

Automatic behavior analysis in tag games: from traditional spaces to interactive playgrounds

Alejandro Moreno¹ · Ronald Poppe²

Received: 12 March 2015 / Accepted: 2 January 2016 / Published online: 18 January 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Tag is a popular children’s playground game. It revolves around taggers that chase and then tag runners, upon which their roles switch. There are many variations of the game that aim to keep children engaged by presenting them with challenges and different types of gameplay. We argue that the introduction of sensing and floor projection technology in the playground can aid in providing both variation and challenge. To this end, we need to understand players’ behavior in the playground and steer the interactions using projections accordingly. In this paper, we first analyze the behavior of taggers and runners in a traditional tag setting. We focus on behavioral cues that differ between the two roles. Based on these, we present a probabilistic role recognition model. We then move to an interactive setting and evaluate the model on tag sessions in an interactive tag playground. Our model achieves 77.96 % accuracy, which demonstrates the feasibility of our approach. We identify several avenues for improvement. Eventually, these should lead to a more thorough understanding of what happens in the playground, not only regarding player roles but also when the play breaks down, for example when players are bored or cheat.

Keywords Automatic behavior analysis · Interactive playgrounds · Entertainment technology · Ambient intelligence

1 Introduction

In children’s playgrounds, tag is one of the most popular games. Players assume one of two roles: tagger and runner. During the game, one or more taggers chase and tag runners. Upon a tag, the roles of the players switch. Players can come and go as they please, and the game itself has no end. However, there are several instances that can disrupt the flow of the game, or outright cause the break-down of play. For instance, having a slow player as a tagger can detract runners from the enjoyment of playing since there is no challenge. Furthermore, the tagger might get frustrated of not being able to tag anyone and might stop playing. There might also be cheating, for example when a player refuses to admit being tagger. Also, confusion over who the tagged is inevitably disrupts the game. Although the presence of these issues during play helps children learn how to cope with them in a safe environment [4, 19, 31], they can become obstacles for play if not dealt with in time.

One way to address them is by building interactive playgrounds [34]: playing environments equipped with an array of sensors and actuators capable of monitoring, actuating, and interacting with the players [29]. Interactive playgrounds are capable of automatically analyzing the behavior of the players, understand the current situation, and manipulate the environment to steer players’ behavior in positive directions. For instance, by using floor projections, we can steer players’ behavior to make the game more engaging for everyone (e.g. giving good players handicaps), to prevent the game from breaking down (e.g. always showing the roles of each player) or to recognize anomalous behavior during games (e.g. cheating or lack of participation). The adaptation of game mechanics could also be used to try to promote social interactions (e.g. get players close to each other), prevent shy children from being excluded from play (e.g. draw arrows

✉ Alejandro Moreno
a.m.morenocelleri@utwente.nl; a.m.moreno@utwente.nl
Ronald Poppe
r.w.poppe@uu.nl

¹ Human Media Interaction, University of Twente, Enschede, The Netherlands

² Department of Information and Computing Sciences, University of Utrecht, Utrecht, The Netherlands

pointing to them) or promoting physical activity (e.g. make the game faster).

In this paper, we address the automatic recognition of players' roles in tag games. Due to its simple rules and widespread familiarity, the game of tag is a good testbed for designing and evaluating vision-based approaches to group behavior analysis in games. We first analyze game sessions of children aged 8–12 playing traditional tag in an open space. After analyzing the data, we identify potentially useful behavioral cues to distinguish taggers and runners. Based on these discriminating cues, we introduce a probabilistic model to automatically classify a player's role. The model considers interactions between players as well as individual cues and global information of the game. We apply this model on recordings of young adults playing tag in a multimodal interactive tag playground (ITP).

The remainder of the paper is structured as follows. Section 2 surveys related work on vision-based behavior analysis, focusing on studies that account for interactions between people and behavior analysis in games. Next, we describe the tag game study carried out in a traditional playground and the analysis of the recordings. Section 4 details our probabilistic role recognition model. We then turn to interactive play. Section 5 introduces our interactive tag playground and the evaluation of the model in the ITP is discussed in Sect. 6. We conclude with a discussion of avenues for future work.

2 Vision-based automatic behavior analysis

Human behavior analysis has proven to be a challenging and interesting problem in computer vision research. Its applications, such as pedestrian tracking or activity recognition (see [1, 33] for overviews) extend to diverse settings such as public spaces [2], political debates or conference rooms [10]. Traditionally, these approaches considered the behavior of individuals as entities isolated from their surroundings. Nonetheless, human behavior is affected by the behavior of people around us [3]. Recently, there has been a shift of focus from individuals towards the analysis of group behavior, for instance to analyze pedestrian movement [32] or determine group activity [6].

2.1 Socially-aware behavior analysis

Tracking pedestrians has benefited greatly from the modeling of social cues [16]. Chen et al. learn elementary groups to infer high level context information to improve the tracking of multiple people [12]. They detect pairwise grouping based on social behavior, which is later used to create a general grouping graph used for the tracking. Yamaguchi et al. also model social factors, but take into account environmen-

tal cues as well to improve their tracking method [39]. Alahi et al. propose the use of social affinity maps (SAM) to predict the destination of people in densely crowded spaces [2]. SAMs are derived from proximity analysis of pedestrians, following observations that social forces are mostly determined by proximity. Ge et al. also propose a tracking method for crowded scenes, but for the tracking of small groups [18].

Accounting for social behavior can also aid in the recognition of individual, pairwise and group activities. Bazzani et al. tackle the identification of groups and regard interactions as important cues [7]. They use a subjective view frustum along with spatial cues to estimate interactions between individuals. In a related study, they also recognize how groups are formed, maintained and dismissed [5]. Many studies have addressed the recognition of group activities such as fighting, walking in groups and queuing [13, 15, 30]. Choi and Savarese present a framework to model some of such collective activities [14]. They estimate not only atomic activities but also pairwise relationships between individuals. Tran et al. also propose an algorithm for group activity analysis that makes use of a grouping method based on social interactions [36]. They cluster people based on the amount of interaction to find relevant groups, and later classify activities. Using a slightly different approach, Chang et al. propose using proximity, not levels of interaction, to define groups in their probabilistic model for scenario recognition [11]. They use a soft-grouping approach with path-based connectivity to define group memberships.

While these social cues are valid in many daily settings, they translate poorly to settings where social conventions are not applicable such as games. Regressing to the sole analysis of individual actions, such as running or jumping, would make it harder to understand joint activities between players or overarching scenarios played out during games. Still, social information is present in game settings, although in the form of player interactions, which several studies have addressed [36, 40].

