A Journey of Computer Vision in Sports:
from Tracking to Orientation-based Metrics

Adrià Arbués Sangüesa

TESI DOCTORAL UPF / 2021

Directores de la tesi
Dr. Coloma Ballester, Dr. Gloria Haro

Department of Information and Communication Technologies

I wish basketball was rocket
science sometimes.

Daryl Morey

Agraïments (Acknowledgements)

Per por a deixar-me algú, mai no he sabut com encarar tot allò que impliqui felicitacions o agraïments. De fet, quan vaig fer 18 anys el 28 de desembre del 2011, la meva estratègia per convidar amics i amigues al meu aniversari va consistir simplement a repassar la llista de la gent que m'havia convidat a mi prèviament a la seva festa i fer-ho recíproc. Aquesta vegada no tinc llista, però tot i això, ho intentaré fer el millor possible. De tota manera, si llegeixes aquesta secció, t'hi sents identificat/da i no hi surt el teu nom, et pots donar per agraït/da.

Abans de res, volia agrair la confiança, la disposició i el tracte de les meves dues tutores, la Gloria i la Coloma, la Coloma i la Gloria (sense cap ordre predeterminat). Tot i que ja existia un biaix que m'assegurava que estaria còmode al vostre grup, les expectatives s'han vist superades i heu fet que el despatx 55.116 sigui per a mi un dels millors aixoplucs que tinc. Gràcies pels mails a hores intempestives, gràcies per no desesperar-vos durant les reunions on començo a disparar tecnicismes de forma inconnexa i gràcies per tots els consells. He après moltíssim amb vosaltres, i no només dins del camp de la visió per computador.

Com no podia ser d'una altra manera, el següent agraïment va dedicat a la meva família, en especial per a mon pare i ma mare, en Jesús

i la Conxa. Tot i que durant els 2 primers anys de tesi no us hagi vingut a veure tant com voldríeu, crec que durant els 2 darrers heu tingut sobredosi d'Adrià. En poc més de dos anys m'heu hagut de cuinar sopetes després que em traguessin 4 queixals del seny de cop, m'heu hagut de fer de xòfers després que em reconstruïssin el turmell defectuós, i m'he confinat amb vosaltres. Han sigut moltíssimes hores junts, i tot i que el planning de Hansel i Gretel no us va acabar de funcionar, no tinc paraules per agrair-vos tot el que heu fet, i les partides del Catán durant els dies de confinament. Sou els millors pare / mare que tinc, i us perdonaré que m'enganyéssiu amb en Baltasar quan era petit. Encara que, per desgràcia, ja no hi siguin, aquesta tesi també va dedicada als meus 4 avis i àvies: la iaia Rosa i els plats de macarrons, l'*abuela* Conxa i els ous amb neula, l'*abuelo* Ramiro i els llagostins, i l'avi Jesús amb els pinyons i el formatge ratllat. Tot i les connexions amb el menjar, tot el que em vau ensenyar i tot el que em vau cuidar va molt més enllà, i em sap greu que marxéssiu tan *d'hora*. Us enyoro. A diferència de moltes persones, jo també tinc un avi no-biològic, en Palle Jacobsen, qui també va saber cuidar-me durant el meu any-i-escaig a Dinamarca; *hygge* és casa teva, Palle, descansa en pau.

Per animar una mica el discurs, ara toca agrair a aquelles persones qui més m'han hagut d'aguantar durant els anys de tesi. A l'estiu del 2017 buscava un pis on viure, i per X o per Y, vaig acabar amb les desconegudes Vero i Neus. Quatre anys més tard, no sé què faria sense elles i sense en Fabiano. L'etapa *Los Viladomat* no l'oblidaré mai, potser la maleeixo quan als 35 tingui mal d'esquena per culpa del sofà infernal, però tota la resta serà un formigueig inevitable, i Montardit (o Bergamo, *donde el aeropuerto*) sempre serà un lloc molt especial. Gràcies per ser-hi, el tito sempre serà vostre, inclús quan tingueu un *lio-de-dia*.

Tot i no haver viscut mai junts, la gent *de sempre* de Vilassar també té els meus *blessings* perpetus: Pere, Fisse, Plando, Ricard, Facund,

Andreu, Sergi, i tota la resta. Tot i que ens coneixem des de l'etapa que pixàvem a l'orinal, i encara que tothom hagi fet el seu propi camí, sou de les persones més incondicionals que conec, i el preu de tenir-vos no es compra ni amb ampolles de Salsa Espinaler. I no patiu, que per molt pesats que sigueu amb les bicicletes, us seguiré estimant.

Agrair a la gent que he conegut a la universitat des del 2011 és molt més complicat. Per començar, tenim les joies de la carrera, els i les 9, *com a la Comunidad del Anillo*, van ser moltes hores d'estudi i moltes samarretes empudegades d'olor carregada de biblioteca i peixera. Ja fa anys que només ens veiem per Nadal, però em vau ajudar una barbaritat a trobar-li la gràcia a les matemàtiques, i a no consolar-me anant "als mínims". Escriure una tesi és el més *empollón* que he fet mai, i és gràcies a vosaltres: Helena C., Tola, Adrià (el bo), Nadia, Guillem, Edu, Andreu i Dani. D'altra banda, durant el doctorat, he conegut altres persones estupendes. Aquí cadascú porta els seus timmings i tothom fa estades a l'estranger, i jo ja em perdo i oblido qui he conegut i quan, però vaja, ho intentaré. Gràcies a tota la gent del IPCV, sobretot a la Maria, Patricia, Lara, Juan i Venkatesh, amb les aparicions de gent com en Samuel, Pierric, Alejandro o Ferran; hem fet menys seminaris dels que pensàvem, però tots els vostres *inputs* han sigut de gran ajuda. Gràcies també al postdoc de confiança, n'Adrián Martín; ens ha quedat clar que no ets un fan dels punts i coma, i que la versió moderna del futbol potser no és tan ràpida com ens pensem, però igualment, merci. L'Arantxa també ha sigut un boníssim descobriment, i és molt millor amiga que jugadora de voleibol (i no perquè sigui dolenta); potser treballo en una cova, i no sóc de fer abraçades, però *deep inside* t'aprecio. I finalment, també posaré dins d'un mateix pack tota la gent del club del dinar a les 13:00: Mireia A., Amelia, Xabier, Mireia M., Juan B., Itziar, Valentin, Pablo... La veritat és que em costarà trobar un ecosistema on es pugui fer una desconnexió total d'una hora i poc, quan sembla que el temps s'aturi. Heu sigut un gran suport durant l'últim tram de doctorat i escoltar-vos sempre s'ha convertit en la millor

font d'inspiració per seguir fent recerca en moments complicats, com un divendres a la tarda; tot i ser uns addictes al cafè, us admiro fort. Així mateix, voldria remarcar la paciència i la immediatesa de totes les persones que treballen a Secretaria Acadèmica, sobretot la Lydia (*aka* bústia) i la Carme. Per acabar el bloc d'agraïments universitaris, volia donar les gràcies al tracte rebut dels i de les alumnes (i de l'Adriana F. amb l'organització de classes); preparar les classes és una feinada, i el dia abans de la meva primera classe (4 d'octubre del 2017) gairebé no vaig poder aclucar els ulls dels nervis. Quatre anys més tard, considero que he après molt de tots i totes, incloent aquells grups virtuals en els quals no he vist cap cara; estic segur que amb el talent que teniu arribareu molt lluny. Moltíssima sort, espero recordar tots els noms i cognoms per si mai ens creuem.

Sense sortir del món de la recerca, si mai em pregunteu quin és l'esdeveniment on més he gaudit i après, ho tindré clar: l'Exporecerca Jove. Després de participar-hi el 2011, he tingut la sort (i la voluntat) de formar part de Magma, Associació per Promoure la Recerca Jove, de la qual només tinc coses bones a dir. Que els joves promoguin la recerca d'altres joves és preciós, i tan sols desitjo que l'Exporecerca se segueixi reinventant. De tota la gent que he conegut dins aquest món, volia donar les gràcies especialment a tres-quatre persones: la primera, l'etern co-director, en Sergi Bonet. La segona, en Manuel Belmonte, que fa 21 anys que és a peu del canó i que, des de la seva modèstia, mai no voldrà que ningú li reconegui la feina feta; ho sento Manuel, però ets una meravella, i no ens cansarem de dir-t'ho. També poso dins d'aquest sac en Tomàs Padrosa. L'última, però no menys important, és en Bruno Gotzens. No vam tenir la sort de parlar gaire cara a cara, però en Bruno sempre serà el director de la XII Exporecerca i el corresponsable principal (amb en Marc Jordana) que ara estigui escrivint aquesta tesi. Des de l'abril de 2011, i gràcies a la gent de Magma, sempre he volgut fer recerca sobre el que més m'apassiona, i això ho he après de gent com en Bruno, que parlava dels animalons amb uns ulls com taronges i una passió espectacular.

Allà on siguis, Bruno, gràcies.

No em vull oblidar de tota la gent que m'ha donat oportunitats d'or dins del món del bàsquet. Per començar, agrair la confiança a en Xavi Pardina, que em va ajudar a fer un salt de gegant dins la meva carrera com a entrenador: l'any a EBA amb el Cornellà va ser un clínic contínu. De la mateixa manera, també vull donar-li les gràcies a en Juan Llaneza per oferir-me la possibilitat d'incorporar-me al F.C. Barcelona, i per tots els debats estadístics; dins del club, també he conegut gent magnífica, com el staff sencer del Cadet B o en Martí Casals, o la gent del futbol com en Javier Fernández.
I finalment, sense cap ordre concret, també volia donar les gràcies a totes aquelles persones que, per motius *variopintos*, han emergit com a bolets a la meva vida: Mar C. i l'illa de Chichimé, *lovely* Andy, Mar B. i la muntanya balla, Judit P. i els iogurts de *papagaya*, Mar F. i Eivissa sencera, Carme-Ferran com a equip Lassie, Adrià I. i Novo Hamburgo, Eva L. i la nostra quedada cada 6 mesos, Ada i els reptes de dades, Marta-Guille i els jocs de taula, Laia B. amb el *Pop-Pop!* i Víctor DLT i el seus berenars d'atleta.

Abstract

Although tracking data have completely revolutionized the whole data science paradigm in sports competitions with the largest economic resources, its use in a European context is still unexplored. In this thesis, three tracking-related contributions are presented in the sports domain. First, the creation of vision-based basketball multi-tracking methods is studied from a single-camera perspective, which could be useful for clubs with low resources or for the recovery of vintage games' tracking. Then, tracking data in the soccer domain is enriched by adding a novel layer of information: player body-orientation, thus complementing 2D location data, which falls short in some scenarios. Finally, the effect of proper orientation is detailed in the most common soccer action: passes. By building passing computational models that express which is the safest pass at a given time, the relevance of orientation is contextualized, hence proving that it is indeed a vital skill for soccer players.

Resum

Tot i que les dades de seguiment han revolucionat el paradigma de la ciència de dades esportiva dins les competicions amb més recursos, el seu ús en un context europeu és encara una incògnita. En aquesta tesi, presentem tres contribucions dins d'aquest camp. Primer s'ha estudiat, a través de la visió per computador, la creació de sistemes de seguiment de jugadors/es de bàsquet utilitzant una sola càmera, el que podria servir per equips amb pocs recursos o per recuperar dades de partits antics. A més, donat que la manca de context és la principal limitació de les dades posicionals, la segona proposta en presenta l'enriquiment amb una nova capa d'informació: l'orientació corporal de jugadors/es de futbol. Finalment, s'ha analitzat l'impacte de l'orientació mitjançant la creació de models computacionals de passades, els quals esbrinen quina és la passada més viable i demostren que l'orientació és una capacitat clau per als jugadors/es.

Preface

The most favorite present I have ever received was an empty squared teacher-notebook that belonged to my parents. I was 9 years old, and although it might seem odd, my main passion was to build basketball rosters on paper based on simple player performance metrics; at that point, I had no clue about any statistical parameters or distributions, but I was truly devoted about naive sports analytics. Apart from basketball, I also made some hand-crafted clustering side-projects with Pokemon toys based on semantic features, but unfortunately, the lack of a universal dataset made me give up.
Once focused only on basketball analytics, I kept improving my amateur General Manager skills, but I could never transfer that knowledge into school-related projects: despite enjoying math and physics (and struggling with the athletic side of sports in physical education), I could not see myself talking about basketball statistics with my peers or teachers. In the secondary school syllabus, there was simply no place for that. When I was 16, the first opportunity of doing research about my main interests emerged out of the blue: during the final high-school thesis (the so-called *Treball de Recerca* in Catalunya), I learned the coding basics and I built a digital coach board application. At that point, my mind clicked, and I realized that, although I would have had a blast taking hypothetical sports analytics courses, I was not ready yet; the potential of sports' advanced statistics goes far beyond from hand-written teacher notebooks, and its basis requires a high understanding level of math, data mining, and sports science, as well as proficient dissemination skills. Therefore, during

my engineering undergraduate studies, I also started coaching while keeping an eye on doing sports-related research. In this context, my bachelor thesis consisted of a couple of Computer Vision algorithms that could automatically detect common basketball infractions, thus being a potentially beneficial tool for officials, and two years later, I trained classical Machine Learning models able to classify basketball plays according to sensor data. By expanding both my engineering and coaching background, the detail level of the published projects improved, hence becoming slightly more relevant contributions to the sports analytics field. Nevertheless, as it usually happens with bachelor / master thesis, the outcome of the presented papers was almost purely theoretical and it did not have a direct practical application that could help players / teams improving their performance. That being said, the nature and the motivation of this PhD thesis are clear: after so many years chasing a research-based career in sports analytics, I could finally spend four years learning and contributing to the research field I have always enjoyed the most. During my thesis, I tried not to stop coaching, and as a matter of fact, I managed to take the assistant coach role in the *Senior* team of Cornellà (*EBA division* 2017-2018) and in the *under-15* Futbol Club Barcelona youth team (2018-2019); being in the actual court definitely made things easier when creating projects from scratch.

Moreover, it has been proved that versatile skills are vital when working in this research field, and not only when training different models of a single specific sport but also when working across sports. For instance, Luke Bornn and Dan Cervone, who are two of the most prestigious sports analysts, switched sports at some point: while Luke has worked with both soccer (AS Roma) and basketball (Sacramento Kings) teams, Dan contributed also to the research fields of basketball and baseball (Los Angeles Dodgers). As a consequence, although I have clearly stated that I am a basketball-based person, I also wanted to contribute to another sports analytics field; bearing in mind that I grew up in a town close to Barcelona whilst watching Ronaldinho, Eto'o and Messi on TV, soccer was an obvious and a safe

choice. Athletes of different sports (especially basketball and soccer) share many contextual features, and ideal research outcomes should be able to generalize properly to different scenarios. For this reason, creating soccer-based models was an enriching experience, where I had to adapt myself into a whole new analytics-related environment, suddenly feeling like the 9-year-old Adrià together with his teacher-notebook once again.

Apart from merging sports knowledge with data mining, another relevant topic I would like to briefly discuss is communication. Since data is a powerful complementary tool for sports teams, establishing a data-driven communication culture is almost mandatory, and although it seems a simple and immediate task, it is rather challenging. Once data sources have been identified and exploited, several handicaps are faced. The first one is strictly related to the lack of numerical-based background evidence: although it might be a huge investment, clubs must be patient when getting started with sports analytics, since few conclusions can be made when the gathered sample size is small; however, data scientists must be the ones to create this non-rushing results-based culture. Moreover, in the vast majority of scenarios, the methodology of the coaching staff has little to do with scientific research groups: apart from being volatile, each one is unique, and there is not a universal communication flow; by learning how to ask the appropriate questions and by understanding the coaches' demands, the connection between analysts and coaches improves. Finally, the last tricky facet of this data-driven culture is the *argot* being used, especially in social networks and media; for instance, currently, there is not a concrete and general definition of *big-data* or *advanced statistics* in sports. These type of concepts are useful to create awareness, but data scientists should be really careful when describing and sharing their work, thus avoiding the use of ambiguous trendy concepts. I truly believe that the power of data in sports is complementary to the coaching staff knowledge and that it can automate some scouting processes, so we should establish a fair communication system where we concisely describe the data sources,

their actual scope, and the possible potential outcomes. Since no one can claim to have the perfect recipe for winning games, the dissemination of new contributions should be explained and shared accordingly. Despite not having previous experience when it comes to sports analytics dissemination, this thesis comes together with a large set of non-academical contributions, where I aimed to create a solid knowledge-plus-communication that could be used by coaches, analysts, and fans to speak the *same* language when it comes to data. First, I published an open-source side-project entitled *BueStats*, which extracts automatic state-of-the-art statistical reports for all female and male teams in FEB Competitions (Spanish Basketball Federation); I also created an open repository of Python Notebooks that explains how to train basic Machine Learning models by using available Euroleague data. Besides, I also approached the complete communication pipeline in the Keynote Talk I gave at the Computer Vision in Sports Workshop at CVPR 2021. At the same time, the main core of this presentation was built from several concepts that were discussed in a set of talks I moderated, entitled *Beyond the 4 Factors*; in these virtual events, several basketball analysts presented their methodology when building metrics / designing graphics for coaches / communicating in a top competitive level, etc. Finally, I have also been pursuing the awareness creation of sports analytics in several talks / clinics / courses, either by presenting general topics or by teaching from scratch how should data be handled in the very first steps (Catalan / Basque / Spanish Basketball Federations or the Catalan Society of Statistics).

Although it might seem a *cliche-ish* ending sentence, I really hope you enjoy / learn while reading this thesis as much as I did when writing it, I really mean it. Finding my own path into sports analytics has been a long and tough road, but it has definitely been worth it. We will see what comes next!

Contents

List of Figures

List of Tables

1 Introduction

Analytics have completely revolutionized the way we understand the game. By switching from a purely intuition-based decision-making process to a hybrid one, where data-driven processes complement the existing know-how, professional teams have changed the existing sports patterns. Although the analytics pioneers emerged in the baseball pitch together with the Moneyball-fever [60], the paradigm shift was also transferred to other sports: it seems that the five well-established basketball court positions do not exist anymore as such [56], Formula 1 cars are constantly outperforming their speed peak whilst optimizing pit stops, and soccer clubs are looking at brand new features, such as player chemistry [7] in order to optimize their line-ups. Moreover, the application of sports science is not only limited to the competition itself, but it is also used in other multidisciplinary facets of the game, such as injury prevention [78] or notable analysis about conditions that might influence player performance [37].
In the past, assistant coaches were the ones in charge of crunching the numbers, but nowadays the *analyst* figure has emerged. We consider the analyst as a data scientist working for a team / organization, but there is not a preassigned department for him/her *a priori*. This analyst might be either included in the team's coaching staff or in the data science department; he/she can even be an outsourced resource directly in touch with the coaching staff or the General Manager (GM) of the club. But apart from defining the analyst skills and its

role, it is vital to understand the core reasons for this abrupt change in the last lustrum. In order to do so, we might have to analyze and answer the following questions: what motivated teams to create data science departments? Why did coaches start delegating certain tasks to analysts?
Both answers converge in similar reasoning: the amount of gathered data in the sports domain has raised substantially; therefore, coaches must focus on their day-to-day scouting tasks, and complex analysis will be conducted concurrently. Although baseline statistics such as box scores have improved and are still a valid resource, what really made a difference in terms of generated data was the inclusion of new data sources. In this sense, the incorporation of Computer Vision (CV) techniques in sports scenarios [113] notably improved both the precision and the quality of gathered data. Chiefly, the inclusion of optical tracking data around 2013 in the most powerful leagues across sports changed the whole context; *i.e.* spatial data, captured by cameras, that indicates the movement of players in the court / field / pitch. For instance, companies such as Stats Perform [103] and Second Spectrum [97] emerged in the basketball domain as the main tracking providers: by installing an array of cameras in the ceiling of stadiums, their methods succeed in tracking players at a notable temporal resolution (25 frames per second). Consequently, the National Basketball Association (NBA) acquired their products, and now all teams in the league benefit from this emergent technology, thus bringing the game to the next level in terms of efficiency.

Tracking is the vehicle that will guide the reader throughout this thesis and the main common factor among its contributions. First of all, and before getting started with the *so-called* advanced statistics, the current need for tracking data in a European basketball context will be studied from scratch from a technical perspective. Once different *low-cost* CV-based suggestions are presented, the intrinsic limitations of tracking are studied. Although pure tracking data are a powerful source able to produce automatic statistics or to train meaningful

predictive models, location information on its own may lack some context, which might reflect the individual willingness of a player to successfully interact with their teammates. Finally, by complementing 2D tracking data with body orientation estimation, the new generation of analytics is presented in this manuscript. However, before jumping into the thesis contributions and the corresponding proposed methods, the upcoming Section aims to disseminate the evolution of the sports analytics' pipeline (in particular, basketball analytics). By contextualizing the relevance of tracking data, the reader will be able to realize the unlocked potential of applications that could be used with this brand-new source of information.

1.1 From the Box Score to Tracking Data

Sports statistics themselves are nothing new, and basic studies were already performed at the beginning of the 20th century. In order to create a historical context, in this Section, **basketball analytics** are discussed from the very beginning. Similar reports could be written for other sports, especially the ones with a notable statistics background, such as football or soccer.

Despite being obvious reasoning, the most important basketball stats are the ones related to scoring. After all, the final outcome of a basketball game is binary and indicates whether a team has won / lost a game, but this flag is always displayed by the final scoreboard, which contains the minimum viable summary of a game and shows the total number of scored points per team. Just by checking this scoreboard, one can already make some assumptions: was the game close or not? According to previous knowledge, was the result the expected one? In order to back up all these guesses, more data are required. Team stats can be a great resource; for instance, rebounding numbers can show if that team dominated the boards and created extra shooting opportunities, assists and turnovers can be useful to see game-control

facets as well as ball-sharing, and by comparing shooting percentages, it can be deduced if a team performed as expected or not. However, team stats are obtained by adding individual contributions up, so dissecting numbers among roster players might be required to get a proper understanding of the overall game course. A clear example of the need for individual performance quantification was provided in March 1990 by Stacey King, former Chicago Bulls' power forward, who claimed: *I will always remember this game as the night that Michael Jordan and I combined for 70 points.* If this sentence is read straight-forward, one may think that both King and Jordan scored 35 points each (or other similar combinations such as 40-30 or 30-40), but the truth is that while Jordan scored 69 points, King only scored a single free throw. The complete summary of individual (plus collective) stats is named **box score** (Figure 1.1), which is a simple spreadsheet being used since 1930 approximately, and their main listed contributions are: minutes, points, field goals (2-point shots, 3-point shots, including both scored / attempted shots), free throws, rebounds (offensive plus defensive), assists, steals, turnovers, blocks, committed / received fouls. Moreover, Dean Oliver created several contextualized statistics in his book entitled *Basketball on Paper* [82], which shows several approaches on how to refine and to normalize box score data in terms of game pace (possessions).

In order to exploit temporal information, *play-by-play / eventing* data are analyzed, which attempts to contextualize the *value* of all events depending on the game phase. Although the concept of *value* might seem ambiguous, a simple example is provided. On paper, all 3-point shots are worth the same in terms of score-board. However, imagine a *Player X*, who scores three long-range shots in the second half of the game while his/her team is trailing by a large margin, and imagine a *Player Y*, who scores a 3-point-shot at the last second of the game when his/her team was trailing by 2. In this situation, the contribution of *Player X* results in innocuous points that do not help his/her team winning the game, while *Player Y*'s game-

	Minutes	Points	2FGM	2FGA	3FGM	3FGA	FTM	FTA	OReb	DReb	AST	STL	TOV	PF
Michael Jordan	50	69	21	31	2	6	21	23	7	11	6	4	2	5
Scottie Pippen	41	7	3	10	0	0	1	2	2	6	7	5	8	5
Horace Grant	40	16	7	14	0	0	2	4	2	3	4	1	2	4
Bill Cartwright	39	9	3	7	0	0	3	4	1	4	1	0	3	5
John Paxon	23	2	1	4	0	0	0	0	0	0	1	0	2	6
B.J. Armstrong	29	6	2	7	0	0	2	2	0	1	3	2	1	5
Stacey King	17	1	0	4	0	0	1	2	1	0	1	2	0	2
Will Perdue	12	0	0	0	0	0	0	0	0	4	0	0	0	3
Charles Davis	11	5	1	2	1	1	0	0	1	1	1	0	1	2
Clifford Lett	3	2	1	1	0	0	0	0	0	0	0	0	0	2
TEAM	265	117	39	80	3	7	30	37	14	30	24	14	19	39

Legend: 2-3FGM /2-3FGA - two-three-point shots made/attempted, FTM/FTA - free throws made /attempted, OReb / DReb - offensive / defensive rebounts, AST - assists, STL - steals, TOV- turnovers, PF - personal fouls. **Bold players**: initial five.

Figure 1.1: Chicago Bulls' box score against Cleveland Cavaliers on March 28th 1990.

winner was crucial. As seen in Figure 1.2, Basketball *play-by-play* summaries include a chronological timeline of events split in three columns: (left / right) listed events regarding *Team A* / *Team B*, together with a (center) temporal timestamp plus the current result at the given timestamp. Apart from including stats-related events (shots / rebounds...), *eventing* data also include game factors, such as substitutions. Even though interpreting *play-by-play* data is way more tedious than simple box scores, several processes have been automated on top of temporal data to get valuable digits, such as *(regularized) plus-minus* / *on-off* / *clutch* statistics.

Once analyzed the main contextual (and temporal) statistics, positional data come into play. Imagine the following situation: *Team A* is playing against *Team B*, that have the ball in the last seconds of a tied game. With contextualized and eventing data, the coach of *Team A* might have some intuition about the player who is going

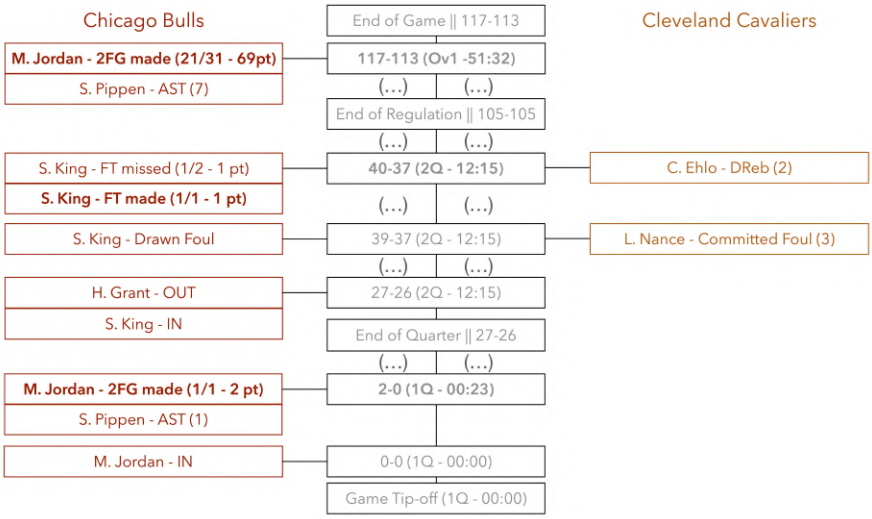

Figure 1.2: Play-by-play summarized sample (Chicago Bulls against Cleveland Cavaliers).

to attempt the final shot. Nonetheless, where is this player going to shoot from? From box scores, the coach might know if a player is likely to attempt a 2- or a 3-point shot, but deeper profiling of shooter types is vital. Among all possible 2-point shots, players might excel in short-, middle-, or long-range situations; similarly, the comfort zone of three-point shots also varies depending on the type of player (central / elbow / corner). In the same way that annotators tag all actions with a timestamp to generate temporal data, they also create **shot charts**, where the 2D location of all shots is tagged. From large samples of shot chart data, and by using visualization tools such as heatmaps, profiles and patterns are easily recognized. In these maps, both the volume of shots and their accuracy can be printed, thus providing the coach with key information regarding the opponents' spacing tactics and characteristics. A couple of shot charts are displayed in Figure 1.3.

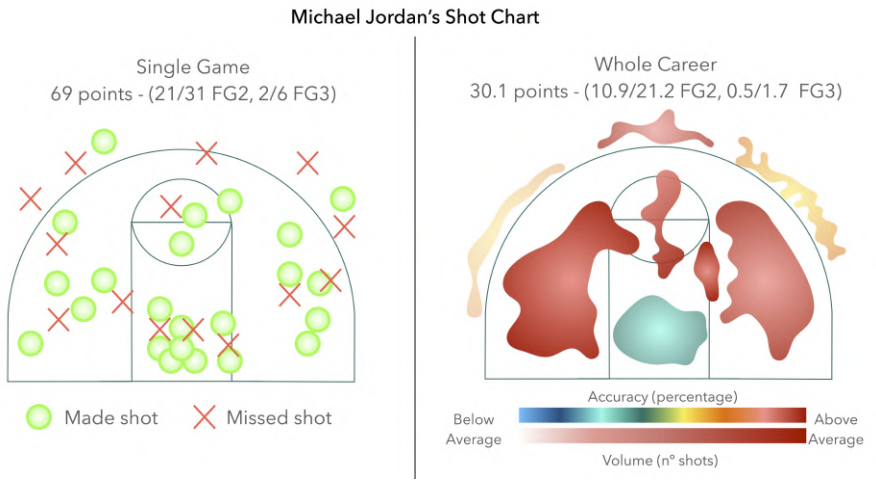

Figure 1.3: Michael Jordan's shot chart data: (left) single game, (right) complete season.

Despite the mix of box score, eventing, and shot chart data results in complete scouting reports, the overall detail level can still fall short in elite scenarios. For instance, when analyzing a 3-point shot, according to the previous sources, its outcome (scored / failed), timing (minute), and 2D location are known, but some details are missing, for instance: was that shot attempted off the dribble or in a static spot-up situation? This question can be answered with **playtype** data, which adds some metadata on top of every event. A summarized individual example of playtype data is shown in Figure 1.4. Nevertheless, who adds all these labels to the given events? Since basketball is a fast sport with a lot of events in a reduced amount of time, and given that table officials are already in charge of tagging each action with its outcome, its timestamp, and (in the case of shots) its location, this task is usually performed off-line. Not so long ago, video-analysis software, such as LongoMatch [64] or NacSports [79], was used to cut video footage into separate specific clips while splitting and exploiting playtype data; normally, assistant coaches

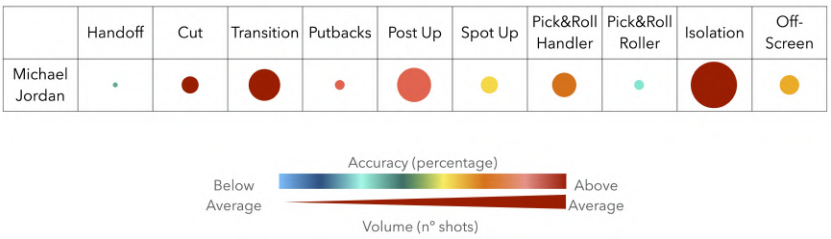

Figure 1.4: Michael Jordan's playtype data.

