TVCalib: Camera Calibration for Sports Field Registration in Soccer

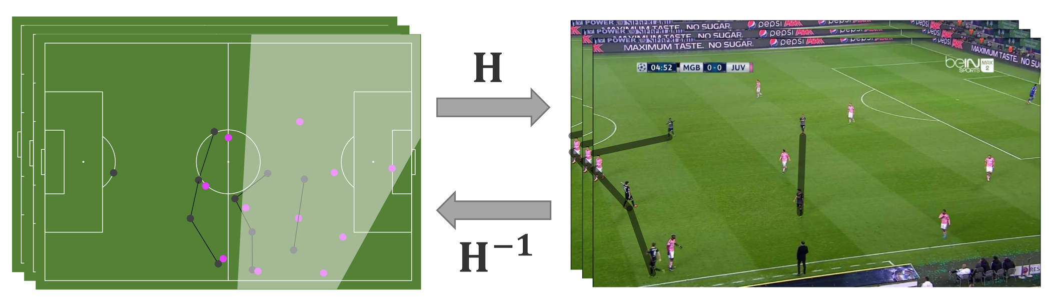
Jonas Theiner[1] Ralph Ewerth[1,2]

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

Motivation - Sports Field Registration in Broadcast Videos

Usually interpreted as the task of homography estimation

- Plane-to-plane mapping
- Broadcast image to bird's eye view and vice versa



Applications

- Augmented reality [Stein et al. TransVizComputGraph'18]
- 3D scene reconstruction [Zhu et al. ECCV'20, Rematas et al. CVPR'21]
- Temporal event detection [Cartas et al. MMSports'22]
- Generation and enrichment of player position data [Theiner et al. WACV'22, Arbues-Sanguesa et al. CVPRW'20]

Related work

Semantic segmentation via CNNs
Keypoint prediction, line segmentation, or area masking

Vanilla approach
Direct Linear Transform (DLT) from point correspondences
Drawbacks:

- Easy-to-detect keypoints can be out of view [Chu et al. CVPRW'22]
- Requires accurate (point) correspondences

Two-step homography estimation: $\mathbf{H} = \mathrm{H}_{\mathrm{init}}\mathrm{H}_{\mathrm{rel}}$

- **Initial estimation**
  - DLT, regression
  - Nearest neighbor retrieval of known camera poses
- **Refinement** as relative image transformation
  - Minimization of the L1 reprojection error
  - Spatial Transformer Networks

Why do we tackle the camera calibration?

- Estimation of actual underlying camera parameters
- 3D sports field registration

Pinhole camera model

The projection matrix **P** projects arbitrary 3D scene coordinates to the 2D image according to:

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

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

Relation to the homography

For planar settings, the homography matrix **H** maps arbitrary 2D scene coordinates to the 2D image and vice versa. We can substitute **P** to **H**, if we set $Z = 0$:

$$\mathbf{H}^{3x3} = \mathrm{KR}^{3x[1,2]}[\mathrm{I}|-\mathrm{t}]$$

Approach for Keypointless Calibration

Estimation of individual camera parameters

- Camera pose (location and orientation)
- Intrinsics (focal length)
- Radial lens distortion coefficients