2.2 Behavior analysis in games

Game settings can vary greatly, from professional sport scenarios, where automatic behavior analysis has been used to aid in understanding team strategies [8, 26], to playground games such as tag, Marco Polo or hide-and-seek, where automatic recognition has been used to monitor children's social skills and diagnose social conditions such as autism [28, 35, 38]. At their most basic level, games require some level of physical exertion (e.g. running). They also need player affiliation (e.g. team or role). Lastly, the actions that players execute are goal-oriented, such as kicking the ball to score a goal or running to avoid a tagger.

Automatic behavior analysis is carried out in two subsequent steps: tracking and classification. Tracking in sports

and games requires a different approach than those used in related fields where motion can be more predictable, such as in pedestrian tracking (see [17] for an overview). In games, the movement exhibited by the players is much more varied. For example, players can have outbursts of speed or a sudden change of direction to perform specific actions like dodging an opponent. Being unpredictable and able to change motion suddenly is often a desirable characteristic.

The tracking can be improved by taking into account the game state. For instance, Lucey et al. show that knowing the role of a player (defender, attacker) can aid the tracking process in field hockey matches [25]. Although teams can adopt many different formations, all are comprised of specific roles and their associated behaviors, which can help limit possible player locations. Moreover, the opposing team players' locations can be used as well, since they need to guard their opponents and thus stay close to them. Liu et al. track basketball players and argue that tracking players using a single model is not optimal [24]. They introduce context game features such as absolute or relative occupancy maps, to model player movements conditioned on the state of the game.

With the players' track information, behavior analysis can be carried out. Kim et al. predict interesting moments in soccer matches based on how the flow of movement converges [22]. They state that the motion of every player is related to the motion of the surrounding players. Even though an individual player's behavior is complex, actions of nearby players can aid in recognizing it. Similarly, Lan et al. recognize activities in field hockey matches by analyzing low-level (i.e. actions) and high-level (i.e. events) information, based on given player locations [23].

Tracking players is not always required to understand games. Lucey et al. track the ball instead of the players in soccer matches [26]. They estimate the amount of ball possession a team has accumulated in any given part of the court to recognize home and away behavior for teams. Completely circumventing the need for tracking is also an option. Motivated by the inherent difficulties in tracking, Khokhar et al. use a spatiotemporal description of the events to classify activities [21]. They present a method for multi-agent activity recognition that extracts motion patterns using optical flow, clusters them, and uses them to build a graph which describes the activity. They recognize activities in American football matches such as middle run and short pass. Similarly, Bialkowski et al. recognize team activities such as penalty corners or face-offs in field hockey without employing tracking [9]. They employ centroid representations or occupancy maps based on player detections.

The research on playground games is limited. Moreno and Poppe [28] proposed a role recognition model that uses pairwise interactions (approach, chase, avoid) between players, and test their approach on simulations of tag games. In this

paper, we also consider tag games, but we present a probabilistic model that estimates the probability of a player being a tagger based on individual, pairwise and global cues. These cues are defined through the analysis of player behavior in traditional tag games. The model is tested on tag game sessions recorded in an interactive installation.

3 Behavior analysis in traditional tag games

One particularly popular and widely known playground game is tag. In its most basic form, players are either runner or tagger. A tagger chases the runners around the playing area to tag them. On the other hand, runners have to avoid being touched by taggers. When a tagger tags (physically touches) a runner, the roles of both players are exchanged. Immediately after a tag, the new runner cannot be tagged for a previously agreed amount of time. This is known as the cool-down period. Tag games can be played almost anywhere, as long as there is enough space to run around. There is no explicit end to tag games and they are typically played until the players are tired or bored. Players are also free to come and go. There are variations to normal tag that provide a finish condition to the game, for instance when those that are tagged are "frozen" or "out".

In order to analyze differences in behavior between the tagger and runner roles, we rely on traditional tag game sessions. We used the Play corpus, a dataset that contains 9 sessions of children playing normal tag.¹ These 9 sessions contain 12 and a half minutes (15,008 frames) of normal tag, with 74 tag events. This amounts to an average time of 10.14 s between two tags. The sessions are supervised by a referee who assigns roles to the players, instructs players to enter or exit the playing area, and stops and starts new sessions. A maximum of 8 children could participate simultaneously in a session. Sessions with different numbers of taggers and runners were recorded. The playground in which the games took place was 7 × 6 m. Sessions were recorded with three cameras, located outside the play area. In addition, there were four Microsoft Kinect sensors placed on the ceiling of the playing area.

The images of the Kinects were stitched and used as the basis for our offline, semi-supervised tracker. Figures 1 and 2 show a frame from the Play corpus recordings after stitching the RGB and depth images from the Kinects. At this point, we are interested only in the positions of the players. Tracking results were propagated automatically and manual input was requested whenever two players were very close. We linearly interpolated missing detections if they were shorter than three

¹ The Play corpus also contains variants of normal tag as well as sessions of another game, pass-a-ball. It can be obtained at <http://hmi.ewi.utwente.nl/playcorpus>.



Fig. 1 RGB image of a frame from the Play corpus

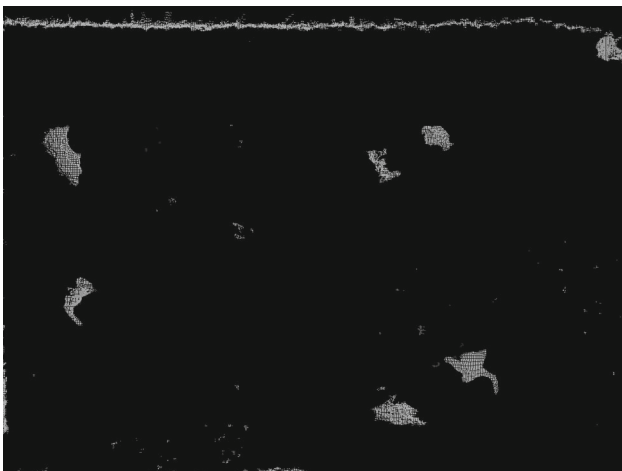


Fig. 2 Depth image of a frame from the Play corpus

seconds. Moreover, we applied a moving median filter to the estimated positions.

Using the RGB cameras' feeds, the role of each player was manually annotated. The process involved one annotator going over the video, frame by frame, while writing down the role of each player. Specific problematic instances where players did not behave appropriately were reviewed several times to make sure the annotation was correct. For instance, a player would tag someone but the other player would not notice, thus the first tagger would resume his tagging role after some time had passed. In cases such as this, the initial tagger was assigned the tagger for the entire duration. This meant going back and forth in the video to see how children reacted to certain tags. Moreover, in some occasions children just cheated and refused to become taggers. The same procedure as before was used in these cases.