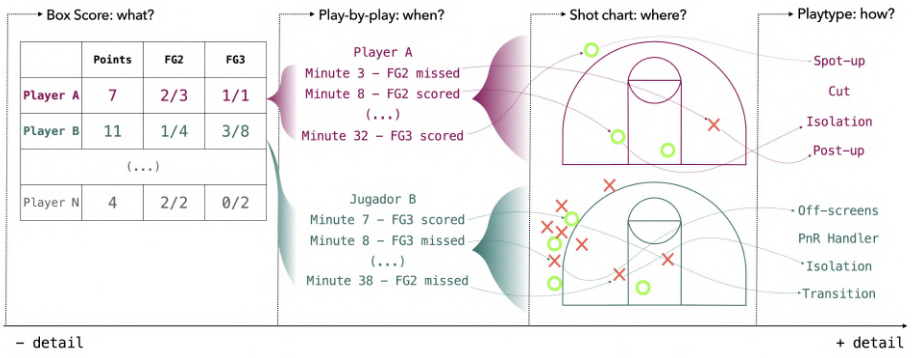

Figure 1.5: Data from different sources are merged in order to build scouting reports.

were the ones in charge of this type of time-consuming tasks. Lately, clubs with economical resources have outsourced these processes to specialized companies like Synergy [109] or InStat [53], that have managed to build competitive teams of professional analysts who cut basketball games from all over the world, thus offering clubs a lot of playtype data without time-consuming shifts. A complete summary of the above-explained statistics is visually displayed in Figure 1.5, where it can be seen how different data sources complement each other, thus reaching fine-grained scouting reports.

Other concerns regarding shot difficulty or defensive performance can also emerge when the level of detail is deeper. On the one hand, by

splitting shots between open / contested, coaches might detect shooting patterns that are hard to spot at first sight. On the other hand, few defensive events are gathered in the above-mentioned basketball data sources: shot chart data are mostly offensive-based, and included data in box scores and *play-by-play* only display few defensive items, such as steals, blocks, or committed fouls. Although both tasks can also be performed through manual labeling, this time establishing visual thresholds is not that simple; for example, it is pretty easy to see (and to label) a pick and roll play that ends up with a scored 2-point shot, but what is the main criterion to decide whether if that shot was contested or not? And more importantly, how can we know if the faced defensive pressure was hard? Both answers go far beyond from playtype, and the most appropriate way to get an automated system that quantifies these facets is through **tracking** data, which are obtained with a camera system (generally placed in the ceiling of the arena) that manages to detect and to follow all players on the court.

By using an accurate tracking system, apart from obtaining advanced performance stats, we can also infer the previous analytics layers. That is, a proper tracking system that gets the position of all court players and the ball at a decent temporal resolution (usually 25 frames per second) can be used to automatically generate: (1) a box score, (2) play-by-play data if the gathered data are synchronized with the game-clock, (3) accurate shot charts, (4) playtype data inferred without time-consuming video editing tasks, and (5) tracking statistics. Inside the latter group, the potential of tracking applications is unlocked and may involve predictive models, defensive accurate metrics, strategies and tactics, strength and conditioning factors for injury prevention (displacements and speed), data visualization tools, etc. Although basketball analytics have been discussed in this Section, note that similar patterns apply to other sports such as soccer, where the inclusion of tracking has been an inflection point for clubs and organizations.

According to the Seth Partnow [1] (former director of basketball research for the Milwaukee Bucks), at the end of the 2019-2020 National Basketball Association (NBA) season, a total of 138 people were working in the team's data science departments across the league; scientists are mainly focused on extracting beneficial numerical insights from tracking patterns, transferring these to coaches, and ensuring that analytics make a difference in the playing court. However, if we compare the number of NBA data scientists with other leagues, there is a huge drop. For instance, in the first division of the Spanish league (Asociación de Clubes de Baloncesto - ACB), although several coaches such as José Angel Samaniego, Lluís Riera, or David Garcia made relevant contributions to the field, there is only one data scientist (Fran Camba). The main difference is simple: in the NBA, the same association was in charge of acquiring the products of tracking companies in 2013 (Stats Perform [103] and SecondSpectrum [97]), thus offering competitive resources to all teams and, as a consequence, revolutionizing the data science departments of all clubs. On the contrary, the Spanish league does not have the required resources to perform this investment, hence limiting the total number of data sources per team; besides, the overall structure of the Spanish (and European) leagues makes it even more challenging. While in the NBA all teams have the same salary cap (around 109 million dollars per team), and all stadiums are somewhat similar, in Europe the lack of a salary cap creates an unbalanced market (team budgets vary from 2 to 40 million dollars), and stadiums differ from each other, hence resulting in camera installation set-ups that are impossible to generalize. Given the economic cost of camera-systems, companies such as RealTrack [91], Catapult Sports [13], or NothingButNet23 [80] emerged, offering sensor-based solutions. Their product is a set of small sensors (mainly with Bluetooth technology) that can fit in many places: while some products are built to be placed on top of the player shorts' cords lace / on the surface of their trainers, oth-

[1] `https://bit.ly/3vUfHvZ`

ers come with a wearable sports bra with space to carry the sensor. Nonetheless, the vast majority of leagues do not allow to play official games with wearables, thus these devices are mainly used for strength and conditioning purposes during practices.
As a summary, it is crystal clear that, within the sport context, there is a strong need for tracking data in order to perform advanced data analyses, hence mutating to a data-driven decision-making culture. However, since only elite organizations have managed to gather and to share this kind of data, there is an even bigger demand to create accessible and *low-cost* solutions that could unblock this bottleneck.

1.2 Manuscript Outline and Contributions

Once stated the relevance of tracking data in sports, this thesis presents several contributions in this research field that, as a whole, constitute a singular Computer Vision journey in sports analysis. In this manuscript, those contributions have been split into three Parts: tracking, body-orientation, and pass feasibility. Besides, before getting started, Chapter 2 provides the reader with some required details / concepts regarding important thesis facets, such as a complete description of pose models, soccer basics, or even a list of open datasets for the sake of reproducibility. The whole flow of the thesis can be seen in Figure 1.6.

In Part I, we approach the creation of automatic basketball multi-tracking methods through a tracking-by-detection fashion in single-camera video footage. Roughly, the presented contributions stem from pose models to detect players; by combining the main output with court filtering and contextual / deep-learning features, notable accuracy is obtained even in challenging cluttered scenarios. The presented trackers show that tracking data can be obtained / estimated without the need for a complex camera setup; in particular, European competitions could benefit from this method, and tracking data

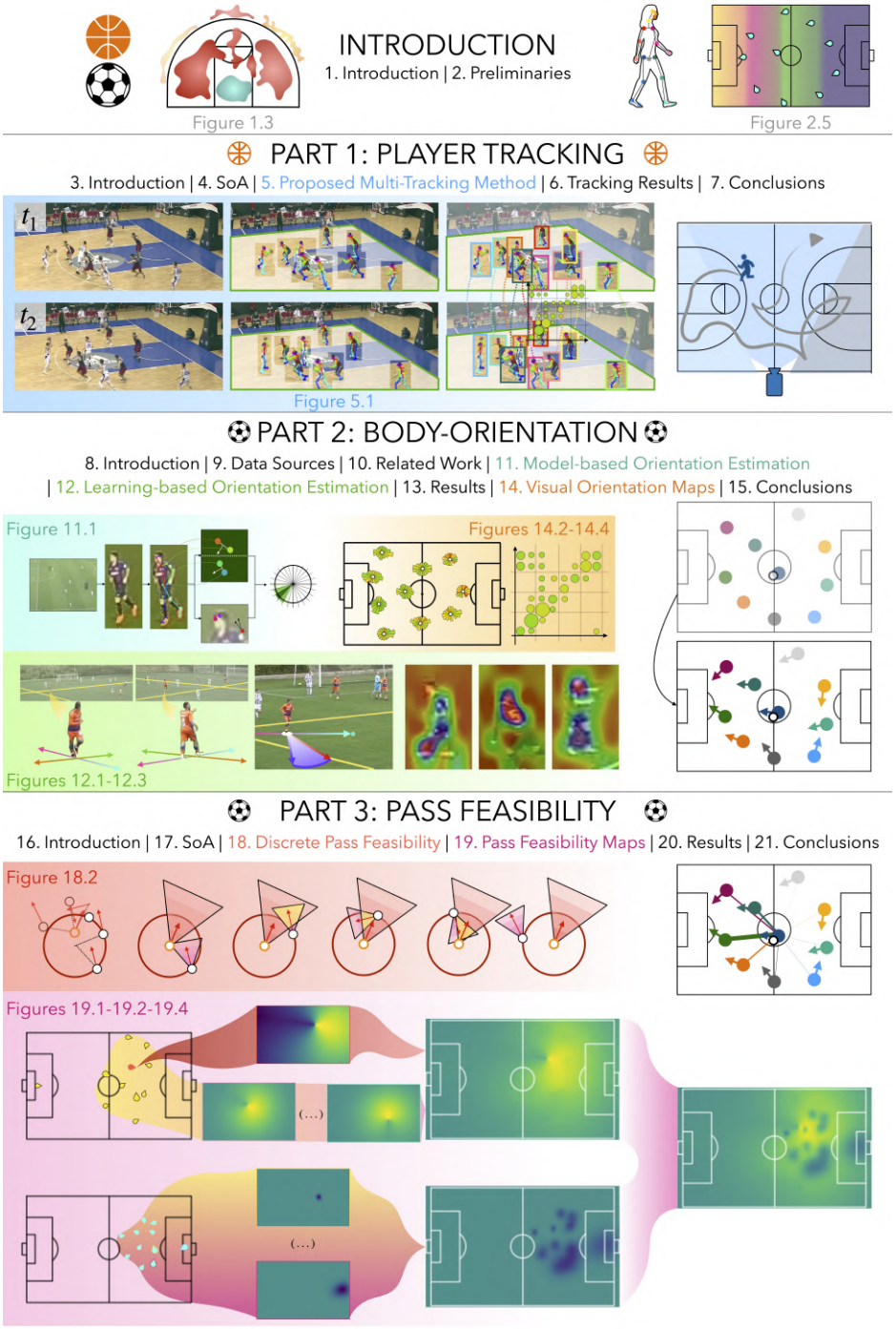

Figure 1.6: Thesis flow. Visual overview of the context of all Parts and Chapters. For a detailed explanation of each visualization, we refer the reader to the corresponding manuscript's Figures.

could also be obtained from vintage games. Regarding this matter, two papers have been published:

- Arbués-Sangüesa A., Haro G., Ballester C., Multi-Person Tracking by Multi-Scale Detection in Basketball Scenarios. *Irish Machine Vision and Image Processing Conference*, 2019.
- Arbués-Sangüesa A., Ballester C., Haro G., Single-Camera Basketball Tracker through Pose and Semantic Feature Fusion. *International Conference of Artificial Intelligence on Sports*, 2019.

In the forthcoming Parts, we present an analysis that shows how 2D tracking data may fall short in some scenarios, since player location on its own might not be powerful enough to describe the current game / player situation and its potential effect in the plays' / events' outcome. In order to enrich tracking data with an extra layer of information, we decided to study the power of player orientation in soccer. First of all, in Part II, two orientation estimation methods are detailed. On the one hand, a model-based estimation of orientation is obtained by projecting players' pose parts into a 2D domain, plus by refining the final estimate with contextual information (ball location). On the other hand, a learning-based approach is also presented; more specifically, the orientation of each player is obtained by leveraging geometric and semantic information contained in player crops via a classification strategy. Regarding orientation estimation, three papers have been published:

- Arbués-Sangüesa A., Martín A., Ballester C., Haro G., Head, Shoulders, Hip and Ball... Hip and Ball! Using Pose Data to Leverage Football Player Orientation. *Sports Analytics Summit (Futbol Club Barcelona)*, 2019.
- Arbués-Sangüesa A., Martín A., Fernández J., Rodríguez C., Haro G., Ballester C., Always Look on the Bright Side of the

Field: Merging Pose and Contextual Data to Estimate Orientation of Soccer Players. *International Conference on Image Processing*, 2020.

- Arbués-Sangüesa A., Martín A., Granero P., , Ballester C., Haro G., Learning Football Body-Orientation as a Matter of Classification. *AI for Sports Analytics Workshop at ICJAI*, 2021.

Once again in the soccer domain, Part III attempts to verify and extend our previous claim by: (1) computing -and proving- the importance of proper orientation in pass events, and (2) building and validating passing tools, which could asses the decision-making process of coaches through a data-driven computational model. In particular, this type of model estimates the most feasible pass at any given time. First, passes are considered as discrete events where there is just one passer and ten potential receivers; later on, this model is extended into a 2D field, where the core's model considers that players can pass the ball towards any field spot (open spaces). By building these tools, orientation-based metrics benefit from interpretability, thus complementing raw orientation data and turning them into insights that could be directly understood by coaches. Both contributions have been documented:

- Arbués-Sangüesa A., Martín A., Fernández J., Ballester C., Haro G., Using Player's Body-Orientation to Model Pass Feasibility in Soccer. *Computer Vision in Sports Workshop at CVPR*, 2020.

- Arbués-Sangüesa A., Martín A., Fernández J., Haro G., Ballester C., Towards Soccer Pass Feasibility Maps: the Role of Players' Orientation. *Under Review in Sports Sciences Journal*, 2021.

At the end of this manuscript, a final Chapter of closure is included, which, apart from wrapping up and detailing the overall thesis conclusions, aims to provide the reader with possible lines of work to be

exploited in a near future; it also seeks to state the overall context of sports analytics at the moment.

1.3 Further Contributions

Apart from the academic publications, as stated in the Preface, this thesis comes along with a set of contributions that fall beyond the research scope of the PhD. Among these: (1) open-source repositories have been shared to provide coaches and analysts with contextualized statistics and machine learning basics; (2) talks have been given / organized in order to disseminate the relevance of data science within the sport context, and similarly (3) several courses and workshops have been directed to a large audience of coaches or statistics enthusiasts. Finally, the obtained awards are also listed.

Open-Source Repositories

- BueStats: Basketball Scrapper + Reporting Tool for FEB teams, `https://github.com/arbues6/BueStats`.
- Euroleague + Machine Learning: clustering, classification, and regression models from scratch, `https://github.com/arbues6/Euroleague-ML`.

Talks

- TedxUPF Talk: *When the idea meets the passion and becomes a project*, `https://youtu.be/gW0Yb779Oq4`, 2018.
- Invited Talk at the 4th Summer School of Deep Learning - UPC: *Tracking basketball players through deep learning features*, 2019.
- Beyond the 4 Factors (I), with Justin Jacobs, Todd Whitehead and Seth Partnow, `https://youtu.be/DKv-1n5OHEc`, 2020.

- Beyond the 4 Factors (II), with Mike Beuoy, Nathan Walker and Andrew Patton, `https://youtu.be/FuUwCMpqkUE`, 2021.
- Catalan Society of Statistics: *Layers of basketball analytics*, `https://youtu.be/QZglqEmur0U`, 2021.
- Keynote Talk at XXII Exporecerca Jove: *From the classroom to the stage* (youth research dissemination), `https://youtu.be/1TP7RbS9mEk`, 2021.
- Keynote Talk at the Computer Vision in Sports Workshop at CVPR: *Bringing Computer Vision to the Court* 2021.

Directed Workshops and Courses

- Catalan Basketball Federation - Technical Committee: Artificial Intelligence in sports (basketball). 2017.
- Basque Association of Basketball Coaches: Pseudo-advanced statistics and Artificial Intelligence. 2019.
- Spanish Basketball Federation - Superior Coaching Course (*CES*). Responsible for the Advanced Statistics module. 2020, 2021.
- ImproveSports - Advanced statistics course. 2020, 2021.
- Fundación La Caixa - Big Data Challenge (EduCoach, plus dataset creation). 2019, 2020, 2021.

Awards

- PhD Workshop (UPF-DTIC 2018) - The Collider mVentures Award.
- PhD Workshop (UPF-DTIC 2019) - EiTIC People's Choice Award.
- ICAIS (2019) - best paper award.

- CVSports (CVPR 2020) - runner-up award.
- #HiloTesis dissemination contest (2021) - national winner.

Media

- Mundo Deportivo (2019) - *Tras el 'tracking' de Michael Jordan*, `https://bit.ly/3uG4I8D`
- Mundo Deportivo (2019) - *Los números revelan el basket del futuro*, `https://bit.ly/3fD2odZ`
- Universitat Pompeu Fabra (2019) - *La inteligencia artificial al servicio del deporte*, `https://bit.ly/3c97bSB`.
- La Vanguardia (2019) - *La UPF controla el partido.*
- Universitat Pompeu Fabra (2020) - *Los modelos computacionales aplicados al fútbol calculan la orientación de los jugadores y predicen el pase más fiable de balón*, `https://bit.ly/3uGC9aM`
- La Vanguardia (2020) - *Crean un modelo que predice cuándo un futbolista puede dar el pase más fiable*, `https://bit.ly/3uDhg0b`.
- Mundo Deportivo (2021) - *Este hilo de Twitter sobre analítica en el deporte ha ganado el premio CRUE*, `https://bit.ly/3fGZqoQ`.
- Universitat Pompeu Fabra (2021) - *Adrià Arbués ha sido uno de los ganadores de la primera edición del concurso #HiloTesis*, `https://bit.ly/2SOxoi9`.

2 Preliminaries

This Chapter aims to provide the reader with some required concepts and details to make the lecture of this thesis more self-contained. First of all, throughout the whole thesis, the same method for estimating human pose in given images will be used, encompassed within the OpenPose library [21]. Therefore, its basis will be detailed together with its evolution and improvements (2014 - 2017). Then, some soccer-related basic preliminaries are explained beforehand, hence contextualizing the obtained results of Part II and Part III. Finally, several open datasets are listed, which might help the reader to reproduce the presented methods.

2.1 Pose Models

Pose Models (or Machines) were proposed in 2014 in The Robotics Institute of Carnegie Mellon University, and several features have been chronologically added to better detect the pose of humans in given frames.
The original article was called *Pose Machines: Articulated Pose Estimation via Inference Machines* [89], and it attempted to solve the challenging scenario where the articulated pose of a human must be estimated from an image; the main difficulties were a large number of degrees of freedom of the underlying skeleton and the large variation of appearance. Until that point, the existing State-of-the-Art

techniques were based on Graphical Models, where dependencies and correlations between part locations were computed. Some simplified models were tree-structured and had a clear double-counting symmetry issue (human parts counted twice for both sides of the body), while the others tried to perform exact inference, but learning the appropriate parameters was close to impossible. It can be then said that there was a complexity versus tractability trade-off. The main contributions of the original paper were: (1) a complete *end-to-end* training scheme of the whole inference procedure, thus avoiding the mentioned trade-off; (2) the merging of richer interaction among multiple variables at a time, plus (3) a novel approach to learning spatial models directly with a modular architecture.

The method concatenated different stages; the output of each stage was a belief map / part (from coarse to fine in terms of stages), which provided an estimate that indicated the probability of each image location to belong to every body part. For the first stage, the input image was windowed (in all possible locations), and image features (color- and gradient-based) were extracted. Then, a multi-class Random Forest predictor was trained for every part of the body, obtaining one belief map / part. Besides, this process was repeated L times for L different levels, hence incorporating a hierarchy by differing in the window size to be used (whole body, full limbs or body parts).
Afterward, context features were computed for every belief output map: (1) patch features told coarse information regarding the confidence of body-neighboring parts (*i.e.* how different were neighbors inside a given window), and (2) offset features expressed precise relative local information (*i.e.* how far the local maxima was from the global one). Context features were concatenated for every level and every part, and a second multi-class classifier was trained with the same pre-computed image features. This process was repeated for T stages, as seen in Figure 2.1. As it can be seen, the predictor was trained, using cross-validation, at each level and stage; in the first step, patches extracted from ground-truth images (with annotated

landmarks) were used, while in deeper stages, the model was trained on top of the concatenation of feature maps and contextual data. Inference could be then performed at test time by extracting features from patches at different levels and refining the estimate through contextual features, hence obtaining confidence maps. The final pose was obtained after picking the maxima of the last confidence map for each part.

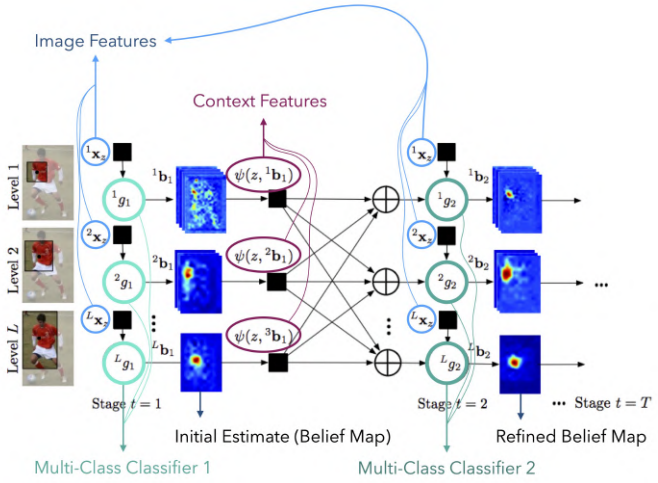

Figure 2.1: Architecture of the first Pose Machines model. Image source: [89].

Two years later, the same authors realized that the presented method could benefit from a convolutional architecture [124]; in their improvement, they learned feature representations for both image and spatial context directly from data, whilst setting a training process able to handle large datasets. Their end-to-end Convolutional Network repeatedly produced 2D belief maps for the location of each part. At each stage, and keeping the previous belief maps as inputs, the receptive field (patch around the output pixel location)

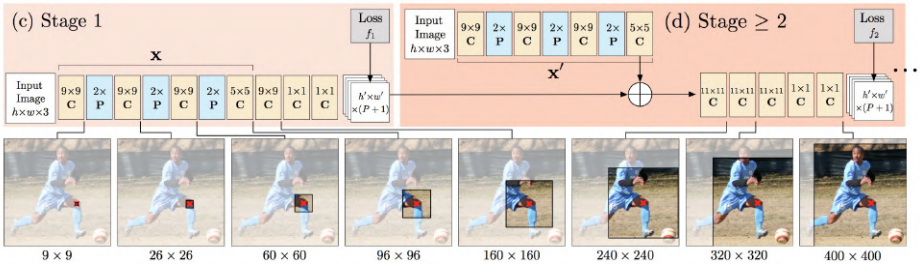

Figure 2.2: Architecture of the second Pose Machine network, including a convolutional architecture. Image source: [124].

was enlarged, hence avoiding having different levels. The main concern could be vanishing gradients, which occur when the strength of the gradient diminishes at each step while backpropagating, and it was addressed through intermediate supervision in the suggested loss function. Besides, implicit spatial dependencies (*i.e* wrist-elbow connection) were learnt requiring neither hand-designed priors nor careful initialization.

The architecture of each stage can be seen in Figure 2.2, where it can be spotted how the size of both convolutional and pooling layers changed at the same time the size of the effective receptive field; the last step of each stage was the evaluation of the loss function, and its output was a vector containing the probabilistic score for each part at each location. This spatial context of easier-to-detect parts could provide strong cues for localizing stronger-to-detect parts in subsequent stages. For instance, if the neck and the right elbow were properly found, but the right shoulder was placed at the left knee, the concatenation of parts would provide enough evidence to displace it in the following stages to its appropriate location. As it can be observed, neither image nor context features were being used.

At this point, the main problem was inferring the pose of multiple people in images, as it has three main challenges, namely: (1) the number of people was unknown at any position / scale, (2) interac-

tions introduced complex spatial interference such as occlusions, and (3) runtime complexity grew with the number of people. For this reason, a new approach including Part Affinity Fields (PAFs) was introduced [10]. Besides, the previous method was based on a top-down-based technique, which has no possible recovery if the person detector fails; the new technique used a Bottom-Up approach, where different PAFs encoded the location and the orientation of limbs, and a novel greedy parsing was computed at a fraction of computational cost.
The method computed at the same time *part confidence maps* and *part affinity fields*, and then, a bipartite matching algorithm was applied to get all part correspondences. Besides, the network was fed with a set of generated feature maps extracted from a VGG-19 Convolutional Network; the architecture of this network can be seen in Figure 2.3.

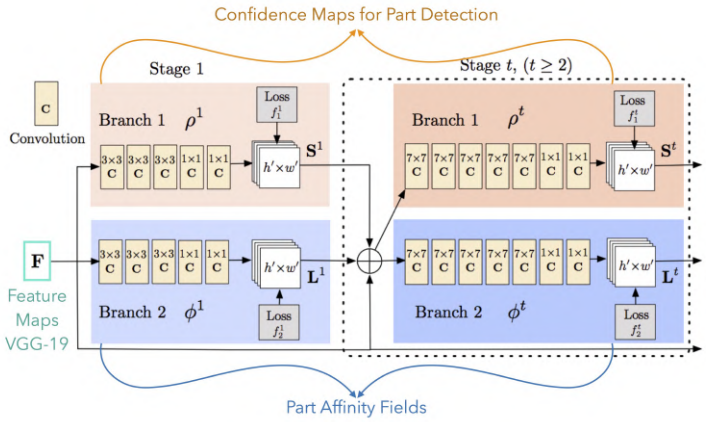

Figure 2.3: Architecture of the final Pose estimation Model (using Part Affinity Fields). Image source: [10].

On the one hand, when computing confidence maps for part detection, the network was fed with: (a) individual maps, and (b) their

ground truth landmarks. In the training process, one peak (maxima) was selected per part and per person; in the testing process, confidence maps were computed, and candidates were obtained with non-maximum suppression. On the other hand, PAFs were computed in order to have a confidence measure for each pair of associations, hence assembling the full body-pose. A 2D vector field for each limb was created, encoding the direction of potential points inside the region with respect to both limb parts. First, using the direction from one part (*e.g.* wrist) to the other (elbow), their corresponding normal vectors and the region of points corresponding to the limb were set; then, each of these points was encoded with the unit vector in the limb direction. During training, ground truth data of all people's limbs were fed into the network; during testing, the line integral over the corresponding PAF was computed, thus measuring the alignment that would be formed (if all points had the same direction, the alignment would be perfect).
Afterward, in order to integrate multi-person detection, non- maximum suppression should be applied together with some weighting for each candidate, but this resulted in a computationally complex problem. The authors suggested a greedy relaxation assuming that pairwise association scores implicitly encoded the global context because of the large size of the receptive field. Bearing in mind that the goal was to indicate which couple of candidates were connected (optimal assignment issue), the problem was reduced to a maximum weight bipartite matching problem, where weights were obtained with the line integral; moreover, knowing that two limbs will not shared a part, the number of choices was reduced. In order to perform a K-dimensional problem, two relaxations were introduced: (1) there was a minimum number of edges, and (2) a decomposition into bipartite matching independent subproblems. Knowing that the presented convolutional network concatenated all features for each part and stage, the relationship between non-adjacent nodes was already taken into account by the network, whilst adjacent nodes were modeled by PAFs. Obtained results outperformed other state-of-the-art

methods by using *gold-standard* MPII and COCO datasets.

2.2 Soccer Basics

Since results of Part II and Part III will contain specific soccer-based language, this Section aims to help readers understand the basics: player positions and game phases. Moreover, please note that, since this thesis has been written in *American English*, the term *soccer* will be used; consequently, the term *football* refers to *American football*.

Game roles / positions:
Within a soccer team, there are mainly four different positions:

- Goalkeeper, the only player that can use his/her hands during the game, and aims to prevent the ball from crossing the goal line.
- Defenders, who constitute the last row of players before the opponents directly face the goalkeeper. Therefore, defenders' main goal is not to allow the other team get past them, either by dribbling or by running to open spaces. In fact, different types of defenders exist:
  - Centre-backs are the most focused ones on defensive stops, and they do not normally take an active offensive role.
  - Full-backs play on both sides of the field (left / right) and face the opponent wingers.
  - Wing-backs also play at the left / right side of the field, and they usually take a more active role on offense.
- Midfielders are a hybrid profile; they play in the middle of the field and are required to combine skills in a large set of game facets, both in the offensive and defensive ends. Despite not

discussing the greedy specifics of each type, midfielders can be divided into center, defensive, attacking, or wide.

- Forwards are the most offensive players and, normally, the ones in charge of scoring goals. Roughly, there are two types of forwards:
  - Strikers (or center forwards) play in a central position, and they are the ones placed in a more advanced location.
  - Wingers play either at the right or at the left side of the field. Their role is also offensive-based, and they usually transfer the ball from wing-backs to the striker.

Figure 2.4 shows all the roles and basic soccer roles / positions in the field. In the central part of the image, a classical 4-3-3 soccer lineup is displayed together with the position-names of each specific spot. Other distribution combinations include tactics such as 4-2-2 or 4-3-2-1, as seen in the small side-fields of Figure 2.4.

Game phases:
Bearing in mind that in a soccer lineup there are mainly 3 rows of horizontally distributed players, by clustering the 2D coordinates of the players in the field, the ball position can be found in three game states or phases:

1. Build-up phase: the ball is located before the first row of players.

2. Progression phase: the ball is located between the first and the second row of players.

3. Finalization phase: the ball is located between the second and third row of players.

All these phases are shown in Figure 2.5 and will be used *a posteriori*.

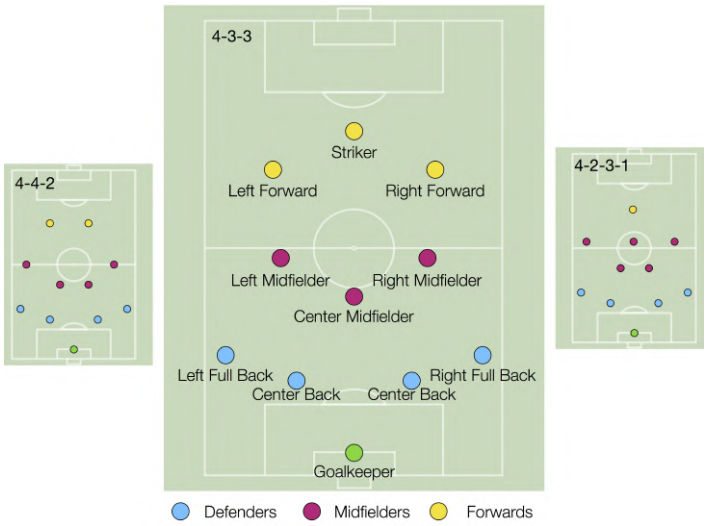

Figure 2.4: Different lineup possibilities: (left) 4-4-2, (center, with specific position names) 4-3-3, and (right) 4-2-3-1. In all these distributions, blue, purple, and yellow dots represent different positions.

