Extraction of Positional Player Data From Broadcast Soccer Videos

Jonas Theiner, Wolfgang Gritz, Eric Müller-Budack, Robert Rein, Daniel Memmert, and Ralph Ewerth

5 min summary

@thejtheiner
theiner@l3s.de

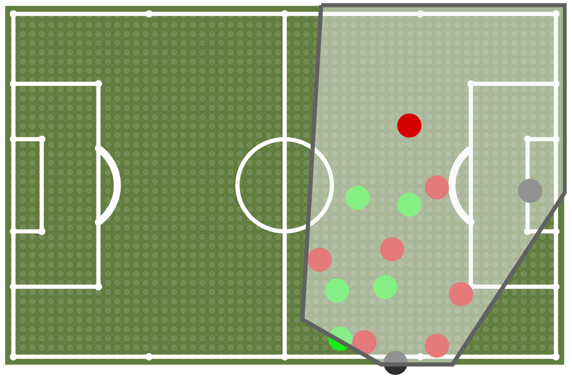
Motivation

2D position of (soccer) players on the pitch are of high interest

- (Automatic) match analysis
- Scouting
- Physiological parameters

… but not always easy to obtain due to

- Financial limitations
- Licencing issues
- Competitive concerns



→ Broadcast TV videos can be assessed more easily

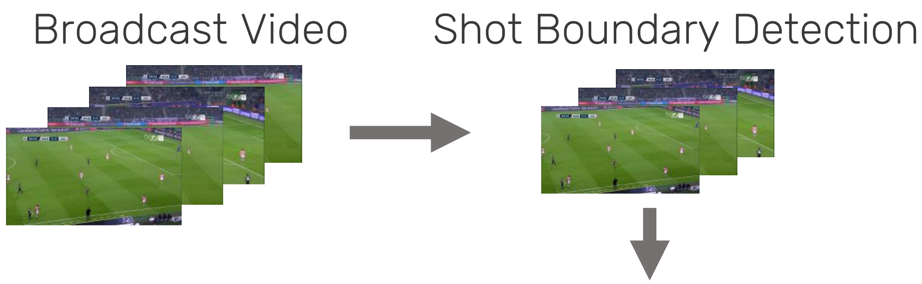
Task: **Player Position Estimation** from pan-tilt-zoom cameras

→ Compound task is not tackled in research
→ Insufficient evaluation of sub-modules regarding real-world applicability
→ Unknown quality of commercial systems

Pipeline

Broadcast Video

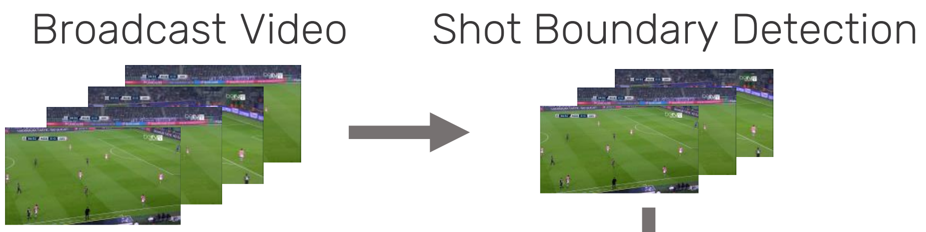
Pipeline

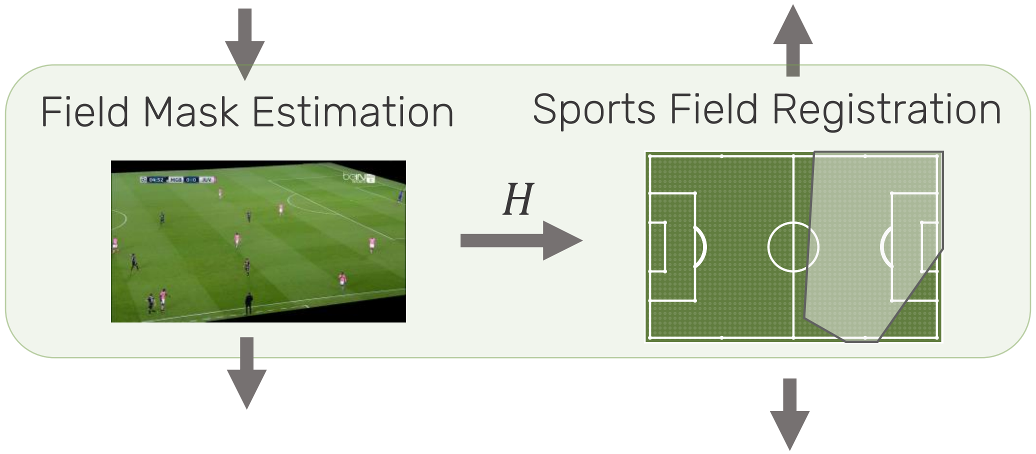


Shot Boundary Detection:

TransNetV2 [1]

Pipeline

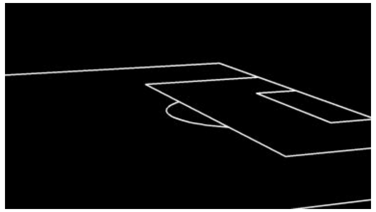




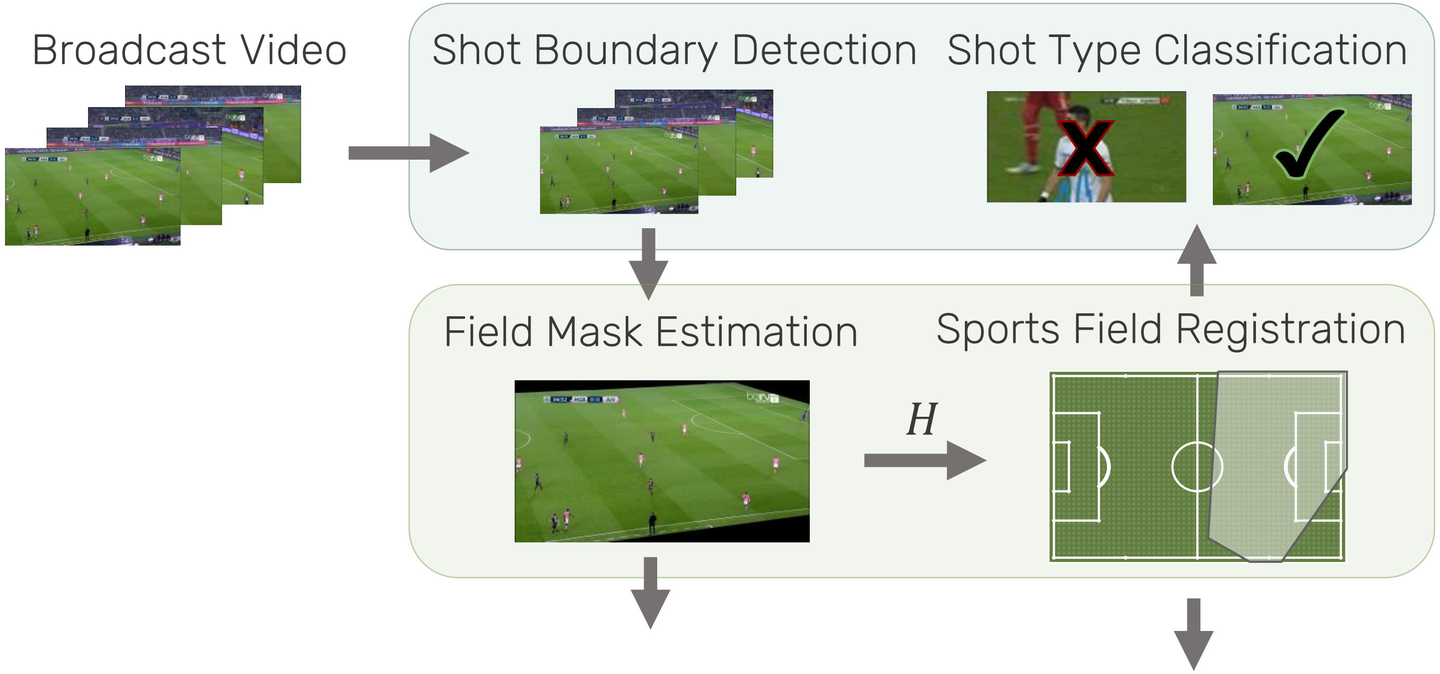
Homography Estimation: Chen and Little [2]