We take a closer look at a number of features that can be derived from the position and movement of the players. We focus on features that appear promising to distinguish

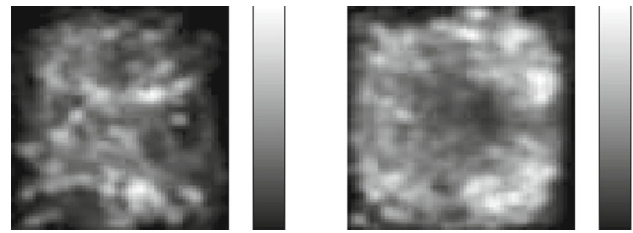


Fig. 3 Occupancy map of players within the playground (*left* tagger, *right* runner)

taggers from runners. For each, we use the position data of the selected sessions in the Play corpus.

3.1 Absolute position

We first analyze the absolute position of the players as we expect differences in where they are within the playground. Taggers, in their attempt to tag runners, should be looking to position themselves such that they can tag people efficiently. A more central position in the playground is therefore likely. Runners, while trying to avoid being tagged, will stay as far away from the tagger as they can. They might be moving especially along the borders of the play area.

Figure 3 shows the mean occupancy of players in the playground, calculated over the sessions. Lighter values correspond to more presence at a specific location. We normalized the values to stretch to the entire color range. We applied a Gaussian filter to reduce the effect of incidental peaks on this stretching and to make the overall pattern more clear. We can see in the figure that the location heat map for taggers and runners are very different, and largely follow our intuition. Taggers tend to operate near the center of the playing area, cutting across it when chasing players. In contrast, runners tend to stay near the borders of the playing area. It seems that the distance of a player to the border, or to the center, of the playground is indicative of the role. This allows us, without having to consider the positions of other players, to make estimates about a player's role given only his position.

3.2 Movement speed

Next, we look into the movement speed of the players when they have a specific role. We expect that taggers should, in general, have a higher speed than runners since they have to chase other players. Runners, on the other hand, can rest when another player is being chased or move slowly away from the tagger while his attention is on someone else. Again, we only consider a single player, without looking at the others in the playground.

We took the inter-frame distance of the positions of each player as a measure for speed. We calculated histograms of

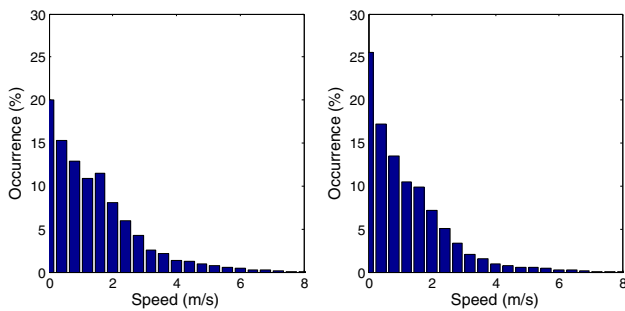


Fig. 4 Frequency histograms of players' movement speed (*left* tagger, *right* runner)

the speed of the taggers and runners individually, shown in Fig. 4. These two histograms are largely similar, except for low speeds. Runners, more than taggers, move at low speeds, including standing still. This is what we expected, as they can take short breaks while the tagger is chasing someone else. It is interesting to notice that taggers also have a significant count for low speed values, which is probably caused by the moments in which they were deciding whom to tag, or to take a rest. Despite the similar movement profiles, taggers do move faster than runners but only slightly so.

Something not directly evident from seeing the graph is that the speed values for runners and taggers were similar within all sessions, even in sessions where the average speed for all players was very low. This could imply that players adjust their speed to match other players. This could be the case when a tagger is slow, and runners do not require much effort to prevent themselves from being tagged.

3.3 Inter-player distance

The third feature under investigation is the distance between players based on their role. Intuitively, we would expect runners to be far away from taggers since they are trying to avoid them. As it happens, we have also noticed that runners tend to group together, either to use others as bait (i.e., hoping the tagger chases them instead) or as protection (i.e., stand behind another player and push them towards the tagger). This would further emphasize the inter-player distance difference between roles, since runners would on average be closer to other runners.

Contrary to what we expected, Fig. 5 shows that there are no large differences when players stand close together. This follows from the fact that taggers are consistently trying to get closer to runners to tag them. Instead of looking at the distance of a player and all other players, we can also only take into account the closest player. These numbers appear in Fig. 6. The difference between the two roles is now more evident, with taggers on average being closer to the closest

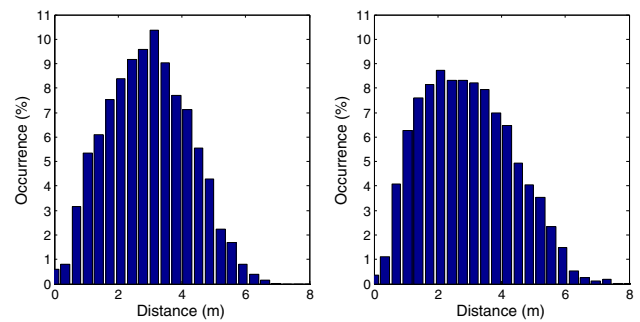


Fig. 5 Frequency histograms of inter-player distances (*left* tagger-runner, *right* runner-runner)

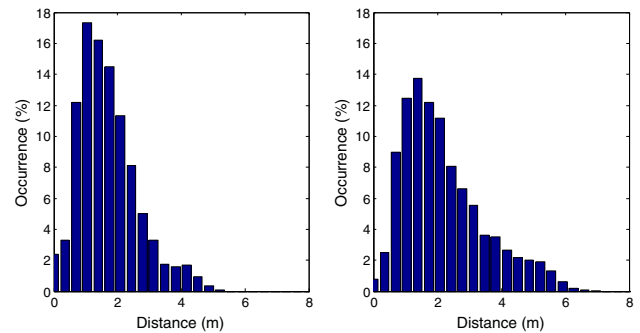


Fig. 6 Frequency histograms of distances of taggers (*left*) and runners (*right*) to the closest runner

runner than runners. Nonetheless, the difference is not marked enough for it to be discriminant.

3.4 Relative movement direction

Finally, we analyze the relative movement direction between players of different roles. We analyze the movement direction of a focus player with respect to the position of another. If this direction is close to zero degrees, the focus player is approaching the other. This behavior is expected to occur more for taggers. Conversely, when the movement is away from the target person, the player might be runner. We are only looking at the relative direction and not the amount of movement.

