Augmented Reality for Sports Spectating and coaching

Authors: Stefanie Zollmann, Tobias Langlotz, Holger Regenbrecht, Chris Button, Wei Hong Lo, Steven Mills

Abstract

Sport spectating and training has changed substantially in recent years. Nowadays match and training-related data are getting captured in huge quantities and qualities. With this increase of available data, there is also a strong need for novel user interfaces and new visualization techniques to present meaningful information. Broadcast media and online content are often used for remote spectating. However, they are not well integrated into the actual events during the sports performance, e.g. a live game. The same is the case for video analysis software that is often used for retrospective analyses of training sessions. Augmented Reality is an interface that focuses on in-situ visualization and comes with the advantage that it integrates the content directly into the field of view of the observer (spectator, coach, manager) which could be beneficial for live events and situations. In this chapter, we will discuss the potential of Augmented Reality for sports spectating and coaching.

Keywords: Augmented Reality, Sports Visualization, Sports Broadcasting, Situated Visualization, Spectators

1. INTRODUCTION

Spectator sports like football, rugby, and cricket have significantly changed the way people watch the games. Fifty years ago, the main option would have been to go to a sports field or stadium to watch the games live. Some more ambitious spectators would have brought their transistor radios to follow sports commentators while watching the game in parallel. In the last decades, many sports have become highly professionalized, stadium capacities were quickly reached, costs for watching games on site increased, media technologies developed, and more and more people are now watching sports events in front of a television set or on portable devices. On one hand, this allowed for different and closer viewing angles, and for the provision of additional, overlaid information display. On the other hand, "couch participation" lacks the atmosphere of a real stadium environment.

In 2020, many fans were excluded from attending live sports and forced into virtual observation because of the global coronavirus pandemic. Interestingly, the hiatus in stadium spectatorship has required sports broadcasters to rethink how they can synthesize the typical experience – e.g., by adding crowd noise - with varying levels of success (Majumdar & Naha, 2020). Many sports clubs and their loyal fan bases have lamented the absence of fans at live sport and the lack of emotion associated with 'crowd-less' broadcasted matches[1]. These recent experiences have clearly illustrated the widening gap that has appeared between the experiences of watching live sport at a venue and watching broadcasted coverage. It would appear that a new age of fandom is rapidly emerging and that sports and broadcasters alike need to seriously consider how best to attract and engage spectators in the future.

In addition, advancements in sports broadcasting and sports media in data capture have allowed broadcasters to retrieve and analyze sports-related information in real-time or near real-time. This information can include team statistics such as ball possession or scores and increasingly include player specific data such as running distances, fouls, and (more recently) heart rate and other physiological readings.

[1] Premier League's home edge has gone in pandemic era: The impact of fan-less games in England and Europe (espn.com)

Also, the access to slow-motion replays and digital animations is becoming more prevalent. Spectators want commentators to review important refereeing decisions or crucial moments in a game from numerous angles and perspectives. This information is processed and visually presented in what has become a key component of sports broadcasting. These developments are ongoing and the next evolution of sports broadcasting is already emerging— Interactive sports broadcasting. This evolution is driven by advances in sensing technology and by the emerging paradigm of Visual Computing which combines Image Processing, Computer Vision, Visualization, Human-Computer-Interaction and Computer Graphics. Thanks to these advances spectators can nowadays experience a soccer game from the perspective of a player (Rematas, Kemelmacher-Shlizerman, Curless, & Seitz, 2018), listen to live dialogue between a rugby referee and players or take a virtual seat in an America's Cup boat without leaving our homes[2]. However, spectators at live sporting events often miss out on this augmented live information. It is only available to remote viewers through broadcast media. This might affect the attractiveness of attending live sports events.

Figure 1 AR interfaces for sports spectating. Left) AR interface on mobile phone using video-see-through AR. Right) User with optical-see-through AR interface (HoloLens) used for sports spectating.

In this chapter, we will discuss the potential and the challenges of using Augmented Reality (AR) as an interface to access sports-related content on-site for sports spectating and coaching. We will discuss how AR as technology can improve the experience for live sports spectators as well as different kinds of sports spectators

[2] https://arl.co.nz/arl-news/221-on-board-the-america-s-cup

(e.g., fans, coaches, officials) could benefit from AR. AR has the potential to augment spectators' views with interactive real-time information and to bridge the widening gap between experiencing live sports events and remote interactive sports broadcasting. The development of mobile AR techniques that precisely overlay digital information onto the on-site spectator's or observer's view could help to increase the attractiveness of attending future live sporting events. AR has the potential to connect producers of sports statistics and event-related information with new audiences (i.e. spectators who currently choose to stay at home go to events and get sports statistics and event-related information and the live experience; coaches, managers, and trainers can view training sessions augmented with real-time and historic and statistical data while observing the performance of the athletes).