- Task: Estimate homography matrix $H = H_{init}H_{rel}$
- Initial guess
  - Nearest neighbor in dictionary with known camera parameters (deep feature retrieval)
  - Synthetic training data
- Refinement as relative image transformation
  → Lucas-Kanade algorithm [4]

- Pix2Pix [3] for semantic segmentation (field mask & edge images)



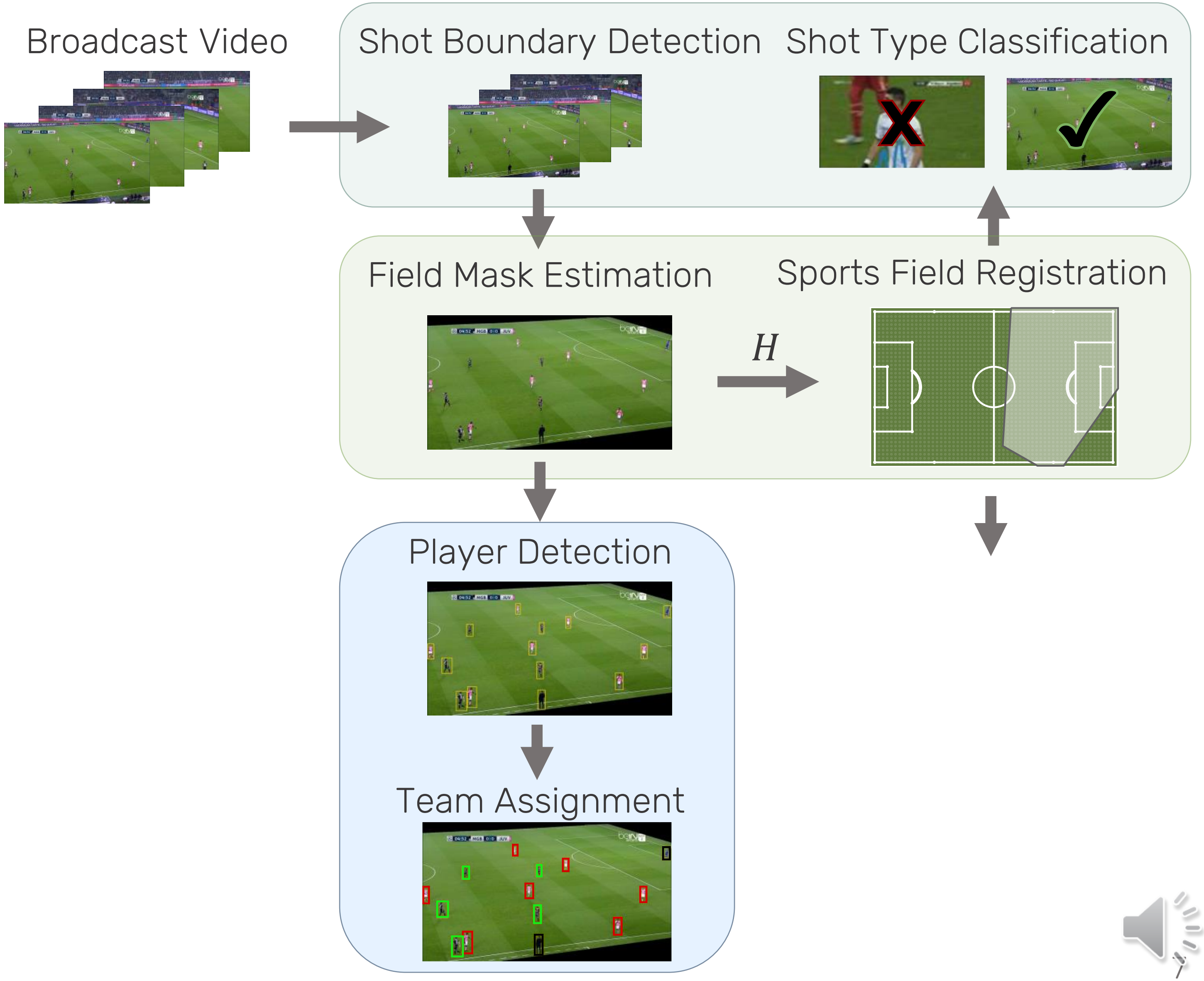
Pipeline



Shot Type Classification:

Tracking of homography changes

Pipeline



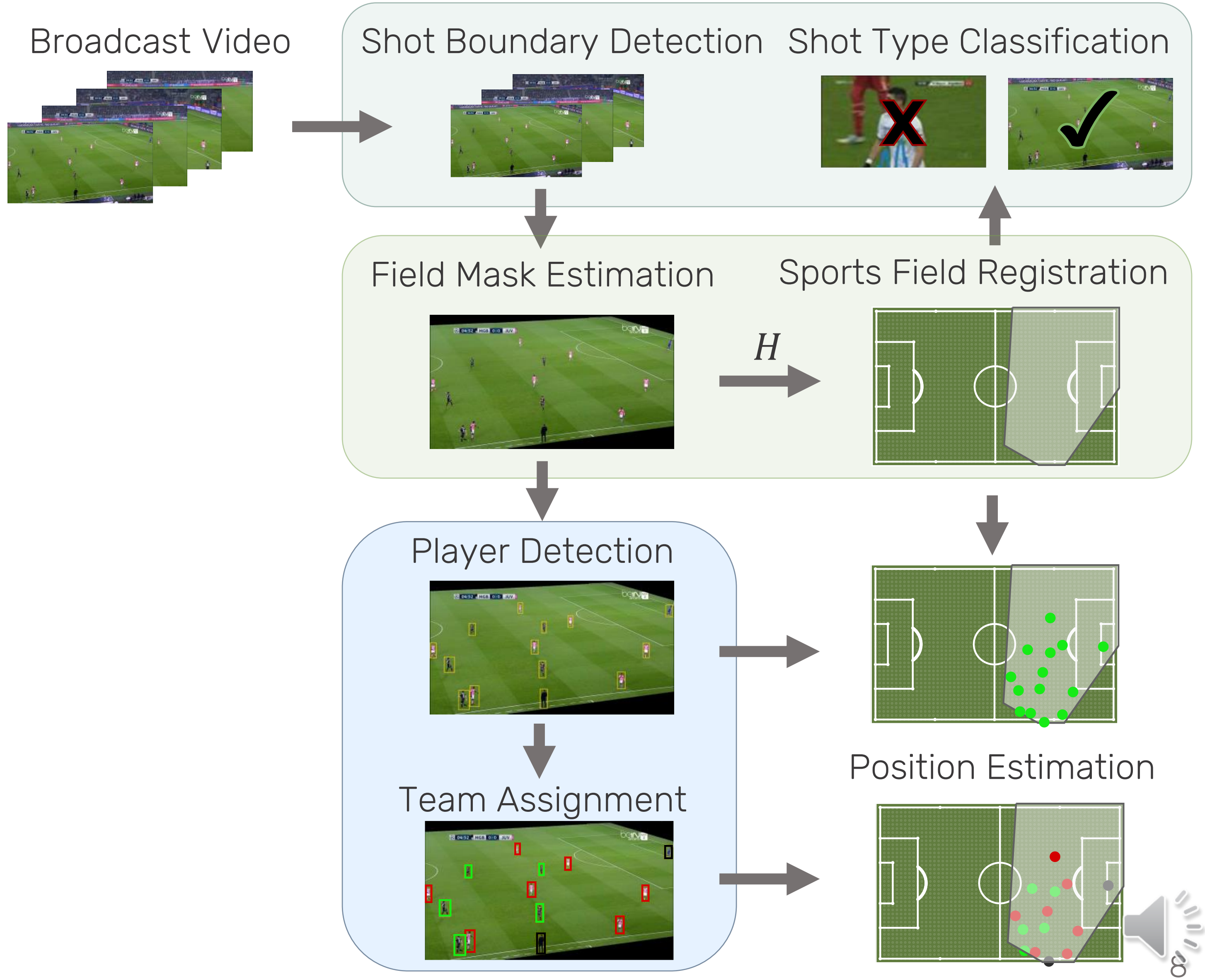
Player Detection:

→ Fine-tuned CenterTrack [5] model

Team Assignment:

→ DBScan with hand-crafted features

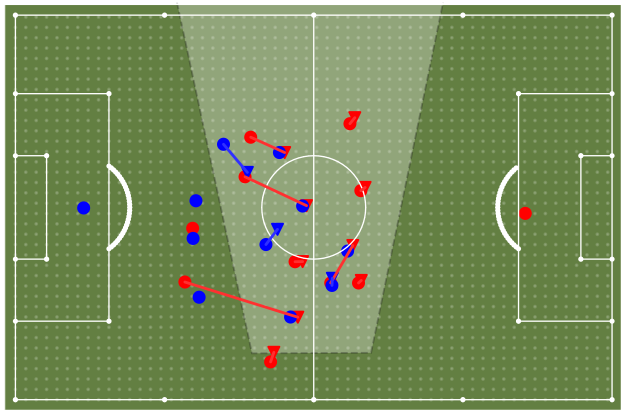
Pipeline


Broadcast Video
Shot Boundary Detection
Shot Type Classification
Field Mask Estimation
Sports Field Registration
H
Player Detection
Team Assignment
Position Estimation

Experimental Setup

Metrics: Compare estimated with ground-truth player positions

- Per-frame error in meters & aggregation per match
- Incorperate errors of individual modules
  - Visible players (ground-truth player mapping)
  - Player detection
  - Team assignment
  - Sports field registration (self-verification)

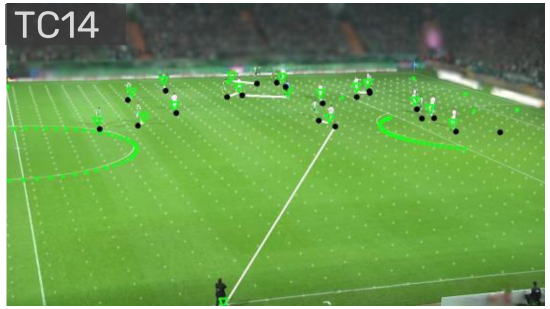


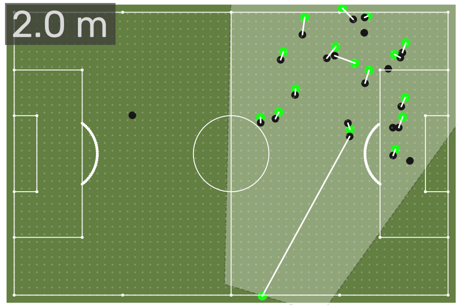
Evaluation Data

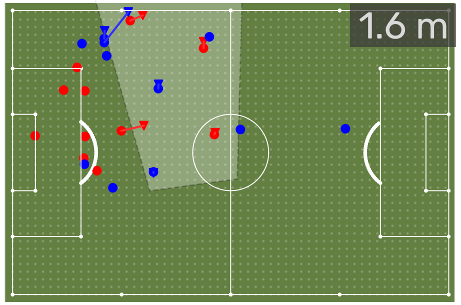
- TV broadcasts with (official) ground-truth positional data → German Bundesliga
- **No overlap to training & validation data** (league, stadium, team)

Experimental Results

- Median error in meters d_{median}
- Fraction of frames within an error < 2m (acc_{2m})







				Team Assignment Constraint			
				No		Yes	
Dataset	sv	pm	Ratio	d_{median}	acc_{2m}	d_{median}	acc_{2m}
TC14			1.00	1.20 m	0.74	1.39 m	0.78
	x		**0.90**	**1.14 m**	**0.79**	**1.34 m**	**0.81**
	x	x	0.79	1.13 m	0.79	1.29 m	0.81
TV14-S			1.00	1.36 m	0.69	2.44 m	0.43
	x		**0.92**	**1.29 m**	**0.73**	**2.34 m**	**0.44**
	x	x	0.75	1.27 m	0.75	2.32 m	0.45

Conclusions

Limitations & Future Work

- Major difficulty: Generalizability of individual modules
- No tracking & player re-identification → temporal consistency

Contributions

1) Transparent baseline with interchangable modules & data
2) How to evaluate the joint task → influence of individual modules

References

[1] Souček, T., & Lokoč, J. (2020). TransNetV2: An effective deep network architecture for fast shot transition detection. arXiv preprint arXiv:2008.04838.

[2] Chen, J., & Little, J. J. (2019). Sports camera calibration via synthetic data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops

[3] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[4] Baker, S., & Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. International journal of computer vision, 56(3), 221-255.

[5] Zhou, X., Koltun, V., & Krähenbühl, P. (2020). Tracking objects as points. In European Conference on Computer Vision (pp. 474-490). Springer, Cham.