Figure 7 shows the angular histograms of relative movement direction for the tagger-runner and runner-tagger combinations, respectively. The histograms show that the relative movement direction of taggers in relation to runners (red) is most often near 0° and falls off quickly. This is precisely what we expected, since taggers have to move towards other players to tag them. The variation in angle is due to taggers predicting the movement of chased players and moving ahead of their path to cut them off. In the calculation of these figures we took into account all runners, which contributes to the wider spread of values for the taggers. In regards to runners, we also find what we were expecting: they tend to

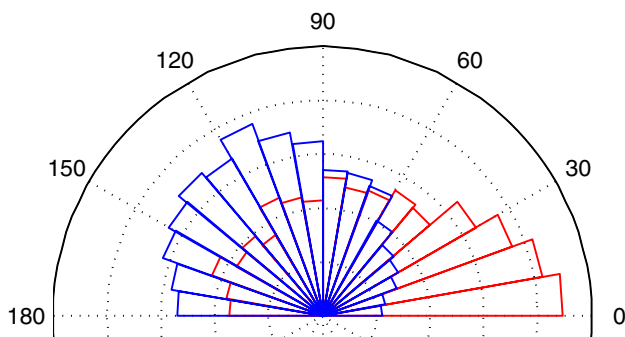


Fig. 7 Angular histogram of the relative movement direction between roles. In *blue* from runner to tagger, in *red* from tagger to runner

move away from taggers at angles in the 90° – 130° range. This, together with the absolute position analysis of runners, leads us to conclude that runners move in circles around the playground, since instead of running in the complete opposite direction of the tagger (180°), they move diagonally, while keeping away from the center of the playing area.

The histograms have been calculated over the same amount of data so we can compare the bins in a pairwise fashion. It can be observed that for angles from 60° and smaller, it is more likely that the player is a tagger. Angles larger than 60° occur more often when the player is a runner. It seems that we can use this cue to distinguish between the two roles.

4 Role recognition model

From the analyses in the previous section, we observe that a player's position within the playground and the relative movement direction differ between the tagger and runner roles. We propose to recognize players' roles by considering these cues. The position is an individual cue, whereas the movement direction is a pairwise cue. We use these cues to define the *boundary response* and the *tagging intention*, respectively. The former is a function that estimates the role of a player based on his location in the playground. The latter is a function that evaluates how likely it is that one player is trying to tag another. We present a probabilistic formulation to determine each player's role individually, by considering these two concepts. At this point, we assume a single tagger and we only look into tagger-runner interactions because runner-runner interactions are not apparent. Finally, we estimate the roles of all players by considering that there is only a single tagger.

Formally, we consider a set of N players, each with $R \in \{t, r\}$ a random variable indicating the tagger and runner role, respectively. We omit the index on the player for clarity. Given a set of observations O , the probability of a

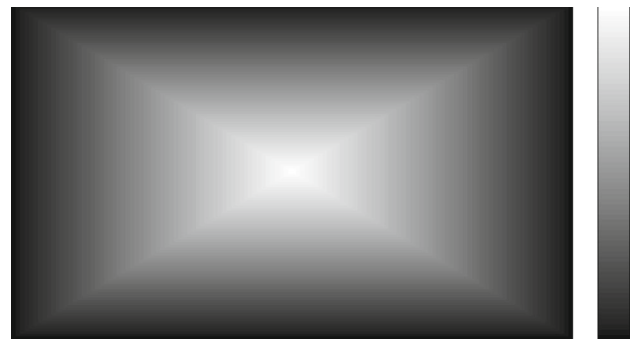


Fig. 8 Boundary response function

player being a tagger follows from Bayes' rule:

$$P(R = t|O) = \frac{P(R = t) \cdot P(O|R = t)}{\sum_{i \in \{r, t\}} P(R = i) \cdot P(O|R = i)} \quad (1)$$

Given that we consider only a single tagger, the prior probability of a player having the tagger role depends on the number of players in the game: $P(R = t) = 1/N$. The likelihood function $P(O|R = t)$ is calculated by the boundary response B and tagging intention I . The normalization term is the sum over all hypotheses, specifically the player being a tagger or a runner. Below we describe the observation term in detail.

4.1 Likelihood function

The likelihood function is estimated using two different functions, B and I . We assume both are independent from each other, so the likelihood function is formulated as follows:

$$P(O|R) = P(B, I|R) = P(B|R) \cdot P(I|R) \quad (2)$$

4.1.1 Boundary response

In the previous section, we showed that taggers have a tendency to stay in the center of the playground, and avoid the borders (see Sect. 3.1). On the other hand, runners prefer to stay near the borders of the playground. Consequently, the boundary response B is defined as a normalized distance function that takes as input the location of a player and the size of the playground, and outputs a response like the one seen in Fig. 8. As such $P(B|R = t) = 1$ when the player is in the center, and $P(B|R = t) = 0$ when he is at the border. Given that we have only two roles, the probabilities for the runner role is reversed, i.e. $P(B|R = r) = 1 - P(B|R = t)$.

4.1.2 Tagging intention

To define the tagging intention of one player towards another, we use the direction of movement of player i in relation to

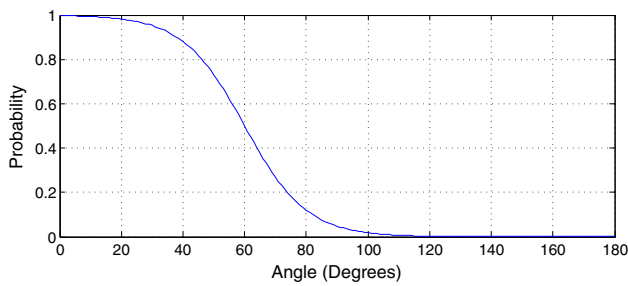


Fig. 9 Tagging intention function $\theta_{i,j}$

player j . Following our observations in Sect. 3.4, we calculate the angle between the vector between i and j and the movement direction of i . On this angle, we obtain $\theta_{i,j}$ by applying a sigmoid function:

$$\theta_{i,j} = 1 / (1 + e^{-a(x-c)}) \tag{3}$$

Here, x is the angle between the two vectors, a the fall-off rate and c the center. Center c is the angle at which the probability for both roles is equal. For smaller angles, in which i is moving more in the direction of j , the probability increases for the tagger role, whereas larger angles lead to an increase for the probability of being a runner. The fall-off rate determines how quickly the probability changes from 0 to 1. Figure 9 shows $\theta_{i,j}$ for $c = 60^\circ$ and $a = -0.1$.

When player i is not, or barely, moving we cannot accurately calculate $\theta_{i,j}$. In this case, we set the tagging intention I to 0.5 as we cannot distinguish between a tagger and runner. We use a conservative speed threshold of 0.2 m/s. Below this threshold, we assign equal probabilities for the tagger and runner roles. When the speed is above the threshold, when i is moving, we define $I = \theta_{i,j}$.

A tagger is typically chasing one out of a number of runners. As $\theta_{i,j}$ considers only a single player, we calculate the probability of a player being a tagger based on the tagging intention as $P(I|R_i = t) = \max_{j \in 1 \dots N, i \neq j} \theta_{i,j}$.