To develop the potential of AR as an interface to sport event-related data, it is important to understand and investigate novel visualization techniques for AR and vision-assisted tracking techniques. In this chapter, we will discuss Visual Computing techniques required to precisely overlay real-time sport-related information onto the on-field action. Vision-based tracking techniques in combination with adaptive visualization methods and local broadcasting of event data to the mobile devices of spectators have the potential to place rich information about sporting events in context for on-site spectators based on their individual location within the venue. In addition to solving computer vision and visualization challenges, there is also a need for a flexible infrastructure for on-site broadcasting of relevant sport-related information to spectators' mobile devices for an on-site mobile AR experience. Through this infrastructure, spectators at live events will be able to receive real-time, sport-related information as well as experience the atmosphere and drama of live sport as it unfolds.

Such an infrastructure will provide additional broadcasting outlets for digital content providers and thus will not only have an impact on spectator experience but also on the content creation industry. AR also has the potential to contribute to a more attractive proposition for sports venues and sports teams. Sports teams significantly benefit from increased numbers of spectators not just financially but also in terms of performance and building community and loyalty (Ludvigsen & Veerasawmy, 2010; Madrigal, 2006; Russell, 1983). This has been vividly demonstrated as sports were disrupted by the coronavirus pandemic and had to consider alternate means to engage their spectators. There are over 5000 major stadiums in the world[3], 300 of which are football (soccer) stadiums with seating capacities exceeding 40,000 each[4] creating a potential spectator base of 17 million users just for football (soccer) alone. When rugby, cricket, track-and-field or spill-over effects to other sports and non-sports-related events – the potential reach of sports stadia easily balloons out to 100 million or more spectators. All these spectators are paying customers and even if only a fraction of the ticket prices can be targeted this is an attractive market for content providers, venue managers, and potential mobile app developers and distributors. While in this chapter we mainly focus on professional (elite) sports events that take place inside sports stadia, there is also potential for such technology for recreational level team sports and beyond into other entertainment domains.

We will first discuss technical requirements and challenges to provide a solid foundation about what is required to create AR applications for on-site spectators based on their location within the venue. We will then discuss options for integrating these AR applications into a flexible infrastructure for on-site broadcasting that aggregates information from different sources and sends it to the spectators' mobile devices. Here we will apply novel techniques for precisely tracking mobile devices within the venue allowing us to place the aggregated information in the spectators' view. Spatial filters and contextual knowledge will help to preserve on-field action while complementing it with relevant information. The outcome of this research will be a new experience for the spectator and a more

[3] http://www.worldstadiums.com/
[4] https://en.wikipedia.org/wiki/List_of_association_football_stadiums_by_capacity.

attractive proposition for the venue. Immediate beneficiaries of this research will be sports spectators and participants, digital content providers and both indoor and outdoor sports venue providers.

2. Augmented REality

Augmented Reality (AR) interfaces extend our view of the real-world environment by visually integrating digital information. With this, we can display additional information that is not physically present. For instance, this digital information can represent non-existing objects, meta-information or hidden information.

In 1997 Azuma identified three major characteristics that describe an AR interface:

1. *Combines real and virtual*
2. *Interactive in real-time*
3. *Registered in 3-D" (Azuma, 1997)*

Back at the time when Azuma formulated these requirements, it was still computationally expensive to implement these characteristics within one system. Expensive and powerful computing devices were required to combine them in one system. This limitation prevented the ubiquitous deployment of AR for conventional users due to high cost, bulky equipment, and limited availability. In recent years, with the increasing computational power of even small devices, omnipresent hardware, such as mobile phones and tablet computers, has grown powerful enough to fulfill Azuma's requirements. These further developments have worked towards a ubiquitous experience of the mixture of physical and virtual information and opened new fields of application, such as entertainment or advertisement, but also various professional applications.

While for decades AR was mainly driven by academic research, recently major investments from the industry have increased the availability of AR toolkits and frameworks. This allows developers with no prior AR expertise to develop AR applications. The reduction of entry requirements to AR development has contributed to many new AR applications that have previously only been described in research work. Examples include AR applications for furniture shopping (e.g.

IKEA Place), games combining virtual and digital elements (e.g. Pokémon Go or Lego), or even AR-supported measuring tools. Most of the applications focus on small-scale environments, usually indoor locations. However, there are a lot of application areas beyond these physical environments that would highly benefit from AR. For example, outdoor trampers and climbers are increasingly using AR topographical mapping technology to help them preview and plan their routes (Wiehr, Daiber, Kosmalla, & Krüger, 2017).

In 2016/17, Pokémon Go became a worldwide phenomenon when it was taken up by millions of users, illustrating the potential of AR applications to be used alongside mobile devices. Mobile AR today has the potential to change users' experience and appreciation of sporting events in the same way as adding computer graphics to television broadcasting did in the '90s. Instead of augmenting the spectator's living room, AR has the potential to bring digital information directly to a spectator at a sports event, enriching the quality and attractiveness of those events. This will foster new forms of content provision and delivery, leading to further innovation and growth.

Pokémon Go's success was based on a novel on-site experience and on the economy of scale of free-of-charge offerings. The success of AR for sports spectating and training will also be based on the quality of the experience, but also on the quality of the content delivery for paying customers. Spectators at sports events are willing to spend a major amount on tickets and are expecting a quality experience for their money. Super Bowl spectators in the US are already paying US$2,500 on average per ticket. The quality of the experience has to match these increasing ticket prices.