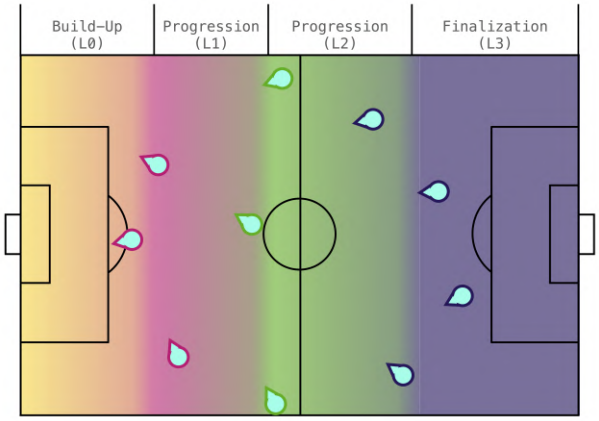

Figure 2.5: The location of the ball with respect to the spatial defensive configuration will indicate the current game phase.

2.3 Open Datasets

In the upcoming Parts, both basketball-based and soccer-based experiments will be performed on top of four private datasets, which contain a notable sample of games, plus high-quality image- and eventing-data. However, for the sake of reproducibility or just in case that the reader wants to perform similar experiments, some open datasets are listed in this Subsection.

- SoccerNet [40], and SoccerNet-v2 [26] include video footage (plus audio) of several soccer games, together with labelled events, thought for performing action spotting challenges.
- SoccerDB [17] is a large-scale soccer database for comprehensive video understanding, also complementary to SoccerNet.
- MetricaSports [73] shared a package of open soccer, including player tracking, eventing data, and images.
- SkillCorner also opened a public soccer dataset [101], which includes 2D tracking data of 9 complete games (without the corresponding frames).
- Despite not including video footage nor tracking, StatsBomb [104] created a public soccer repository with eventing data, together with game logs. Similarly, Pappalardo *et al.* [83] also contributed with a large collection of soccer eventing data, including logs of 7 different soccer competitions.
- When it comes to basketball, few datasets have been publicly shared; the first one includes data from SportsVU, which contains a set of tracking data from NBA games [77] corresponding to the 2015-2016 season, and the second one, created by Mike Beuoy, includes a filtered dataset about individual NBA shooting curves [5]. Apart from professional NBA data, multi-camera amateur datasets have also been shared [90; 128; 39].

What is more, even though open data have not been shared yet, other companies such as SciSports [24; 7], Sport Logiq [96], Stats Perform [99; 107] or Genius Sports [88] made a huge investment in research groups (in some cases, in collaboration with academia) whilst publishing their findings.

Part I:
Player Tracking

If we want machines to think, we need to teach them to see.

Fei-Fei Li

3 Introduction

The inclusion of tracking data has been a key ingredient in the most powerful sports competitions since several professional clubs and organizations started digging data deeper while creating research departments. These departments are in charge of bringing valuable numerical insights to the field / court, thus providing the coaches with novel sources of information that could potentially boost the performance of a player / the whole team. However, since tracking data come at the cost of complex and expensive setups, a strong economical investment has to be made, not only in terms of infrastructure but also in terms of personnel. In the case of basketball, the only competition that sends tracking data to teams is the NBA, where all clubs have the same salary cap (around 109 million dollars per team). Nonetheless, the economical conditions of the NBA have little to do with other competitions; generally, leagues suffer from notable differences budget-wise, thus creating an unbalanced market where the cost of some products is a tiny / large portion of the clubs' resources. Still in the NBA scenario, Stats Perform [103] and Second Spectrum [97] are currently the official tracking providers of the NBA, and their setup consists of an array of more than 10 overhead cameras. Nevertheless, since none of these cameras are used *a posteriori* for television broadcasting, their installation is solely thought for tracking purposes.

The main goal of the first Part of this thesis is to study the viability of automatic tracking systems on top of European basketball video footage. As opposed to the technology being used in the NBA, the upcoming Chapters attempt to build a solid and automatic tracking baseline by only using video footage obtained from the main TV broadcasting camera, which does not involve an extra cost for the league or the club. By making tracking data accessible to European clubs, GM's and coaches could definitely benefit from a data-driven assessment in several decision-making processes (*e.g.* hiring or scouting), hence potentially boosting the performance of the team. Moreover, by creating single-camera multi-trackers based on broadcasting cameras, apart from gathering data from new games, tracking from vintage games' footage could be gathered.

Given the broadcasting-camera video feed, a tracking-by-detection algorithm is adopted: first, the court is identified, and right after, potential players are detected and outliers are filtered out. In the feature extraction process, by quantifying how much do players resemble in different frames, a similarity matrix is obtained. By maximizing the similarity among instances across frames, bounding boxes are matched. Several types of features for establishing the similarities are evaluated:

- Geometrical features, which involve normalized distances (in frame coordinates) between detected targets.
- Visual features, which quantify how different bounding boxes look alike by comparing color similarity metrics in different small neighborhood patches.
- Deep learning features, which are obtained by post-processing the output of a convolutional layer in a CNN.

Besides, we show that the combination of the whole feature extraction process with camera stabilization techniques helps improve the trackers' overall performance, reaching over 68% accuracy in terms

of *Multiple Object Tracking Accuracy* metric. In particular, the implemented camera stabilization method is based on homography estimation and leads to compensated camera-motion sequences, where displacements of corresponding players in consecutive frames are considerably reduced.

The rest of this Part is divided into the following Chapters: in Chapter 4, the state-of-the-art regarding sports tracking is detailed, including both raw-tracking methods and potential applications and metrics that can be built on top of this kind of data. Then, Chapter 5 presents the proposed multi-tracker, including all the required steps: court filtering, player detection, feature extraction, and matching. The obtained results are presented in Chapter 6. Conclusions are drawn in Chapter 7, where future lines of work are also suggested.

4 State-of-the-Art (Tracking and Applications)

In this Chapter, the main state-of-the-art regarding sports multi-tracking methods is detailed, together with several applications that can be built on top of this kind of data.

For clarification purposes, the state-of-the-art techniques of this Chapter are split into the following groups:

1. **Tracking Methods** designed for sports sequences are analyzed in order to compare how players can be tracked from different points of view.

2. **Spatial Analysis of Plays** (basketball-based). Based on tracking data, articles included in this group try to reach a high-level understanding of complex concepts, such as ball movement or spacing strategies.

3. **Metrics Quantification** (basketball-based). In this Section, different basketball-based advanced statistics are quantified stemming from tracking data. Articles are divided into **offensive**, **defensive** and **rebounding** metrics.

4. **Deep Predictions** (basketball-based), which aim to predict a solution of simulated ghosting scenarios by training models with tracking data.

4.1 Tracking Methods

Multi-object tracking has been and is still a very active research area in CV. One of the most used tracking strategies is the so-called tracking by detection, which involves a previous or simultaneous detection step to identify the desired targets in the given scene, and posterior matching across frames. Some of these works, like the one presented by Girdhar *et al.* [41], used a CNN-based detector with a tracking step, while others were based on global optimization methods. Among them, a joint segmentation and tracking of multiple targets was proposed by Milan *et al.* [74], whereas Henschel *et al.* [49] presented a full-body detector and a head detector that were combined to boost the performance; similarly, Doering *et al.* [29] combined Convolutional Neural Networks (CNNs) and a Temporal-Flow-Fields-based method to exploit temporal information. Another family of tracking methods, which achieves a good compromise between accuracy and speed, is based on Discriminant Correlation Filters. More concretely, features are extracted first, and then correlation filters are used. To obtain these features, several approaches have been used, such as hand-crafted methods or deep-learning-based ones (*e.g.* [87]). Results improve when the feature extraction process is learned in an end-to-end fashion, such as in the work by Wang *et al.* [121]. Similarly, Brasó *et al.* [8] also trained a graph-based differentiable framework that was not only used when extracting features but also in the final association step. In this context, a complete overview of different approaches together with their corresponding state-of-the-art results, plus future lines of research, can be found in the survey by Ciparrone *et al.* [18].

Although the vast majority of trackers are presented from a frame-domain perspective (*i.e.* follow certain targets across frames), recent contributions have succeeded when computing 3D multi-object tracking [125], which might be really useful in the fields of robotics or autonomous driving. Besides, in the case of tracking humans, apart from following their position on the screen, extra layers of in-

formation, such as their pose, can be included in the model's target; for instance, Ning *et al.* [81] presented a computationally inexpensive method based on a siamese graph convolutional network that estimated pose properly and achieved a notable performance when matching instances across frames.

The implementation of tracking methods has a large set of potential direct applications, each of them with their corresponding challenges. Among all the researched fields, sports is a demanding one, because cluttered scenarios produce partial or total occlusions and require really precise algorithms. Besides, as it will be detailed throughout the whole manuscript, the acquisition of tracking data in sports can provide really meaningful statistics to the coaching staff. Therefore, sports is a highly defying but extremely rewarding domain for new vision-based tracking methods, not only because of its challenges but also because of its potential use *a posteriori*. The remaining articles of this Section will be chronologically listed and are related to sports-dependent tracking methods. Nonetheless, before getting started, we refer the reader to the book written by Moeslund *et al.* [114] for a general survey of all existing CV techniques applied to sports sequences. This book does not only contain a deep study on player and ball tracking, but other current commercial applications are also detailed, such as camera calibration or broadcast enhancements, including player modeling and analysis of motion players. Another detailed survey about tracking was published by Manafifard *et al.* [69], where the main soccer-based existing methods were analyzed together with the proposed solutions to common challenges, such as field detection or occlusion resolution.

In order to provide some historical context, the very beginning of sports tracking is enclosed in the following contributions. In 2006, Perse, Er *et al.* created one of the first basketball tracking methods to perform data analysis *a posteriori* [85; 30]. With a 2-camera configuration setup in the ceiling of the arena, a method was de-

signed in order to help planning training sessions based on players' movements. Their method created a play-designer module, which contained a playbook of stored templates with different plays. Then, the phase of the game (offensive / defensive / time-out) was estimated by clustering the distribution of players on the court with a Gaussian Mixture Model [105]. Afterwards, the small-scale parts of the game were found: by dividing the court into 9 sections, basic events were used in order to define the player motion on the court. Finally, recognition was performed by using the stored templates in the play-designer. Although their dataset was not huge, their results were consistent; nevertheless, there was no ball information and the algorithm did not have the possibility of learning new plays on its own. Two years later, Fleuret *et al.* presented one of the most-cited publications in the *tracking in sports* research field [36], which used video footage from different cameras to track individuals through a probabilistic occupancy map. Their method used background subtraction to estimate the probability of occupancy at each spot of the plane, and a generative model was applied: ideal synthetic images were created by modeling humans as rectangles at each spot where a potential person could be identified. Occupancy probabilities were then re-approximated by using the Kullback-Leibler divergence; basically, this method computed how the probability of a synthetic image changed when comparing with the initial prediction. Once found marginal probabilities, individuals were tracked by combining color and motion cues together. This method differed from previous techniques –which pretended to perform detection at every single frame– by computing the global optima of scores when having a long sequence of frames. Seven years later, authors from the same research group combined the above-mentioned method with a novel technique to automatically track the ball in team-sport sequences [68]. The main challenge to be solved was the complex interactions that happen when players perform any kind of action involving the sphere. Since classical 2D circle detection algorithms might fall short to track the ball, their method modeled ball tracking as a graphical problem,

where (a) position, (b) state, and (c) available image evidence were quantified at each step; its goal was to maximize an energy function by computing feature correlation and temporal smoothness. Besides, a physical model was used to impose some constraints on ball motion: for instance, they introduced a prior that expressed how zero acceleration had to be taken into account if the sphere was in a *flying free* state. Their results outperformed existing contributions using some basketball datasets such as APIDIS [23].

In terms of action spotting, Ramanathan *et al.* [90] published a method to recognize events and key actors in multi-person videos by detecting the focus of attention of different basketball plays. The goal of this research was to amend the lack of a universal method to emphasize attention or include key actors in sport sequences. Once labeled a large set of plays, they extracted features for every class, including both scene and particular player information; right after, a deep learning framework was used to classify. To properly track the players, the Lucas-Kanade tracker [6] was implemented in combination with a bipartite graph, which was used for matching. Their event detection method was done through a sliding window technique that displayed attention with a heat-map. Results outperformed some state-of-the-art methods, and their dataset was shared publicly. However, the number of classes was simplified to a few similar plays (*i.e* 2-points shot success / failure, 3-point shot success / failure). Still in the basketball field, two datasets have been recently shared. On the one hand, the authors of [39] created a solid annotated dataset containing street-ball footage and compared state-of-the-art trackers to their approach, which was based on joint detection and embedding. On the other hand, Wu *et al.* [128] used a relatively more expensive setup, and created a dataset where cameras were placed at human height and contained overlapped regions among them. In the latter, the association step between frames was achieved by a clustering method that computed metrics among tracklets.

Apart from basketball trackers, state-of-the-art tracking methods do exist in other disciplines. For instance, in the case of soccer, Kim *et al.* [58] succeeded in tracking multiple targets by approaching the matching process as a multiscale foreground-sampling problem, where dissimilarity metrics could express how much did detections resemble. More recent contributions in the soccer domain included self-supervised methods, as the one proposed by Hurault *et al.* [52], which can be used in challenging video footage; by training two networks in a student-teacher fashion, notable performance was achieved even with low-resolution players. Other tracking solutions have been published from a general sports perspective using multi-camera setups, such as the ones presented by Zhang *et al.* [131], which stemmed from a deep player identification, or by Liang *et al.* [61], which adapted a complete k-shortest framework in order to perform the matching process.

4.2 Spatial Analysis of Plays

In the paper written by Lucey *et al.* [65], the authors analyzed how teams managed to have open shots in order to improve shooting percentages. The motivation of this paper emerged when checking the statistics of the NBA teams, as the authors realized that there is a notable drop in shooting percentages when attempting pressured shots (almost a 15% decrease in some cases). First, their algorithm assigned a role (position) to every player at the beginning of the action. Then, the different factors that might affect when attempting a shot were checked from a more analytic point of view; these included features such as the closest distance from a defender, the shooter's speed, the number of dribbles, or the number of seconds the player kept the ball before shooting. Finally, different plays were retrieved using tracking data, which clustered similar plays into permutations from the original one (the exact same action will not occur twice in a game). Extracted results showed that one of the most relevant fea-

tures to get an open shot was the defending switches that might occur during the game, which generate mismatches[1]. Although it is rather a statistics-based paper, this project also aimed to extract relevant information from tracking data to improve the understanding of the game.

With Sports VU raw tracking data, Wang and Zemel [119] designed an algorithm to classify a closed-set of plays using Recurrent Neural Networks (RNN), with the purpose of generating detailed reports with a high-level basketball understanding. Their approach turned tracking data into pictorial representations in order to deal with an image classification problem. Positions of the players were estimated by comparing their shooting tendencies and frequencies in different positions in the court (*e.g* an exterior player usually moves behind the 3-point line and attempts more long-range shots than an interior one), and they built an *anytime* prediction system, as one same play might change due to defensive strategy. Their results (expressed with *top-1 accuracy*) seemed to be promising, but the system was thought for a particular team in a specific season, so it was not automatically tuned to any kind of team.

Miller and Bornn [76] made another relevant contribution. They organized a large set of plays by grouping structural similarities, as they observed that there was not an efficient scouting method for professional basketball teams. Their goal was achieved through: (1) segmentation of short plays to shorter manageable segments (modeled with Bezier curves), (2) possession modeling by adapting topic models, and (3) a bag-of-words structure. Finally, having clustered data with nearest-neighbors algorithms, different types of analysis were done. Although the attached videos showed promising results, no numerical evidence was displayed. This work was an improvement

[1] A mismatch occurs when a big player has to guard a small one or vice versa. *A priori*, in these situations, big players can take advantage of their strength, whilst small ones can take advantage of their speed.

of a previous contribution of the same authors (Miller *et al.* [75]), where the actions occurring in a basketball court were analyzed by a point process factorization based on intensities.

A low-cost approach towards play recognition was introduced by Arbués-Sangüesa *et al.* [2]. In this work, small sensors gathered tracking data of youth basketball practice; after the corresponding parsing processes, a classical Machine Learning model was trained (in a supervised way) in order to classify a closed-set of 5 basketball plays; the extracted feature vector contained 56 characteristics of that play, relevant from the coach point-of-view (initial display, distance among players, speed...). Results showed 98% accuracy using cross-validation and principal component analysis [55] in order to avoid overfitting, but the set of plays being used contained roughly 100 instances, so there is still plenty of room for improvement in this field. This method could be transferred to European leagues with fewer resources.

Finally, Bornn *et al.* [72] focused on a really challenging problem: do teams have an identity that could be described from displacement patterns? Having Sports VU tracking data, three experiments were performed: (1) to recognize a team from a single possession, (2) to recognize a team from the complete set of possessions of a game, and (3) to see what the star players' spacing impact is. The deep trajectory network took as input the stack of all player trajectories during a series of time, and two 1D convolutional layers were used (together with two corresponding pooling layers). Besides, in the first couple of experiments, the origin of trajectories was always considered to be the ball, so it could be said that the input vector was a set of distances with respect to this *anchor*. Results showed that identifying a team from a single possession was really difficult (24% accuracy), but bringing together the set of all game possessions improved team recognition to 95%. In the third experiment, the anchor was switched to the star player, which was manually selected; this case tried to ex-

emplify, for instance, how different it was to identify Golden State Warriors from a single possession with the presence of Stephen Curry on the court, and accuracy reached 43% for a single possession. Besides, the same network generalized to other sports such as football.

4.3 Metrics Quantification

In this Section, different articles containing novel metrics to quantify intangible basketball aspects will be summarized in a chronological way. First, offensive metrics will be explained; then, defensive ones, and finally, rebounding or other techniques.

One of the first interesting offensive metrics was introduced by Kirk Goldsberry [42], who presented new visual and spatial analytics to determine who was the best shooter in the NBA. The problem he tried to solve was that the league leader in field goals percentage (measured by dividing the number of scored shots by the total number of attempts) tends to be a *center* who takes no mid- / long-range shots; therefore, the goal was to define a metric to determine who was the player that shot better from as many court spots as possible. His system was built on top of a composite shot-map for all the shots attempted in 5 different seasons (2006-2011), finding 1284 unique shooting cells. Then, spread parameters were defined and weighted by their distance to the basket (number of cells with acceptable accuracy), thus favoring those players that attempted long-shots with high reward (3 points). This metric definitely penalized those *centers* that did not take risky shots, and provided a robust knowledge on how well players shots. Obtained rankings proved to be precise, as those coincided with the opinion of basketball journalists when talking about the top-5 shooters in the league.

Cervone *et al.* [14] presented a new way to mathematically model how

good the decision-making process of players during a game possession in real-time was since *basketball IQ* is one of the most important features when GM's seek for new player hirings. The authors defined *Expected Possession Value* (EPV) as a metric that expressed the expected points to be scored / received at any moment; then, having the position of the player driving the ball, they modeled his/her added value by dividing eventing data into macrotransitions (shoot / pass / turnover) or microtransitions (basic movement). With EPV metrics, two applications were shown: (a) a ranking of the NBA players who made better decisions and (b) an equation to measure the shot satisfaction, which could help to identify selfish attitudes. Both applications showed adequate results and proved that EPV models are a promising baseline when building data-driven tools.

Although neither CV nor machine learning was applied, based on the previous work of Goldsberry [42], Marty tried to add more dimensions to understand why shooters miss [70]. In this high-resolution method, three players attempted a total of 22 million shots, and not only their position in the court but also the interaction with the rim were gathered with the Noahlytics system (a sensor placed above the rim). The main goal was to obtain a deep analysis of right-left (and left-right) deviations when shooting from several positions and angle values to correct flat shots. Results helped to indicate where players should charge for the rebound (*i.e.* in a left-corner 3-point shot, they should generally charge the right side, close to the baseline), but the analysis of this technique was still naive, as the method was tested only with three different players of different shooting percentages.

Goldsberry and Weiss [43] attempted to quantify defensive metrics of NBA basketball games. The motivation emerged from the isolation of defensive concepts in NBA box scores, where only defensive rebounds, steals, and blocks are annotated. Their contribution was called *the Dwight Effect*, and they wanted to prove that the leader of the league in blocks might not be the best defender, but the player

who changes the shooter's behavior and efficiency more often. In this article, and using Sports VU tracking data once again, they first separated frequencies and effectiveness of different kinds of shots of every player in the NBA; then, they computed the *basket proximity*, which is the balance between the percentage in field goals and the number of avoided shots when a certain interior player contests the shot. Afterward, *shot proximity* was estimated by checking how often an interior player was close to a shot attempt. Their results were meaningful from the point of view of a basketball coach, as a single metric summarized several factors regarding rim protection. However, this quantification was restricted to interior players. In order to complement this work, Franks *et al.* [38] presented new defensive metrics for exterior players, including the *Volume Score*, which contained the magnitude of shot attempts in front of a certain defensive player, the *Disruption Score* expressing the effectiveness of those shots and *Counterpoints*, which indicated who was responsible for contesting a certain shot. This analysis was based on: (1) modeling the evolution of defensive matchups (different swaps when defending a team) over the course of possession as a Markov Model, and (2) the posterior computation of the mentioned metrics using logistic regression plus predicting the *a priori* efficiency of a shot.

Another interesting quantifiable defensive metric was introduced by McIntyre *et al.* [71], who analyzed how NBA teams defended ball screen situations considering 4 different options (over, under, trap or switch). Their goal was to quantify not only which were the most repeated strategies but also the most efficient ones. This contribution enabled novel analysis of defensive strategies using Sports VU tracking data. Their method had a validation set that comprised manual annotations of ball screen situations of 6 different basketball games (a total of 199 instances). Then, using an algorithm based on pairwise distances within players, 270853 ball screen situations were tested, obtaining 69% accuracy on three classes (*traps* could not be included because of a small number of samples); besides, the defensive effort /

strategy of the teams was shown, which provided interesting metrics to identify the most aggressive teams in the NBA. If the validation set had been larger, greater accuracy would have been obtained, which could have lead to a robust system to be used in professional games.

Reinforcement learning was later introduced by Wang *et al.* [118] in order to check if it was worth practicing double-team defense in NBA games. Generally, all NBA teams have a *Star Player* who takes a lot of shots. Coaches prepare special defenses for this kind of players: at some point, instead of opting for an individual defense (five defenders on five offensive players), coaches might want to try the alternative of double-team defense: two defenders guard the *star player* when he/she has the ball, and the other three defenders try to occupy spaces and to contain the remaining four. The risk is obvious: if the double-teamed player manages to give a good pass, it will be an easy offensive situation for the opponents. Once analyzed all possessions of NBA teams during one season, double-teams were detected by a simple rule-based on timings and distances of defensive players. Then, reinforcement learning was applied: within this framework, an agent observed the current state (offensive situation), chose an action (double-team or not), and transitioned to another state according to a probability distribution and a Markov Decision Process. Besides, the authors introduced a policy to affect the decisions made by the agent, which included game conditions; for instance, it might not be a good idea to double-team the best passer in the last seconds of a tied game. Finally, the agent received an instantaneous reward (the other team did score or not). Their results ranked the best and worst double-team defensive pairs and teams, apart from analyzing the defense of all teams against the Cleveland Cavaliers in the 2016-2017 season, which had one of the most dominant NBA players (LeBron James) at that time. However, their conclusions were not that clear, as the trained model suggested to double-team James less and to apply this defense to worse players, but there were not enough data of double-team defenses over *non-star* players.

Besides, other metrics were also introduced to contextualize rebounds with the purpose of numerically identifying whether a player captures a rebound all alone or grabs it after hustling with other players. Maheswaran *et al.* [66] deconstructed the rebound by checking the factors that influenced this type of action. First, they filtered Sports VU tracking data to end up only with rebound observations and they built a heat-map with all these locations (around 11000 instances). Right after, rebound location probabilities were checked given the shot position (distance and angle); from these regions, another heat-map was built, containing the coordinates where the ball decreased from 8 feet, which indicated the potential rebound location. Given the position of all players, the presented model aimed to predict who had more chances to catch the rebound as the action went forward. Their results showed that in mid-range shots, the probabilities of grabbing an offensive rebound were low and that there was not a significant directional bias depending on the shot location.
The same authors [67] extended their contribution by analytically decomposing the rebound into three concrete factors. *Positioning* (modeled with a Voronoi region) was used to see the position of a player when: (a) there was a shot and (b) few seconds after it. These coordinates helped to indicate the player's intention: he/she could either try to capture an offensive rebound (also known as *crashing*) or he/she could retreat to a defensive position. The second factor was *Hustle*, which told if a player was able to create a rebound opportunity despite not being at the best initial spot. Finally, *Conversion* estimated if a certain player allowed others to grab rebounds when he/she had the best positioning; that is, if a player captured easy rebounds or not. Once again, their results were shown in different rankings and coincided with the experts' opinions. However, these same experts could argue that *Positioning* might not be a skill, but a matter of luck or other factors.

Furthermore, Wiens *et al.* [126] conducted more concrete research

to analyze only offensive rebounds, trying to quantify the trade-off between two strategies: attacking the offensive rebound (*crashing*) and retreating to a defensive position. Having filtered Sports VU tracking data and gathered only offensive rebound situations after mid- / long-range jumpshots, a reaction time was established. Specific metrics were defined: *odds ratio* (probability of a good event to occur) and *net gain*, which indicated the possibility of scoring having grabbed the offensive rebound combined with the possibility of preventing the other team to score having retreated on the defensive end. Once modeled *threat neutralization* (how effective the defensive transition in terms of pairwise distances between players is), results showed that *crashing* is a risky strategy, and an early *threat neutralization* limits the negative impact of transitions. Anyway, this article should be tested again with the inclusion of more data, as it only had the strategies of 12 teams (and few observations were obtained for some of them).

4.4 Deep Predictions

Another trend within sports analytics has been the need of obtaining data-driven answers from *what-if* scenarios, which might help *a priori* in terms of strategizing games or competitions.

On the one hand, the work by Seidl *et al.* [98] introduced the concept of *ghost defenders*. This research emerged from the need of coaches to design perfect plays, which is one of the most difficult challenges for them, as it is almost impossible to take all details into account: the score, the remaining time, the players that have to be on the court, the type of defense that the other team may perform... The first module of this work was an interface, which worked on any digital surface; once the coach sketched which play does he/she wanted, an animation could be seen with the movements of all players on

the screen. Then, a deep learning model predicted how the defense would adapt to that play at that given scenario (ghost defenders), and finally, the same program suggested similar plays that could optimize the outcome. The network to be implemented was a Recurrent Neural Network of variable length sequences, where individual trajectories were set as the input and were modeled using a two-layer long short-term memory architecture (LSTM). They determined, at a given point, where the player should go next given a specific role. It has to be mentioned that each player had his/her own policy, and those were trained by computing the distance between predicted and real positions on existing Sports VU tracking data. Results proved to be promising, and even complex basketball concepts such as weak-side helps were being taken into account; besides, given that enough data of each team were fed into the network, characteristics for each particular game were considered.

On the other hand, the work of Sandholtz and Bornn [95] aimed to analyze the new trends of NBA games, with teams shooting more long-range shots than ever. In a basketball possession, the ball-handler might choose within several possible actions: passing, shooting, dribbling... These states could be modeled with Markov Processes, but the main problem was that transition probabilities are not stationary: *i.e.* if there is only one second left in the shot clock, the shooting probability is almost 1, so a policy was introduced to take the environment into account. Once again, there was a reward function to be maximized by altering the policy; this function simply answered the question "*how many points do we expect to get after player X, in a state S, decides to take action Z?*". The main problem was then how to model the shot probability, as it is a latent skill that depends as well on the position of the shot; the authors proposed a Bayesian logistic regression model that clustered players with similar shooting characteristics. Results, which were obtained on top of data from the 2017-2018 season, showed that if teams shot 20% fewer contested mid-range shots at the end of each possession and took more open

long-range shots at the beginning, the total scoring points per game of all teams of the NBA would improve. Another experiment showed that, if teams shot 90% fewer mid-range shots, the expected outcome would decrease, as these shots would have not been properly selected.

5 Proposed Multi-Tracking Method

In this Chapter, the proposed single-camera multi-tracking algorithm is presented, which aims to track multiple players at a time in (single view) basketball footage. In order to track all players on the basketball court, a tracking-by-detection approach is used. First, the court is filtered by merging line detectors with basic segmentation or color filters. Then, pose models are used to detect the players in the image; moreover, the underlying location of body parts is later used to extract features, either from a visual or DL perspective. Finally, the matching process is also detailed, which associates detections across frames.

