Extraction of Positional Player Data From Broadcast Soccer Videos

Jonas Theiner[1], Wolfgang Gritz[1], Eric Müller-Budack[2], Robert Rein[3], Daniel Memmert[3], and Ralph Ewerth[1,2]

[1] L3S Research Center, Leibniz University Hannover, Hannover, Germany
[2] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
[3] Institute of Exercise and Sport Informatics, German Sport University Cologne

WACV WAIKOLOA, HI JAN 4-8

Problem Definition

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

- (Automatic) match analysis
- Physiological statistics generation
- Scouting

... but not always easy to obtain (e.g., calibrated multi-cameras, sensors)

- Financial limitations
- Licensing issues
- Competitive concerns

- Broadcast TV videos can be assessed more easily

- **Real-world 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

Contributions

1. Transparent baseline with interchangable modules & data
2. Comprehensive experimental evaluation
   - Evaluation of individual modules
   - Identify the influence of errors to subsequent modules
   - Comparison with ground-truth positional data (joint task)

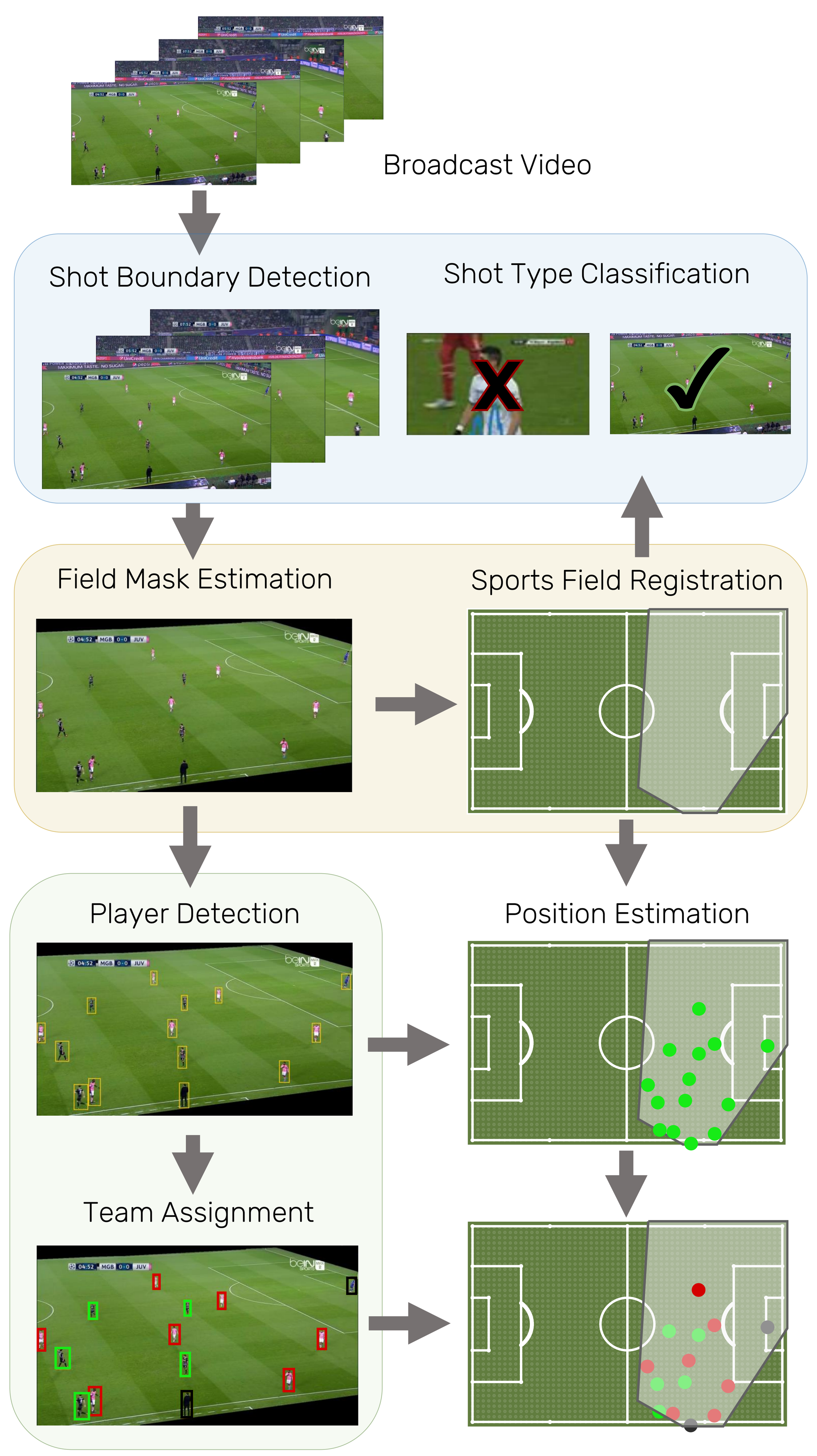
Player Position Estimation Pipeline

- Shot Boundary Detection: TransNetV2 [1]
- Shot Type Classification: → Tracking of homography changes

- Homography Estimation: Chen and Little [2]
  - Task: Estimate homography matrix $H = H_{init} H_{rel}$
  - Pix2Pix model [3] for segmentation (field mask & edge images)
  - Initial guess:
    - Nearest neighbor in dictionary with known camera parameters
    - Deep feature retrieval
    - Synthetic training data
  - Refinement as relative image transformation (Lukas-Kanade algorithm [4])

- Player Detection: Fine-tuned CenterTrack [5]
- Team Assignment → DBScan with hand-crafted features

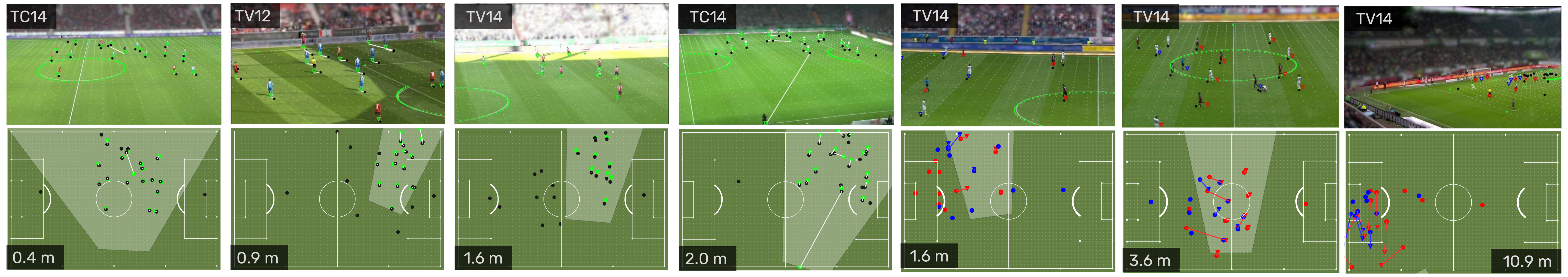
Fig. Pipeline: Player Position Estimation