4.2 Role classification

Recall that we assume a single tagger and an arbitrary number of runners. We classify the roles of the players by estimating their probability of being a tagger using Eq. 1, and select the player with the highest probability as the tagger. The other players are assigned the runner role.

5 Tag games in interactive playgrounds

We now move from traditional tag games to interactive play. Before evaluating the role recognition model in Sect. 6, we introduce the Interactive Tag Playground (ITP) that was used



Fig. 10 Students playing tag in the ITP

in our evaluation. The ITP is a multimodal interactive installation where people can play tag games. The playground tracks players in the playing area and displays a colored circle around their feet as they move (Fig. 10). The color of the circle is indicative of the role: red for taggers, blue for runners.

When the game begins, one random player is assigned the tagger role. Instead of physically tagging other players, a tag occurs when the tagger’s circle collides with a runner’s circle. Once this happens, the roles of the players switch, as well as the colors of their circles. Additionally, the former tagger has a two second cool-down period where he cannot be tagged. If the tagger leaves the playing area, another player is chosen as the tagger. An important characteristic of the ITP is that, since it knows the location of the players and their roles at all times, it generates automatically annotated game log files.

In this study, the game projections are used solely to inform players of their roles. However, they can also be used to enhance the classic game of tag by, for instance, adding power-ups that give special bonuses such as bigger circles or tag protection [37]. The color of the circles could also be used to give players interesting information. For example, good tag players could start glowing, bad players could have dim-colored circles that wobble, or players that have not been tagged in a while could get a yellow circle that emits particles. Importantly, these technological interventions still allow players to show the kind of physically active and social behavior shown in traditional tag [27].

It seems redundant to recognize player roles given that these are assigned by the ITP. Eventually, we are interested when the observed player behavior deviates from what we regard as proper tag behavior. Given a good model, this occurs when the recognized roles differ from the assigned roles. In this case, players might be doing something else such as cheating or not be actively taking part in the game. We will address this in future work.

The ITP features an online top-down, multi-person tracker that tracks the position of the players. It uses Kinect depth

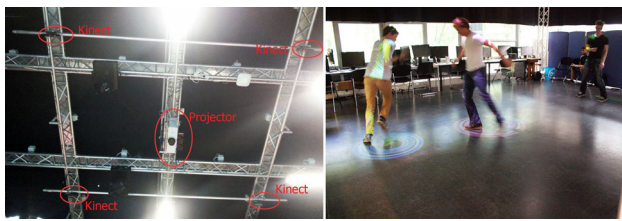


Fig. 11 Setup of the interactive tag playground

images as input. Color information is discarded because the ITP operates in a dark environment for the projections to be visible. These projections partly overlay colors on the players, which makes it harder to track them reliably using color information. To make the tracking more robust, each player wears a wireless gyroscope (YEI 3-space) in a strap around the chest. These additional measurements help to prevent tracking errors when detections are merged or lost, as described later.

5.1 Setup

The playground is composed of four Kinects and a single projector located in the ceiling. The Kinects are arranged in a grid-like setup, 4.0 m apart, while the projector is located in the center of the grid (Fig. 11). The ceiling of the playground sits at 5.3 m. At this height, the ITP is able to track people in a 7.0×6.0 m area, but only project into a 6.0×3.3 m area, which defines the effective playing field.

5.2 Player detection

To detect the locations of players in the playground, we first apply a threshold to the depth images from the Kinects to remove the floor and potential small objects. Since we know the exact height at which the Kinects are located, the threshold can be set simply by taking into account the players' heights. The resulting depth images contain depth values for the head and shoulders region, but typically also contain the arms. To promote the head region, we filter the images with an approximation of the Mexican Hat filter, a Difference of Gaussians kernel. This filter gives higher weight to Gaussian-like objects such as the head-shoulder region, while non-Gaussian objects such as stretched arms will receive low values. Importantly, when two players are very close, their outlines merge. The Mexican Hat filter can still identify multiple peaks, each corresponding to a head, in such a region. Given that tag is a physical game in which there is a considerable amount of physical contact, these processing steps are essential. Based on the filtered image, we select the highest values using a threshold. The corresponding locations typically correspond to the heads of the players.

Next, the locations of the players are mapped onto real-world coordinates relative to each Kinect. We apply this procedure for each Kinect individually. Since we know the physical disposition of each Kinect and their distances to each other, we can map the Kinect-based real-world coordinates to playground-based real-world coordinates. Since the Kinects' fields of view partly overlap, we check for detections that originate from different Kinects but are within 0.5 m of each other, measured from the center of the detection. If this is the case, we assume that they belong to the same person and merge them.

5.3 Player tracking

Given the detected players, we track them to ensure that the identities of the players are maintained over time. We use Kalman filters as these are straightforward and we observed they are well capable of tracking running players. The motion of each player is modeled as a combination of location and speed. A prediction is made at each time step. Based on the player detections, we correct the model. Essentially, we label the player according to the Kalman filter that makes the best prediction, calculated as Euclidean distance between detection and prediction.

When a new detection is found, we create a temporary track to assign it to. After five frames (approximately 0.25 s), we validate the track's stability to reduce tracks created from noisy observations. If the detection assigned to this track has been visible in three out of the five frames, the track is maintained. To delete tracks, we take a similar approach. If a track is not assigned a detection, we keep the track alive for 15 frames (approximately 0.75 s). During this time, if the track is not assigned any detection, it is deleted. Using this approach, we can handle occasional missed detections, while still preserving tracking accuracy.

When a merged tracks splits, for example when two physically close players start moving in different directions, we additionally use the gyroscopes to identify the players. We compare the direction of the movement obtained from the player detection to that read from the gyroscopes. We apply this procedure continuously. In each frame, if the difference between Kinect and gyroscope estimates exceeds a certain threshold, we assume a mislabeling has taken place. In this case, we check for another player's gyroscope that shows the same inconsistencies with its assigned track's movement direction, and swap the identities to correct the assignment if it has lasted more than 45 frames (2.25 s). This allows the system to recover automatically from errors in track assignment. This is a rare event, and typically is solved swiftly after a complex situation, for example when multiple players bump into each other.

6 Experimental results

In this section, we evaluate the role recognition model on tag sessions recorded in the ITP. The assignment of roles in the ITP is handled automatically. We can therefore easily obtain both the position and the ground truth for the role of each player. In the next section, we explain how the recognition of the roles from the positions helps to better understand behavior in the playground. First, we describe the data used in the evaluation. Next, we present and discuss our findings.

6.1 iTag corpus

In the ITP, we recorded the iTag corpus consisting of interactive tag sessions. In total, we recorded 14 sessions with a total duration of slightly over one hour. The length of each session varied as the participants were allowed to play as much as they wanted, but typically lasted around four minutes. In total, we recorded 73,902 frames of tag game (at 20 frames per second), in which a total of 682 tags occurred. In each session, three players played simultaneously. The players were young adults aged 20–30. At the beginning of each session, while the game was explained, players were equipped with wireless gyroscopes inside jogging strap holders. These straps did not limit the movement of the players. Each sensor was assigned to the corresponding player’s track.