5.1 Court Filtering

Individual frames belonging to basketball footage include much more content apart from all 10 players and the ball, such as fans, real or tv-synthetic scoreboards, bench players, coaches... Consequently, before getting started with tracking methods, a pre-processing stage is required to delimit the region of interest where the desired targets (players on the court) can be found. In particular, our approach consists of segmenting the court region, which is a rectangular area whose projection to the camera results in a trapezoid. Thus, the filtering

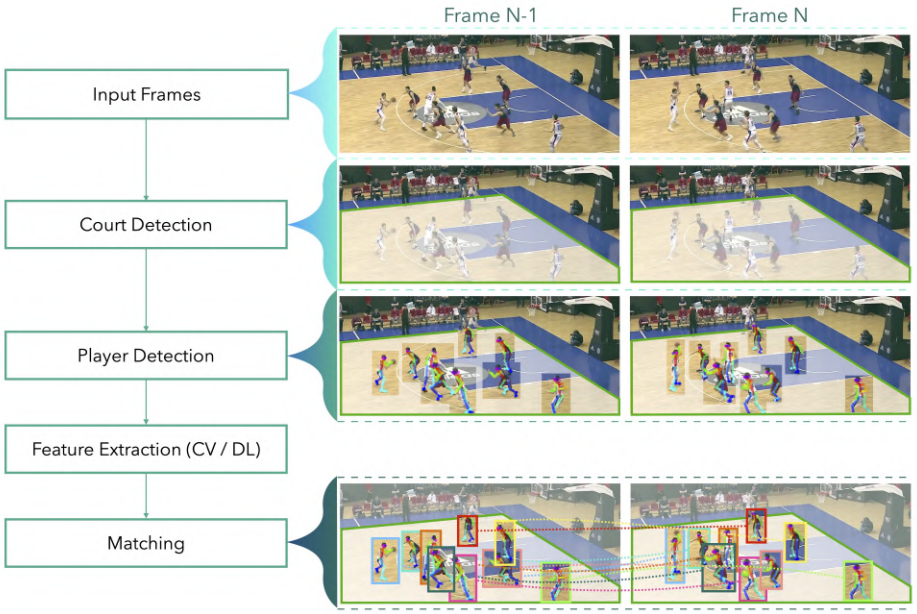

Figure 5.1: Generic Pipeline: for each frame, players are detected (through pose models) and tracked (via feature extraction and matching).

challenge is reduced to the identification of visible court boundaries in the image: the sidelines and baselines (from 1 to 4 depending on the camera's point of view). Frequently, some of these court boundaries are only partially visible due to occlusions, or even not visible at all, as shown in Figure 5.1.
The method starts by detecting all the line segments in the image using a fast and robust parameter-less method [116]. Right after, dominant lines, *i.e.* lines with the longest visible parts, are estimated employing a voting procedure. Those lines will correspond, in general, to the sidelines / baselines or, in cases of strong occlusions, to court lines parallel to the sidelines / baselines. The strength of the vote of each line is proportional to the sum of detected segments' length on the line, as seen in Figure 5.2, where the detected segments are displayed in yellow. Given that in broadcasting sequences only one baseline (or none) can be seen at a time, and that even in cases that both sidelines are in the field of view of the camera one of them may appear occluded by the public (*e.g.* Figure 5.3), the purpose is to find a *horizontal* dominant line (either a sideline or its orientation) and a *vertical* dominant one (a baseline or its orientation). Horizontal lines are considered to be the ones which intersect the image at the left and right boundaries (Figure 5.2(a)), while vertical ones intersect in one of the following pairs of image sides: top-left, bottom-left, top-right or bottom-right (examples in Figure 5.2(b)-(c)). In order to find the location of court boundaries, the playing area is pre-segmented and the set of lines with a dominant orientation that better delimits the court is selected through an iterative process. However, an important facet has to be taken into account once the dominant orientations are found: the detected segments used to determine the dominant orientation might not be part of the desired baseline / sideline. That is, the mentioned set of lines might not only contain the first candidate but also all the other parallel candidates that could potentially fit. While in the case of the baseline lines are distributed from the top to the bottom of the image, when dealing with sidelines the line distribution goes from left to right. Two differ-

ent solutions are proposed to pre-segment the court in two different professional basketball scenarios: (a) European, and (b) NBA games. For NBA games (Figure 5.3-top), the scenario is challenging, because there is almost no space between sidelines and fans. In order to find the horizontal boundaries, instead of checking for color components, Conditional Random Fields [134] is applied at a coarse resolution to find the total area of the regions in the image domain containing people. Once having this rough estimation, an iterative algorithm is applied to delimit court boundaries: at the very beginning, two line candidates with the dominant orientation are placed at the top and bottom of the image; then, for each iteration, these lines are iteratively moved towards the middle until convergence. In each iteration, the product of the following percentages is computed: (a) people-pixels above the top line, (b) people-pixels below the bottom line, and (c) non-people-pixels below the top line and above the bottom one. If there is a drop in either the first or the second percentage, the position of the corresponding line is fixed; convergence is reached when both lines stop moving. Potentially, in the horizontal court limits, the product of these three terms will correspond to a maximum, meaning that there is a large contribution of people pixels above and below the top and bottom line respectively (corresponding to fans), and a small contribution in between (corresponding to the court with a maximum of 10 players plus 3 officials). Once both baselines are set, and having masked the original image, the best vertical candidate is found in the same way but scanning only from left to right or from right to left, depending on the situation. In European games (Figure 5.3-bottom), court surroundings usually share the same color, and fans sit far from team benches. For this reason, a basic color filter (in the HSV colorspace) is created; for each possible line candidate, the contribution of pixels that satisfy filter conditions is checked at both right-left (vertical) or above-below (horizontal) sides of the tested candidate. The horizontal and vertical candidates with the highest response will be then considered as court limits.

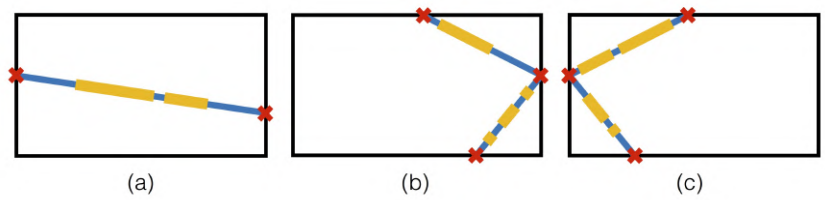

Figure 5.2: Line contributions with potential detections (and occlusions): (a) sidelines, (b)-(c) right-left baselines, respectively.

Figure 5.3: Court detection results in different scenarios: (top row) NBA, and (bottom) European games

5.2 Player Detection

As mentioned, the presented tracker is based on multiple detections in each individual frame. More concretely, the implemented method relies on pose models techniques [89; 124; 11] stemming from an implementation of OpenPose [21] (more details about pose models can be found in Section 2.1). Given a basketball frame, the output of the main inference pose function returns a 25×3 vector for each player, with the position (in screen coordinates) of 25 keypoints, which belong to the main biometric human-body parts, together with a confidence score. Note that there might be situations where specific parts might not be detected, resulting in unknown information in the corresponding entry of the pose vector of the whole skeleton. In addition, 26 heatmaps are returned, indicating the confidence of each part being at each particular pixel. By checking all the parts' positions and taking the minima and maxima XY coordinates for each detected player, bounding boxes are placed around the respective players as displayed in Figure 5.4.
Since the presented person detection method does not include priors such as a maximum number of players on the basketball court or information on both team uniforms, the set of detected people might include some referees (an example is displayed in Figure 5.5).

Besides, in order to ease the tracking of the players, an additional camera stabilization step to remove the camera motion can be incorporated. Taking into account that its inclusion represents extra computations, an ablation study is provided in Chapter 6 to discuss the extent of its advantages. When enclosed, the camera stabilization method and implementation proposed by Sánchez *et al.* [94], based on homographies, is used.

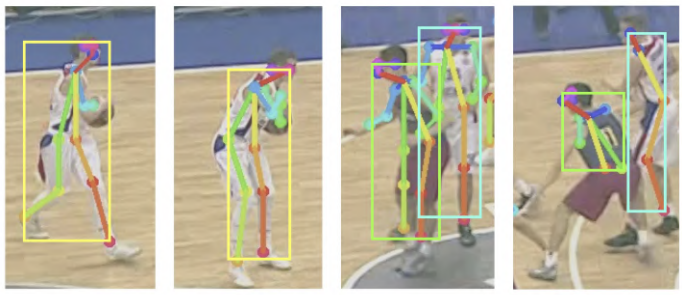

Figure 5.4: Detected parts with the corresponding bounding box.

5.3 Feature Extraction

Once bounding boxes are obtained across frames, the forthcoming step should consist of assigning individual tracks to each one; nonetheless, prior to that, all boxes must be characterized. With the purpose of quantifying this process, different approaches can be used whilst extracting features. For the remaining part of this Section, B_{t_1} and B_{t_2} are considered as two different bounding boxes, detected at t_1 and t_2 respectively.

5.3.1 Geometrical Features

A classical approach can be used to measure distances or overlapping between bounding boxes in different frames. If the temporal resolution of the video is not coarse, it can be assumed that players' movements between adjacent frames are not large; for this reason, players can be potentially found at a similar position in screen coordinates in short time intervals, so the distance between bounding boxes' centroids can be used as a metric. That is, given $\mathbf{x}_{B_{t_1}}$ and $\mathbf{x}_{B_{t_2}}$ as the centroids of two bounding boxes, the normalized distance

Figure 5.5: Obtained results in adjacent frames, where all players (and referee) inside the playing court are properly detected (bounding box) and tracked (color identifier).

between centroids can be expressed as

$$C_d(B_{t_1}, B_{t_2}) = \frac{1}{\sqrt{w^2 + h^2}} \|\mathbf{x}_{B_{t_1}} - \mathbf{x}_{B_{t_2}}\|, \tag{5.1}$$

where w and h are the width and the height of the image domain. Another similar metric that could be used is the intersection over union between boxes, but due to the fact that basketball courts are usually cluttered and players move fast and randomly, it was discarded.

5.3.2 Visual Features

Distances might help distinguish basic correspondences, but this simple metric does not take into account key aspects, such as the jersey color (which team do players belong to) or their skin tone. For this reason, a color similarity is be implemented in order to deal with these situations. Moreover, in this specific case, knowing that body positions are already obtained, fair comparisons can be performed, where the color surroundings of each body part in t_1 will be only compared to the neighborhood of the same body part in another bounding box in t_2. Nevertheless, it has to be pointed out that only the detected pairs of anatomical keypoints in both B_{t_1} and B_{t_2} (denoted here as $\mathbf{p}_1^k$ and $\mathbf{p}_2^k$, respectively) will be used for the computation. The color and texture of a keypoint can be computed by centering a neighborhood around it. That is, let $\mathcal{E}$ be a squared neighborhood of 3×3 pixels centered at $\mathbf{0} \in \mathbf{R}^2$. Then,

$$C_c(B_{t_1}, B_{t_2}) = \frac{1}{255|S|\,|\mathcal{E}|} \sum_{k \in S} \sum_{\mathbf{y} \in \mathcal{E}} \|I_{t_1}(\mathbf{p}_1^k + \mathbf{y}) - I_{t_2}(\mathbf{p}_2^k + \mathbf{y})\| \tag{5.2}$$

where S denotes the set of mentioned pairs of corresponding keypoints detected in both frames, and $|S|$ and $|\mathcal{E}|$ the cardinal of S and $|\mathcal{E}|$, respectively.

5.3.3 Deep Learning Features

Deep Learning (DL) is a widely-explored research field with many possible applications, such as classification, segmentation, or body pose estimation. The basis of any DL model is a deep neural network formed by several layers, which serve to predict values from a given input. Convolutional Neural Networks (CNN) are special cases in which weights at every layer are shared spatially across an image, and their impact results in a reduced number of required parameters, hence gaining robustness to image transformations. Then, a CNN architecture is composed by several kinds of layers, being convolutional layers the most important ones, but also including nonlinear activation functions, biases, etc. This type of layer computes the response of several filters by convolving with different image patches. The associated weights to these filters, and also the ones associated to the non-linear activation functions, are learnt during the training process (in a supervised or unsupervised way) in order to achieve maximum accuracy for the concrete aimed task. It is well known that the first convolutional layers will produce higher responses to low-level features such as edges while posterior layers correlate with mid-, high- and global-level features associated with more semantic attributes.

In the presented experiments, the popular VGG-19 network [100] is used for feature extraction, initialized with weights trained on the ImageNet dataset [27]. The original model was trained for image classification, and its architecture consists of 5 blocks with at least 2 convolutional layers, and 2 fully-connected layers at the end that output a class probability vector for each image. The network takes as input a $224 \times 224 \times 3$ image, and the output size of the second convolutional layer of each block is shown in Table 5.1.

In order to feed the network with an appropriately sized image, a basic procedure is followed as seen in Figure 5.6: considering that player boxes are usually higher than wider, and having the center of

	Width	**Height**	**Nº Filters**
b2c2	112	112	128
b3c2	56	56	256
b4c2	28	28	512
b5c2	14	14	512

Table 5.1: Output size of VGG-19 convolutional layers. In the first column, b stands for block number and c stands for the convolutional layer number inside that block.

the bounding box, its height H_{B_t} is checked. Then, a squared image of $H_{B_t} \times H_{B_t} \times 3$ is cropped around the center of the bounding box; finally, this image is resized to the desired width and height (224 and 224, respectively). In this way, the aspect ratio of the bounding box content does not change.
However, extracting deep learning features from the whole bounding box introduces noise to the feature vector, as part of it belongs to the background (*e.g.* court). Therefore, features are only extracted in those pixels that belong to detected body parts, resulting in a quantized vector with a length equal to the number of filters. Moreover, apart from its length, the obtained output from a convolutional layer is smaller in terms of width and height with respect to the input image, since pooling operations are applied throughout all the network architecture. Therefore, the original 2D location of each body part in the image domain has to be resized according to the convolutional output shape in order to find the corresponding downscaled location. Note that the final vector is normalized with L2 norm. An example using the 10th convolutional layer of VGG-19 is shown in Figure 5.7, where a $1 \times (25 \times 512)$ vector is obtained.

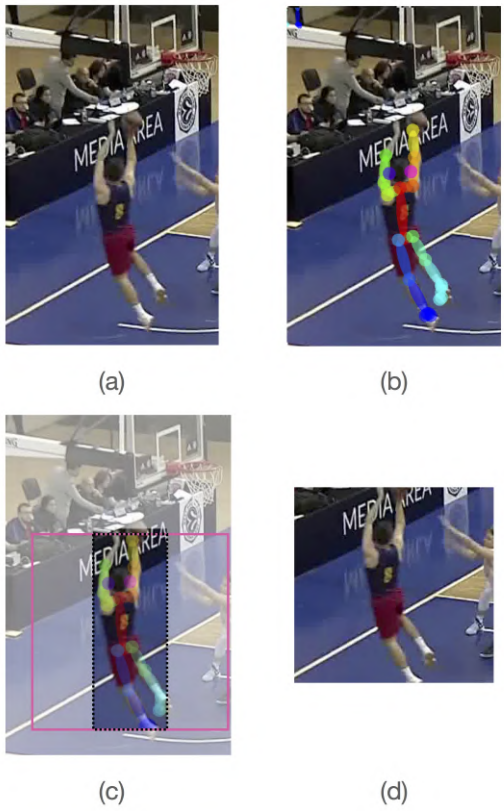

Figure 5.6: Player and Pose Detection: (a) image patch centered around a detected player, (b) detected pose through pretrained models, (c) black contour: bounding box fitting in player boundaries, pink: bounding box with default 224×224 pixels resolution, (d) reshaped bounding box to be fed into VGG-19.

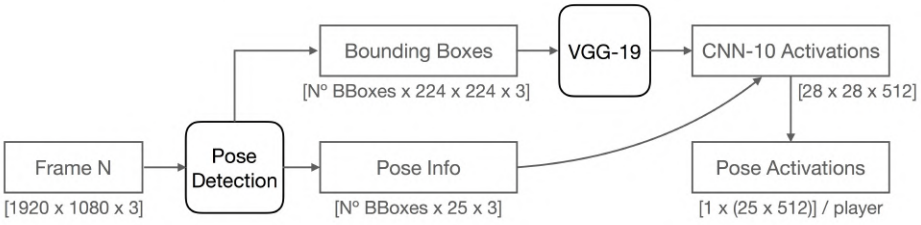

Figure 5.7: Feature Extraction of all body parts using the 10th convolutional layer of a VGG-19 network.

Once all boxes have their corresponding feature vectors, the metric defined in [122] is used to quantify the cost error; in particular, the cost between two feature vectors $f_{t_1,k}^{y_{t_1}}$ and $f_{t_2,k}^{y_{t_2}}$, belonging to bounding boxes detected in t_1 and t_2 respectively, can be defined as:

$$\mathrm{CE}(f_{t_1,k}^{y_{t_1}}, f_{t_2,k}^{y_{t2}}) = \frac{\exp(-f_{t_1,k}^{y_{t_1}} \cdot f_{t_2,k}^{y_{t2}})}{\sum \exp(-f_{t_1,k}^{y_{t_1}} \cdot f_{t_2,k}^{y_{t2}})} \tag{5.3}$$

where k corresponds to the particular body part and y_{t1} and y_{t2} to the pixel position inside the neighborhood being placed around the keypoint. Therefore, the total cost when taking all parts into account is defined as:

$$C_{DL}(B_{t_1}, B_{t_2}) = \frac{1}{|S|} \sum_{k \in S} \min_{\substack{y_{t1} \in \mathcal{E} \\ y_{t2} \in \mathcal{E}'}} (\mathrm{CE}(f_{t_1,k}^{y_{t_1}}, f_{t_2,k}^{y_{t2}})) \tag{5.4}$$

where $\mathcal{S}$ corresponds, once again, to the set of detected parts in both frames, and $\mathcal{E}$ and $\mathcal{E}'$ correspond to the set of pixels in the neighborhood placed around each keypoint.

Nevertheless, two important remarks have to be pointed out:

1. Some of the detected pose parts have a low confidence associated value; note that the confidence range goes from 0 to 1. Since in the given dataset a mean of 16.2 parts are detected per player, the ones with lower confidence can be discarded while

preserving proper performance. In particular, all parts with lower confidence than 0.3 are not taken into account when extracting features. Hence, the subset $\mathcal{S}$ in Equations (5.2) and (5.4) considers all detected parts in both bounding boxes that satisfy the mentioned confidence threshold.

2. Convolutional layer outputs (as implemented in the VGG-19) decrease the spatial resolution of the input. Since non-integer positions are found when downscaling parts' locations (in the input image) to the corresponding resolution of the layer of interest, the features of the $N \times N$ closest pixels at that layer are contemplated. Then, the cost will be considered as the most similar feature vector to the $N \times N$ target one given. In Tables 6.3 and 6.4 a discussion on the effect of the approximate correct location is included.

5.4 Matching

Having quantified all bounding boxes in terms of features, a cost matrix containing the similarity between pairs of bounding boxes is computed by combining the different extraction results. In the presented experiments, the following weighted sum of different costs has been applied:

$$C(B_{t_1}, B_{t_2}) = \alpha C_{Feat1}(B_{t_1}, B_{t_2}) + (1-\alpha) C_{Feat2}(B_{t_1}, B_{t_2}) \qquad (5.5)$$

where C_{Feat1} refers to C_d given by (5.1), C_{Feat2} refers either to C_{DL} in (5.4) or C_c in (5.2) and $\alpha \in [0, 1]$. From this matrix, unique matchings between boxes of adjacent frames are computed by minimizing the overall cost assignment:

1. For each bounding box in time t_N, the sorted list association costs (and labels) among all the boxes in t_{N-1} is stored in an $A_{t_N, t_{N-1}}$ matrix.

2. If there are repeated label associations (*i.e.* two or more boxes in t_N associated to the same box of t_{N-1}), a decision is made in terms of cost:

   - If the cost of one of the repeated associations is considerably smaller than the others (by a margin of more than 10%), this same box is matched with the one in the previous frame.
   - If the cost of all the repeated associations is similar (within a range of 10%), the box with the largest difference between its first and second minimum costs is set as the match.
   - In both cases, for all boxes that have not been assigned, the label of their second minimum cost is checked too. If there is no existing association with that specific label, a new match is set.

3. In order to provide the algorithm with more robustness, the same procedure described in steps 1 and 2 is repeated with boxes in t_N and t_{N-2}. This results in an $A_{t_N,t_{N-2}}$ matrix.

4. For each single box, the minimum cost assignment for each box is checked at both $A_{t_N,t_{N-1}}$ and $A_{t_N,t_{N-2}}$, keeping the minimum as the final match. In this way, a 2-frame memory tolerance is introduced into the algorithm, and players who might be lost in one frame can be recovered in the following one.

5. If there are still bounding boxes without assignments, new labels are generated, considering these as new players that appear on the scene. Final labels are converted into unique identifiers, which will be later used in order to compute performance metrics.

6 Tracking Results

In this Chapter, a detailed ablation of quantitative tracking results is provided and discussed, comparing all the above-mentioned techniques and combinations (types of features, the inclusion of memory, camera stabilization...). Besides, the content of the gathered dataset is explained.

A dataset of 22 European single-camera basketball sequences has been used. Original videos have a full-HD resolution (1920×1080 pixels) and 25 frames per second, but in order to reduce the computational cost, only 4 frames are extracted per second. The included sequences involve static plays of offensive basketball motion, with several sets of screens / isolation play; moreover, different jersey colors and skin tonalities are included. The court is a European one for all situations, and there are no fast break / transition plays. The average duration of these sequences is 11.07 seconds, resulting in a total of 1019 frames. Ground-truth data has been manually obtained for the given sequences, containing bounding-boxes over each player and all three referees (taking the minimum visible X and Y coordinates of each individual), plus their corresponding identifier, in every single frame (when visible); this results in a total of 11339 boxes.

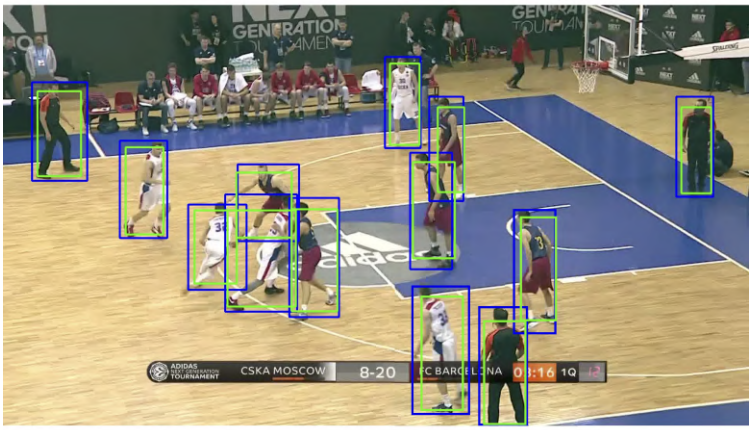

Figure 6.1: Player Detections (green boxes) together with its ground truth (blue boxes).

	Precision	**Recall**	**F1-Score**
Open Pose	0.9718	0.9243	0.9470
YOLO	0.8401	0.9426	0.8876

Table 6.1: Detection Results

6.1 Quantitative Results

Although it is not part of this Part's contribution, a quantitative assessment of the detection method is shown in Table 6.1, where its performance is compared to the state-of-the-art YOLO network [92]; for a fair comparison, only the *person* detections within the court boundaries are kept in both cases. These detections can be seen in Figure 6.1 with their corresponding ground truth boxes.
From now on, all quantitative tracking results will be expressed in terms of Multiple Object Tracking Accuracy (MOTA), which is defined in [4] as:

$$MOTA = 1 - \frac{\sum_t fp_t + m_t + mm_t}{\sum_t g_t},$$

Layer	b2c2	b3c2	b4c2	b5c2
MOTA	0.5396	0.5972	**0.6369**	0.6321

Table 6.2: MOTA results obtained with $\alpha = 0$ in (5.5), C_{Feat2} equal to C_{DL}, and by extracting DL features in the output of different convolutional layers.

where fp_t, m_t, mm_t and g_t denote, respectively, false positives, misses, mismatches and total number of ground truth boxes over all the sequence.
Another meaningful tracking metric that has been computed as well is the Multiple Object Tracking Precision (MOTP), which can be defined as:

$$MOTP = \frac{\sum_{i,t} IoU_{i,t}}{\sum_t c_t},$$

where $IoU_{i,t}$ and $\sum_t c_t$ correspond to the intersection over union between two boxes, and to the sum of correct assignments through the sequence, respectively. The detected bounding boxes for all the upcoming experiments are the same ones (thus the intersection with ground-truth bounding boxes does not change either), and knowing that the total number of instances is large, the MOTP results barely change in all presented combinations of techniques: it remains 0.6165 ± 0.0218.

Starting only with DL features (that is, $\alpha = 0$ in (5.5) and C_{Feat2} equal to C_{DL}), Table 6.2 shows the obtained MOTA results. As mentioned, a pre-trained VGG-19 architecture is used, and features are gathered and post-processed after every second convolutional layer in each block. The best MOTA results are obtained with the output of the fourth block, corresponding to the 10th convolutional layer of the overall architecture. For the remaining tests, all DL features will be based on this layer, which has an output of size $28 \times 28 \times 512$.
Having used a random-search grid [3], Table 6.3 shows the most significant MOTA results for a non-stabilized video sequence. In this

experiment, a comparison between Geometrical and DL features is shown, thus displaying the performance on their own as well as its best-weighted combination. Besides, as explained in Subsection 5.3.3, when extracting DL features, three different tests have been performed regarding the neighborhood size around each pose part. As it can be seen in Table 6.3, DL features outperform Geometrical ones, especially in the case of a 2x2 neighborhood. By combining them, and by giving more weight to the DL contribution, results are improved in all cases, thus indicating that the two types of features complement each other. In Table 6.4 the same experiments are shown, but this time using a stabilized video sequence. In this case, the performance of geometrical-based features outperforms the DL-based ones, but as mentioned, these metrics will drastically drop if the included dataset sequences contain fast camera movements (or even large pannings).

From both Tables 6.3 and 6.4 it can be deduced that the best filter size when extracting DL pose features is a 2×2 neighborhood. *A priori*, one might think that a 3×3 neighborhood should work better, as it is already including the 2×2 one, but a 3×3 spatial neighborhood in the output of the 10th convolutional layer is equivalent to a 24×24 real neighborhood around the specific part in the image domain. Accordingly, adding these extra positions will include court pixels, thus resulting in noise-prone feature vectors, and as a result, non-meaningful matches.

Apart from comparing Geometrical and DL features through C_d and the different mentioned C_{DL}, the effect of Visual features (color similarity C_c, explained in Subsection 5.3.2) is checked too. Moreover, tracking results have been compared with [74], which is a generic state-of-the-art tracking method by Milan *et al.*; in all tests, our ground-truth detections have been used, thus starting off with the same conditions. In Table 6.5, the best-weighted combinations in terms of MOTA / MOTP are shown for a non-stabilized and a stabilized video sequence. In both cases, DL features outperform color ones by a 3% margin. The combination of all Geometrical, Visual, and DL features outperforms the rest of the techniques but just by

0.2%, which comes at a cost of computation expenses. Besides, obtained results show how existing literature methods, not trained *a priori* for sports sequences, perform notably in terms of MOTA, but there is a large drop in MOTP due to the implicit challenges of cluttered courts.

Neighborhood	α	1-α	**MOTA**
—	1	0	0.5689
1x1	0	1	0.5923
1x1	0.3	0.7	0.6289
2x2	0	1	0.6369
2x2	0.2	0.8	**0.6529**
3x3	0	1	0.6171
3x3	0.3	0.7	0.6444

Table 6.3: Non-stabilized results obtained from only 4 video frames per second.

Neighborhood	α	1-α	**MOTA**
—	1	0	0.6506
2x2	0	1	0.6369
1x1	0.6	0.4	0.6752
2x2	0.55	0.45	**0.6825**
3x3	0.7	0.3	0.6781

Table 6.4: Stabilized results, with the same 4 video frames per second and weights as in Table 6.3.

Combination of Features	**MOTA**	**MOTP**
Geometrical + Visual	0.6233	0.6185
Geometrical + VGG	0.6529	0.6276
Geometrical + Visual [Stab]	0.6583	0.6225
Geometrical + VGG [Stab]	0.6825	0.6197
Geometrical + VGG + Visual [Stab]	**0.6843**	0.6238
Joint Track. + Segm. [74]	0.6714	0.3375

Table 6.5: Effect of Visual and Deep Learning features in combination with Geometrical ones.

In order to break down and to evaluate the contribution in MOTA of every single pose part, Table 6.6 is displayed; these results have been obtained with a 2x2 neighborhood around parts, and without combining with Geometrical features. As it can be seen, there are basically three clusters:

1. Discriminative features, above a 0.35 MOTA, that manage to track at a decent performance only with a 1×512 feature vector / player. These parts (shoulders, chest, and hip) belong to the main shape of the human *upper-torso*, and it coincides with the jersey-skin boundary in the case of players.

2. Features that fall within a MOTA of 0.20 and 0.35, which are not tracking players properly but their contribution might help the discriminative ones to achieve higher performance. These parts include skinned pixels of basic articulations such as elbows, knees, and ankles.

3. Concrete parts that have almost no details at a coarse resolution, thus resulting in low MOTA performance. Eyes could be an example: although people's eyes have many features that make them discriminative (such as shape, color, pupil size, eyebrow's length), players' eyes in the dataset images do not embrace more than a 3x3 pixel region, and all of them look the

same shape and brown or darkish. This results in poor tracking results when checking only for these parts.

Given the mentioned clusters, 3 different tracking tests have been performed by taking only some parts into account, in particular, and in terms of MOTA:

1. Taking the top-6 parts (over 0.35 MOTA).
2. Taking the top-12 parts (over 0.2 MOTA).
3. Taking the top-20 parts (over 0.1 MOTA).

Results are shown in Table 6.7, where it can be seen that the second and third clusters complement the top ones, while the bottom-5 parts actually contribute to a drop in MOTA. The drawback of this clustering is that it requires some analysis that cannot be performed in test time, and different video sequences (*i.e* different sports) might lead to different part results.
Finally, the obtained effect after the inclusion of memory between t_N and t_{N-2} is shown in Table 6.8. By comparing the extracted features across three consecutive frames, the obtained MOTA results improve by a margin larger than 5%; a particular scenario that benefits from memory inclusion are missed players in a single frame, which are successfully recovered due to preserved features of t_{N-2}.

A qualitative visual detection and tracking result (obtained with the best combination of Geometrical + Deep Learning features without camera stabilization) is displayed in Figure 6.2, where players are detected inside a bounding box, and its color indicates their ID; as it can be seen, all 33 associations are properly matched except a missed player in the first frame and a mismatch between frames 2 and 3 (orange-green boxes).

Part	**MOTA**
Chest	0.5349
L-Shoulder	0.4726
R-Shoulder	0.4707
R-Hip	0.3961
Mid-Hip	0.3956
L-Hip	0.3867
L-Knee	0.3156
R-Knee	0.3062
L-Elbow	0.2862
R-Elbow	0.2545
R-Ankle	0.2418
L-Ankle	0.2407
L-Toes	0.1935
R-Toes	0.1920
L-Ear	0.1348
L-Heel	0.1259
L-Wrist	0.1235
R-Heel	0.1126
L-Mid-Foot	0.1116
R-Wrist	0.1111
R-Mid-Foot	0.0964
L-Eye	0.0916
Nose	0.0771
R-Eye	0.0655
R-Ear	0.0677

Table 6.6: Individual Part Tracking Performance, obtained with $\alpha = 0$ in (5.5) and C_{Feat2} equal to C_{DL}.

Min. MOTA	Nº of Parts	Total MOTA
>0.35	6	0.6105
>0.20	12	0.6412
>0.10	20	**0.6423**
all	25	0.6369

Table 6.7: Clustering Part Results ($\alpha = 0$ and $C_{Feat2} = C_{DL}$) without stabilization.

	MOTA	MOTP
Geometrical + VGG (No Memory)	0.6237	0.6086
Geometrical + VGG (Memory)	**0.6825**	**0.6138**

Table 6.8: Tracking performance with the inclusion of memory.

Figure 6.2: Obtained tracking and pose results in three consecutive frames, where each bounding box color represents a unique ID.

7 Conclusions

Once stated the current unbalanced situation regarding basketball tracking data, where only clubs in the NBA benefit from this type of resource, the first Part of this thesis has consisted in exploring the viability of single-camera trackers. To this end, we have proposed a method to automatically track multiple targets (players). Roughly, the presented tracking methods have been built from the following four main steps:

1. Court filtering, that limits the boundaries where players may be located. This segmentation has been achieved through different approaches depending on the type of video footage. Apart from using line segment detection: (a) in the case of European courts, a simple color filter has been applied, and (b) in crowded NBA games, an iterative approach through coarse CRF has been used.