Content providers for added-value sports broadcasting experiences are seeking new and innovative forms of delivery. AR as an interface for spectators will give them the opportunity to offer new forms of end-user experiences. They will be able to deliver these directly to end-users, making them less dependent on established TV and internet broadcasters. Multi-national corporate companies like Intel want to disrupt the market by offering immersive virtual reality (VR) sports experiences for the living room (e.g. Project Alloy) requiring billion-dollar investments in capturing the live events. However, "You need to experience sports events live and they just aren't the same when you're not in the stadium" says Intel's CEO Brian Krzanich[5] indicating that VR in the living room alone might not completely replace stadium attendance.

3. Technical requirements

Augmented Reality (AR) comes with a set of technical requirements for allowing it to be embedded into a user's view. These requirements are closely related to the characteristics of AR interfaces set by Azuma (Azuma, 1997).

In this section, we discuss those in particular 1) tracking and registration methods, 2) how to combine virtual and real information as well as 3) interaction techniques in the context of sports spectating and coaching.

A. Registration

In order to visualize any digital content in an AR interface, the first prerequisite is having suitable registration and tracking techniques available. Registration in the context of AR describes how virtual objects are aligned to the real world and how to assure that the digital data is correctly placed in relation to real-world objects. Often, the registration can consist of a localization approach coupled with tracking technology. The localization approach thereby computes an initial spatial relationship between the user's device and the real world. Tracking technology then supports the process of continually estimating and tracking this spatial relationship when the user moves away from a known position.

In AR, there are different options to achieve a correct registration, varying from simple maker-based registration techniques (Kato & Billinghurst, 1999) to tracking methods that use natural-features (Wagner, Reitmayr, Mulloni, Drummond, & Schmalstieg, 2008) sensor fusion (Schall et al.,

[5] https://www.si.com/edge/2017/01/11/future-virtual-reality-merged-sports-intel-ces-2017.

2009) to localization-based approaches (Ventura & Hollerer, 2012). All these technologies use different approaches to achieve the same goal; aligning the virtual data in relationship to the physical world for a coherent AR visualization.

Marker-based and natural-feature-target based techniques are often used for indoor AR applications. For outdoor applications, they are usually not an option due to larger working environments and environmental influences. To achieve a reliable registration in unknown outdoor environments, often more sophisticated sensors are integrated into the AR setup or sensor-fusion approaches are used (Schall, Zollmann, & Reitmayr, 2013).

However, when it comes to sports spectating, one of the main challenges is that the area where we want to place content is quite large and dynamic. This creates a major challenge for a lot of the traditional registration techniques as well as the ones that target large open outdoor areas. For computing and placing digital overlays in the spectator's field of view, we need to compute the spatial relationship between the spectator's view and the event site and often also with regards to the digital content. Suitable AR interfaces are mobile phones or AR glasses (e.g. MS HoloLens[6]). To track the view we must solve two challenges: 1) The localization challenge in which we need to compute an accurate pose describing the position and rotation of the AR interface within the physical stadium. 2) Once this pose information is determined, we need to keep track of the movements of the AR interface with respect to the stadium. We call this the tracking challenge. Solving both, localization and tracking, allows us to register digital overlays in the spectator's view using the AR clients.

i. Localization

[6] HoloLens: https://www.microsoft.com/en-us/hololens

The main goal of the localization step is to compute the position and orientation of the AR interface (mobile phone or AR glasses) with regard to 3D models of the playfield or the venue. In our previous work, we investigated different methods for localizing a user within such an environment (Zollmann, Langlotz, Loos, Lo, & Baker, 2019). Options vary from user-guided methods to automatic localization methods.

User-guided Localisation

A user-guided localization method comes with the advantage that the user is in control of the registration technique and can provide immediate feedback. In addition, such methods have the advantage that they work independently from venues and do not require any additional dataset capture of the environment. However, they come with an additional workload for the user as the user needs to make sure they aligned the model accurately with the environment. During a lot of sports events, spectators are seated and have an allocated seat number, similarly, coaches or sports officials might select a predefined area with a good view over the playfield. If a seat number or a specific location is known beforehand, this can serve to provide a rough estimate of the user's position. Using this estimate we can provide an AR overlay of the sport venue which can be refined by the user by interactively aligning the overlay and sports ground (Figure 1, Left). This approach requires a spatial mapping of all seats in the stadium which is also not always easy (e.g. not always numbered seats). If such a spatial mapping is not available another option is to use a traditional perspective-n-point solution that requires the user to align known 3D marks and 2D points in the AR client's view to compute the pose of the spectator's device. While both approaches produce suitable results, the usability is affected by the field of view of the AR client.

For instance, usually not all the corners of the sports field are visible to users, while other markings such as lines or advertisements are not always reliable (being badly visible or sometimes only roughly marked).

Besides that, sensor data of the mobile device can be used to specify a location on a map. The user is then required to align their view of the event site to the mapping interface. The success rate of this method depends heavily on the accuracy of the device's sensor data. Furthermore, GPS data is not always reliable in sports venues with large stands and roof structures. The disadvantage of the user-guided methods is that they put a major task load onto the user, and their ability to perform an accurate alignment will then, later on, have an impact on the user experience.