Contributions

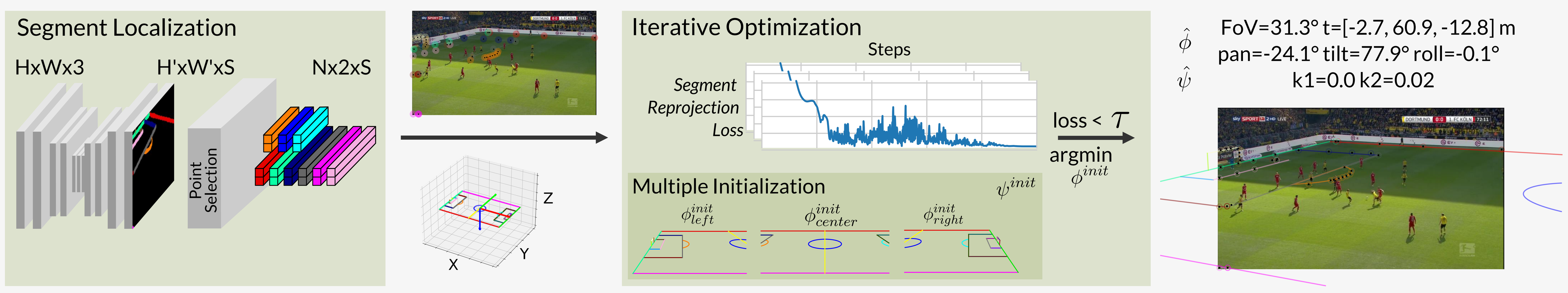
- Differentiable ***segment reprojection loss*** function
- Novel pipeline for sports field registration

Sports field segments as calibration pattern

- Points on individual lines
- Points on individual circles

Iteratively minimize the *segment reprojection loss*

- Point-line and point circle-distances at image space
- Gradient-based solver



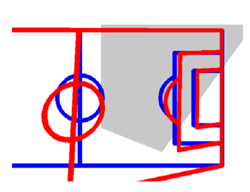
Experiments

Evaluation

- Issue: Absence of ground truth camera parameters
- Image annotations are the only ground truth we know

Image reprojection error via ACC@t [Giancola et al. MMSports'22]
from annotated points at several thresholds in pixel.

Projection error via $\mathrm{IoU}_{\mathrm{part}}$
calculates the binary IoU of the visible part of the projected area (bird's eye view) of an estimated and a manually annotated homography matrix.



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

Ablations studies

- Homography estimation vs. calibration
- Impact of segment localization
- Multiple camera location initialization
- Self-verification
- Lens distortion correction

Qualitative results on *SoccerNet-Calibration* dataset

[Giancola et al. MMSports'22]

Evaluating the homography on *WorldCup2014* dataset

[Homayounfar et al. CVPR'17]

Calibration	Segmentation	ACC@5	ACC@10	ACC@20	$\mathrm{IoU}_{\mathrm{part}}$ (mean)	$\mathrm{IoU}_{\mathrm{part}}$ (median)
H		54.1	82.9	92.4	100.0	100.0
TVCalib	Ground Truth	62.7	84.9	95.5	96.1	97.1
Chen & Little CVPR'19	Ground Truth	61.2	82.4	90.6	95.2	97.3
TVCalib	Predicted	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

Conclusions

Reprojection error

Superior results compared to reimplemented approaches

- On two benchmark datasets
- Both when using ground truth and custom segmentation

Projection error

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

Limitations

- Local minima due to gradient-based optimization
- Importance of a **good camera parameter initialization**

References

- Chu et al. '*Sports Field Registration via Keypoints-Aware Label Condition*', CVPRW'22
- Giancola et al. '*SoccerNet 2022 Challenge Results*', MMSports@MM'22 (https://github.com/SoccerNet/sn-calibration)
- Homayounfar et al. '*Sports Field Localization via Deep Structured Models*', CVPR'17
- Chen & Little '*Sports Camera Calibration via Synthetic Data*', CVPRW'19
- Cartas et al. '*A Graph-Based Method for Soccer Action Spotting Using Unsupervised Player Classification*', MMSports@MM'22
- Theiner et al. '*Extraction of Positional Player Data from Broadcast Soccer Videos*', WACV'22
- Sangüesa et al. '*Using Player's Body-Orientation to Model Pass Feasibility in Soccer*', CVPRW'20
- Rematas et al. '*Soccer on Your Tabletop*', CVPR'18
- Zhu et al. '*Reconstructiong NBA Players*', ECCV'20
- Stein et al. '*Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis*', Trans. Vis. Comput. Graph.'18

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

Source code

mm4spa.github.io/tvcalib

Forschungszentrum · L3S · Research Center

Funded by
Federal Ministry of Education and Research

MM SPA Multimodal Analysis for Sports Analytics