2. Player detection, obtained through a pre-trained model, OpenPose, able to detect not only the location but also the pose of multiple humans.

3. Feature extraction. In particular:

    - Geometrical features have been obtained in terms of pairwise distances between players across frames (in pixels).

- Visual features benefit from the previously estimated pose and characterize color features from small neighborhoods around key body parts.
- Similarly, deep learning features have been extracted by combining pose information with the output of convolutional layers of a VGG-19 network.

4. Finally, player detections have been matched according to their associated features by solving a cost minimization problem.

Having gathered a dataset from scratch, and having labeled more than 11k ground-truth bounding boxes, an ablation study –in terms of tracking metrics, such as MOTA and MOTP – has been included to justify all the choices of the presented tracker.
Several conclusions can be extracted from the presented experiments:

- First of all, detections of OpenPose proved to work better than other state-of-the-art networks such as YOLO [92], reaching a final 0.947 F1-Score.

- DL features outperformed Visual ones when combining them with Geometrical information; in particular, the obtained MOTA boost is +0.03. On the contrary, the combination of all of them has not implied a significant performance boost (only +0.0018 MOTA).

- In the case of VGG-19, DL extracted features from the 10th convolutional layer have provided the best accuracy; moreover, placing a 2x2 neighborhood around downscaled body parts (instead of single pixels) has improved the tracking performance.

- Classical CV techniques such as camera stabilization have improved the overall method's performance, but it might have related drawbacks, such as the incapability of generalization to all kinds of camera movements.

- When extracting pose features from convolutional layers, those body parts that are not distinguishable at a coarse resolution (*e.g.* nose, ears, wrists, heels...) have had a negative effect on the overall performance. That is, instead of considering all 25 default body-parts, a combination of the most important ones (for instance chest, shoulders, and hips) might generalize better.

7.1 Future Work

Despite obtaining promising results, tracking data require being really precise in order to train models on top of it and to build tracking reports that can be interpreted by coaches or GM's. The presence of missed targets, or even miss-detections, results in noise-prone tracking reports that cannot be used to draw valuable conclusions. Moreover, in the cluttered basketball scenario, there are a lot of situations where miss-detections occur with ease and are crucial for the play's outcome. For example, the most common basketball plays nowadays are based on *ball-screens*, where 2 offensive players, who are wearing a similar jersey, aim to create a scoring opportunity through a screen in a really small space; moreover, two (or even three) defenders are also standing in the same space, which generates notable partial-occlusions of pose parts. The decision-making process in this type of plays needs to be precisely tracked at high confidence (over 0.95 MOTA). Therefore, the presented method should be a solid baseline to be combined with other synchronized video footage *a posteriori*. Actually, the effect in terms of accuracy when adding more cameras to the existing set-up should be studied, thus finding out the minimum number of cameras to get decent results. Hopefully, by including video footage from 3-5 existing broadcasting cameras (not only the main one), accurate tracking data could be gathered without increasing the overall cost. Besides, the presented method should be able to handle fast basketball situations; for instance, during offen-

sive transitions, players (and the ball) move fast from one side of the court to the other, thus involving large camera panning that cannot be handled when stabilizing the sequence.
Another line of research could include the refinement of the court-filtering process; so far, the presented approach properly segments the playing area, but we lack specific coordinates or labels that could indicate where each corner / relevant part of the court is. By having at least 4 correspondences between the image and a given template, tracking data could be expressed not only in terms of bounding boxes (pixels) in the given image but also in 2D court-coordinates. Moreover, geometrical features, which consist of pairwise distances between detections, could also be expressed in terms of court coordinates instead of pixels. This strategy, which involves computations in the 2D space, will be followed in the upcoming Parts of this manuscript.
Another alternative that could be beneficial in terms of accuracy would be switching from a tracking-by-detection approach to an *end-to-end* training process. Given the lack of ground-truth labeled data, unsupervised approaches could be helpful to tackle this challenge as a self-supervised method [52]. Some experiments were performed by fine-tuning existing networks such as Unsupervised Deep Tracking [120], but the obtained results did not seem to generalize to basketball players. Generally, state-of-the-art tracking networks are able to confidently track single targets within non-cluttered scenes (or less challenging, at least); however, the process of turning the problem into multi-class classification results in a notable fall in terms of the model's performance.
Apart from orientation-based metrics, which will be detailed in the next Parts of this manuscript, other applications could be built using the content of bounding boxes. For instance, action recognition models would definitely improve raw tracking data, by indicating not only where a player is located, but also detailing what he/she is doing (*i.e.* running, shooting, dribbling, jumping...).

Part II:
Orientation Estimation

NBA is a very competitive league,
so whatever can give us an
advantage, we try to keep it.

Ivana Seric

8 Introduction: Beyond Tracking

Until this Part, the importance of tracking methods has been contextualized, and as mentioned, with this brand-new data source, the overall structure of data-science sports departments has changed completely. Despite the unlocked potential of tracking data, exhaustive post-processing techniques have to be applied in order to use their insights. The outcome of tracking a complete game of any sport is a large file that cannot be understood anyhow at first sight; players, coaches, or analysts cannot draw conclusions from raw data as if they were analyzing, *e.g.*, a simple box score. Consequently, sports data scientists are continuously attempting to build automatic applications from tracking (and eventing) data that could be easily interpreted. Among many other applications, with this type of tracking-based tools, coaches and analysts: (1) can study the effect of tactical strategies, (2) are able to split the player performance into different game phases, or (3) can estimate / back-up intangible statistics, thus finding out who is the player that adds more value to a team instead of the one with the highest number of scored goals. These models are currently the holy grail of soccer analytics and they are solely based on tracking data, but... Are 2D tracking data powerful enough to encompass all types of events? Do we have unexploited metadata in the image that could improve the model's performance? Luckily, yes; if 2D tracking data are used on their own, not enough evidence is obtained to determine if a player is in a favorable condition of prop-

erly acting during the play, since external factors such as the player's own pose and orientation are crucial. In fact, coach Pep Guardiola often explains how elder people claim that, while in yesteryear soccer you had to control the ball, then look and turn around, and finally, make the pass, in today's faster version of soccer, players need first to look and orient correctly before controlling and passing the ball. Nonetheless, body orientation is a yet-little-explored area in sports analytics research. For the sake of clarification, and despite being an inherently ambiguous concept, player orientation is defined in this thesis as the projection (2D) of the normal vector placed in the center of the upper-torso of players (3D).

An existing type of tool that could benefit from the inclusion of body-orientation could be Expected Possession Value (EPV) models. Given that the main reward of soccer players is to score a goal, and knowing that this type of action is a rare event, Fernandez *et al.* [34] extended the previous basketball-based work of Cervone *et al.* [14] by creating an EPV framework that values player actions. The main objective of this metric is to predict an expected value of scoring / receiving a goal at a given time in any field position, based on a spatial analysis of the whole offensive and defensive setup at that moment; more concretely, in pass events, having a passer P, an EPV map can be computed for each field position $x \in \mathbb{R}^2$, which estimates the above-mentioned expected-value if P passes the ball to x. The main EPV model consists of different likelihood components, especially emphasizing a passing probability model. Ultimately, the inclusion of body-orientation into EPV models or the creation of computational passing models will be studied, but first, this second Part aims solely to estimate body-orientation from soccer video footage, thus enriching tracking data with an extra variable. Note that from now on, for both Part II and Part III, all conducted research will be focused on soccer instead of basketball, since better datasets were available for our purpose in terms of tracking and orientation ground-truth data (extracted from EPTS-held devices). Two different types

of orientation estimation methods are proposed:

- On the one hand, a model-based approach is proposed. It relies on the combination of pose models and 3D vision techniques. Furthermore, extra steps such as enhancing image quality – through a super-resolution network –, a coarse skeleton corroboration, and a final refinement based on the ball position improve the overall performance. Validated results show less than 30 degrees of median absolute error per player.

- On the other hand, a learning-based approach that does not depend on pose models is proposed as well. Instead, a VGG-19 network [100] is fine-tuned to classify players' bounding boxes into different orientation bins; the network benefits from an angle compensation strategy and a cyclic loss function. Our results show a mean absolute error of fewer than 12 degrees.

Moreover, three novel types of orientation maps are proposed in order to make raw orientation data easy to visualize and understand, thus allowing further analysis at team- or player-level. More concretely, *OrientSonars* integrate player orientation and show how players are oriented during pass events. *Reaction* Maps show how players move during the pass, by comparing their orientation at the beginning and at the end of the event. Finally, *On-Field* Maps merge and compare the pure body orientation of players with their relative orientation with respect to the offensive goal.
The rest of this Part is organized as follows: since several sources of data will be employed, Chapter 9 provides a complete definition of the given datasets, their corresponding domains, and completion techniques. Later, Chapter 10 details the existing related work regarding generic pose and gaze orientation. Proposed methods are then described in Chapters 11 and 12, where the model- and the learning-based approaches are described, respectively. While numerical results can be found in Chapter 13, Chapter 14 suggests several visual orientation maps. Finally, Chapter 15 lists the final conclusions regarding orientation estimation.

9 Data Sources and Completion

Before introducing the proposed method, a detailed description of the required materials to train this model is given. Similarly, since we are going to mix data from different sources, their corresponding domains should be listed as well:

- **Image-domain**, which includes all kinds of data related to the associated video footage. That is: (*i1*) the video footage itself, (*i2*) player tracking and (*i3*) position of the field's corners. Note that the result of player tracking in the image-domain consists of a set of bounding boxes (as the output of the methods of Part I), expressed in pixels; similarly, corners' location is also expressed in pixels. In this research, full HD resolution (1920 x 1080) is considered, together with a temporal resolution of 30 frames per second.

- **Sensor-domain**, which gathers all pieces of data generated by wearable EPTS devices. In particular, data include: (*s4*) player tracking, and (*s5*) orientation data. In this case, players are tracked according to the universal latitude and longitude coordinates, and orientation data are captured with a gyroscope in all XYZ Euler angles. In this work, sensor data were gathered with RealTrack Wimu wearable devices [91], which generate GPS / Orientation data at 100 / 10 samples per second, respectively.

- **Field-domain**, which expresses all variables, such as (*f1*) player tracking, in terms of a fixed two-dimensional football field, where the top-left corner is the origin.

In the upcoming Chapters of this same Part and also in Part III, three different datasets are used:

- A complete dataset, in which all variables (*i1*, *i2*, *i3*, *s4*, *s5*, *f1*) are available. Note that both image- and sensor-data include unique identifiers, which are easy to match by inspecting a small subset of frames. In particular, our complete dataset contains a full game of F.C. Barcelona's Youth team recorded with a tactical camera that contains almost no panning and without zoom; this dataset will be named youthFCB$_{DS}$.

- An orientation-based dataset, where only part of the information is available (in particular, *i1*,*s4*,*s5*). More specifically, our orientation-based dataset contains a full preseason match of CSKA Moscow's professional team, recorded in a practice facility (without fans) with a single static camera that zooms quite often and has severe panning. Similarly, this second dataset will be named CSKA$_{DS}$. Furthermore, the intersection of field lines and field corner coordinates in the image were manually identified and labeled in more than 4000 frames of CSKA$_{DS}$ (1 frame every 45, *i.e.* 1.5 seconds), with a mean of 8.3 ground-truth field-spots per frame (34000 annotations). In order to estimate the missing pieces (*i2, i3*) and match data across domains, a sequential pipeline is proposed in Section 9.2.

- A tracking-based dataset, which contains data from both the image- and the field-domains (*i.e. i1*, *i2*, *i3*, *f1*), but no sensor data. In this research, the tracking-based dataset, named FCB$_{DS}$ contains data from 9 games of the professional F.C. Barcelona during the 2019-2020 season; by filtering eventing data, around 6000 event passes have been gathered among 12 different players. Note that this dataset will not be used to

assess the performance of orientation estimation methods. Instead, its main purpose is to bring together a large number of passing events in order to create orientation-based data visualization tools once orientation has already been estimated; furthermore, as it will be detailed in Part III, this large dataset will be also used in Part III to evaluate pass feasibility.

9.1 Homography Estimation

Since the reference system of the image- and the sensor-domain is not the same, corners' positions (or line intersections) are used to translate all coordinates into a 2D template representing the field-domain. On the one hand, obtaining field locations in the sensor-domain is pretty straightforward: since the sensor's gathered coordinates are expressed with respect to the universal latitude / longitude system, the corners' locations are fixed. By using online tools such as the *Satellite View* of Google Maps, and by accurately picking field intersections, the corners' latitude and longitude coordinates are obtained. On the other hand, corner's positions in the image-domain (in pixels) depend on the camera shot and change across the different frames; although several literature methods [15; 20] can be implemented in order to get the location of these field spots or the camera pose, our proposal leverages homographies computed from manual annotations. From now on, the homography that maps latitude and longitude coordinates into the field will be named H_{SF}, whereas the one that converts pixels in the image into field coordinates will be named H_{IF}. The complete homography-mapping process is illustrated in Figure 9.1.

For the sake of clarification, in the field-domain, it can be assumed that 0^o / 90^o / 180^o / 270^o are the corresponding orientations of players facing towards the right / top / left / bottom sides of the fields, respectively, as shown in Figure 9.2. Moreover, as it will be detailed

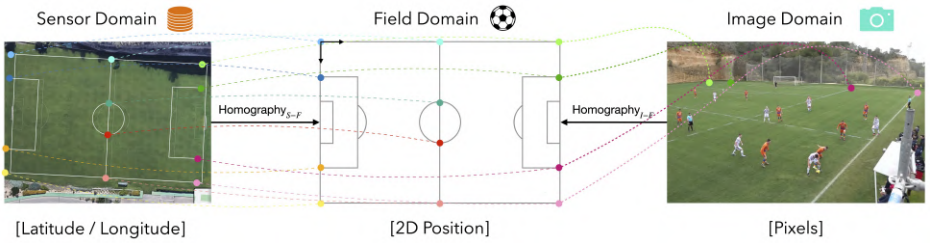

Figure 9.1: Several domains are merged in the upcoming parts. (left) Sensor-, (middle) field-, and (right) image-domain. By using corners and intersection points of field lines, the corresponding homographies are used to map data across domains into one same reference system.

afterward, similar angles will be clustered into orientation bins for both the presented methods (orange and yellow lines in Figure 9.2). As it will be enclosed afterward, in our methods and experiments, we will consider either 12 bins (corresponding to the 12 angular regions limited by the orange lines in Figure 9.2) or 24 bins (corresponding to the union of orange and yellow lines). Chiefly, when sorting, the first / last orientation bins will always correspond to the ones including 0^o / 360^o, respectively. Last but not least, throughout the remaining parts of this thesis, the pink camera of Figure 9.2 will be referred to as the *reference* camera, in which there is no panning and the viewing direction points to the center of the field; in some given scenarios, if the camera pose does not coincide with the reference camera, some compensation will have to be applied.

9.2 Automatic Dataset Completion

In this Section, the complete process to convert an orientation-based dataset into a complete one is described. It has to be remarked that the aim is to detect players in the image-domain and to match

them with sensor data, hence pairing sensor orientation with bounding boxes in the image. Note that this procedure has been applied to CSKA$_{DS}$, which did not contain ground-truth data in the image-domain. The proposed pipeline is also displayed in Figure 9.3.

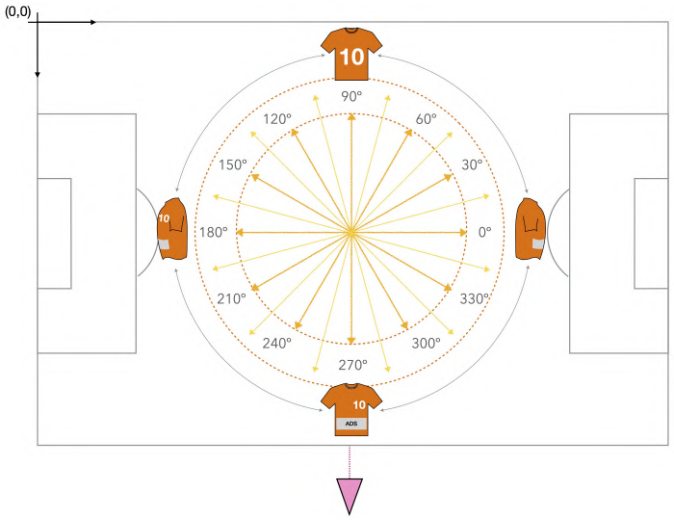

Figure 9.2: Orientation references in the field-domain. Besides, since similar orientations will be clustered into bins, their portions are shown as well.

- **Player Detection**: the first step is to locate players' location in the image. In order to do so, literature detection models can be used, such as OpenPose [11] (used in this research) or Mask R-CNN [48]. Once identified all different targets in the scene, detections are converted into bounding boxes. Note that this step does not exploit any temporal information across frames.
- **Jersey Filtering**: since sensor data are only acquired for one specific team, approximately half of the detected bounding boxes (opponents) are filtered out. Given that the home /

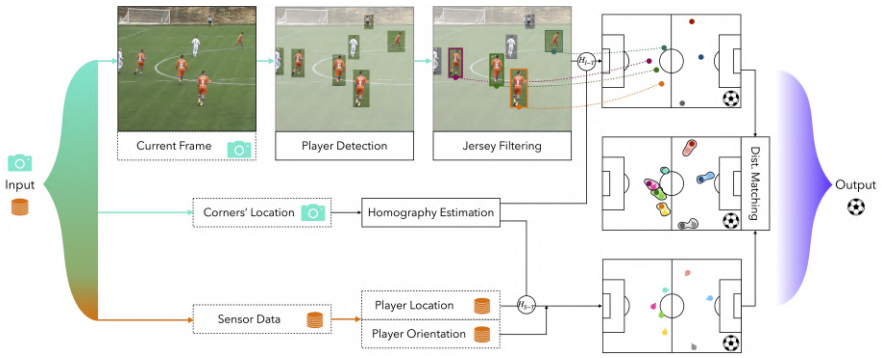

Figure 9.3: Proposed pipeline to match sensor orientation data with bounding boxes. Different input sources are merged: (top, image-domain) video footage, which is used for player detection and jersey filtering; the resulting bounding boxes are later mapped into the field-domain. (middle, image-domain) Corner's location, which is used for building the corresponding mapping homographies, and (bottom, sensor-domain) ground-truth data, which are also mapped into the field-domain. Finally, players in the 2D-domain are matched through pairwise distances.

away teams of football matches are required to wear distinguishable colored jerseys, a simple clustering model can be trained. Specifically, by computing and by concatenating quantized versions of the HSV / LAB histograms, a single 48-feature vector is obtained per player. Having trained a K-Means model, with $K = 3$, boxes with three different types of content are obtained: (1) home team, (2) away team, and (3) outliers.

- **Mapping**: in order to establish the same reference system for both sensor and image data, all tracking coordinates are mapped into the field-domain. More specifically, corner-based homographies H_{SF} and H_{IF} are used; in the latter, since we are dealing with bounding boxes, the only point being mapped

for each box is the middle point of the bottom box's boundary.

- **Matching**: once all points are mapped into the field-domain, a customized version of the Hungarian method [59] is implemented, thus matching sensor and image data in terms of pairwise field-distances.

10 Related Work

Since, to the best of our knowledge, there are no existing contributions to infer body-orientation of players in the sports domain, this Chapter aims to detail several related works regarding pose and gaze orientation. Note that all contributions are clustered by different fields of research, and within each field, papers are listed chronologically.

Estimating the pose of athletes from sports video footage is nothing new. Since these methods can yield to the extraction of valuable analytics that can numerically answer many coaching concerns, the estimation of 2D / 3D pose has been split into many contributions, each one solving particular sports challenges from the CV perspective. Starting with the 2D pose estimation, in 2013, Fastovets *et al.* [31] presented a combination of inference algorithms and probabilistic prior models to extract athlete pose estimation directly from TV sports footage. The authors exploited spatio-temporal data in a graph fashion to ensure the consistency of joints across frames. That same year, Hayashi *et al.* [47] focused on team sports, but their work was limited to the head and upper body pose estimation method while using low-resolution footage. By detecting and by tracking the player's head and pelvis (without temporal information), an estimation of the 2D spine was obtained, and together with its orientation, a 3D spine pose was computed. The presented results of both pa-

pers showed how the head- and upper-body-pose was leveraged for individual players, resulting in 3D visualizations that could be then studied by coaches. Two years later, the latter authors [46] extended their previous work by training a poselet-regressor that produced an accurate estimation of the spine pose. By introducing priors based on the spine angle, several body classifiers were trained, which output a coarse orientation value of the upper body; note that, once the head region of each player was detected, the contribution of the presented regressor was to estimate the relative pelvis location for each target. Although presented results (soccer and football footage) showed a promising baseline, orientation was estimated in the image-domain (*perceived orientation*); therefore, the challenge of estimating the absolute player orientation in the field-domain is still unsolved. Another head-pose contribution was made by Chen *et al.* [16], who also attempted to estimate it through low-resolution frames, since surveillance cameras may have to operate in scenarios where privacy protection is required. Similarly, their contribution stemmed from a feature extraction process that combined histogram of oriented gradients features and gradient-based ones, and the resulting vectors were used in order to train and test a regressor. Obtained results were validated with Kinect-based data, which captured depth and allowed a further inspection of the captured frames; the final estimations contained less than 13 degrees of mean-absolute error in all roll, yaw, and pitch angles. The last sports-based 2D pose estimation is the one proposed by Sypetkowski *et al.* [110], who focused on soccer data across several types of video footage, including both high and low-resolution videos / frames. Their deep convolutional network showed a notable capability of generalization when retraining models with unseen data.

By extending 2D to the third dimension, the estimation of the 3D human pose was the main goal of Akhter *et al.* [1]. By gathering (and sharing) a large capture motion dataset, the existing pose priors that are assumed to work on joint limits were updated, thus describing the limits of human joints when it comes to movement-related de-

grees of freedom. However, their main contribution was a detailed parametrization of body poses that allowed the 3D pose estimation using redundant information. Their results worked fine for general non-sport datasets, but when dealing with cluttered sports scenarios (such as the Leeds sports pose dataset), manual annotations were required to gather accurate 3D pose data. In order to avoid the tedious need for extensive manual joint annotation, Sumer *et al.* [108] proposed a self-supervised learning approach that was based on pose embedding and incorporated spatio-temporal data to learn pose similarities. The proposed architecture was a siamese convolutional network, which provided training labels that were later double-checked by a curriculum learning step. Besides, the model benefited from repetitive poses, which might be used to detect outlier joints. Results were also tested on challenging sports datasets, and the overall pose seemed to generalize properly to unseen bounding boxes. Zhang *et al.* [132] went far beyond the widely-practiced sports, and presented a multi-view dataset that included images from less popular disciplines, such as dancing or martial arts. Apart from providing at least 3 color views and their corresponding depth maps – plus calibrated ground-truth poses –, which could be used to estimate the 3D pose, the authors also provided baseline pose estimation results using state-of-the-art models (and exploiting temporal information). Their drawn conclusions detailed that, while discriminative models performed better when large sets of data were being used to train the model, generative models were more robust to extreme poses. In 2019, another 3D pose estimation contribution was published by Bridgeman *et al.* [9], who attempted to correct the most common inconsistencies of this type of model, and succeeded in tracking 3D skeletons through the association of 2D poses between different cameras in a greedy fashion. Apart from detecting joints and merging data from several cameras, priors – when it comes to joints and limbs – were also considered to correct the existing false positives. Moreover, their method was sports-based, and the model had to be able to: (a) properly handle cluttered scenarios, and (b) reduce the pro-

cessing time as much as possible without dropping accuracy; in fact, this paper also considered the 3D pose estimation of multiple targets at the same time. The need for fast and efficient models was also considered by Zhang *et al.* [130], who proposed a light model based on fast pose distillation learning. By leveraging pose data with simple architecture, the pose structure was post-processed by a strong teacher network that was in charge of refining all the obtained outputs into solid estimations. This novel approach resulted in the abolition of the existing compromise in the trade-off that relates accuracy and efficiency.

Apart from obtaining the 2D / 3D pose of players, other interesting challenges in sports involve the identification of player's jersey numbers, which might be a key factor to determine the player orientation and was studied by Liu *et al.* [63]. In this work, given that in traditional soccer footage the camera shot changes drastically with panning and zooming, and given that players keep turning around, an R-CNN network was trained to exploit player body cues. Roughly, the presented model classified the bounding boxes' pixels into (a) background, (b) player, or (c) digit; the latter were fed into another classifier, which stabilized the given input and ended up inferring the number of the given box. Results were validated with real soccer-match data, and the method outperformed existing number recognition models. Note that, by using the segmented image as a network input, an unsupervised clustering model could be trained to distinguish between front- / back- / side-poses among players.

Once pose has been estimated, the design of pose-based tools, which can help coaches improve the performance of players / teams during practices or games, is sport-dependent. Stemming from a purely basketball-based dataset, Felsen and Lucey [32] aimed to find correlations between different types of shots and the body position of the shooter. Their motivation was to complement the existing 2D Sports VU tracking data, because when taking only spatial coordinates into

account, some relevant information might be missed. Their method included a quantification of the involved anatomy in a three-point-shot and a machine learning module, where a model was trained both to identify open / tough shots and to attribute correlations by comparing them. Furthermore, the authors also performed a deep analysis of the shooting parameters of the best NBA shooter at that time (Stephen Curry, 2015-2016 season), and found out that, although there were many biometric correlated factors in open / tough shots, those cannot be generalized into a single model, as Curry had a notable percentage from long-range, but he attempted more tough shots than the vast majority of players. When considering other sports, Zecha *et al.* [129] predicted the motion of pose kinematics and dynamics for an automatic swimming athletic performance assessment. In this paper, the authors worked with challenging aquatic footage, where the corresponding left / right pose parts can be swapped easily and some of them might be partially or totally occluded. By defining a cost function in a graph fashion and by using integer linear programming, the labels of body parts (mainly shoulders, arms, and leg parts) were constantly double-checked. Meanwhile, Zhi *et al.* [135] dealt with the estimation of both individual and collective key pose recognition, which has great value for strength and conditioning coaches. By using a deep neural network, the authors managed to collect data from weightlifting high-resolution footage and classified the obtained frames into normal / abnormal scenarios, thus limiting the potential region of interest. Consequently, the key pose was extracted.

Finally, even though body orientation is claimed to be more meaningful in the sports context than gaze orientation, a brief literature review from the latter topic could also be helpful to obtain meaningful player insights. First, Kellnhofer *et al.* [57] presented a model to estimate 3D gaze *in the wild*, together with a large-scale gaze-tracking dataset (Gaze360). Given the diverse nature of the shared dataset (indoors, outdoors, camera shots...), their proposed gaze model ex-

ploited temporal information and outperformed state-of-the-art results. Moreover, the authors also tested their model with benchmark ones using a cross-dataset self-supervised adaptation, hence proving that the trained model did not overfit. Also in 2018, Fischer *et al.* [35] provided robust gaze estimations under natural conditions, hence solving several challenges such as fast lightning changes. What is more, the authors also suggested a solution so that the accuracy would not drop while the detected target moved further away from the camera. By capturing a solid training dataset with eye-tracking glasses, and once applied semantic image inpainting to make the train images resemble the test ones, a deep convolutional network was trained. Cross-dataset evaluations were also performed, and results showed that not a lot of accuracy was compromised when the human-camera pairwise distance increased.

11 Model-based Orientation Estimation

In this Chapter, the proposed model-based approach to estimate orientation from soccer players is detailed. Roughly, this method uses pose models, contextual information, and 3D vision techniques to obtain orientation data directly from video footage.

Our model-based orientation method benefits from two different kinds of orientation estimation: pose data and ball position, which will be detailed in Sections 11.1 and 11.2, respectively. An overall pipeline of the method is presented in Figure 11.1. The output of all these individual estimations produces both a numerical orientation result and a confidence value. More concretely, orientation is measured in degrees and discretized into 24 probability bins using the reference system previously displayed in Figure 9.2. Those 24 bins correspond to the 24 angular regions limited by the union of orange and yellow lines. While the orientation value indicates the bin with higher probability, the confidence value is used as a prior to quantify, in an inversely proportional way, how many other neighboring bins have a non-zero probability. More concretely, the proposed method outputs a probability density function (pdf) of the estimated orientation, from which we define player orientation as the angle corresponding to the maximum of the obtained pdf, and its confidence, defined as the inverse of the pdf support. Nevertheless, the aforementioned pdf is refined by incorporating contextual information about

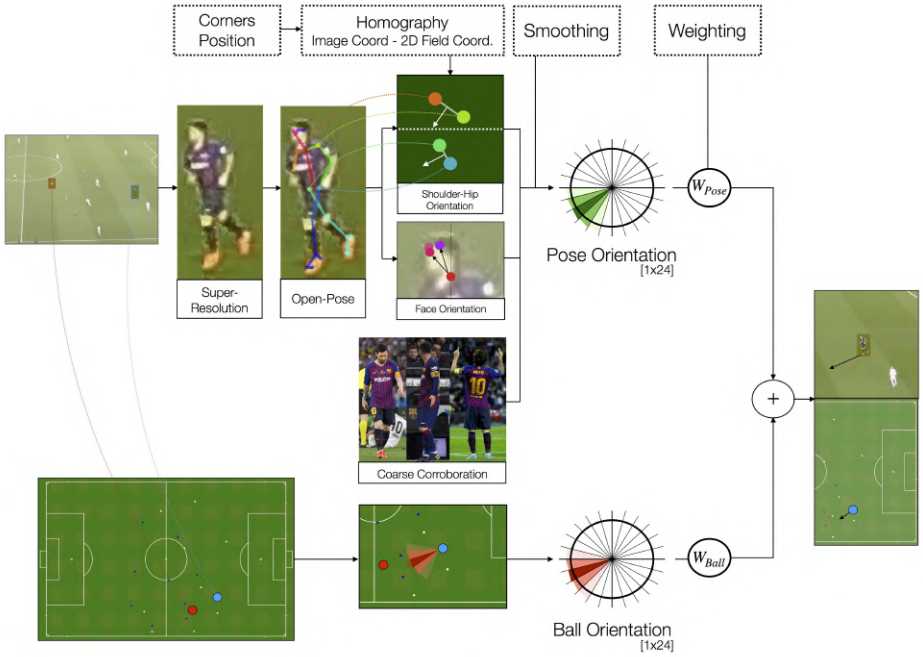

Figure 11.1: Proposed pipeline. On the one hand, pose orientation is found by combining a super-resolution network, OpenPose and 3D vision techniques (plus a coarse validation); on the other hand, ball orientation is also computed. Finally, both pdf's are merged into a single final orientation estimation.

the position of the ball (Section 11.3) to output the final orientation of each player.

