-011-0833-Y/TABLES/9>
- [4] Xina Cheng, Norikazu Ikoma, Masaaki Honda, and Takeshi Ikenaga. 2018. Simultaneous physical and conceptual ball state estimation in volleyball game analysis. *2017 IEEE Visual Communications and Image Processing, VCIP 2017* 2018-Janua (2018), 1–4. <https://doi.org/10.1109/VCIP.2017.8305067>
- [5] Varuna De Silva, Mike Caine, James Skinner, Safak Dogan, Ahmet Kondoz, Tilson Peter, Elliott Axtell, Matt Birnie, and Ben Smith. 2018. Player Tracking Data Analytics as a Tool for Physical Performance Management in Football: A Case Study from Chelsea Football Club Academy. *Sports* 6, 4 (2018). <https://doi.org/10.3390/sports6040130>
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2010. ImageNet: A large-scale hierarchical image database. *Institute of Electrical and Electronics Engineers (IEEE)*, 248–255. <https://doi.org/10.1109/cvpr.2009.5206848>
- [7] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, and P. L. Mazzeo. 2009. A semi-automatic system for ground truth generation of soccer video sequences. In *6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009*, 559–564. <https://doi.org/10.1109/AVSS.2009.69>
- [8] Seyed Abolfazl Ghasemzadeh, Gabriel Van Zandycke, Maxime Istasse, Niels Sayez, Amirafshar Moshtaghpour, and Christophe De Vleeschouwer. 2021. DeepSportLab: a Unified Framework for Ball Detection, Player Instance Segmentation and Pose Estimation in Team Sports Scenes. In *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*. BMVA Press, 379. <https://www.bmvc2021-virtualconference.com/assets/papers/0735.pdf>
- [9] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Journal of Machine Learning Research*, Vol. 9, 249–256. <http://www.iro.umontreal>
- [10] Peter J. Huber. 1964. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* (mar 1964). <https://doi.org/10.1214/aoms/1177703732>
- [11] Paresh R. Kamble, Avinash G. Keskar, and Kishor M. Bhurchandi. 2019. Ball tracking in sports: a survey. *Artificial Intelligence Review* 52, 3 (2019), 1655–1705. <https://doi.org/10.1007/s10462-017-9582-2>
- [12] Taeone Kim, Yongduck Seo, and Ki Sang Hong. 1998. Physics-based 3D position analysis of a soccer ball from monocular image sequences. *Proceedings of the IEEE International Conference on Computer Vision* (1998), 721–726. <https://doi.org/10.1109/ICCV.1998.710797>
- [13] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- [14] K. C. Amit Kumar, Pascaline Parisot, and Christophe De Vleeschouwer. 2011. Demo: Spatio-temporal template matching for ball detection. In *2011 5th ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2011*. <https://doi.org/10.1109/ICDSC.2011.6042940>
- [15] Mikel Labayen, Igor G. Olaizola, Naiara Aginako, and Julian Florez. 2014. Accurate ball trajectory tracking and 3D visualization for computer-assisted sports broadcast. *Multimedia Tools and Applications* 73, 3 (dec 2014), 1819–1842. <https://doi.org/10.1007/s11042-013-1558-x>
- [16] Christoph H. Lampert and Jan Peters. 2012. Real-time detection of colored objects in multiple camera streams with off-the-shelf hardware components. *Journal of Real-Time Image Processing* 7 (2012). <https://doi.org/10.1007/s11554-010-0168-3>
- [17] Adrien Maglo, Astrid Orcesi, and Quoc-Cuong Pham. 2022. A ResNet-18 Network to Estimate the Ball Diameter in Basketball Images. <https://github.com/DeepSportradar/2022-winners-ball-3d-localization-challenge/blob/main/doc/BasketBallSizeEstimation.pdf>
- [18] Andrii Maksai, Xinchao Wang, and Pascal Fua. 2016. What players do with the ball: A physically constrained interaction modeling. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-Decem. 972–981. <https://doi.org/10.1109/CVPR.2016.111>
- [19] Lucas Antônio Monezi, Anderson Calderani Junior, Luciano Allegretti Mercadante, Leonardo Tomazeli Duarte, and Milton S. Misuta. 2020. A Video-Based Framework for Automatic 3D Localization of Multiple Basketball Players: A Combinatorial Optimization Approach. *Frontiers in Bioengineering and Biotechnology* 8 (2020). <https://doi.org/10.3389/fbioe.2020.00286>
