A Unified Taxonomy and Multimodal Dataset for Events in Invasion Games

Henrik Biermann[1], Jonas Theiner[2], Manuel Bassek[1], Dominik Raabe[1], Daniel Memmert[1], and Ralph Ewerth[2,3]

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

Motivation & Aim

- large interest in automatic detection of sports events
  - different objectives: sports science & ML communities
  - practicioners' objectives are result orientated
- lack of consistent definitions and datasets
  - e.g., action spotting vs. temporal boundary localization
  - incomplete descriptions of games
  - lack of fine-grained approaches
  - lack of public datasets, especially for multimodal data
- We aim to provide a framework for the development and evaluation of automatic event detection models

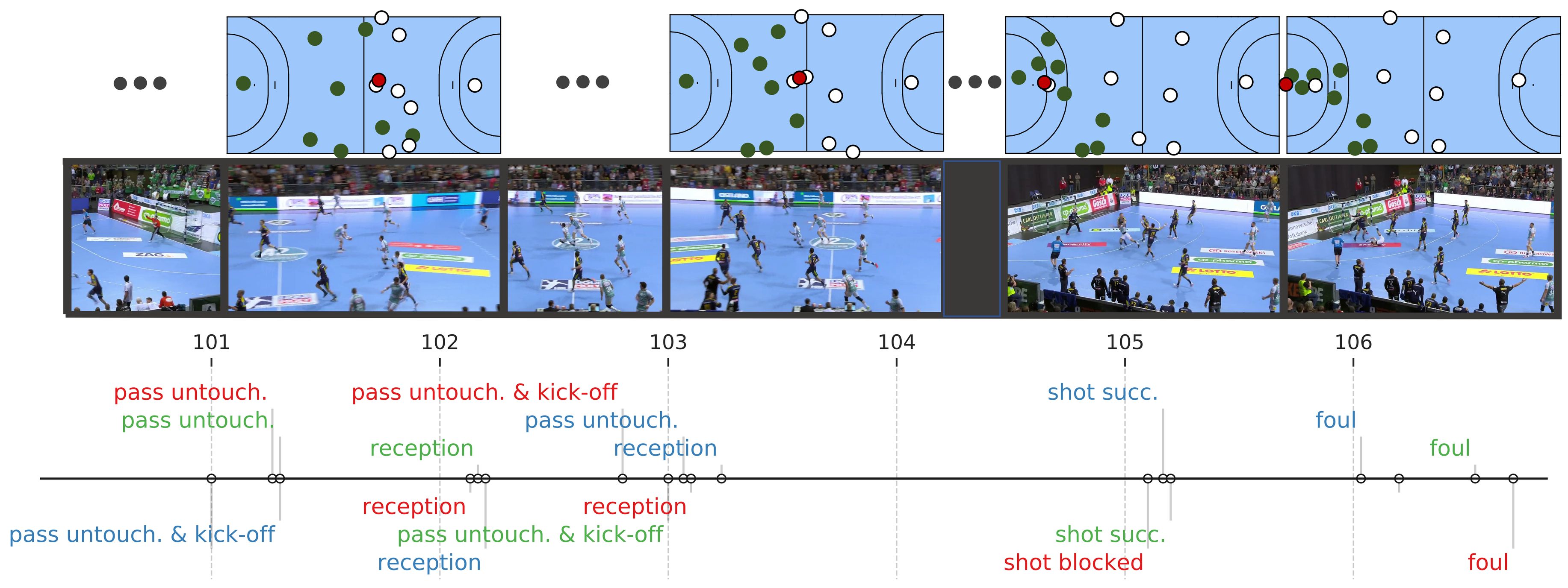

Fig. 1: Example annotations from several human annotators on EIGD-Handball using our taxonomy. Despite uncertainties regarding the concrete event type, the annotated timestamp often aligns. A mapping back to shared properties such as the motoric skill (e.g., ball release) leads to higher levels of agreement.

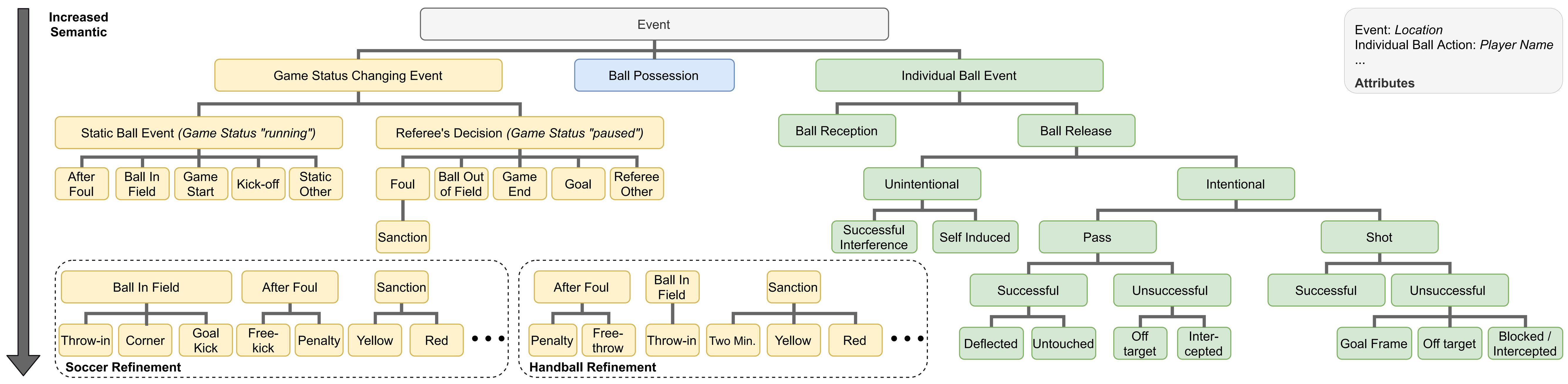

Fig. 2: Unified hierarchical base taxonomy for invasion games with exemplary refinements for soccer and handball.