11.1 Pose Orientation

Estimating orientation from pose data is a key ingredient of this method, and uses pre-trained models and 3D vision techniques in order to obtain a first orientation estimation of each player. Given

temporally-smoothed bounding boxes of players, a combination of super-resolution and pose detection techniques is applied to find the pose of every player. Both the left-right shoulders and the left-right parts of the hip will be considered as the main upper-torso parts. By projecting these parts in a 2D space, the normal vector between these points is extracted (Figure 11.2(b)). A detailed description of this method is given in the upcoming Subsections.

11.1.1 Pose Detection

Having the bounding boxes for all visible players in each frame, the OpenPose library [21] is used to extract the pose of every single individual (we refer to Section 2.1 for details of pose models). However, detecting the pose of players in sports scenarios is always challenging given the frequent occlusions and fast movements that lead to motion blur. Moreover, the average resolution of bounding boxes around players in Full-HD frames is around 15×50 pixels. Hence, small image crops are not always properly processed by OpenPose, resulting in a null set of landmarks. For this reason, a super-resolution network is previously used to preprocess bounding boxes and enhance the image quality instead of a simpler interpolation technique. More concretely, the applied model is a Residual Dense Network (RDN) [12; 133].

11.1.2 Angle Estimation

Once the pose is extracted for each player, the coordinates (and confidence) associated with the upper-torso parts are stored to estimate the pose orientation. From the output of OpenPose, the coordinates of the main upper-torso parts are found in the image domain. By using H_{IF} (Section 9.1), the left-right pair parts (either shoulders or hips) can be mapped into a 2D field, thus obtaining, as seen in Figure 11.2(a), a first insight about each player's orientation. Basically, in the case of 24 clustered orientation bins, the player can be inclined

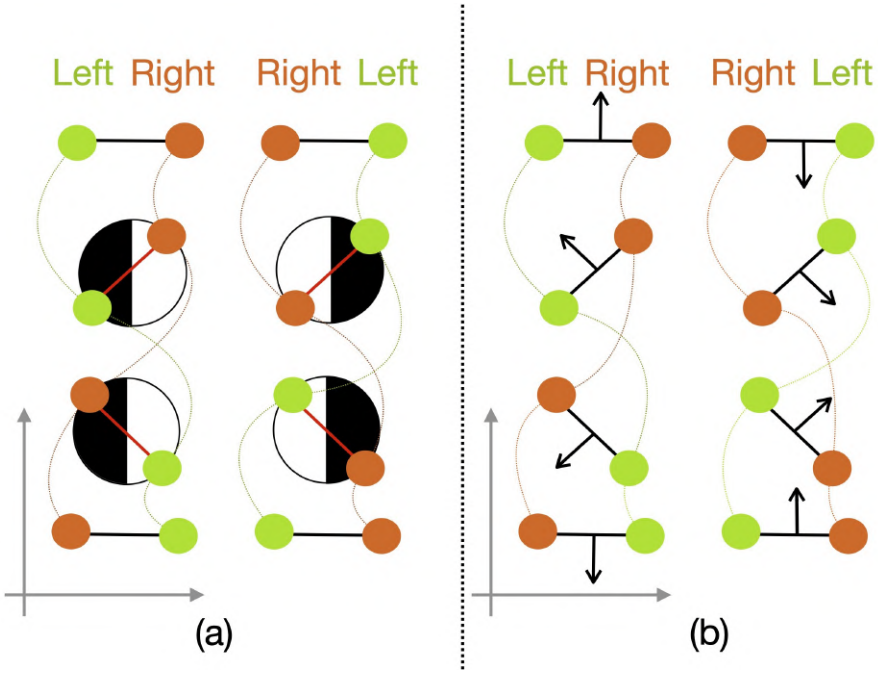

Figure 11.2: (a) Different 2D combinations of left-right mapped parts; (b) same combinations with normal vectors.

towards the right (0-90º, 270-360º, bins 0-11) or the left (90-270º, bins 12-23) side of the field. From now on, this first binary estimation, which indicates if the orientation belongs to the first or second half of the orientation histogram, will be called *LR-side* parameter. Figure 11.3 shows in more detail how pose orientation is estimated: first, left-right shoulders and hips are mapped via the estimated homography (H_{IF}) into the 2D space; then, *LR-side* booleans (LR_{Sh}, LR_{Hi}), angles (α_{Sh}, α_{Hi}) and confidences (C_{Sh}, C_{Hi}) are obtained, where the suffixes Sh and Hi stand for shoulders and hips, respectively. The associated confidences are the product of OpenPose's individual shoulder and hips confidences, respectively. However, OpenPose might fail detecting either the left or the right hip parts / shoulders; while in the case of a missing hip part, the middle-hip position

is used as a substitute, when a shoulder is missing, the chest position is picked. Then:

1. If LR_{Sh} and LR_{Hi} agree, both individual confidence values are checked: in case $C_{Sh} > C_{Hi}$, α_{Sh} is considered as the pose orientation estimation and C_{Sh} its confidence. If not, α_{Hi} and C_{Hi} are selected.

2. Otherwise, if $|C_{Sh} - C_{Hi}|$ is smaller than a threshold (set to 0.4 in our results), the player's face direction is checked. In the image domain, the difference among the X positions of all face parts and the player's neck is computed. If most of the parts move towards the origin of the X axis (Figure 11.3(c)), the player's *LR-side* will be left; otherwise, the player's *LR-side* will be right.

Then, given the final pose orientation estimation α_P and its related confidence C_P, a Gaussian probability distribution is located around it, with effective support size

$$N_P = \max\left(\left\lfloor N_{bins}\left(\frac{1-C_P}{2}\right)\right\rfloor, 1\right), \tag{11.1}$$

centered at

$$\text{or}_P = \begin{cases} \left\lfloor \frac{\alpha_P}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor & \text{if } \frac{\alpha_P}{360/N_{bins}} < 18 \\ \left\lfloor \frac{\alpha_P}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor - N_{bins} & \text{if } \frac{\alpha_P}{360/N_{bins}} \geqslant 18 \end{cases} \tag{11.2}$$

where the second element of the sum is an offset that compensates the bin order. The output vector of this orientation estimation will be denoted as H_P.

11.1.3 Coarse Orientation Validation

Despite the notable performance of Open Pose, image quality problems (*e.g.* blurry or really small players) are challenging scenarios where the estimated players' pose might be flipped 180º: this

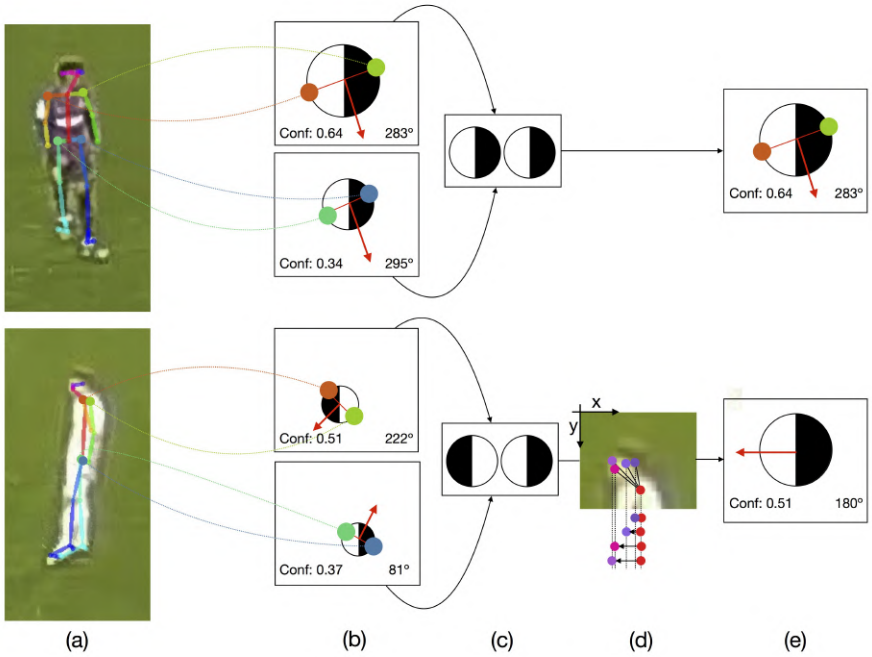

Figure 11.3: Pose orientation estimation: (a) OpenPose output and its (b) mapped 2D coordinates. (c) Side check between shoulder and hip parts, plus, if required, (d) face direction double-check. Right after, (e) a final estimation is obtained.

is, the right-left shoulders (or hip parts) of the corresponding player are swapped. Inaccurate detection of the player pose results in huge errors while estimating the pose angle, as the actual normal vector is the opposite of the predicted one, thus introducing errors that might oscillate between 120º and 180º. In order to double-check the pose orientation estimation and to ensure that the upper-torso normal vector is computed in the correct direction, a Support Vector Machine model has been trained to classify three types of coarse orientations: front-, side- and back-oriented players (see Figure 11.4):

- Front-oriented players are the ones whose upper-torso is pointing straight to the camera position. These players usually have an orientation between 200 and 340 degrees (red class in Figure 11.4(d)), and chest-jersey advertisements can be easily spotted.
- Side-oriented players are the ones placed almost completely perpendicular with respect to the camera. These players usually have an orientation that can vary from 160 to 200 degrees (if the *LR-side* parameter points left) or from 340 to 20 degrees (*LR-side* pointing right and blue class in Figure 11.4(d)).
- Back-sided players are the ones whose upper-chest is pointing in the opposite direction with respect to the camera position. These players usually have an orientation between 20 and 160 degrees (yellow class in Figure 11.4(d)), and the number of the player in the backside of the jersey is very visible.

Two characteristics are concatenated in the feature vector: color features in the Hue-Saturation-Value color space (histogram of 36-18-18 bins in the respective channel) and geometrical properties (pixel-wise distances between the 4 upper-torso coordinates). Having the position of the upper-torso parts, obtained from pose keypoints, the above-mentioned features are only computed inside the defined trapezoid, hence discarding misleading features such as the color of the field. Therefore, this model is used after estimating player pose orientation, with two main possible outputs:

- The resulting angle estimation coincides with the coarse classification (*e.g.* a player oriented towards 90 degrees according to pose orientation classified as front-oriented). In this case, the final pose orientation does not change from the previous estimation.

- The player's pose orientation does not match the output of the coarse classification model (*e.g.* a player oriented towards 90 degrees according to pose orientation classified as a back-sided player). In this situation, the final pose orientation will be the opposite angle of the previously computed normal vector (+180°).

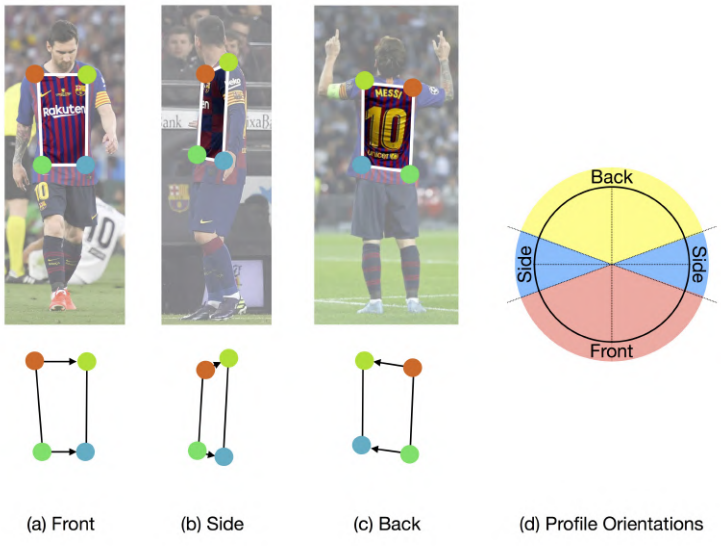

Figure 11.4: (a) front-, (b) side-, and (c) back-oriented players with their (d) corresponding potential pose orientation.

11.2 Ball Orientation

The other performed estimation is related to the position of the ball. Logically, players close to the ball tend to be strongly oriented towards it, while players placed far away may not have to be duly oriented. Hence, having all pairwise distances and the corresponding angles, the orientation of players with respect to the ball can be estimated. Then, for a given player at (P_x, P_y), in a moment where the ball is at (B_x, B_y), and an angle of β degrees between player-ball, the effective support size of the related pdf is:

$$N_B = \frac{N_{bins}}{4}\left\lfloor 1 - \frac{MD - \sqrt{(P_x^2 - B_x^2) + (P_y^2 - B_y^2)}}{MD}\right\rfloor + \frac{N_{bins}}{8}, \quad (11.3)$$

where MD is a maximum distance that regularizes how far a player can be from the ball without being influenced by it; this parameter is set to $\frac{w}{6}$ in practice, where w indicates the field width. Then, the central bin with the highest weight is:

$$\text{or}_B = \begin{cases} \left\lfloor \frac{\beta}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor & \text{if } \left\lfloor \frac{\beta}{360/N_{bins}} \right\rfloor < 18 \\ \left\lfloor \frac{\beta}{360/N_{bins}} + \frac{N_{bins}}{4} \right\rfloor - N_{bins} & \text{if } \left\lfloor \frac{\beta}{360/N_{bins}} \right\rfloor \geqslant 18 \end{cases} \quad (11.4)$$

Once again, the outcome of this estimation is a discrete probability vector, called from now on H_B; the overall process can be spotted in Figure 11.5.

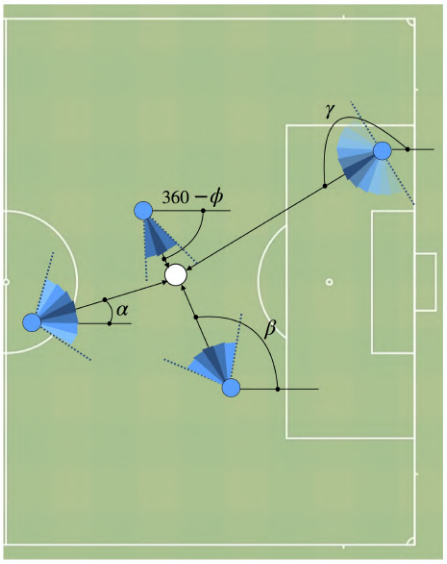

Figure 11.5: Orientation computation with respect to the ball of 4 different players, considering both the angle (direction) and the distance (magnitude).

11.3 Contextual Merging

Once both histograms are obtained, a simple weighting is performed between them, thus merging pose and ball orientations. In particular:

$$H_{\mathrm{TOT}} = wH_P + (1-w)H_B, \tag{11.5}$$

with $w \in [0,1]$; more concretely, several values of w will be tested in Section 13.1. Ultimately, the orientation θ of each player is the central value of the bin H_{TOT} with the highest weight, namely:

$$\theta = \mathrm{argmax}\left(H_{\mathrm{TOT}}\right) \cdot \frac{360}{N_{bins}} + \frac{360/N_{bins}}{2} \tag{11.6}$$

In terms of visualization, orientations can be displayed in the 2D field; starting from a 2D point (P_b), which indicates the position of a given player, another point (P_e) can be projected at a given distance T_θ having an orientation of θ degrees; as a result, the vector joining P_b and P_e will have the estimated orientation and a length proportional to the estimated confidence T_θ. Moreover, having both points, the same coordinates can be mapped back into the original frames by multiplying P_b and P_e by the inverse homography (H_{F-I}), thus showing the orientation vector in the video frame.

12 Learning-based Orientation Estimation

In this Chapter, another approach to estimate orientation directly from bounding boxes is detailed. In this case, instead of approaching this challenge with the combination of CV methods, a learning-based fashion is used.

The presented learning-based orientation stems from a fine-tuning process where a state-of-the-art network is trained with bounding boxes and their corresponding ground-truth orientation, obtained through ETPS-held devices. More concretely, the method compensates angles *a priori* (Section 12.1), and uses a VGG-19 architecture (Section 12.2); what is more, by including a cyclic loss function (Section 12.3), and a thoughtful training setting (Section 12.4) the overall generalization capability of the model improves. Note that with this proposed learning-based strategy, there is no need anymore to compute the player's pose, and instead, orientation can directly be obtained with the raw content inside a bounding box.

12.1 Angle Compensation

The apparent orientation of each player is influenced by the current image content, which is drastically affected by the camera pose and its orientation. This means that, if a bounding box of a particular player is cropped without taking into account any kind of

field reference around him/her, it is not possible to obtain an absolute orientation estimation. As displayed in Figure 12.1 (top), the appearance of three players oriented towards the same direction (0 degrees) can differ a lot. Since the presented classification model only takes a bounding box as input, we propose to compensate angles *a priori*, thus assuming that all orientations have been obtained under the same camera pose; *i.e.* the *reference* camera, described in Section 9.1. For instance, if the full chest of a player is spotted in a particular frame, its orientation must be approximately 270 degrees, no matter what the overall image context is.

In order to conduct this compensation, as seen in the bottom row of Figure 12.1, the orientation vector of the player is first mapped into the field-domain. Then, the *apparent zero-vector* is considered in the image-domain; for the *reference* camera, this vector would point to the right side of the field whilst being parallel to the sidelines. By using H_{IF}, the *apparent zero-vector* is mapped into the field-domain, and the corresponding compensation is then found by computing the angular difference between the mapped *apparent zero-vector* and the reference zero-vector in the field-domain. According to Figure 12.1, this difference indicates how the orientation vector differs from the *apparent zero-vector*.
Formally, for a player i with non-compensated orientation α'_i at position $P_i = (P_{i,x}, P_{i,y})$ and being the (unitary) *apparent zero-vector* Z described by $(1, 0)$, another point is defined towards the zero direction:

$$P_i^0 = P_i + Z = (P_{i,x} + 1, P_{iy}) \tag{12.1}$$

Both points P_i and P_i^0 are mapped into the field domain by using H_{IF}, thus obtaining their 2D position F_i and F_i^0, respectively. The final compensated angle is then found as:

$$\alpha_i = \alpha'_i - \angle(\overrightarrow{F_i F_i^0}), \tag{12.2}$$

where $\angle$ expresses the angle of the vector $\overrightarrow{F_i F_i^0}$ with respect to the reference zero-vector.

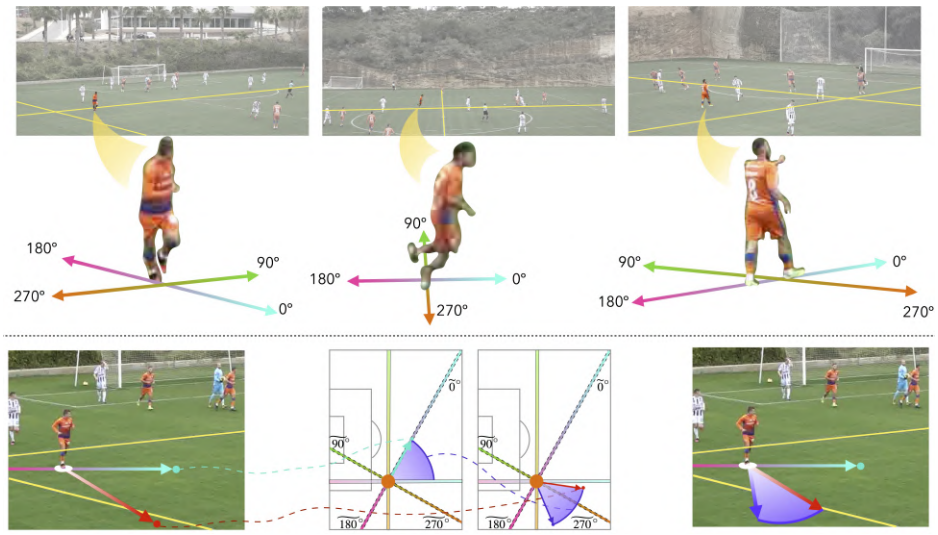

Figure 12.1: (Top) Three players oriented towards 0º can look really different depending on the camera pose and orientation. (Bottom) Proposed technique for angle compensation: (left) detected player together with his orientation {red} and *apparent zero-vector* {cyan}; (middle-left) mapped *apparent zero-vector* in the field-domain {dashed axes - apparent reference system, continuous axes - absolute reference system} (middle-right) Applied compensation on the original orientation {purple}; (right) resulting compensated absolute orientation {purple}.

12.2 Network

Once all bounding boxes have an associated compensated body-orientation value, the model is set to be trained. In this work, orientation estimation has been approached as a classification task, where each bounding box is classified within a certain number of orientation bins. As it will be detailed in Section 13.2, orientation data are grouped into K bins, each one containing an orientation range of $360/K$ degrees. Consequently, the above-mentioned bounding boxes in the image-domain were automatically labeled with their corresponding class according to their compensated orientation. Another reason for grouping similar angles into the same class is the noisy raw orientation signals generated by the EPTS devices.
The chosen network to be fine-tuned is a VGG-19 [100]; this type of network has also been used as a backbone in existing literature methods such as OpenPose [11]. However, in order to further analyze and to justify our choice, alternative results are shown in Section 13.2 when using DenseNet [50]. The original architecture of VGG-19 is composed of 5 convolutional blocks – each one containing either 2 or 4 convolutional layers –, and a final set of fully connected layers with a probability output vector of 1000 classes. For the presented experiments, as seen in Figure 12.2, the architecture adaptation and the proposed method consists of:

1. Changing the dimensions of the final fully-connected layer, thus obtaining an output with a length equal to the desired number of classes.

2. Freezing the weights of the first couple of convolutional blocks.

3. Re-training the convolutional layers of the third block and the fully connected layers of the classifier.

4. Omitting both the fourth and fifth convolutional blocks.

Figure 12.2: Proposed architecture for fine-tuning a VGG-19 according to the main blocks of the original network.

By visualizing the final network weights with Score-CAM [117] (Figure 12.3), it can be spotted how the most important body parts regarding orientation (upper-torso) are already being vital for the sake of classification after the third block. In fact, the responses of the fourth block do not provide useful information in terms of orientation. Therefore, omitting blocks 4 and 5 is a safe choice to have an accurate model whilst decreasing the total number of parameters to be trained.
Let us finally remark that image values in bounding boxes are converted into grayscale, thus improving the overall capability of generalization, since the model will not be learning the specific jersey colors as happened with the coarse corroboration of the model-based method (Subsection 11.1.3). In terms of data augmentation, brightness, and contrast random changes are performed for all boxes in the training set.

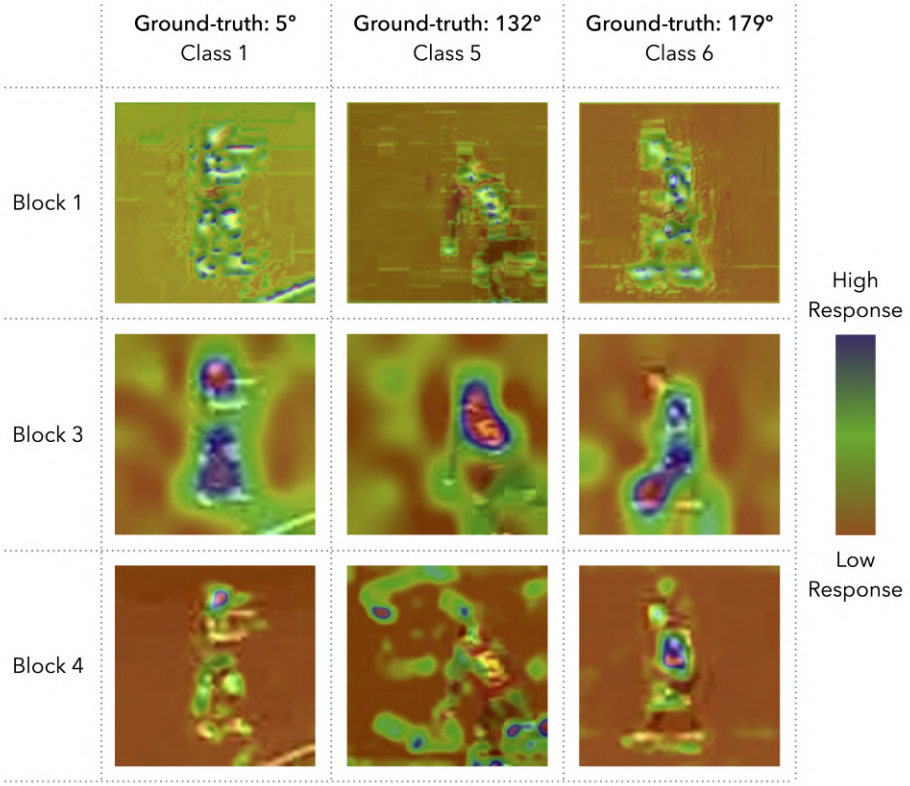

Figure 12.3: Obtained ScoreCam [117] responses. While the 1st block mainly responds to edges and shapes, the 3rd one has a high response over the players' upper-torso. The last row shows how the 4th block learns specific features that have little to do with orientation.

12.3 Cyclic Loss

An important aspect of the training process is the definition of the loss function. *A priori*, state-of-the-art loss functions such as binary cross-entropy could be a valid resource, but in general classification scenarios, the order and the distance within classes are not taken into account. Nonetheless, in this particular scenario, we have K ordered-cyclic classes and a distance between them that can be well-defined. Besides, in this classification problem, since similar orientations have been grouped into bins, enforcing a one-hot encoding is not the best solution. For example, if $K = 12$ and each orientation bin encompasses 30 degrees, imagine a player P_1 oriented towards 31º and another P_2 oriented towards 59º; both players are included in the second bin, which encompasses all orientations between 30-60. With one-hot encoding, it would be assumed that since both P_1 and P_2 are in the second bin, both of them have the same orientation (45º). However, alternatives such as soft labels [28] can describe the players' class as a mixture; in the given example, the soft labels of P_1 / P_2 would indicate that these players are right between the first-second / second-third bins, respectively. The other challenge to be solved is the need for this loss function to be cyclic, as the first bin (number 1, 0-30º) and the last one (12, 345-360º) are actually really close.
Let $\{b_1, b_2, ..., b_{12}\}$ be the set of orientation classes and let
$\chi = \{r_1, r_2, \ldots, r_{12}\}$ be the set such that each r_j denotes the central angle of bin b_j, for all $j \in \{1, \ldots, 12\}$. Then, for a player i with compensated ground-truth orientation α_i, the soft labels representing the ground-truth probability distribution are defined as the vector y_{ij} with coordinates:

$$y_{ij} = \frac{exp(-\phi(\alpha_i, r_j))}{\sum_{k=1}^{K} exp(-\phi(\alpha_i, r_k))}, \text{ for } j = 1, \ldots, 12 \qquad (12.3)$$

where ϕ is the cyclic distance between the ground-truth player's orientation α_i and the angle corresponding to the jth bin, r_j, defined

by:

$$\phi(\alpha_i, r_j) = \frac{\min(|\alpha_i - r_j|, 360 - |\alpha_i - r_j|)^2}{90}. \tag{12.4}$$

Let us denote as x_i the estimated probability distribution of orientation of player i obtained by applying the softmax function to the last layer of the network. Finally, our loss is the cross-entropy between x_i and the ground-truth soft labels y_i.

12.4 Training Setting

As mentioned in Chapter 9, several datasets have been used throughout this whole thesis. In this case, we are only interested in using those two datasets including orientation data, *i.e.* youthFCB$_{DS}$ and CSKA$_{DS}$, which were recorded under different camera shot conditions. Consequently, as seen in Figure 12.4, the content inside both bounding boxes differs a lot: while in youthFCB$_{DS}$ players are seen from a tactical camera and have small dimensions, players in CSKA$_{DS}$ are spotted from a camera that is at almost the same height as the playing field, thus resulting in big bounding boxes. Although all bounding boxes are resized as a preprocessing stage of the network, the raw datasets suffer from concept drift [123].
The proposed solution in this thesis is to build an unbalanced-mixed training set; that is, merging bounding boxes from both datasets with an unbalanced distribution in the train set, whilst using the remaining instances from youthFCB$_{DS}$ and CSKA$_{DS}$ on their own to build the validation and the test set, respectively. In particular, the presented experiments of Section 13.2 have been carried out with a 90-10 distribution in the training set; that is, the model should be able to generalize to both different games despite having almost no data from one of the games. For each class, a total of 4500 bounding boxes are included in the training set, where 4000 of them are obtained from youthFCB$_{DS}$ and the 500 remaining ones are gathered from CSKA$_{DS}$. While the validation set includes 500 bounding boxes from

youthFCB$_{DS}$, the test set is built with the same number of instances from CSKA$_{DS}$.

Figure 12.4: Resized bounding boxes of both datasets; several artifacts can be spotted in youthFCB$_{DS}$ (*e.g.* JPEG, ringing, aliasing).

13 Orientation Estimation Results

In this Chapter, results obtained with both presented methods are detailed. Note that, in order to validate the obtained results, sensor data have been used: more concretely, youthFCB$_{DS}$ is employed in the model-based approach, and both youthFCB$_{DS}$ and CSKA$_{DS}$ are exploited in the learning-based one.

The obtained classification results will be shown in terms of angular difference for both methods (and confusion matrices in the latter). Nonetheless, it has to be remarked that, when clustering orientations as bins, an intrinsic error is being introduced: assuming that each bin contains a spectrum of d degrees and that a player classified in bin k has an orientation that corresponds to the central value of the bin, players who have been properly classified may have an associated absolute error up to $d/2$ anyway. Note that in the model-based method, there are 24 orientation bins ($d/2 = 7.5$), whilst in the learning-based one, it is limited to 12 ($d/2 = 15$).

13.1 Model-based Results

Bearing in mind that OpenPose detected upper-torso parts in **89.69%** of the given image crops, the following metrics were validated

with sensor data:

- **Coarse orientation validation**: as explained in Subsection 11.1.3. a classifier was trained from scratch, using geometrical- and color-based features, in order to distinguish players facing front, back, or sidewards. 14000 players were manually labeled with a tag corresponding to one of the three classes; by randomly splitting it into train and test (80-20), 85.91% accuracy was obtained. The main limitation of this method is the need for different trained models for different teams since feature vectors include histogram data, which directly depend on the jersey colors; tuning a new model takes up to 10 hours of manual procedure. Therefore, in the gathered dataset, only F.C. Barcelona players wearing the red-blue jersey from the 2018/2019 and the 2019/2020 seasons were included.

- ***LR-side*:** this metric shows the accuracy of the *LR-side* parameter (detailed in Section 11.1.2), which indicates if a player is facing the left or the right side of the field. Considering a sequence of duration T and being i_t an individual player in a total of NP_t players in frame t, pose orientation α_{i_t}, and the corresponding ground-truth orientation ω_{i_t}, this metric can be computed as:

$$\mathrm{LR}_{\mathrm{acc}} = \frac{\sum_{t=0}^{T} \sum_{i_t=0}^{NP_t} LRV_{i_t}}{\sum_{t=0}^{T} NP_t}$$

  where:

$$LRV_{i_t} = \begin{cases} 1 & \text{if } |\alpha_{i_t} - \omega_{i_t}| < |\alpha_{i_t} + 180 - \omega_{i_t}|, \\ 0 & \text{otherwise.} \end{cases}$$

  LR-side performance reached **96.57%** accuracy.