				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

Tab. 1: Comparing ground-truth positions with estimated player positions: Results regarding median error (d_{median}) in meter and fraction of frames with an error of less or equal than 2 meters (acc_{2m}). Ratio indicates how many frames are kept for evaluation after applying different criteria (system output: only with *sv*).

Fig. 2: Qualitative results: Top row: Green triangles correspond to the estimated positions of players; team assignments are colored read and blue.



Experimental Results

Evaluation data

- TV broadcasts with synchronized official positional data
  - German Bundesliga (different saisons)
  - Standard TV view & *tactical cam* (smaller focal length & no cuts)

→ No overlap to training & validation data (league, stadium, team)

How to compare with ground-truth positional data?

- Mapping between visible and actual player positions
  - Solve linear-sum-assignment problem
  - Tolerate minor errors
    - Player detection
    - Team assignment
    - Ground-truth player mapping
    - → Per-frame aggregation: 80%-percentile
  - Cover larger errors from sports field registration
    - Self-verification (sv) criteria
    - Player mismatch (pm) criteria

Metric: Per-frame error in meters with aggregation per match (Tab. 1)

Key Findings

- Major difficulty: generalizability of individual models
  → Sports field registration & team assignment
  → Fail when test data is sligthly out of training distribution
  → Need for more training data or more robust algorithms
- How to evaluate the overall task
  → Influence of individual modules

Future Work

- Player tracking & re-identification
- Automatic team-performance analysis
  - With in-complete (visible players) data
  - With erronnous data

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.
.

Contact

- {theiner, gritz}@l3s.de
- {eric.mueller, ralph.ewerth}@tib.eu