6.2 Results

We apply the role recognition model to the sessions in the iTag corpus. Despite the different age group (children aged 6–8 and young adults aged 20–30) and play area size, we assume that our observations, and the model that was based on it, generalize to the iTag corpus. In the model, there are two parameters in the calculation of the tagging intention: fall-off rate a and center c . We set c to 60° , following our observations in Sect. 3. Here, we evaluate different values for a and measure the overall accuracy. The results of Eq. 1 can be somewhat noisy, for example when a player moves at very low speeds. Therefore, we run a median filter with a window size of 7 frames (0.35 s) on these probabilities. After this temporal smoothing, we select at each frame the player with the highest probability of being the tagger.

The results for different values of a are visualized in Fig. 12. These numbers are calculated over all frames in the iTag corpus. As can be seen, the differences between settings are generally small. Values of a closer to zero result in a more gradual decrease in tagger probability, see Eq. 3. In the figure, we observe that the best accuracy is obtained for $a = -0.1$. In the analyses in the remainder of the paper, we will use this value.

We summarize the results for the tagger and runner roles in Table 1. The confusion matrix appears in Table 2. Our

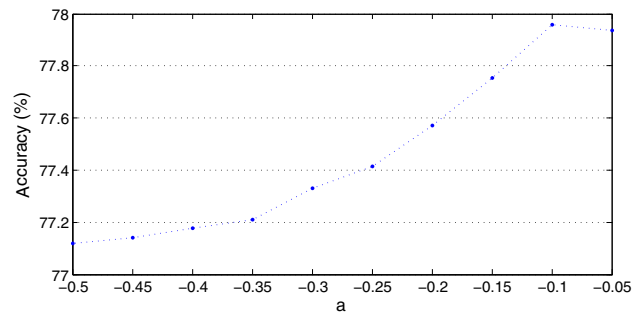


Fig. 12 Variation of the model’s accuracy with respect to parameter a

Table 1 Results of the role recognition model over all sessions in the iTag corpus

	Precision (%)	Recall (%)	Accuracy (%)
Tagger	67.07	66.29	77.96
Runner	83.29	83.78	

Table 2 Confusion matrix over all tag sessions in the iTag corpus

	GT tagger	GT runner
Guessed tagger	45,191	22,191
Guessed runner	22,984	114,585

GT ground truth

role recognition model is able to determine the roles with a 77.96 % accuracy. The baseline accuracy is $5/9 \approx 55.56\%$ overall but there is a difference in baseline for each of the roles. Given that there is one tagger at each moment, guessing the tagger in these three-player sessions would give a baseline of 33.33 %. However, guessing that each player is a runner would lead to an accuracy of 66.67 % but without any tagger identified. This is also reflected in the results, which show better performance for runners, both in precision and recall.

To understand the relative importance of the boundary response and the tagging intention, we also tested the model with an adapted likelihood function. When using only boundary response B , the accuracy of the role classification is 69.52 %. For tagging intention I only, this increases to 76.59 %. This shows that the interactions between players are more informative than their locations for recognizing their roles. This is not surprising since the interactions are directly related to the expected role behavior, while location only gives a coarse approximation of the role. Nonetheless, the location does add useful information as shown by the slightly higher accuracy when using both features.

6.2.1 Analysis of different sessions

When looking at the individual sessions, the precision, recall and accuracy remain relatively stable as seen in Fig. 13. Ses-

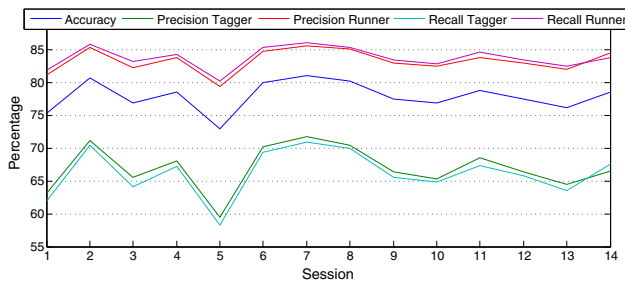


Fig. 13 Role recognition results for all sessions in the iTag corpus

sion 5 scored the lowest accuracy of all sessions. Closer analysis of the videos revealed that the players initially did not rely on the automatically assigned roles, but rather used physical tag despite the prior explanation of the game. Halfway through the session, they started using the assigned roles. Although the role classification model uses only behavioral cues to classify the roles, the role information provided by the ITP is used as ground truth for the classification. This means that the ground truth for the first half of the session is essentially incorrect, which affects the model's performance. Overall, the results between sessions are not too different even though there were notable differences in skill level, roughness and enthusiasm between the different groups of players. We therefore believe that the approach is general enough to suit a broad audience.

6.2.2 Temporal analysis

Tag games are dynamic and there are several phases to be identified. Taggers chase runners, go for a tag, the roles switch and the process starts again. The behavior cues that we identified are most pronounced when a tagger chases a runner. Once the tag has been made, the behavior is somewhat less pronounced. Taggers have to switch from chasing to running and the opposite is true for the runners. We hypothesize that our recognition accuracy around tags is therefore lower. To this end, we have calculated the average accuracy around a tag, which we identified from the automatic annotations as a change in roles.

Figure 14 shows the average probability variation for the two seconds before and two seconds after a tag. The figure shows both the probability values of being a tagger for the old tagger (that becomes a runner) and for the new tagger. Slightly before the tag, the probability for the old tagger starts to decrease, which is probably due to the player reducing his speed and trying to avoid a full-on collision with the other player. We see that this probability further decreases after the tag, which makes sense as the roles are then switched and the old tagger needs to flee from the new tagger. For the new tagger, we see the probability increase, but only after a second. The new tagger first needs to realize he is tagged and then localize a target.

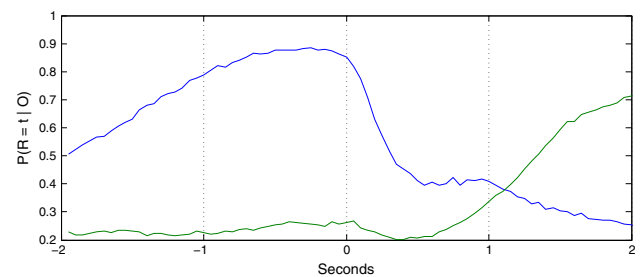


Fig. 14 Average probability of being a tagger over a four second window around a tag of the former (*blue*) and new (*green*) tagger

Table 3 Results on the iTag corpus with the approach proposed in [28]

	Precision (%)	Recall (%)	Accuracy (%)
Tagger	67.35	13.22	69.00
Runner	69.12	96.80	

Since we select the player with the highest probability of being a tagger as the guessed tagger, this delay of over a second lowers the overall accuracy of our recognition model. In the iTag corpus, there is a tag on average every 5.4 s. A delay of over a second therefore has a significant effect on the performance. These observations therefore motivate the introduction of a model that takes into account the different phases of a tag game. Still, given that the behavior is less pronounced, it will be more difficult to make correct guesses on who the tagger is. Alternatively, we could rely more on the estimated tagger probabilities. Apparently, a drop in probability for the current tagger occurs before the increase in probability for the next tagger.