In terms of orientation estimation, the error between the obtained orientation (α_{i_t}) results and ground-truth data (ω_{i_t}) can be computed

w	$(1-w)$	*MEAE*	*MDAE*
0	*1*	35.33	31.59
1	*0*	29.98	27.75
0.3	*0.7*	33.77	29.87
0.7	*0.3*	**29.78**	**27.66**

Table 13.1: MEAE and MDAE given different weights.

with their angular error. By using a random search grid [3], results in Table 13.1 indicate the error margin of different tests, showing the performance of each individual orientation estimation and their best mixture. As it can be observed, ball orientation produces the less accurate predictions; actually, pose orientation outperforms this prediction by a notable margin. These individual results prove that pose orientation needs to be heavily weighted while merging both estimations: by setting w to 0.7, the mean absolute angle error (MEAE) is reduced to **29.78º** and the median absolute angle error (MDAE) to **27.66º**.

13.2 Learning-based Results

The results of six different experiments are shown in Table 13.2. More concretely: (1) t_{12} and (2) t_{24} use a VGG-19 architecture that classifies into 12 and 24 orientation bins, respectively, both trained with compensated angles; (3) t_{12nC} uses the same network as in t_{12} but trained without angle compensation, and (4) t_{12den} uses a DenseNet architecture – fine-tuning of the fourth dense block – that performs a 12-bin classification, (5) t_{12CE} uses binary cross-entropy instead of the proposed cyclic loss, and (6) t_{12CV} shows the performance of the model-based method (this time with 12 bins as well). Table 13.2 contains the mean absolute error (MEAE) and the median absolute error (MDAE) of the estimated angles in each experiment.
As it can be spotted, the test of 12 classes is the one providing the

	MEAE_v	MDAE_v	MEAE_t	MDAE_t
t_{12}	17.37	9.90	**18.92**	**11.60**
t_{24}	**13.13**	**7.70**	24.34	13.01
t_{12CE}	22.34	17.00	28.98	23.00
t_{12nC}	21.47	14.16	31.75	24.54
t_{12den}	15.22	10.46	25.27	17.29
t_{CV}	-	-	38.23	32.09

Table 13.2: Obtained results in all experiments, expressed in terms of the mean / median absolute error, both in the validation and test set.

most reliable test results in terms of generalization; in particular, classifying orientation into 24 classes produces better results in the validation set, but seemingly, the model overfits and learns specific features that do not generalize properly. Moreover, the model benefits from the cyclic loss implementation, as binary cross-entropy introduces errors both in the validation and in the test set due to the unknown distance between classes and the non-cyclic angular behavior. Actually, the obtained boost with this cyclic loss is displayed in the confusion matrices of Table 13.1. The addition of angle compensation also proves to be vital, especially in the test set, where the corresponding video footage (CSKA_{DS} contained a lot of panning and zooming). Besides, the performance of DenseNet does not seem to generalize either; however, it is likely that with an exhaustive trial-error procedure of freezing weights of particular layers and performing small changes in the original DenseNet structure, this architecture should be able to generalize as well. Finally, it can be spotted how the presented learning-based outperforms the model-based, which has been tested this time without the SVM in charge of the coarse corroboration.

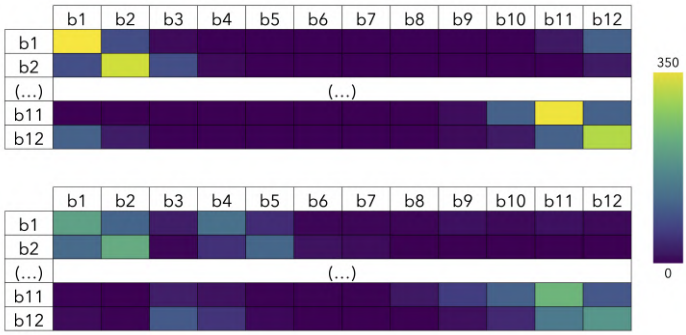

Figure 13.1: First and last rows of the obtained confusion matrix (test set) when using the (top) proposed cyclic and (bottom) binary cross-entropy as a loss function (t_{12} and t_{12CE} respectively).

14 Visual Orientation Maps

In this Chapter, the effect of body orientation is visually analyzed in soccer passes; more concretely, passing events from FCB_{DS} have been used for this analysis.

Before getting started, note that this Chapter aims to provide purely orientation-based insights out of the obtained estimations, so no models are being trained on top of body-orientation data yet. For the rest of this Chapter, three types of orientations are considered:

1. Orientation of the receiver in a Pass Event: this value quantifies the orientation of a potential receiver right at the moment when the passer kicks the ball.

2. Orientation of the receiver in a Reception Event: this value quantifies the orientation of a player who is receiving the ball at that precise moment.

3. Orientation of the passer in a Pass Event: this value quantifies the orientation of the player kicking the ball when performing a pass.

Moreover, the following performance statistics are used in order to evaluate the impact of body orientation in the observed passes:

1. Pass success / accuracy, which indicates if the pass was successful or not; this is, if the potential receiver has actually received

the ball. This metric can be used to get an overall picture of orientation, but there might be a lack of context: an easy pass between two defenders is valued the same way as a difficult assist that ends up in a goal. Besides, a failed pass might happen due to multiple circumstances, such as a bad pass, a bad reception, or a remarkable performance of a defender.

2. Added EPV [34], which quantifies the contribution of each action by modeling the conditional probability of scoring / receiving a goal at a given time and a given scenario. EPV is computed both at the Pass Event and right after the Reception Event; the difference between these two values will indicate the added contribution of the receiving player and exemplifies what happens after receiving the ball. For instance, a player might receive the ball appropriately but he/she might lose it due to a disadvantageous orientation, resulting in an EPV drop.

In order to introduce context in the mentioned visualizations, different phases of the offensive plays are evaluated individually as well (introductory soccer-based details are given in Section 2.2). Bearing in mind that in a soccer lineup there are mainly 3 rows of horizontally distributed players, their orientation can drastically change depending on the context: if an almost-static defender is carrying the ball, strikers will not be strictly oriented towards it, but if a midfielder is generating a play in the offensive court, forwards will be highly influenced by his/her position.

14.1 OrientSonars

PassSonars have recently gained a lot of popularity in soccer analytics; this kind of map is used to display the passing frequency and the accuracy of players in different directions inside the field, just by taking 2D information from these. In this article, OrientSonars are proposed, which integrate player orientation and show how players

are oriented during pass events. In this display, the following size-color codification is adopted: the radius of each portion in the map quantifies the volume of passes at a particular orientation, while the color displays their associated accuracy. OrientSonars can be performed at two levels:

- **Individual level**: simple visual reports of each player can be built by combining different OrientSonars in the 3 above-mentioned possible events. These visualizations can be useful to spot specific details when scouting a particular player. An example is shown in Figure 14.1, where the main orientation characteristics of Ivan Rakitic are shown.

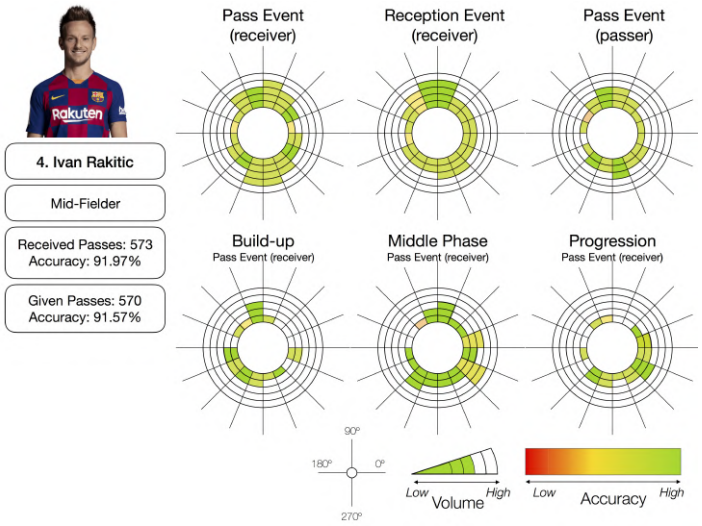

Figure 14.1: OrientSonar of Ivan Rakitic, showing his performance in pass events (both passing and receiving the ball) and reception ones, as well as different offensive phases. Accuracy is expressed with pass accuracy and color encoding, while portion size indicates the passing volume.

In this specific example, it can be seen that Rakitic, as a midfielder, has a strong duality receiving passes when oriented completely backward (270º) and upward (90º), as he has to receive passes from defenders (backward) and organize the forwards at the same time. In particular, Rakitic excels in reception events when the orientation oscillates between 67.5 and 112.5 degrees, which matches the most natural reception orientation for right-footed players. Moreover, game phases indicate that Rakitic is oriented towards defenders in the build-up phase (especially the left-side ones), but when the ball is carried towards the middle of the court, he is also oriented towards the offensive goal, thus potentially generating passes to forwards.

- **Team level**: as individual performances might be biased towards specific team tactics, the whole picture of the corresponding lineup has to be evaluated as well. In this map, the individual OrientSonar of all players is placed at the average position of every single individual. An example can be seen in Figure 14.2, where accuracy and added-EPV are compared, and Figure 14.3, where different game phases can be distinguished.

Several conclusions can be drawn from these maps: from Figure 14.2, it can be inferred that the pass success might not be the best accuracy metric to be used when comparing all kinds of players, mainly because defenders perform many non-risky passes among themselves, while forwards receive the ball in fewer situations (and often under the pressure of defenders) with higher risk and potential reward. For this reason, defenders in Figure 14.2(a) have a lot of high-accuracy bins, and forwards receive fewer passes at a lower accuracy rate. This situation swaps when checking EPV: on the one hand, defenders add less real value to the play, and on the other hand, offensive players have some portions with high contribution when they receive in advantageous situations (facing slightly upwards).

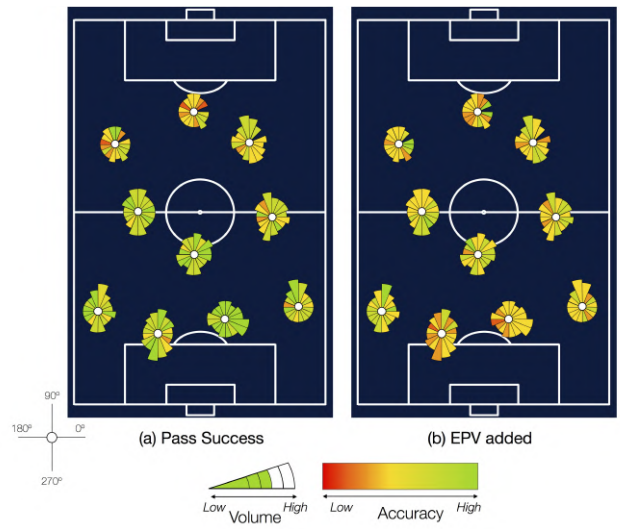

Figure 14.2: OrientSonar of the whole team during Pass Events as receivers, displayed with different accuracy metrics: (left) pass success metrics, and (right) added EPV.

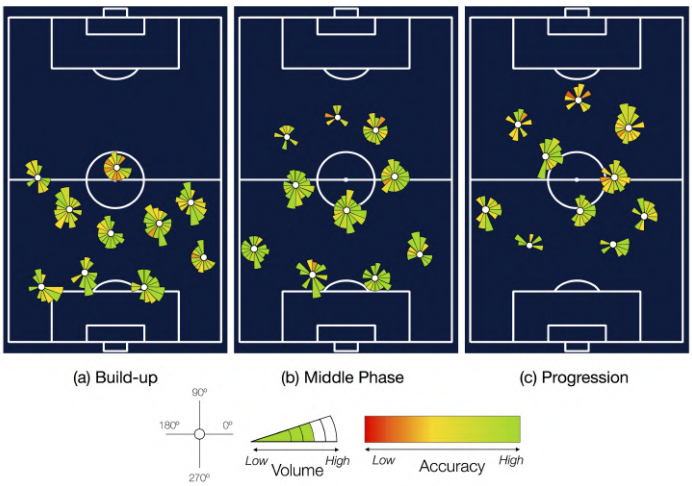

Figure 14.3: OrientSonar of the whole team in all three game phases: (a) build-up, (b) middle, and (c) progression.

Moreover, EPV peaks do not appear in random clusters: instead, a notable increment of EPV can be observed when specific couples of players interact. For instance, when the striker receives the ball from approximately the position of the right forward or vice versa, the team not only keeps the ball but also creates potential goal opportunities. The same pattern is repeated with the center- and left-midfielder. Besides, orientation patterns may be useful to distinguish the dominant player side: left-sided players (i.e. left full back) tend to be oriented towards the middle of the field, so right-side clusters have a higher volume of passes (and vice versa). From Figure 14.3, the interaction of players is even more detailed according to the context: in the build-up phase, midfielders are oriented towards defenders, waiting for the ball in order to generate an offensive play. In the middle phase, the same midfielders have the highest relevance in terms of volume, distributing the ball in potentially advantageous situations; meanwhile, strikers look for open spaces, and rarely receive the ball backward (except the right-forward in the given example). Finally, in the progression phase, two possible player roles can be distinguished: while some players are oriented towards regions with high risk but a notable potential reward, the rest occupy safe positions that allow them to move back to another medium phase if required without losing control of the ball, hence generating new offensive opportunities.

14.2 Orientation Reaction Maps

Although there are many different types of soccer passes, the behavior of players during the ball displacement is crucial for the outcome of that specific play; defenders are always trying to anticipate, so offensive players must orient and move accordingly before getting tackled. Orientation reaction maps show how players move during the pass, by comparing the orientation at the beginning (X-axis) and at the end (Y-axis) of the event; once again, the color represents accu-

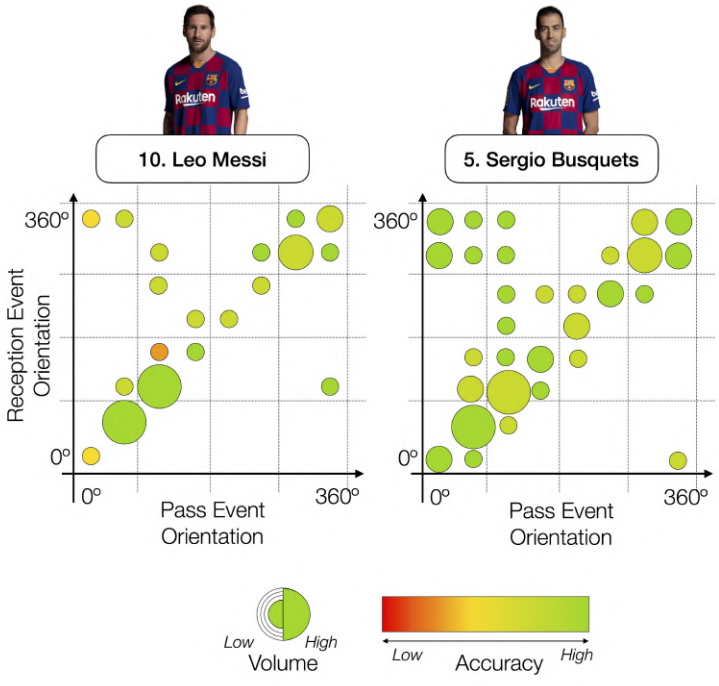

Figure 14.4: (left) Leo Messi – (right) Sergio Busquets reaction maps. The X axis represents the orientation of the player at the pass event, and the Y axis the one in the reception. Accuracy has been expressed with pass success.

racy, and the dot area expresses the volume. If a player keeps his/her orientation, the resulting map will just have dots in a diagonal line; on the contrary, if a player rotates while receiving, off-diagonal dots will appear in the graph. Figure 14.4 shows the orientation reaction maps of Messi (right forward) and Busquets (central midfielder).

Once again, the visual outcome differs for those players who occupy different positions. In the given example, Messi has a main diagonal line with some outliers, as he receives many passes from players who are in front of him (facing backward) when he is running towards the goal (huge blobs in the 75-105º), as well as straight passes from

midfielders when he is facing backward (270-315º). Meanwhile, Busquets has more dots in his map, mainly because he receives passes from many different positions; besides, being close to the exact middle point of the field makes things even trickier, as he has defenders trying to tackle him from several positions, thus forcing him to move even more to find a safe spot. As it can be seen in the map, Busquets has a remarkable performance in every single orientation situation, especially in the right-side clusters. In conclusion, there is not an optimal reaction map, and comparisons have to be performed by contextualizing the player position in the field together with his/her individual skills, as it is difficult to establish similarities among players with different characteristics. Instead, this tool could be a great resource for comparing players or even to keep track of youth players' progress.

14.3 On-Field Orientation Maps

Despite being the goal the most important part in soccer games, all the previous displays were only based on the orientation of the player at given non-goal events. Hence, proposed on-field orientation maps merge information and compare the pure body orientation of players with their relative orientation with respect to the offensive goal. On-field maps can be extracted at a player-level, as seen in Figures 14.5, 14.6, 14.7, where both left-right full-backs (Alba – Semedo) are compared together with of a midfielder (Arthur) in terms of pass reception. In these visualizations, the X-axis represents the orientation with respect to the offensive goal (being 0-90 on the left side and 90-180 on the right side), and the Y-axis represents the orientation of the player. In this type of map, it is even more distinguishable how players are clustered depending on their position. Despite the difference in spatial performance, the visualizations of Alba and Semedo show almost symmetric results for left- and right-sided players. While Alba is completely restricted to the left side of the court (0-45º in

goal orientation), Semedo tends to deviate his orientation more towards the middle part of the court, which results in regions with an EPV drop. This particular scene shows one of the main differences between experienced players and the rest: in this case, bearing in mind that Jordi Alba has been on the team for 8 seasons in a row, it is reasonable to conclude that he has already found his comfort zone in court, where he manages to fit all his skills without the need of taking unnecessary risks. Apart from the fullback comparison, the plot of Arthur shows that this type of midfielder operates on both the central sides of the court at more or less the same frequency; although orientation performance could seemingly be the same when checking pass accuracy, EPV can help detect complex patterns. In the given example, Arthur adds higher EPV contributions when he is placed on the right side of the court, especially when receiving in a backward orientation (most likely from a defender); nevertheless, this type of conclusion has to be again contextualized with different prior information (*e.g.*, a player who just started playing in a new spot for the first time in the season and needs some adaptation).

Figure 14.5: On-field orientation map of Alba as a receiver in Pass Events, evaluated both with (left) pass accuracy and (right) EPV.

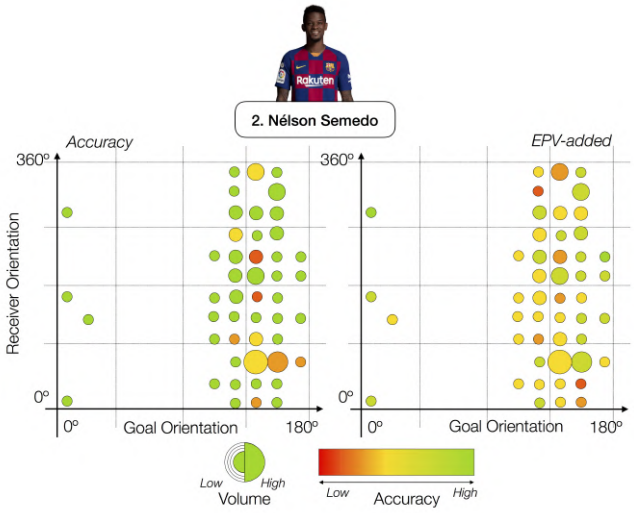

Figure 14.6: On-field orientation maps of Semedo as a receiver in Pass Events, evaluated both with (left) pass accuracy and (right) EPV.

Figure 14.7: On-field orientation maps of Arthur as a receiver in Pass Events, evaluated both with (left) pass accuracy and (right) EPV.

15 Conclusions

In this Part, two novel techniques to compute soccer players' orientation from soccer monocular footage have been presented.
On the one hand, the model-based method combines two different orientation estimators: pose-based and ball-based. While pose orientation is obtained by projecting OpenPose output on a 2D space and computing the normal vector to the projected torso, the orientation of all players with respect to the ball is also taken into account to refine the former estimation. Having mapped both orientation estimations into probability vectors, a simple weighting is performed and an individual angle (in degrees) is obtained for each player. Results have been tested and validated with professional soccer matches; although the associated video footage was not optimal, the results are promising. 96.6% accuracy is obtained in left-right side orientations, and a median absolute error of 26.80º is achieved.
On the other hand, the learning-based model learns how to classify players' crops into orientation bins. The core of this method combines a VGG structure with frozen and re-trained layers, an angle compensation strategy to get rid of the camera behavior, and a cyclic loss function based on soft labels that take the intra-class distance into account. The obtained results outperform the model-based ones; more concretely, a median absolute error of 11.60 degrees in the test set is achieved. Moreover, since complete datasets are difficult to gather, a sequential-based pipeline has also been proposed, which

fuses data from different domains in order to establish the ground truth orientation of the player (sensor-domain) in each bounding box (image-domain). The main limitation of the learning-based model is that only two different games have been used in the given dataset, as ground-truth sensor data (together with high-quality frames) are difficult to obtain. Nonetheless, this research shows that even with unbalanced training sets it is possible to train a model with a notable generalization capability that already outperforms the model-based method, hence promising results should be obtained with a more varied and balanced dataset in terms of different games.

Being the output of this research a set of raw time-based numerical orientations, different types of visualizations have been proposed in order to create tools that can provide coaches with useful orientation insights about their players. In particular, OrientSonars, and Reaction and On-Field maps illustrate the volume of given / received passes at each given orientation together with the corresponding accuracy at given game phases.

Future lines of research will be addressed in Chapter 21 at the end of Part III, since that part is closely related, in terms of applications, to the purposes of the current Part II. Nonetheless, in terms of improving the presented orientation estimation methods, novel training approaches could be explored. For instance, by training a network with two stacked inputs, such as individual player bounding boxes and the complete frame, and by merging their features, the angle compensation strategy could be avoided, since the network could already learn visual clues about the current camera shot.

Part III:
Pass Feasibility

A trained human eye sees the game a lot better than numbers, but the numbers see all the games, and that is a big advantage.

Dean Oliver

16 Introduction: Exploiting Orientation Data

In Part II, tracking data have been enriched with body orientation data, and even some purely-based orientation visualizations have been detailed, but... Does orientation really make a difference? Namely, does a pass between two properly oriented players have higher chances of reception? And if so, how much?

Once player body-orientation is gathered with the learning-based method presented in 13.2, this Chapter exploits these types of data in the context of soccer and provides quantitative and qualitative evidence, which prove that orientation is indeed a vital skill in a large set of situations. More concretely, we show that the inclusion of body-orientation benefits passing models in different key aspects:

- By modeling the passer's field of view, the potential receiving candidates are filtered out, since the passer does not generally move the ball towards a receiver outside his/her viewing area. In fact, the study of the human field of view has a long history [127; 25; 22], and the main associated outcomes explain how the humans' field of view splits into central and peripheral vision. In these articles, it is shown how humans deal with cognitive load in these areas, and among all the presented experiments, a large performance boost can be spotted when handling tasks within the visible spectrum (central vision, mostly). In the case of soccer, the so-called *no-looking-passes* are a rare and

a risky event, so players generally pass to those of their open teammates who can be spotted within their field of view.

- Given the current speed of soccer, receivers who are not properly oriented are prone to lose the ball easily because of bad control. By running towards prolific field spots with the appropriate orientation, receivers can control the ball without decelerating. Therefore, among all receiving candidates, the player is more likely to pass to the best-oriented one.

- Roughly, by gathering orientation data, it can be deduced whether defenders are approaching the offensive team while running forward or backward, hence drastically changing the location of open spaces surrounding the defender's back-side. By bringing together the orientation of all defenders, the big-picture of the current event is obtained, which shows the most dangerous field spots caused either by faulty movements or by an inappropriate set of orientations.

In particular, we present two different contributions that stem from body orientation data:

- A novel computational model to compute pass feasibility between a specific passer and the set of potential receivers. This model returns who is the safest ball-receiver at a time; throughout this Chapter, we will refer to this approach as the *discrete states model*. By merging positional data from both teams, and by adopting a geometrical solution to estimate the orientation fit between the passer and all receivers, notable accuracy in terms of Top_N measures is obtained. Moreover, the inclusion of orientation proves to boost the obtained accuracy, not only in terms of the presented feasibility model but also within the context of existing EPV models.

- Since soccer is a dynamic team-sport where open spaces play an important role, an extension of the discrete model is also

provided, thus resulting in 2D *pass feasibility maps*, which estimate the safest receiving spots in the field. As a matter of fact, orientation proves to be crucial when modeling field-of-view and correct positioning of players, since it limits the potential receiving area of all candidates. Different proposals are given to evaluate the proposed pass feasibility map; previous results are outperformed when orientation is included as a map feature.

Note that along this part, the notion of feasibility is used instead of the term *probability*, since the outcome of the pass event highly depends on the decision-making process of the passer, who is the one in charge of kicking the ball towards his/her chosen location. Therefore, the main goal of discrete states / feasibility maps is not to show the probability of a player passing to a particular player / location, but instead, to compute how safe it is, *a priori*, to pass the ball towards each receiver / field-spot based on the position and orientation of all the players in the field.

The rest of this Part is organized as follows: in Chapter 17 the state-of-the-art regarding passing tools and models is described. Later on, the discrete-states model is explained in Chapter 18, while its extension to a pass feasibility map is detailed in Chapter 19. The obtained results of both presented methods are shown and discussed in Chapter 20; last but not least, conclusions regarding the inclusion of orientation in this kind of applications can be found in Chapter 21.

17 State-of-the-Art (Passing Maps and Tools)

This Chapter reviews different soccer pass maps and automatic tools for analyzing pass events that have been proposed in the existing literature.

First of all, in order to clarify the contributions of other researchers, we would also like to emphasize and explain the different types of passing tools that have been used in previous literature. Given that there is not a universal term for the vast majority of them, their definition according to our understanding is provided:

- Pass Networks (*e.g.* [84]) show the interaction of players in terms of passing patterns in a graph style. The output of this map is the probability of a player passing to another one based on prior knowledge.
- Pass Maps (*e.g.* [86]) show where players pass to in a 2D template by using complementary tools such as arrows. For instance, it can display the most common pass directions of each player.
- Pass Probability Maps (*e.g.* [33]) predict where a player will pass in a given situation and at a given time.
- Pass Feasibility Maps predict which is the safest passing spot at every given moment in a given situation. Note that this kind of map is the one described in Chapter 19.

In 2015, Gyarmati and Anguera introduced a novel automatic extraction method to categorize the different passing strategies of soccer teams [44]. Their approach consisted of a dynamic time warping algorithm to identify spatial passing patterns given the ball 2D coordinates, thus reducing by a large margin the amount of time that has to be spent analyzing video clips when performing scouting-related tasks. Using this technique, high-level passing statistics were obtained, such as the frequency of a team running a particular play as well as its accuracy. Using a dataset from the first Spanish Division, results were split into several teams, hence showing its direct application in the soccer domain.
One year later, in order to quantify and evaluate different types of soccer passes in a given sequence or strategy, Gyarmati and Stanojevic proposed *Q-pass* [45]. Normally, pass accuracy is evaluated just by a straight-forward percentage that only takes into account if the receiver gets (or does not get) the ball. Since there exists a large set of possible passes (*e.g.* easy ones between centrals against tough ones between a central and a forward), this metric usually falls short. Therefore, *Q-pass* quantifies the quality of soccer passes given their initial conditions, such as field spot or defensive pressure. The main finding of the analysis of *Q-Passes* was that sometimes it is worth taking the risk and committing turnovers, because in the end, if a risky pass is successfully delivered in an advantageous field spot, a clear goal opportunity is generated.
Also in 2016, a relevant contribution was made regarding the quantification of risk and reward in passing events by Link *et al.* [62]. The authors proposed a real-time approach able to estimate the *dangerousity* in a generic soccer scenario similar to the above-mentioned EPV models [14; 34]. In that article, the concept of pass density was introduced, where the location of offensive players and the faced defensive pressure was modeled for every field spot. The passing risk was later suggested as an accuracy passing metric, that could complement the binary existing one. Similarly, Power *et al.* [86] also aimed to measure the risk-reward trade-off in those types of events by

approaching it as a regression problem. By using high-level features (micro, tactical, and formation) on a large dataset, many regressors were trained for specific situations (different types of passes). The authors extracted passing individual statistics such as *passing plus minus* or *passes received added*, which provided coaches with meaningful insights that indicated which players created the most profitable passes in particularly risky field locations.

Another statistical model thought for measuring the passing ability of players was proposed in [111] by Szczepanski *et al.*, where different passes were classified depending on several factors such as: location, time since the previous pass, type of pass, game time, current result, etc. Given these diverse scenarios, passes were examined under different conditions: control, passing player pressure, distance, receiving player pressure, or familiarity. By statistically modeling every single parameter based on tracking data and game knowledge, predictions were made, hence assessing whether passes were likely to be received or not. Consequently, the passing ability could be obtained by measuring how better / worse a player performs in terms of passing given a particular situation and comparing him/her to the rest of the league. Despite the dataset being used by the authors was old (Premier League 2006/2007), results were encouraging and its core should generalize to modern soccer.

From a physics-based point of view, and by quantifying the concepts of interception and control time, Spearman *et al.* also presented a passing model [102]. By treating pass events as Bernoulli trials, meaningful insights were obtained, such as pass value or data-driven predictions that indicated, for each event, who was the most likely receiver. Predicting where the ball should go during passes was analyzed by Hubácek's *et al.* [51], who refined the previous work made by Vercruyssen *et al.* [115]. In their paper, the authors proposed a deep learning architecture that assessed who was the most likely player to receive in each event. By building a feature vector of 13 dimensions, including game knowledge of both the offensive and defensive team, spatial relations were converted into convolutional filters, resulting in

a vector of receiving probability for each potential receiver. Using 200 random sequences, 0.55 Top_3 accuracy was obtained, outperforming previous methods. Similarly, Fernández *et al.* also proposed a deep learning architecture in *SoccerMap* [33], which used high-frequency spatiotemporal data in order to output probability surfaces. Their model was able to assess the decision-making process of players as well as identifying potential passing options and their associated risk.

18 Discrete Pass Feasibility

In this Chapter, we propose a discrete computational model to estimate the most plausible ball player pass at any given time based on game factors: player orientation, location, and faced defensive pressure.