Figure 2: Localization methods for AR interface. Left) User-guided localization requiring the user to input seat location and provide alignment to a digital pitch map. Right) Automatic model-based localization using a 3D model of the stadium.

Automatic Localization

In many cases, sensor-based localization (e.g. GPS) does not deliver the accuracy required as urban structures may affect the satellite-based localization. Another option is through localization that uses a computer vision method. This approach uses known, fixed features from the environment and thus potentially has greater accuracy. An option for using such image features is given by the fact that a lot of sports venues display advertisements. These come with the advantage that they are very distinguishable from the rest of the environment and that they can be used for automatic recognition and pose estimation. So-called natural feature tracking targets use an approach where the device camera extracts image features and uses them for computing positioning information.

Image features often represent 2D image points that are easily recognizable for computer vision algorithms such as corners or other prominent image regions. Popular choices for feature detection are the Scale-invariant Feature Transform (SIFT) or the more efficient Features from Accelerated Segment Test (FAST) (Rosten, Porter, & Drummond, 2010)). We then perform pose estimation by using those detected features and their counterparts in a reference image along with the known dimensions of the advertisements. AR Software Development Kits such as Vuforia[7] support natural feature tracking targets and make it easy to manage multiple targets at the same time. In our previous work, we explored these options and obtained promising results for printed targets that are placed on the playfield. It is important to note that this approach strongly depends on the type and size of the advertisement and the position within the stadium and performance will be different if the used image target is far away or contains bad features according to SIFT or similar approaches.

For a lot of professional sports venues, there are 3D models of the environment available. In these cases, model-based approaches that use a 3D model of the environment for computing the spatial relationship between the device and playfield are an option (Figure 2, Right). In our previous work, we investigated such methods with regards to their suitability for usage in AR (Baker, Zollmann, Mills, & Langlotz, 2019). The initial findings showed that these methods are location-dependent. Often, we find repetitive structures as well as dynamic elements that create difficulties to these model-based approaches. Alternatively, for several types of sports, it is possible to make use of the line-markings as a reference. This has been demonstrated for static broadcast cameras before and has also been used within AR interfaces (Skinner & Zollmann, 2019).

ii. Tracking

The localization methods described in the previous section allow for an initial global alignment between the user's device and the real environment. After initialization, tracking methods allow for a continuous update of this relationship in order to

[7] https://developer.vuforia.com

compute the camera pose required for placing the AR content. Simultaneous Localization and Mapping (SLAM) approaches are often used as tracking technology in AR. SLAM is a tracking algorithm that has its origin in robotics and simultaneously computes a map of the environment while at the same time using it for localizing a device or camera (Davison & Murray, 2002). Within the mapping part of the algorithm often a 3D map is computed from 2D image features such as SIFT. These image features are identified and tracked in consecutive frames seen by a camera. Their 2D location in the image is then used to compute 3D rays and to compute the 3D position of those 2D image features using triangulation. The 3D position information is stored in a map that is then used for computing the positioning (and orientation) information of the device or camera. While traditional robotic SLAM approaches use stereo cameras or additional sensors (e.g. laser), for mobile AR applications monocular approaches have been developed to support single camera devices.

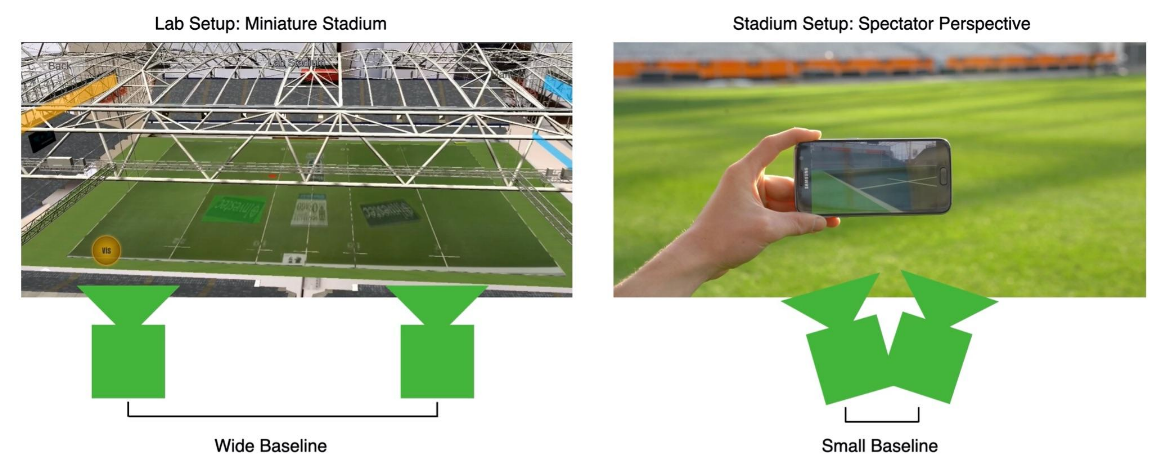

Figure 3: Wide Baseline compared to Small Baseline for computing 3D information using Structure-from-Motion (SfM). Left) In a lab environment, it is easy to fulfill the requirement of using a wide baseline for localization as the user performs large translation movements in relation to the 3D stadium model. Right) In an actual stadium environment, spectators often maintain a fixed position and perform only rotational movements which are not ideal for SfM due to the small baseline.

