WACV WAIKOLOA HAWAII JAN 3-7 • 2023

Leibniz Universität Hannover

TVCalib: Camera Calibration for Sports Field Registration in Soccer

Jonas Theiner[1] Ralph Ewerth[1,2]

[1] L3S Research Center, Leibniz University Hannover, Hannover, Germany
[2] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

theiner@l3s.de
ralph.ewerth@tib.eu

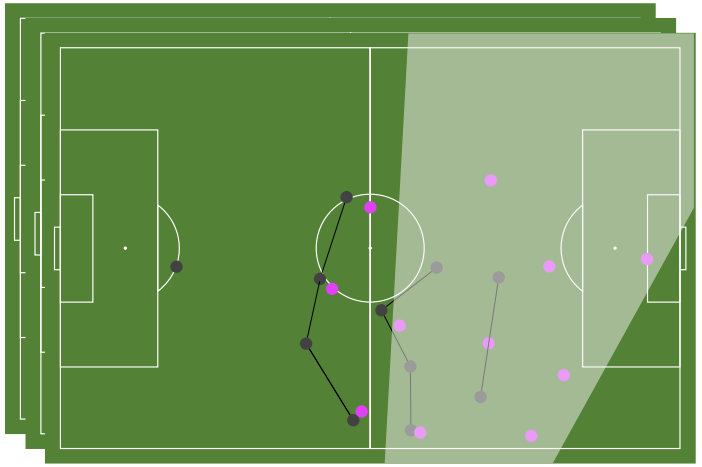
mm4spa.github.io/tvcalib

Forschungszentrum · Research Center L3S

Funded by
Federal Ministry of Education and Research

MM SPA
Multimodal Analysis for Sports Analytics

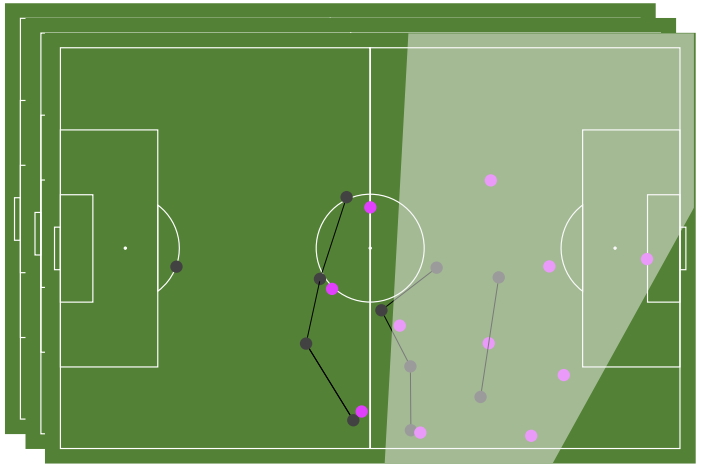
Sports Field Registration in Broadcast Videos



Sports Field Registration

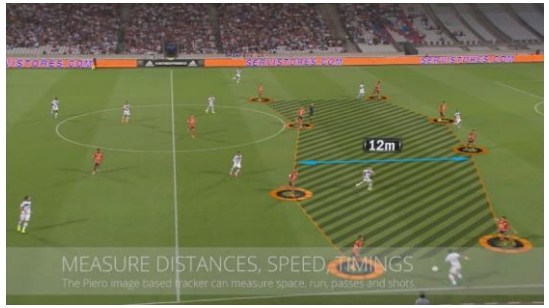
- Usually interpreted as homography estimation $\mathbf{H}$
  - Plane-to-plane mapping
  - Broadcast image to bird‘s eye view and vice versa

Sports Field Registration in Broadcast Videos



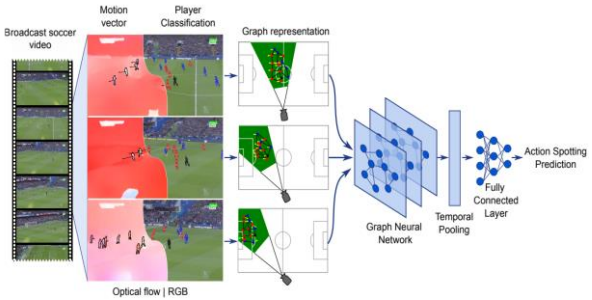
Sports Field Registration

Augmented reality



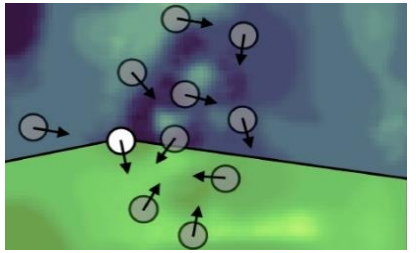
[Fischer et al. 2019]