Contributions

I) A unified **taxonomy** for (on-ball) events in **invasion games** (Fig. 2)
- team sports that all share the goal of sending a certain object to a target
- the team that achieves this goal more often in a certain time span wins

II) *Events in Invasion Games Dataset (*EIGD) for soccer (-S) and handball (-H)
- 5 sequences à 5 minutes from 5 games
- 8 domain experts (5 for soccer and 3 for handball)
- Timestamp annotations at the finest hierarchy level
- Annotator with the highest performance is ground truth

Evaluation

- choice of metrics is crucial for real-world application
- possible bias in common evaluation metrics
- how to match events from multiple annotations? (Fig. 1)
  - nearest neighbor (NN) matching: Possible many-to-one mapping
  - sequence consistency (SC): forced 1-to-1 mapping based on temporal anchors
    - e.g., from the static-ball event to referee's decision
    - apply if and only if the number of events within a sequence matches, and allows for conclusive comparison but discards part of the data (additional metric)

Human Performance

Dataset	Metric	Game Status Change Event	Intentional Release	Pass	Pass Intercepted
	Tolerance τ	6.04 s	0.44 s	0.44 s	0.44 s
EIGD-Handball	Number of Events	135.7 ± 23.0	841.0 ± 26.5	778.3 ± 26.6	6.7 ± 0.9
	Mean Exp.[%] (Prc=Rec)	NN: 78.9 SC: 90.0 (40)	NN: 92.9 SC: 83.3 (33)	NN: 92.6 SC: 81.8 (35)	NN: 45.0 SC: 100 (45)
	Mean Exp.[%] (Prc=Rec)	NN: 78.9 SC: 90.0 (40)	NN: 92.9 SC: 83.3 (33)	NN: 92.6 SC: 81.8 (35)	NN: 45.0 SC: 100 (45)
EIGD-Soccer	Number of Events	113.4 ± 3.4	500.0 ± 7.0	487.8 ± 7.2	58.8 ± 18.9
	Mean Exp.[%] (Prc=Rec)	NN: 95.0 SC: 98.7 (78)	NN: 96.2 SC: 93.3 (48)	NN: 96.1 SC: 93.4 (48)	NN: 60.0 SC: 85.0 (33)
	Mean Exp.[%] (Prc=Rec)	NN: 95.0 SC: 98.7 (78)	NN: 96.2 SC: 93.3 (48)	NN: 96.1 SC: 93.4 (48)	NN: 60.0 SC: 85.0 (33)

Tab. 1: Average human performance for two metrics (NN, SC) given unique tolerance areas around representative events

I3D Baseline (Automatic Solution)

Ball Reception	Shot	Pass Successful
0.44 s	0.44 s	0.44 s
821.0±10.7	62.7±1.2	65.7±25.8
NN: 45.6 SC: 0 (0)	NN: 43.2 SC: 0 (0)	NN: 46.4 SC: 0 (0)
NN: 93.9 SC: 0 (0)	NN: 41.0 SC: 0 (0)	NN: 91.5 SC: 0 (0)

Tab. 2: I3D baseline performance (see caption Tab. 1)

Shared Parent Event	Event	EIGD-S	EIGD-H
Ball Possession Change		171	136
Ball Reception		923	2268
Ball Release		1531	2470
Pass		1346	2292
	Intercepted	83	14
	Off Target	175	9
	Successful Deflected	24	6
	Successful Untouched	1064	2263
Shot		31	175
	Blocked/Intercepted	17	61
	Goal Frame	0	8
	Off Target	8	12
	Successful	6	94
Unintentional		154	3
	Other	74	0
	Successful Interference	80	3
Referee Decision		142	252
	Ball Out of Field	101	21
	Foul	32	114
	Goal	3	86
	Other	5	10
	Two Minutes	n.d.	8
	Yellow	1	13
Static Ball Action		121	207
	Corner	11	n.d.
	Free-Kick	29	84
	Game Start	1	2
	Goal-Kick	18	n.d.
	Kick-Off	1	87
	Other	0	6
	Penalty	1	11
	Throw-In	60	17

Tab. 3: Dataset distribution

Findings

- hierarchy is beneficial to tackle uncertainties (evaluation + application)
- multiple metrics are required for evaluation
- general similarities between handball and soccer
  - exceptions: more fluent transition of game status events in handball leads to uncertainty

Discussion

- Is there a theoretical or practical upper limit for model performance?
- What magnitude of error is tolerable (for a given task or application)?
- How to exploit multimodal information (positions + video) for event detection?

Contact
- h.biermann@dshs-koeln.de
- theiner@l3s.de

Dataset: https://github.com/mm4spa/eigd