- [20] Pascaline Parisot and Christophe De Vleeschouwer. 2011. Graph-based filtering of ballistic trajectory. In *Proceedings - IEEE International Conference on Multimedia and Expo*. <https://doi.org/10.1109/ICME.2011.6012126>
- [21] Pascaline Parisot and Christophe De Vleeschouwer. 2019. Consensus-based trajectory estimation for ball detection in calibrated cameras systems. *Journal of Real-Time Image Processing* (2019). <https://doi.org/10.1007/s11554-016-0638-3>
- [22] Gopal Pingali, Agata Opalach, and Yves Jean. 2000. Ball tracking and virtual replays for innovative tennis broadcasts. *Proceedings-International Conf. on Pattern Recognition* 15, 4 (2000). <https://doi.org/10.1109/ICPR.2000.902885>
- [23] Jinchang Ren, James Orwell, Graeme A. Jones, and Ming Xu. 2008. Real-time modeling of 3-D soccer ball trajectories from multiple fixed cameras. *IEEE Transactions on Circuits and Systems for Video Technology* 18, 3 (2008), 350–362. <https://doi.org/10.1109/TCSVT.2008.918276>
- [24] Hubert Shum and Taku Komura. 2004. A spatiotemporal approach to extract the 3D trajectory of the baseball from a single view video sequence. Technical Report. 1583–1586 pages. <https://doi.org/10.1109/icme.2004.1394551>
- [25] Hugo Silva, André Dias, José Almeida, Alfredo Martins, and Eduardo Silva. 2011. Real-Time 3D Ball Trajectory Estimation for RoboCup Middle Size League Using a Single Camera. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7416 LNCS (2011), 586–597. https://doi.org/10.1007/978-3-642-32060-6_50
- [26] Brian Skinner and Stephen J. Guy. 2015. A Method for Using Player Tracking Data in Basketball to Learn Player Skills and Predict Team Performance. *PLOS ONE* 10, 9 (09 2015), 1–15. <https://doi.org/10.1371/journal.pone.0136393>
- [27] Jonas Sköld. 2015. Estimating 3D-trajectories from Monocular Video Sequences Estimating 3D-trajectories from Monocular Video Sequences. (2015).
- [28] TensorFlow. 2022. (feb 2022). <https://doi.org/10.5281/ZENODO.5949169>
- [29] Gabriel Van Zandycke. 2020. Image and Signal Processing Group (UCL) | Softwares: <https://sites.uclouvain.be/ispgroup/Softwares/DeepSport>.
- [30] Gabriel Van Zandycke. 2021. DeepSport dataset: <https://www.kaggle.com/gabrielvanzandycke/deepsport-dataset>.
- [31] Gabriel Van Zandycke and Christophe De Vleeschouwer. 2019. Real-time CNN-based segmentation architecture for ball detection in a single view setup. *MMSports 2019 - Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports, co-located with MM 2019* (2019), 51–58. <https://doi.org/10.1145/3347318.3355517>
- [32] Gabriel Van Zandycke and Christophe De Vleeschouwer. 2022. 3D Ball Localization From A Single Calibrated Image. , 3471–3479 pages. <https://doi.org/10.1109/CVPRW56347.2022.00391>
- [33] Gabriel Van Zandycke, Vladimir Somers, Maxime Istasse, Carlo Del Don, and Davide Zambrano. 2022. DeepSportradar-v1: Computer Vision Dataset for Sports Understanding with High Quality Annotations. In *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports (Lisboa, Portugal) (MMSports '22)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3552437.3555699>
- [34] Christophe De Vleeschouwer, Fan Chen, and Damien Delannay. 2008. Distributed video acquisition and annotation for sport-event summarization. *NEM summit* (2008). [http://hdl.handle.net/2078.1/90154http://apidsis.org/Dataset/all/metadata/NEM\[_\]Summit\[_\]APIDIS08\[_\]final.pdf](http://hdl.handle.net/2078.1/90154http://apidsis.org/Dataset/all/metadata/NEM[_]Summit[_]APIDIS08[_]final.pdf)
- [35] Yukun Yang, Min Xu, Wanneng Wu, Ruiheng Zhang, and Yu Peng. 2018. 3D Multiview Basketball Players Detection and Localization Based on Probabilistic Occupancy. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*. 1–8. <https://doi.org/10.1109/DICTA.2018.8615798>