Event detection



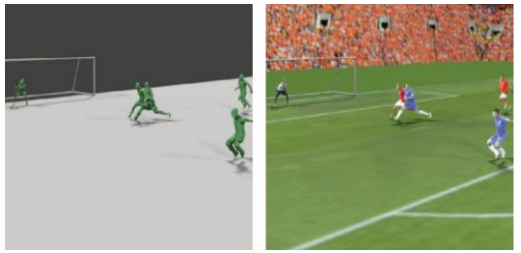
[Cartas et al. MMSports'22]

Generation & enrichment of player position data



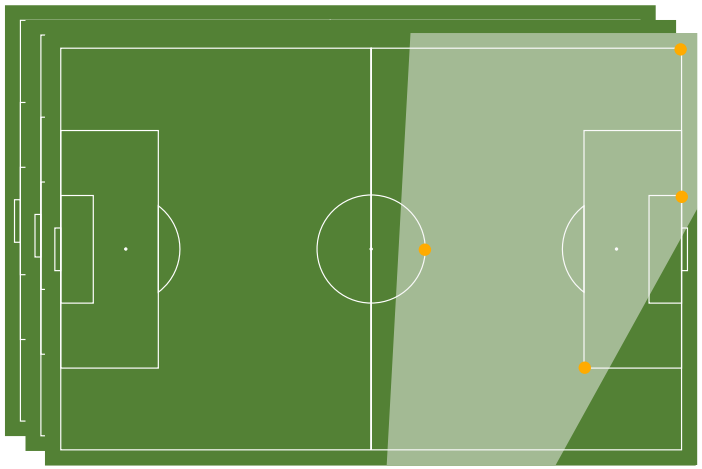
[Sangüesa et al. CVPRW'20]

3D scene reconstruction



[Ramatas et al. CVPR'18]

Sports Field Registration in Broadcast Videos



$\mathbf{H}$

Sports Field Registration

$\mathbf{H^{-1}}$

- Vanilla approach: Direct Linear Transform (DLT) from point correspondences
  - Requires accurate (point) correspondences
  - Easy-to-detect keypoints can be out of view [Chu et al. CVPRW'22]

Related Work

Semantic segmentation

- Keypoint prediction
  [Chu et al. CVPRW'22] [Shi et al. WACV'22] [Nie et al. WACV'21]
- Line segmentation [Chen & Little CVPRW'19]
- Area masking [Chen & Little CVPRW'19] [Sha et al. CVPR'20]

Initial estimation

- DLT [Nie et al. WACV'21] [Citraro et al. Machine Vision and Applications'20]
- Regression of $\mathbf{H}$ [Chu et al. WACV'22] [Jiang et al. WACV'20]
- Nearest neighbor retrieval of known camera poses

  [Shi et al. WACV'22] [Sha et al. CVPR'2] [Nie et al. WACV'21]

Focus

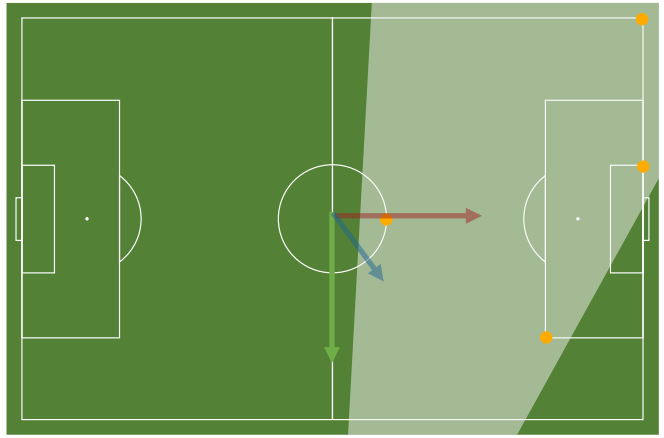
- Homography estimation
- Two-step estimation: $\mathrm{H} = H_{init} H_{rel}$