Modern SLAM approaches have shown promising results in larger-scale environments (Mur-Artal & Tardos, 2017) and proprietary AR SDKs such as ARKit[8]

[8] ARKit: https://developer.apple.com/augmented-reality/

and ARCore[9] combine the data from motion sensors and SLAM approaches in Visual Inertial Odometry approaches for more robust AR tracking. We investigated some of these existing Visual Inertial Odometry approaches for their feasibility for AR in larger sports venues (Zollmann et al., 2019). Our initial tests turned out to work reliably enough for developing early prototypes and to demonstrate AR to sports spectators. However, we also noticed that the specific movement patterns for sports spectators (static position with mainly rotational movements) and location (large open environments with only minimal parallax) are conflicting with some assumptions traditionally made for SLAM trackers. In particular, the requirement of having a wide baseline between video frames for computing 3D information from 2D images via Structure from Motion (SfM) poses a challenge (Figure 3). A wide baseline allows triangulating image rays of known image features from two camera images (Figure 3, Left). However, mostly stationary users within a large environment such as a stadium create the problem of not having enough translational movement for a large enough baseline (Figure 3, Right). In order to address these challenges, we investigated the suitability of spherical SfM. For this purpose, we applied a spherical SfM method that computes the absolute pose based on a spherical movement constraint assuming all device motion can be represented by movements on a sphere. These approaches showed improvements when comparing them to traditional SLAM approaches. within the use case of a sports venue (Baker, Ventura, Zollmann, Mills, & Langlotz, 2020).

B. Visualization

Once we computed the exact camera position and viewing direction (camera pose) of the user or the device, we can use AR visualization techniques to overlay graphical representations of event-specific information into the field of view of the user. For a video see-through implementation using a mobile phone, a simple overlay of such information can be implemented by rendering a 3D model or annotation on top of the video image captured by the device's camera. For optical see-through displays such as the HoloLens, a 3D model would be displayed by rendering the 3D representation on the display. The combination of 3D content and view of the actual

[9] ARCore: https://developers.google.com/ar

environment would then happen in the optical combiner of the device. For both, video see-through and optical see-through, the camera pose information is used to correctly align the 3D data with the view of the user's environment.

While such simple AR overlays are often used in various AR prototypes, they are often subject to perceptual issues (Zollmann, et al., 2020). In particular, Augmented reality information visualization for complex environments such as sports venues is challenging since the environment, as well as the presented content is constantly changing. While content presentation in professional sports broadcasting relies on extensive manual editing to create a consistent visual output, this is not viable for on-site AR applications. There is often no control over the user's viewing perspectives, spectators and coaches will observe the player's action from various viewpoints and locations. Digital content placed on top of real-world objects might interfere with those real-world objects as well as with other digital elements. This is problematic as it is difficult to control the real environment where content is to be placed. In addition, the devices' screen sizes are often smaller compared to the devices that spectators use for consuming broadcast, e.g. larger TV screens.

These challenges can lead to cluttered presentations or occlusion of important parts of the view (e.g. placing an information label on top of a player). In order to address cluttered presentations in AR, research has been conducted previously on how image analysis and filtering can support automatic content placement and view management. For instance, image analysis techniques such as edge extraction and saliency computation have been used for placing content in AR application in a way to avoid overlapping items and occlusions as well as misalignment of digital items and real-world objects (Rosten, Reitmayr, & Drummond, 2005; Sandor, Cunningham, Dey, & Mattila, 2010; Zollmann, Grasset, Reitmayr, & Langlotz, 2014).
An option for addressing the challenges of small screen sizes is to make optimal use of the environment surrounding the user by using situated visualization techniques that analyze the user's environment for suitable content placement areas (Langlotz, Nguyen, Schmalstieg, & Grasset, 2014).

The visualization strongly depends on the type of input. In order to present sports event-related data to the user in an AR interface, different data sources have to be combined and transferred into a common 3D coordinate system. It is important that this coordinate system allows for putting the data in reference to the localization and tracking approaches. This is particularly a challenge as data sources can include 2D and 3D models (e.g. stadium models or static line overlays), information from player tracking that allows showing player paths or labels with information about a specific player. In addition, sports event-related information can also be provided by commercial sports databases sometimes referring to 2D or 3D coordinates, but sometimes also being related to a specific event or a team. In this case, it is important to define a spatial component relevant to the specific data. In our previous work, we demonstrated a system that integrates these different data sources for this purpose (Zollmann et al., 2019). In order to do so, we identified three main content categories for visualization in AR, these include **a) Player-based, b) Team-based and c) Game-based**. Examples for player-related data would be player names (Figure 4, Right) as well as physiological data such as heart rates. Team-based visualizations include visualization of data that corresponds to a specific team, such as ball possessions or field events. Game-based visualization includes the visualization of data that is relevant to the overall game or training event showing game rules, instructions, or hints such as offside that are not specific to a team or a player. In addition to these previous categories, an AR interface can also be used for visualizing crowd-based data. Crowd-based visualization could help to make the event more collaborative and share information between users on-site. This can include interactive games or entertainments for the crowd during breaks or even for emergency situations (e.g. evacuations) (Lo, Zollmann, Regenbrecht, & Loos, 2019).