6.2.3 Comparison to related work

Finally, we compare our proposed approach with that of [28]. Their approach takes into account interactions between all possible pairs of players. The results obtained on the iTag corpus with their model, with their best recognition threshold of 0.8, are summarized in Table 3. The accuracy of the model is 69.00 %, which is 9 % lower than the role recognition model presented in this paper. Changing this threshold to 0.3 increases the accuracy to 71.28 %, with increased precision and recall for taggers (61.57 and 36.32 %, respectively) at the expense of a lower accuracy for the runner role. Even though pairwise interactions are considered, no individual cues nor knowledge about the number of taggers is used. The tagger's recall is specially hurt by this, because when the tagger is not actively chasing someone, the model does not have additional information to make a correct decision. They also use a proximity function to limit the extent to which the interactions are taken into account. While nearby interactions seem more important, we noticed that chasing also takes

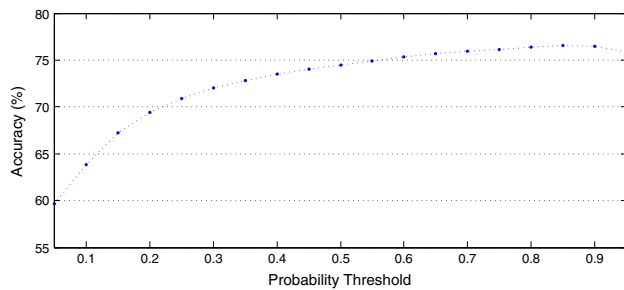


Fig. 15 Variation of the model's accuracy with respect to the tagger probability threshold

place at a greater distance. An additional proximity term in our likelihood function (Eq. 2) led to a lower accuracy. This also follows from the analyses of the Play corpus described in Sect. 3.

Currently, our model assumes that there is a single tagger. If we would drop this assumption, in line with [28], we could identify all players with values for Eq. 1 over 50 % as taggers. This would lead to a 74.47 % accuracy. It should be noted that the recall for taggers is lower (63.68 %), which is due to the bias towards the runner role. Instead of having the threshold at 50 %, we evaluated a range of thresholds. Figure 15 shows that the best accuracy is obtained with a threshold of 85 %. This means that a player is only classified as tagger if the probability of being a tagger is at least 85 %. Obviously, the recall for taggers is even lower in this case. This comparison shows that the knowledge that there is a single tagger in the playground helps in the classification of the players. Notably, the recall for taggers is much higher.

7 Conclusions and future work

We have introduced a novel probabilistic model for the recognition of player roles in tag games. The model is derived from the behavior analysis of children's tag game sessions in the Play corpus. In this dataset, up to eight children played tag in an open space, the traditional setting for this game. We obtained the positions of each of the players and manually annotated their roles: tagger or runner. The analysis of this dataset shows that the location of players in the playground, and the orientation in which they move relative to other players, carries discriminating information about their roles. These two features are used to define the boundary response and tagging intention functions, which are used to formulate our probabilistic model.

The model is tested on the iTag corpus which consists of over one hour of tag game sessions of young adults in the Interactive Tag Playground (ITP). The ITP is an interactive installation that tracks players using an online multi-modal tracker that combines Kinect depth cameras and gyroscope information. A ceiling-mounted projector displays colored

circles around the feet of the players. The colors differed for the tagger and runner roles. Instead of relying on physical touch, a tag occurred when the circle of the tagger overlapped that of a runner. The tag sessions in the iTag corpus consisted of three players playing tag for around four minutes. Overall, our model is able to recognize roles with a 77.96 % accuracy, over a baseline of 55.56 %. In line with the higher baselines, we achieved higher precision and recall values for runners. These results were stable across game sessions.

Closer analysis revealed that misclassifications were more often made around tags. The recognition of the tagger showed a delay, which is mainly due to the less pronounced behavior around a tag. The players have to adapt to their new roles, and quickly change their direction of movement. We envision an extension of the model which takes into account the different phases in tag games, e.g. chasing, tagging and reversing roles. Different behavioral cues could be used for each of these phases. The current model considers only two discriminant cues found from the corpus analysis, but more complex cues could be found after a thorough analysis.

The recognition of roles in a playground that assigns these roles seems redundant. However, we are interested in knowing when the observed gameplay deviates from the model. We expect that these moments occur when the estimated roles are different from the assigned roles. In these cases, the players might be tired or bored. Also, players might be cheating or suffer from behavioral disorders. If the role recognition is used online, we could compare the role information provided by the ITP with the behavior exhibited by players during the game. For instance, the ITP can calculate the time a player has been a tagger and aid him if it seems he is having problems tagging someone [37]. If, besides this, the player's behavior is not the expected one as recognized by our model (e.g. actively trying to tag other players), the ITP could change the interactions to restore engagement or notify caretakers of the issue.

We are very interested to see how the model would perform under different settings, such as with different numbers of players and multiple taggers. It would also be interesting to test whether the proposed recognition approach could prove valuable for the analysis of other social interactions in which patterns based on proximity, location and movement between people play a role. For instance, in other game installations such as the one presented in [20], the model could be used to detect social interactions between opponent players to evaluate additional player characteristics. For now, we would like to test our model with children playing in the ITP. To do this, however, the problem of having a limited playing area needs to be addressed, especially if we want to test with more varied player configurations. Additionally, the current model lacks any temporal considerations which could aid in the classification.