Refinement as relative image transformation

- L1 reprojection [Shi et al. CVPRW'22]
- Lucas-Kanade algorithm [Chen & Little CVPRW'19]
- Spatial Transformer Networks

  [Sha et al. CVPR'20] [Chen & Little CVPRW'19]

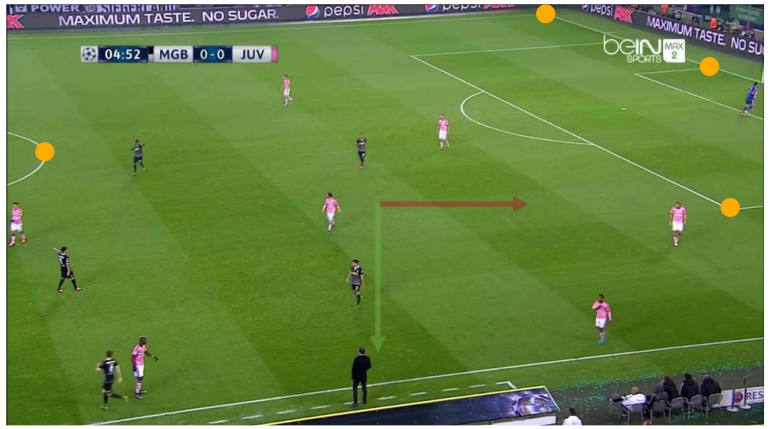
Why do we tackle camera calibration?



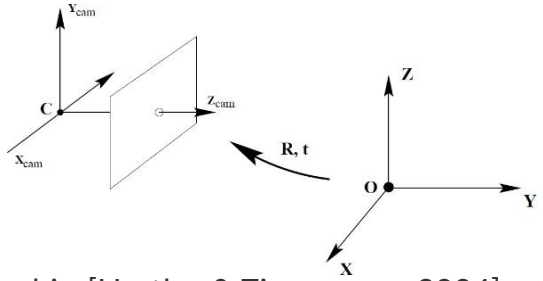
$$\mathbf{H}[X_i, Y_i, 1]^T$$

$$\mathbf{P}[X_i, Y_i, Z_i, 1]^T$$

$$\mathbf{H}^{-1}\,[x_i, y_i, 1]^T$$



Pinhole camera model



Graphic: [Hartley & Zissermann, 2004]

$$\mathbf{P}^{3x4} = \mathrm{KR}[\mathrm{I}|-\mathrm{t}] = \overbrace{\begin{bmatrix} flx & shear & u_0 \\ 0 & fly & v_0 \\ 0 & 0 & 1 \end{bmatrix}}^{\text{intrinsics}} \overbrace{R \begin{bmatrix} 1 & & -X_c \\ & 1 & -Y_c \\ & & 1\ -Z_c \end{bmatrix}}^{\text{extrinsics}}$$

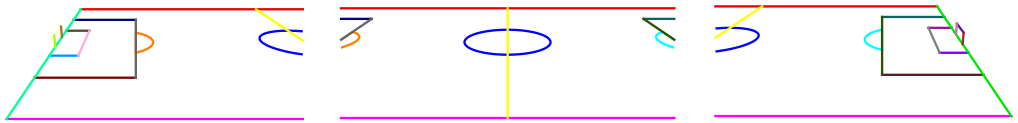
$$\mathrm{R}^{3x3} = \mathrm{R}_z(roll)\mathrm{R}_X(tilt)\,\mathrm{R}_z(pan)$$