Figure 4: Visualization of different types of data in AR interface for sports spectating for Rugby. Left and Middle) Game-based information such as tackles visualized as an overlay onto the playfield. Right) Player-based information such as labels rendered as AR overlay.

In addition, data can be defined as spatially anchored. This means the data to be presented already comes with a spatial coordinate that allows the AR interface to put it into the correct spatial location. Examples would be 3D models of game-related items, such as goalposts, or 2D representations of game-related objects such as line-markings. For training purposes, this could also include the rendering or a 3D crowd to create a more live event-like setting for the players during a training session.

One particular challenge relevant to using AR as an interface for sports spectating and training is that some information traditionally delivered for TV broadcasting comes in close to real-time (e.g. with a latency of seconds). This is no problem for TV production as there will be a small delay when sending the content, but the combination of information and footage is in sync. However, for AR, real-time overlays are required as the user is directly presented with their view of the environment. An offset between the action on the field and the presented data will be directly visible. To compensate for this and to adjust for delays in data delivery that are unavoidable, indirect AR can be a suitable solution. Indirect AR is an alternative implementation of an AR interface that uses a previously captured representation of the environment such as a panorama (Wither, Tsai, & Azuma, 2011). In contrast to direct AR where the overlay is shown on top of the live camera feed of the device, within the indirect AR interface, digital information can be overlaid on top of either a panoramic image, panoramic video, or a rendering of the scene. The advantage of this alternative interface is that one is able to replay a scene (thus relaxing the latency issue for some of the content) as we are not relying on the live camera feed while still keeping the immersiveness of the AR interface as the device is still fully tracked and responding to the users' movements(Zollmann et al., 2019). While this form of AR is only of limited value for replacing live augmentation of an entire event, it can be used for replays or the visualization of time-critical content.

C. Interaction

Once the localization, tracking, and visualization requirements are addressed, spectators can access event-related data on-site during a game or a training session. However, often there is a lot of different content to access, so one of the remaining questions is how will spectators interact with this content? While mobile AR interfaces such as a video see-through AR app on mobile allow for using standard user interface elements, head-worn implementations require alternative ways of interacting with the content. In particular, gesture-based and speech-based input has gained a lot of interest in recent years and has been integrated into commercial devices such as the HoloLens. In feasibility tests with our early prototypes, we noticed some usage patterns of mobile AR applications for sports spectating. In particular, that the AR interface would not be used continuously but rather in situations when users want to access specific information. Holding up a mobile phone for the whole duration of a match seems to be unfeasible. This is in contrast to the usage of head-worn devices. While these devices are still too bulky for casual users, they have the potential of better and more seamless integration of the interfaces as there is no explicit need for the user to switch on the device to look up any information. While speech input seems to be unfeasible for usage during large sports events, gesture input could still be an option for live sports events.

4. Opportunities of AR for sports spectating and Coaching

Augmented Reality interfaces provide a set of benefits and potential for both sports spectators and to inform coaches for training. Sports spectators will benefit from the option of accessing data on-site in a similar way they are used to consuming content at home in front of the television. Sports spectators are already used to having overlays of graphics content on top of broadcast footage. Embedding a similar visual representation into their own perspective of the game on-site has the potential to deliver additional statistics and information that can be helpful for game understanding. This also has the potential to create more engagement and excitement during a sports event supporting fans but also sports teams. Instead of simply attending a match, AR provides a more interactive experience. Bringing digital content into the right spatial context has the potential to reduce the mental workload that is required when accessing additional data such as from traditional fan-focused mobile applications and web interfaces. Spectators can explore game-relevant statistics in the actual context of the action on the field. For instance, a user could tap on a player and a 3D label with player statics will appear next to the player, heat maps that visualize game relevant information will directly be overlaid on the pitch or team relevant statistics will appear on the site of each team. This will make it much easier to directly access relevant data.

Similarly, training staff and other professional personnel involved in sports team training and coaching will benefit from the ability to access data directly within the field of action. During a training session, a coach or manager could use AR as an interface to directly access additional stats about a player. Current session-related data such as events or historical data from previous matches could be displayed on top of the field action instead of using a 2D interface. If player tracking or ball tracking data is available this could be used to compare the performance of one player to another one. Players could use that to revise their own performance.

5. Conclusion

In this chapter, we discussed the requirements, challenges, and opportunities that arise from using AR as an interface for sports spectating and training. The main idea of AR as an interface to sports event-related data and content is to bring information to users on-site by overlaying it on top of their view of the actual environment. Examples are labels attached to players that display names and additional information such as physiological data or performances from previous events, or the visualization of game-related information such as heat maps or explanations relevant to the game development.