The evaluation in the ITP can be regarded as a first investigation into interactive tag. When moving from traditional

playgrounds to interactive spaces, many opportunities for more engaging play arise, not only for tag. We plan to enhance the game experience for different types of children games. For now, we have explored the possibility of enhancing one of such games: tag.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. *ACM Comput Surv* 43(3):A16
- Alahi A, Ramanathan V, Fei-Fei L (2014) Socially-aware large-scale crowd forecasting. In: *Proceedings of conference on computer vision and pattern recognition*, pp 2211–2218
- Argyle M, Dean J (1965) Eye contact, distance, and affiliation. *Sociometry* 28(3):289–304
- Barnett LA (1990) Developmental benefits of play for children. *J Leisure Res* 22(2):138–153
- Bazzani L, Cristani M, Murino V (2012) Decentralized particle filter for joint individual-group tracking. In: *Proceedings of conference on computer vision and pattern recognition*, pp 1886–1893
- Bazzani L, Cristani M, Paggetti G, Tosato D, Menegaz G, Murino V (2012) Video analytics for business intelligence, chap. *Analyzing groups: a social signaling perspective*. Springer, New York, pp 271–305
- Bazzani L, Cristani M, Tosato D, Farenzena M, Paggetti G, Menegaz G, Murino V (2012) Social interactions by visual focus of attention in a three-dimensional environment. *Expert Syst*
- Beetz M, von Hoyningen-Huene N, Kirchlechner B, Gedikli S, Siles F, Durus M, Lames M (2009) ASpoGAMo: automated sports game analysis models. *Int J Computer Sci Sport* 8(1):1–21
- Bialkowski A, Lucey P, Carr P, Denman S, Matthews I, Sridharan S (2013) Recognising team activities from noisy data. In: *Proceedings of conference on computer vision and pattern recognition workshops*, pp 984–990
- Bousmalis K, Mehu M, Pantic M (2013) Towards the automatic detection of spontaneous agreement and disagreement based on non-verbal behaviour: A survey of related cues, databases, and tools. *Image Vision Comput* 31(2):203–221
- Chang MC, Krahnstoeber N, Ge W (2011) Probabilistic group-level motion analysis and scenario recognition. In: *Proceedings of ICCV*, pp 747–754
- Chen X, Qin Z, An L, Bhanu B (2014) An online learned elementary grouping model for multi-target tracking. In: *Proceedings of conference on computer vision and pattern recognition (to appear)*
- Cheng Z, Qin L, Huang Q, Jiang S, Tian Q (2010) Group activity recognition by gaussian processes estimation. In: *Proceedings of ICPR*, pp 3228–3231
- Choi W, Savarese S (2014) Understanding collective activities of people from videos. *Pattern Anal Mach Intell* 36(6):1242–1257
- Choi W, Shahid K, Savarese S (2011) Learning context for collective activity recognition. In: *Proceedings of conference on computer vision and pattern recognition*, pp 3273–3280
- Cristani M, Raghavendra R, Del Bue A, Murino V (2013) Human behavior analysis in video surveillance: a social signal processing perspective. *Neurocomputing* 100:86–97
- Enzweiler M, Gavrilu D (2009) Monocular pedestrian detection: survey and experiments. *Pattern Anal Mach Intell* 31(12):2179–2195
- Ge W, Collins R, Ruback R (2012) Vision-based analysis of small groups in pedestrian crowds. *Pattern Anal Mach Intell* 34(5):1003–1016
- Hughes F (2010) *Children, play, and development*. Sage
- Jensen MM, Rasmussen MK, Mueller F, Grønbaek K (2015) Designing training games for soccer. *Interactions* 22(2):36–39
- Khokhar S, Saleemi I, Shah M (2013) Multi-agent event recognition by preservation of spatiotemporal relationships between probabilistic models. *Image Vision Comput* 31(9):603–615
- Kim K, Grundmann M, Shamir A, Matthews I, Hodgins J, Essa I (2010) Motion fields to predict play evolution in dynamic sport scenes. In: *Proceedings of conference on computer vision and pattern recognition*, pp 840–847
- Lan T, Sigal L, Mori G (2012) Social roles in hierarchical models for human activity recognition. In: *Proceedings of conference on computer vision and pattern recognition*, pp 1354–1361
- Liu J, Carr P, Collins RT, Liu Y (2013) Tracking sports players with context-conditioned motion models. In: *Proceedings of conference on computer vision and pattern recognition*, pp 1830–1837
- Lucey P, Bialkowski A, Carr P, Morgan S, Matthews I, Sheikh Y (2013) Representing and discovering adversarial team behaviors using player roles. In: *Proceedings of conference on computer vision and pattern recognition*, pp 2706–2713
- Lucey P, Oliver D, Carr P, Roth J, Matthews I (2013) Assessing team strategy using spatiotemporal data. In: *Proceedings of the international conference on knowledge discovery and data mining*, pp 1366–1374
- Moreno A, van Delden R, Poppe R, Reidsma D, Heylen D (2015) Augmenting traditional playground games to enhance game experience. In: *Proceedings of the international conference on intelligent technologies for interactive entertainment*
- Moreno A, Poppe R (2013) “You’re It!”: Role identification using pairwise interactions in tag games. In: *Proceedings of conference on computer vision and pattern recognition workshops*, pp 657–662
- Moreno A, van Delden R, Poppe R, Reidsma D (2013) Socially aware interactive playgrounds. *IEEE Pervasive Comput* 12(3):40–47
- Ni B, Yan S, Kassim A (2009) Recognizing human group activities with localized causalities. In: *Proceedings of conference on computer vision and pattern recognition*, pp 1470–1477
- Pellegrini AD (2009) *The role of play in human development*. Oxford University Press, Oxford
- Pellegrini S, Ess A, Van Gool L (2010) Improving data association by joint modeling of pedestrian trajectories and groupings. In: *Proceedings of ECCV*, pp 452–465
- Poppe R (2010) A survey on vision-based human action recognition. *Image Vision Comput* 28(6):976–990
- Poppe R, van Delden R, Moreno A, Reidsma D (2014) *Playful user interfaces, chap. Interactive playgrounds for children*. Springer, New York, pp 99–118
- Rehg JM, Abowd GD, Rozga A, Romero M, Clements MA, Sclaroff S, Essa I, Ousley OY, Li Y, Kim C, Rao H, Kim JC, Presti LL, Zhang J, Lantsman D, Bidwell J, Ye Z (2013) Decoding children’s social behavior. In: *Proceedings of conference on computer vision and pattern recognition*, pp 3414–3421
- Tran K, Gala A, Kakadiaris I, Shah S (2014) Activity analysis in crowded environments using social cues for group discovery and human interaction modeling. *Pattern Recognit Lett* 44:49–57
- van Delden R, Moreno A, Reidsma D, Poppe R, Heylen D (2014) Steering gameplay behavior in the interactive tag playground. In: *Proceedings of European conference on ambient intelligence*, pp 145–157

38. Wang P, Abowd GD, Rehg JM (2009) Quasi-periodic event analysis for social game retrieval. In: Proceedings of ICCV, pp 112–119
39. Yamaguchi K, Berg AC, Ortiz LE, Berg TL (2011) Who are you with and where are you going? In: Proceedings of conference on computer vision and pattern recognition, pp 1345–1352
40. Zhu Y, Nayak NM, Roy-Chowdhury AK (2013) Context-aware modeling and recognition of activities in video. In: Proceedings of conference on computer vision and pattern recognition, pp 2491–2498