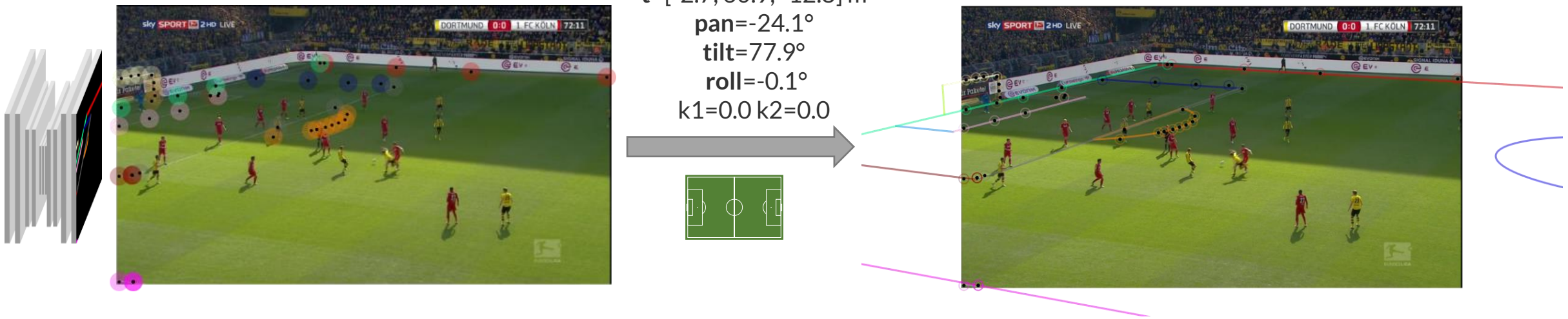
For planar settings ($\mathrm{Z} = 0$): $\mathbf{H}^{3x3} = \mathrm{KR}^{3\mathrm{x}[1,2]}[\mathrm{I}|-\mathrm{t}] = \mathbf{P}^{3\mathrm{x}[1,2,4]}$

TVCalib – Main Idea

- **Estimation of underlying camera parameters** (extrinsics, intrinsics, potential lens distortion)
- Sports field segments as calibration pattern (points on lines, points on circles)

- Iteratively minimize the *segment reprojection loss*
  - Point-line and point-circle distances
  - Gradient-based solver
  - Initialize from multiple views





Experiments

Benchmark Datasets

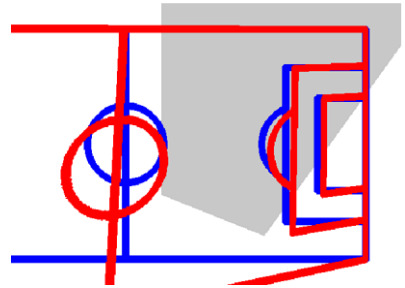
- *SoccerNet-Calibration* [Giancola et al. MMSports'22]*
- *FIFA WorldCup2014* [Homayounfar et al. CVPR'17]

Metrics

- Unavailability of ground-truth camera parameters
- (Image) **reprojection error** for individual segments: ACC@t [px] = TP / (FP+FN+TP)

[Giancola et al. MMSports'22]*

- **Projection error** (reconstruction): IoU_{part}
  - Binary IoU for the visible part of the projected field
  - Estimated and manually annotated homography matrix



[Citraro et al. Machine Vision and Applications'20]

* https://github.com/SoccerNet/sn-calibration

Evaluating the Homography on *WorldCup2014* Dataset

		(image) reprojection accuracy			projection accuracy	
Calibration	Segmentation	ACC@5	ACC@10	ACC@20	IoU_{part} (mean)	IoU_{part} (median)
H		54.1	82.9	92.4	100.0	100.0
TVCalib	GT	62.7	84.9	95.5	96.1	97.1
[Chen & Little CVPR'19]	GT	61.2	82.4	90.6	95.2	97.3
TVCalib	Pred	38.8	69.1	89.4	95.3	96.6
[Chen & Little CVPR'19]	Chen & Little	35.8	66.3	84.4	94.6	96.3
[Jiang et al. WACV'20]	Jiang et al.	36.9	62.9	81.5	95.2	97.1
[Shi et al. WACV'22]	Shi et al.				96.6	97.8
[Chu et al. CVPRW'22]	Chu et al.				96.0	97.0

- Quality of annotated matrices **H** introduces bias [Homayounfar et al. CVPR'17]
- Superior results compared to reimplemented approaches
- Projection error: Similar results compared to state-of-the-art approaches

Conclusion

Limitations

- Local minima due to gradient-based optimization
- Optimization of an image reprojection error only

Future Work

- Usage of temporal information
- End-to-end learning approach

project website & source code

theiner@l3s.de
ralph.ewerth@tib.eu

mm4spa.github.io/tvcalib

Forschungszentrum · Research Center L3S

Funded by
Federal Ministry of Education and Research

MM SPA
Multimodal Analysis for Sports Analytics