To use AR as an interface for sports spectating and training, several requirements need to be addressed such as tracking and localization, visualization, and interaction. In this chapter, we gave a brief introduction to Augmented Reality and discussed solutions for addressing the main requirements. We also discussed the opportunities that arise when using AR as an interface for sports spectating and coaching. However, it is important to note that there are several remaining challenges. For instance, the success of AR interfaces is tightly related to the improvement of display technology— in particular, when it comes to head-worn displays, there is a need for reducing the weight and improving the acceptability of wearing hardware and sensors. There is a need for higher accuracy for localization and tracking to make sure that displayed content appears in the correct position. This is tightly connected to the development of better and additional sensors as well as the improvement of tracking algorithms. Also, current requirements with regards to power consumption are often a challenge for long-term use of AR interfaces, as the requirements for tracking algorithms and rendering often need high processing power. Using this technology for a complete sports event is still a challenge. In addition, social and ethical aspects need to be considered carefully, e.g. to address privacy concerns around and issues around continuous use of AR interfaces and with large crowds. Addressing all these challenges to provide an AR-enhanced experience for sports spectators remains a challenge but clearly not an insurmountable one.

Acknowledgement

We would like to thank Animation Research Ltd and in particular Sir Ian Taylor and John Rendall for their valuable input and support. We also thank the Forsyth Barr Stadium, the Highlanders and Otago Rugby (ORFU), for the opportunity to do on-site testing with our AR prototypes, and OptaPerform for their support. In addition, we would like to thank Mike Denham and Craig Tidey from the School of Surveying at the University of Otago for their support in surveying the stadium. The work discussed in this chapter is part of the ARSpectator project supported by an MBIE Endeavour Smart Ideas grant in New Zealand.

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. Retrieved from http://nzdis.otago.ac.nz/projects/projects/berlin/repository/revisions/22/raw/trunk/Master's

Baker, L., Ventura, J., Zollmann, S., Mills, S., & Langlotz, T. (2020). SPLAT: Spherical Localization and Tracking in Large Spaces. In *IEEE Virtual Reality (IEEE VR)*.

Baker, L., Zollmann, S., Mills, S., & Langlotz, T. (2019). Softposit for Augmented Reality in Complex Environments: Limitations and Challenges. In *International Conference Image and Vision Computing New Zealand* (Vol. 2018-Novem). https://doi.org/10.1109/IVCNZ.2018.8634761

Davison, A. J., & Murray, D. W. (2002). Simultaneous localization and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 865–880. https://doi.org/10.1109/TPAMI.2002.1017615

Kato, H., & Billinghurst, M. (1999). Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System, 85. Retrieved from http://dl.acm.org/citation.cfm?id=857202.858134

Langlotz, T., Nguyen, T., Schmalstieg, D., & Grasset, R. (2014). Next-generation augmented reality browsers: Rich, seamless, and adaptive. *Proceedings of the IEEE*. https://doi.org/10.1109/JPROC.2013.2294255

Lo, W. H., Zollmann, S., Regenbrecht, H., & Loos, M. (2019). From lab to field: Demonstrating mixed reality prototypes for augmented sports experiences. In *Proceedings - VRCAI 2019: 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*. https://doi.org/10.1145/3359997.3365728

Ludvigsen, M., & Veerasawmy, R. (2010). Designing Technology for Active Spectator Experiences at Sporting Events. *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 96–103. https://doi.org/10.1145/1952222.1952243

Madrigal, R. (2006). Measuring the multidimensional nature of sporting event performance consumption. *Journal of Leisure Research*, *38*(3), 226–267.

https://doi.org/Article
Majumdar, B., & Naha, S. (2020). Live sport during the COVID-19 crisis: fans as creative broadcasters. *Sport in Society*. https://doi.org/10.1080/17430437.2020.1776972
Mur-Artal, R., & Tardos, J. D. (2017). ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, 1–8. https://doi.org/10.1109/TRO.2017.2705103
Rematas, K., Kemelmacher-Shlizerman, I., Curless, B., & Seitz, S. (2018). Soccer on Your Tabletop. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00498
Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: a machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(1), 105–119. https://doi.org/10.1109/TPAMI.2008.275
Rosten, E., Reitmayr, G., & Drummond, T. (2005). Real-time video annotations for augmented reality. *Advances in Visual Computing*. Retrieved from http://link.springer.com/chapter/10.1007/11595755_36
Russell, G. W. (1983). Crowd Size and Density in Relation To Athletic Aggression and Performance. *Social Behavior & Personality: An International Journal*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=8651254&lang=pt-br&site=ehost-live
Sandor, C., Cunningham, A., Dey, A., & Mattila, V. V. (2010). An Augmented Reality X-Ray system based on visual saliency. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)* (pp. 27–36). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5643547
Schall, G., Wagner, D., Reitmayr, G., Taichmann, E., Wieser, M., Schmalstieg, D., & Hofmann-Wellenhof, B. (2009). Global pose estimation using multi-sensor fusion for outdoor Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2009)* (pp. 153–162). IEEE. https://doi.org/10.1109/ISMAR.2009.5336489
Schall, G., Zollmann, S., & Reitmayr, G. (2013). Smart Vidente: Advances in Mobile Augmented Reality for Interactive Visualization of Underground Infrastructure. *Personal Ubiquitous Comput.*, *17*(7), 1533–1549. https://doi.org/10.1007/s00779-012-0599-x
Skinner, P., & Zollmann, S. (2019). Localisation for Augmented Reality at Sport Events. In *International Conference Image and Vision Computing New Zealand*. https://doi.org/10.1109/IVCNZ48456.2019.8961006
Ventura, J., & Hollerer, T. (2012). Wide-area scene mapping for mobile visual tracking. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 3–12). IEEE. https://doi.org/10.1109/ISMAR.2012.6402531
Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., & Schmalstieg, D. (2008). Pose tracking from natural features on mobile phones. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2008)* (pp. 125–134). IEEE. https://doi.org/10.1109/ISMAR.2008.4637338
Wiehr, F., Daiber, F., Kosmalla, F., & Krüger, A. (2017). ARTopos - Augmented reality terrain map visualization for collaborative route planning. In *UbiComp/ISWC 2017 - Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive*

and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. https://doi.org/10.1145/3123024.3124446
Wither, J., Tsai, Y.-T., & Azuma, R. (2011). Indirect augmented reality. *Computers & Graphics*, *35*(4), 810–822. https://doi.org/10.1016/j.cag.2011.04.010
Zollmann, S., Grasset, R., Reitmayr, G., & Langlotz, T. (2014). Image-based X-ray Visualization Techniques for Spatial Understanding in Outdoor Augmented Reality. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (pp. 194–203). New York, NY, USA: ACM. https://doi.org/10.1145/2686612.2686642
Zollmann, S., Langlotz, T., Loos, M., Lo, W. H., & Baker, L. (2019). Arspectator: Exploring augmented reality for sport events. In *SIGGRAPH Asia 2019 Technical Briefs, SA 2019*. https://doi.org/10.1145/3355088.3365162
Zollmann, S., Langlotz, T., Grasset, R., Lo, W. H., Mori S. and Regenbrecht H. (2020), "Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2986247.

Bios

Stefanie Zollmann is a Senior Lecturer at the University of Otago in New Zealand. Before, she worked at Animation Research Ltd on XR visualization and tracking technology for sports broadcasting and urban planning. She worked as a post-doctoral researcher at the Institute for Computer Graphics and Vision (Graz University of Technology) where she also obtained a PhD degree in 2013. Her main research interests are XR for sports and media, visualization techniques for augmented reality, but also include capturing for XR and immersive experiences.

Tobias Langlotz is an Associate Professor at the University of Otago. Tobias was previously a senior researcher at the Institute for Computer Graphics and Vision (Graz University of Technology, Austria) where he also obtained his PhD. Tobias' main research interest is Vision Augmentations and Computational Glasses utilizing AR technology, spontaneous interaction for wearable AR systems, and nomadic mobile telepresence solutions, where he works at the intersection of HCI, Computer Graphics, Computer Vision and Ubiquitous Computing.

Dr Holger Regenbrecht is a computer scientist and Professor at the Department of Information Science at the University of Otago, New Zealand. He received his doctoral degree Dr. Ing. from Bauhaus University Weimar, Germany. Holger has been working in the fields of virtual and mixed reality for more than 20 years. His work spans theory, concepts, techniques, technologies, and applications. Holger's research interests include human-computer interaction, applied computer science and information technology, presence and telepresence, ubiquitous, immersive, and collaborative augmented reality, three-dimensional user interfaces, psychological

aspects of virtual and mixed reality, computer-aided therapy, and VR learning and collaboration.

Professor Chris Button is the Dean of the School of Physical Education, Sport and Exercise Sciences. He gained his PhD in Motor Learning at Manchester Metropolitan University in 2000 and then worked at Edinburgh University before moving to Otago in 2003. The processes by which humans learn to move skilfully is the central theme for Chris' research. Chris is passionate about movement and he uses models from sport and physical activity to explain skill acquisition. His research outputs cover a broad range of topics including the assessment of movement competence in children, individual differences, and sports technology. The theory of ecological dynamics, which underpins Chris' research, is featured in several textbooks he has co-authored. Over the last ten years, Chris has been working closely with Water Safety New Zealand to improve the learning of water safety skills amongst New Zealanders.

Wei Hong Lo is currently a PhD student at the University of Otago, majoring in the field of Augmented Reality Visualization and Human-Computer Interaction (HCI). Before his PhD studies, he graduated with a BSc (Hons.) Software Engineering in the University of Nottingham Malaysia Campus, where he obtained an interest in the field of HCI and visualization. He then proceeded as a research assistant and was involved in projects such as developing interactive exhibits for museums and exhibitions before starting his PhD.

Steven Mills is an Associate Professor at the University of Otago, where he gained his PhD in 2000. Between being a student and an academic at Otago he worked in a variety of commercial research and development roles and as a lecturer at The University of Nottingham. His interests lie in computer vision, particularly 3D reconstruction from images and applications with cultural and heritage value.