To build the discrete model, we will attribute each potential receiver a feasibility score obtained by defining appropriate estimations that take into account player orientation and the configuration of the offensive and defensive team in the 2D field at that time. Intuitively, it stems from the fact that, in a pass event, there are 10 potential candidates of the same team who might receive the ball, each one of them holding a particular orientation with respect to the passer and at a certain position in the field; besides, the defensive team also needs to be taken into account.
Let $u(\cdot, t)$ be a color video defined on $\Omega \times \{1, \ldots, T\}$, where $\Omega \subset \mathbb{R}^2$ denotes the image frame domain and $\{1, \ldots, T\}$ is the set of discrete times. Given a time t, our method first considers the visible players in $u(\cdot, t)$ (*i.e.*, visible players in the image frame at time t) together with their body orientation. In this case, the detection of the players is given but, alternatively, as explained in Section 9.2, other approaches and detectors can be used, such as, *e.g.*, [93; 19; 54]. From now on, the position and orientation of the players will be considered over a 2D field template. To simplify the notation, the dependence on t of the considered elements will be omitted. Let P denote the 2D position

in the template field of the player with the ball at time t who is going to execute the pass. Let $\{R_i,\, i = 1, \ldots, I\}$ and $\{D_k,\, k = 1, \ldots, K\}$ denote, respectively, the 2D position in the field of the visible teammates of P, and the current defenders at time t, with $I \leq 10, K \leq 11$. The former ones constitute the set of visible receivers of the ball at time $t + \Delta_t$, being Δ_t the duration of the pass.

Let H_i denote the hypothesis that player P is going to pass the ball to receiver R_i. The main idea is to define a feasibility measure which is grounded on three elements: (a) the body orientation of every player together with (b) the pressure of the defenders D_k, both on P and R_i, and (c) the relative position of R_i with respect to P. Then, the most feasible ball pass $\hat{H}$ is computationally selected as the one maximizing

$$\hat{H} = \arg\max_i F(i), \tag{18.1}$$

where $F(i)$ is the feasibility of the event pass in H_i, which can be defined as

$$F(i) = F_o(i)F_d(i)F_p(i), \tag{18.2}$$

where $F_o(i)$, $F_d(i)$, and $F_p(i)$ stand for the orientation, defenders and proximity scores, respectively, defined later in this Chapter. Finally, it must be stated that all feasibility measures are obtained right at the moment when the passer P kicks the ball.

18.1 Orientation

One of the aspects that drastically affects the outcome of a pass is the players' body-orientation. If a player is relatively close to the passer and without being defended, he/she might still not be able to receive the ball properly if he/she is facing away. For a given pass event, the orientation of each player is computed in a window of $\pm Q$ frames with respect to the exact pass moment t. The median value of these $2Q+1$ observations is considered as the player orientation in

the event at time t. In practice, a window of 5 frames is used in 25 fps videos. Once obtained this estimation, an orientation-based pass feasibility measure is proposed, which takes into account geometrical quantities and outputs a score of how well a player is oriented in order to receive the ball. In order to take only the orientation information into account (proximity between players will be considered in the 3rd feasibility measure, as seen in Subsection 18.3) all potential receivers R_i are placed at the same distance with respect to the passer whilst preserving the original angle in the 2D field between the passer P and each receiver R_i. Note that this angle is only related to relative position and not to player body orientation. This step is illustrated in Figure 18.1.

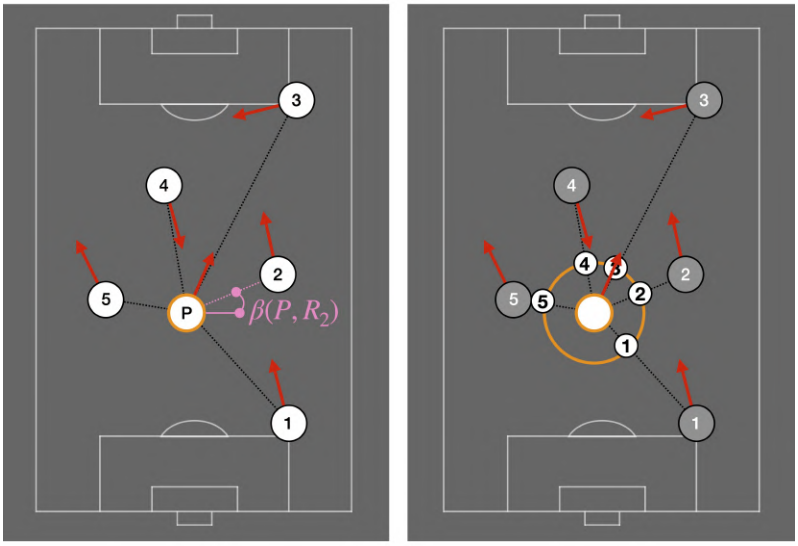

Figure 18.1: In order not to take pairwise distances into account while computing orientation feasibility, all players are moved towards an equidistant distance (unit circle).

Once all potential receivers are placed at an equidistant distance $Z > 0$ with respect to the passer, the body orientation of all players, expressed as $\phi(P)$ and $\phi(R_i)$ for the passer and the receiver i, respec-

tively, is considered (it corresponds to red vectors in Figures 18.1 and 18.2). Intuitively, $\phi(P)$ provides an insight of the passer field of view, and by setting a range of $\pm\psi^{\text{o}}$ with respect to the passer body orientation, an approximate spectrum of the passer field of view is obtained. By setting $\psi^{\text{o}} > 0$ to a fixed value (set to 30 degrees), an isosceles triangle with the two equal sides of length $2Z$ is defined (see Figure 18.2). This triangle is denoted by T_P and imposes a limit to the region where the player can pass the ball. The same procedure is repeated for $\phi(R_i)$, with the triangle T_{R_i} indicating the field of view of the receiver, which shows in which directions he/she can get a pass from; the length of the two equal sides of triangle T_{R_i} is set to Z. Figure 18.2 displays some possible scenarios. We claim, and numerically verify in Section 20.1.1, that the weighted area of the intersection of triangles T_P and T_{R_i} gives a measure of how easy it can be for a player to receive a pass in the given orientation configuration: no intersection indicates the inability to get it, whilst partial or total intersection indicates a proper orientation fit. Accordingly, orientation-based feasibility is defined as

$$F_o(R_i) = \frac{1}{c} \int_{T_P \cap T_{R_i}} \left(e^{-\mathrm{d}(P,x)} + e^{-\mathrm{d}(R_i,x)} \right) dx \tag{18.3}$$

where $c > 0$ is a normalizing constant and $\mathrm{d}(a, b)$ denotes the Euclidean distance between a and b normalized so that the maximum distance in the field is 1.

Let us first discuss the terms in (18.3), namely, both exponential functions. The intrinsic geometry of the triangle has an obvious limitation when it comes to shape intersection: considering the vertex that coincides with the passer position as the triangle beginning, triangles contain a large portion of the area in regions placed far from their beginning. Hence, the values inside the computed triangles are weighted according to their relative position with respect to the triangle beginning, fading out in further positions. This effect can be seen as different color opacity in the triangles displayed in Figure 18.2. Finally, the reasoning for setting different triangle heights is

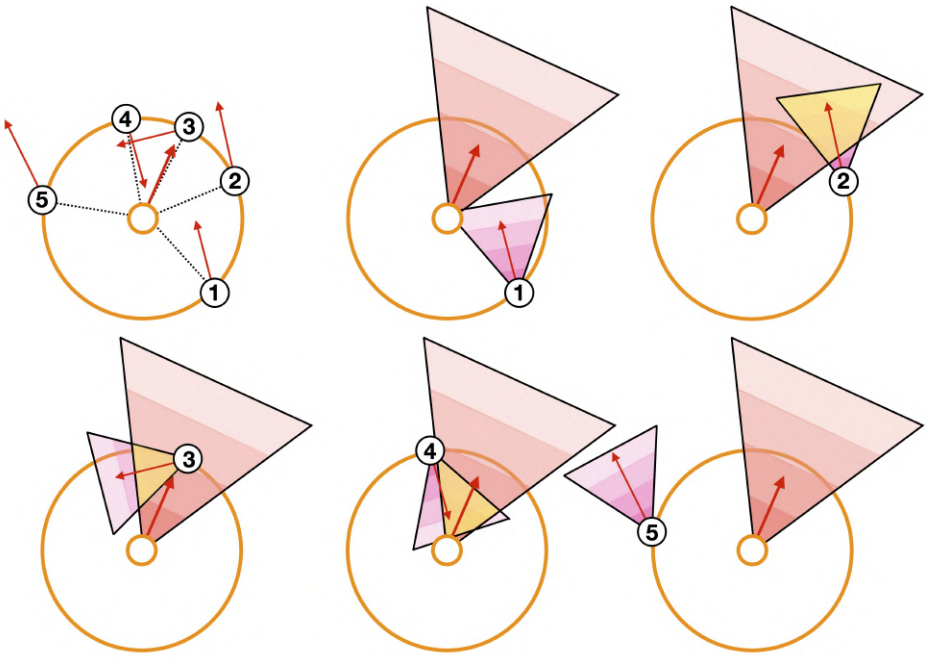

Figure 18.2: Individual scenarios of intersection given the relocated players of Figure 18.1. As it can be seen, the top-right player is the best oriented candidate to receive the ball.

that, if both passer' and receiver' associated triangles had the same height, players that are located behind a passer who is not looking backward would intersect notably, despite being a non-feasible pass (like in the top-centered example sketch of Figure 18.2).

18.2 Defenders Position

Apart from considering all visible players of the offensive team, the behavior of the defenders, $\{D_k\}_k$, is continuously changing the decision-making process. Even if a player is near the passer and properly oriented, the probability of receiving the ball can be really

low if he/she is properly guarded; however, it is hard to define how well a player is being defended at a time. Considering only passing events, defenders close to the line that connects the passer with the receiver (passing line) are the ones in a more advantageous position to transform a pass into a turnover. Let us denote by $\beta(P, R_i)$ the angle in the 2D template field between the passer P and the receiver R_i (see Figure 18.1), and by $\beta(P, D_k)$ the one between the passer P and defender D_k. Using this angle, the proposed defenders-based feasibility will take into account two feasibility scores: (a) the feasibility $\mathrm{F}_{d,P}(R_i)$ of passing in the direction of $\beta(P, R_i)$ and (b) the feasibility $\mathrm{F}_{d,R}(R_i)$ of receiving the ball from P. For the first case, the distance and the angle of all defenders with respect to the passer is computed. Therefore, the definition of the feasibility measure $F_{d,P}(R_i)$ depends on the Euclidean distances of the closest defenders with respect to the passer:

$$F_{d,P}(R_i) = \exp\left(-\frac{1}{J}\sum_{k\in\mathcal{N}_P} w\left(\beta(P, D_k), \beta(P, R_i)\right)\left(1 - \mathrm{d}(P, D_k)\right)\right) \tag{18.4}$$

where $\mathcal{N}_P$ denotes the set of the J nearest neighbor defenders from P, according to the weighted distance d_w, defined as

$$\mathrm{d}_w(P, D_k) = w(\beta(P, D_k), \beta(P, R_i))\,\mathrm{d}(P, D_k) \tag{18.5}$$

where $\mathrm{d}(P, D_k)$ denotes the normalized Euclidean distance between P and D_k. Finally, the weights w are defined as

$$w(\beta(P, D_k), \beta(P, R_i)) = \begin{cases} 0.25 & \text{if } \alpha < 22.5^{\mathrm{o}} \\ 0.5 & \text{if } 22.5^{\mathrm{o}} \leq \alpha < 45^{\mathrm{o}} \\ 2 & \text{otherwise} \end{cases} \tag{18.6}$$

where $\alpha = |\beta(P, D_k) - \beta(P, R_i)|$ (modulus 360^{o}). In practice, we take $J = 3$.

Function w is used to model that defenders close to the passing line (and thus with an associated small ω value) entail a higher risk for that specific pass. This whole procedure can be seen in the left side of Figure 18.3, where the three closest defenders are highlighted for two hypothetical passes.
For $\mathrm{F}_{d,R}(R_i)$, the same procedure is repeated with respect to the receiver; however, in order to have two independent quantities, the J nearest neighbors considered when computing $F_d(P)$ are discarded. Hence, $\mathcal{N}_{R_i}$ is the set of the J nearest neighbor defenders from R_i (according to d_W) belonging to $\mathcal{N}_{\mathcal{P}}^{\mathcal{C}}$, *i.e.*, the set of the visible defenders at time t that are not in $\mathcal{N}_P$. The feasibility to receive the ball from a given angle can be expressed as:

$$F_{d,R}(R_i) = \exp\left(-\frac{1}{J}\sum_{k\in\mathcal{N}_{R_i}} w\left(\beta(R_i, D_k), \beta(P, R_i)\right)\left(1 - \mathrm{d}(R_i, D_k)\right)\right) \tag{18.7}$$

The right part of Figure 18.3 shows a graphical example, where the top closest weighted defenders are found with respect to the receiver once discarded the closest defenders found when computing $F_{d,P}(R_i)$ (Figure 18.3). To conclude, the defender's feasibility is defined as $F_d(R_i) = F_{d,P}(R_i)F_{d,R}(R_i)$, and it is a measure of how likely the event of passing to a particular player is, given the defensive spatial configuration.

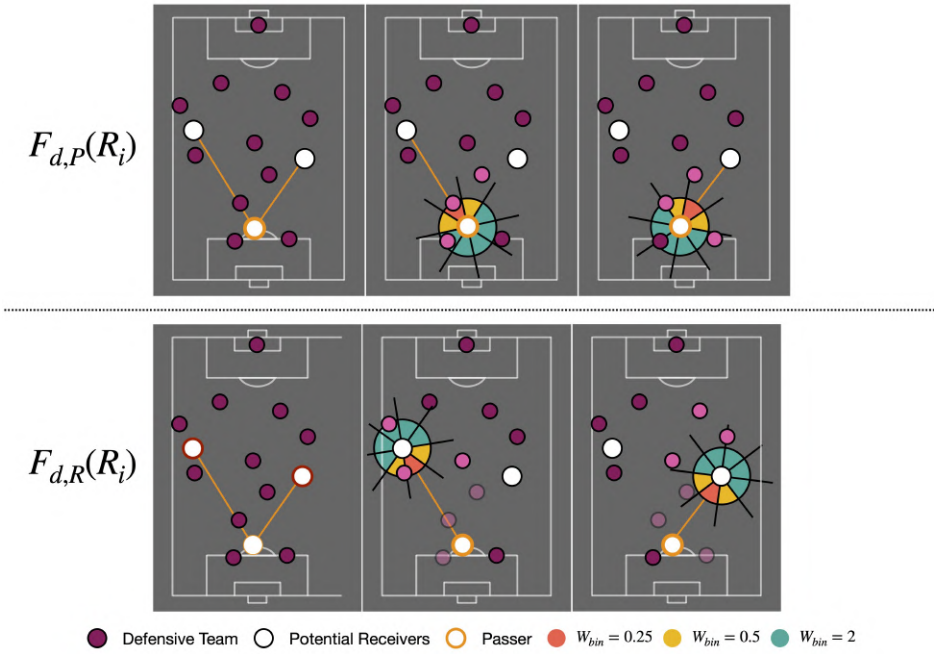

Figure 18.3: Computation of $F_{d,P}(R_i)$ and $F_{d,R}(R_i)$ for two different potential receivers. For both cases, (left) general setup, plus detection of the 3 closest weighted defenders in the scenario of the (middle) left-sided and (right) right-sided player.

18.3 Pairwise Distances

Finally, the position in the 2D field also affects the passing options, as players placed closer to the passer have a higher probability of receiving the ball. For this reason, the feasibility of receiving the ball based on pairwise distances or proximity can be defined as inversely proportional to the distance by:

$$F_p(R_i) = \exp\left(-\mathrm{d}(P, R_i)\right) \tag{18.8}$$

18.4 Combination

Once all three independent feasibility measures are computed, Equation (18.2) is proposed to combine them. Notice that a low feasibility value in one of the three features (orientation, defenders, or distance) indicates that the pass has a high associated risk, no matter what the other values are.

19 Pass Feasibility Maps

In this Chapter, the previously-explained discrete model is extended into a 2D pass feasibility map whilst preserving orientation as a key feature in the method's core.

Since raw orientation data cannot be used by coaches to provide the team with new strategies and tactics, and given that discrete feasibility values limit the location of potential receptions, pass feasibility maps get the complete picture of pass events. Once again, by default, our model assumes that the most feasible receiving candidate is the (a) best oriented, (b) closest and (c) less defended one. Orientation hereby plays a vital role both in the offensive and the defensive team. On the one hand, the orientation of offensive players is included in the offensive map M_O (defined in Section 19.1), which estimates the orientation fit between the passer and all potential receivers regardless of the defensive setup. On the other hand, the defensive map M_D (defined in 19.2) is in charge of modeling the defensive pressure in every field spot, thus finding out which defenders are creating larger open spaces in their surroundings. The proposed feasibility map is defined by combining both the offensive M_O and defensive M_D contributions as:

$$M(\mathbf{x}) = \kappa M_O(\mathbf{x}) + M_D(\mathbf{x}), \tag{19.1}$$

where $\mathbf{x}$ denotes a position in the 2D field and $\kappa \geq 1$ is a scalar parameter that balances the given weight of the attackers' informa-

tion in the final feasibility map, *i.e.* a larger / smaller value of κ would favor riskier / more conservative passes. An example is displayed in Figure 19.1: yellow regions belong to the field-zones where safe passes can be attempted, as offensive players are in favorable conditions (proper orientation and location) to get the ball; instead, blue zones represent field parts where the defensive team is likely to recover the sphere.

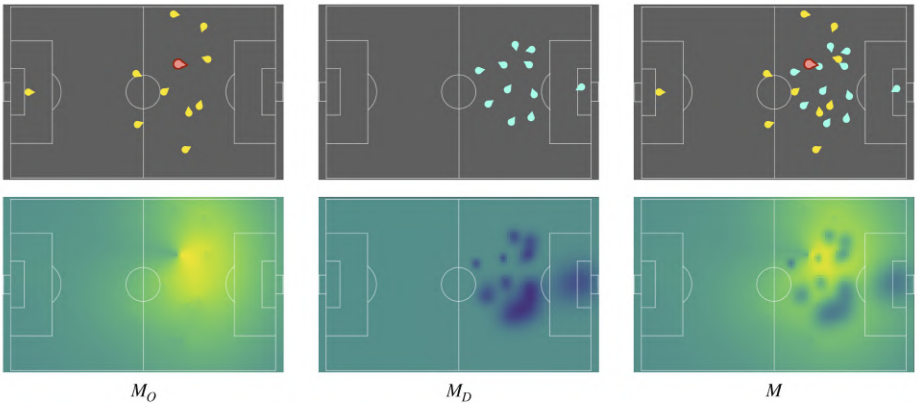

Figure 19.1: Procedure to obtain a feasibility map for a given pass event. We can see the spatial configuration (plus orientation) of: (top-left) the players in the offensive team, being the red player the one carrying the ball, (top-center) the players in the defensive team, and (top-right) their combination in one same display. The final feasibility map M (bottom-right) is obtained through the aggregation of the offensive map M_O (bottom-left) and the defensive map M_D (bottom-center). Note that yellowish regions are the ones with higher associated feasibility (safer passes towards these directions), and bluish regions are the dangerous parts of the field.

19.1 Offensive Team Modeling

During pass events, offensive players can adopt two roles: passer or receiver. On the one hand, the passer controls the ball and the overall situation; after all, his/her decision-making process will make the ball move in a particular direction. On the other hand, receivers' goal is to facilitate the passer's decision by moving towards field locations where the probability of scoring / receiving a goal is potentially maximized / minimized. For instance, if the passer is surrounded by a lot of defenders in a dangerous position, the other offensive players need to create easy passing lines to avoid committing a turnover; on the other hand, if the passer is in an advantageous situation in the offensive side of the field, receivers need to find open spaces and to create scoring opportunities. Therefore, the orientation fit between the passer and the receiving candidates (plus their spatial distance) is expected to influence the outcome of passes. In our proposal, the orientation fit is modeled through two maps, namely, the Receiver Map M_R, which includes location and orientation data of all 10 potential receivers, and the Passer Map M_P that only includes data from the passer. Both maps are then combined into the offensive map M_O as seen in Figure 19.2, which is defined as:

$$M_O(\mathbf{x}) = M_P(\mathbf{x})M_R(\mathbf{x}). \tag{19.2}$$

19.1.1 Receiver Map

Given a receiver j at position $\mathbf{r}_j$ with orientation α_{Rj} , the aim of this map is to model his/her *receiving area*, which expresses in which part of the field the current receiver is likely to get the ball. For a single receiver, the proposed receiver map, at a generic position $\mathbf{x}$ in the 2D field, is built on Gaussian functions and defined as

$$M_{Rj}(\mathbf{x}) = \underbrace{\exp\left(\frac{-\left\|\mathbf{x}-\mathbf{r}_j\right\|^2}{\sigma_R^2}\right)}_{g_{R_j}(\mathbf{x})} \underbrace{\exp\left(\frac{-(\angle(\mathbf{x}-\mathbf{r}_j)-\alpha_{Rj})^2}{\sigma_a^2}\right)}_{g_{aRj}(\mathbf{x})}, \tag{19.3}$$

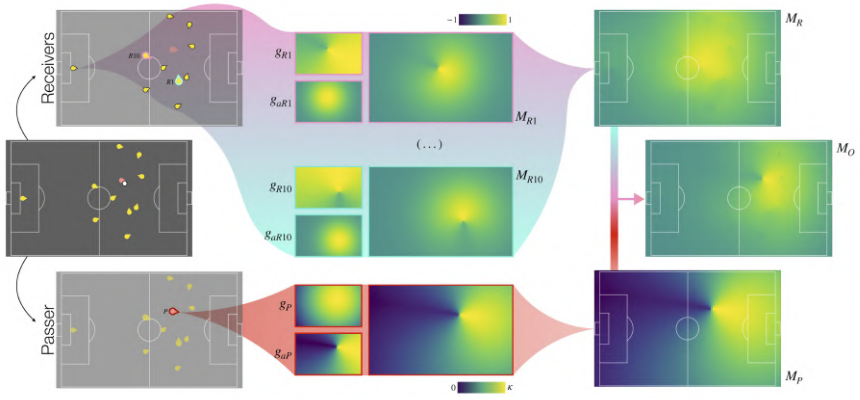

Figure 19.2: Offensive team map modeling M_O. (left column) Given an initial 2D setup (locations plus orientation) of all the players: (top and middle row) a receiver map M_R is created by adding together all receivers' contributions, and it is later combined with (bottom row) the passer map M_P, which takes into account his/her field of view.

where $\sigma_R, \sigma_a > 0$, thus resulting in a function that decreases as point $\mathbf{x}$ moves away from $\mathbf{r}_j$ and as its orientation with respect to $\mathbf{r}_j$ differs from α_{Rj}. The difference of angles $\angle(\mathbf{x} - \mathbf{r}_j) - \alpha_{Rj}$ is computed in modulo 2π. A detailed explanation of parameters σ_R and σ_a is later provided in Section 19.3.
Please also note that g_{Rj} and g_{aRj} in Equation 19.3 express the two priors to define who is the *best-positioned* candidate, *i.e.* location and orientation, respectively. The final receiving map is defined as:

$$M_R(\mathbf{x}) = \sum_{j=1}^{N_R} M_{Rj}(\mathbf{x}), \tag{19.4}$$

where N_R is the number of potential receivers. As seen in Figure 19.2, maximum values can be found in M_R, where receiver's contributions overlap; besides, it can also be seen how both individual priors g_{Rj} and g_{aRj} are displayed in the middle column.

19.1.2 Passer Map

Given the passer at position $\mathbf{p}$ and orientation α_P, the proposed passer map relies again on Gaussian functions and is defined as

$$M_P(\mathbf{x}) = \underbrace{\exp\left(\frac{-\|\mathbf{x}-\mathbf{p}\|^2}{\sigma_P^2}\right)}_{g_P(\mathbf{x})} \underbrace{\exp\left(\frac{-(\angle(\mathbf{x}-\mathbf{p})-\alpha_P)^2}{\sigma_a^2}\right)}_{g_{aP}(\mathbf{x})}, \quad (19.5)$$

where $\sigma_P > 0$, and once again g_P and g_{aP}, are related to the location and orientation of the player respectively (bottom row of Figure 19.2).

19.2 Defensive Team Modeling

The role that defenders' orientation plays in the decision of the passer is significantly different. As it has been mentioned, the goal of some offensive players is to detect open spaces and occupy them before the defenders do; these offensive players are usually oriented, since the beginning of the event, towards the open space. Nonetheless, defenders' orientation switches more often: although they start facing the opponent nose-to-nose (moving backward), they might have to turn around and recover back to defense at some point (moving forward). Now, instead of talking about a *receiving area*, the concept of *influence area* will be used, expressing the part of the field that is being *controlled* by each defender. Besides, when modeling defenders, two other aspects have to be remarked:

- The defender's current position is crucial for the pass outcome. Nevertheless, the danger of each defender in his/her surroundings depends on the ball-defender pairwise distance. In short passes, defenders who are close to the ball have almost no reaction time, so the potential influence area should be limited to a small spatial neighborhood; nonetheless, during long passes, the reaction time is way larger, thus resulting in a broad influence area.

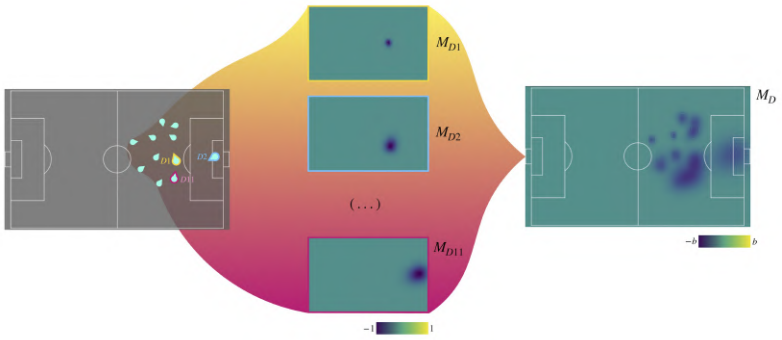

Figure 19.3: Procedure to model the defensive team map M_D, based on the aggregation of individual reach contributions M_{Di}. As it can be seen, each individual reach is proportional to the pairwise ball-defender distance and it points towards the player's orientation.

- The orientation of defenders is an important factor in order to model the shape of the influence area: in particular, a function defined by two parts with smooth contour lines is used. The first part corresponds to the front side of defenders (with respect to their orientation), which is the one they attempt to control and reach. The second part falls behind defenders; since moving backward is slower than running forwards, its influence area is reduced.

More concretely, given a defender i at position $\mathbf{d}_i$ plus orientation β_i, the proposed defensive pass map is defined as:

$$M_{Di}(\mathbf{x}) = \begin{cases} -\exp\left(\frac{-\|R_{\beta_i}(\mathbf{x}-\mathbf{d}_i)\|^2}{\sigma_{D_i}^2}\right) & \text{if } (R_{\beta_i}(\mathbf{x}-\mathbf{d}_i))_x < 0, \\ -\exp\left(\frac{-(\frac{1}{2}(R_{\beta_i}(\mathbf{x}-\mathbf{d}_i))_x^2+(R_{\beta_i}(\mathbf{x}-\mathbf{d}_i))_y^2)}{\sigma_{D_i}^2}\right) & \text{otherwise,} \end{cases} \tag{19.6}$$

where R_{β_i} denotes a rotation of angle β_i and $(.)_x$, $(.)_y$ denote the x and y coordinates, respectively; Figure 19.4 illustrates this map for

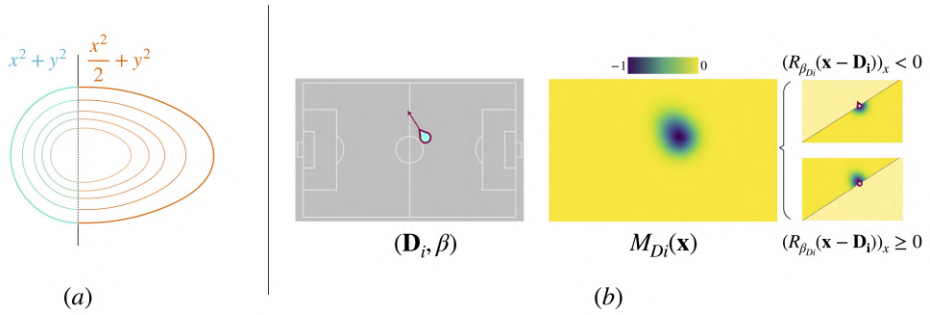

Figure 19.4: (a) Level lines of an individual defensive contribution, M_{D_i}, without taking orientation into account; (b) final individual defensive map for a particular player with orientation β.

a certain orientation β_i. Moreover, $\sigma_{D_i} > 0$ depends on the pairwise distance between the ball and the defender, and characterizes the mentioned influence area (check Section 19.3 for a detailed explanation).
Finally, the sum of all defenders' contributions is aggregated into one single defensive map M_D:

$$M_D(\mathbf{x}) = \sum_{i=1}^{N_D} M_{Di}(\mathbf{x}), \tag{19.7}$$

where N_D is the number of current defenders in the field (an example is displayed in Figure 19.3). Notice also that the exponential functions of the defensive team are negative, as opposed to the positive values of the offensive team.

19.3 Parameter Choice and Discussion

This Section discusses the parameter choice in order to balance the offensive and defensive contributions within a similar range and reach and, thus, aiming to build an easy-to-interpret visual resource for coaches and analysts. The optimization of these parameters has

been approached as the maximization of Top1 and Top3 accuracy in pass events, as it will be later defined in Section 20.2; by using random-search [3], the values for the described parameters have been set.

19.3.1 Offensive Gaussian Size

The first two parameters to be adjusted are σ_R and σ_P, which can be found in g_{Rj} and g_P, from (19.3) and (19.5), respectively; both parameters aim to model, inside the player's field of view, the spatial *reach* of every offensive player. In particular:

- For the passer (σ_P in g_{Rj} (19.3)), this reach introduces a prior on how far he/she is likely to pass the ball.
- For all the potential receivers (σ_R in g_P (19.5)), this reach estimates the viable *receiving area* around each player. Since the receiver has higher chances of getting the ball at his/her current position or nearby, the map's values vanish when moving away from the receiver.

Logically, the passer must have a larger reach, since a strong kick can move the ball far away at a much faster speed than the player's average velocity. The effect of different choices of σ_R and σ_P can be observed in Figure 19.5. After the optimization step, σ_R^2 and σ_P^2 are fixed to 10^3 and 10^4 respectively, which seem reasonable values, since they provide the receivers / passer with a 15- / 50-meters reach respectively.

19.3.2 Offensive Angle Compensation

The main aim of σ_a, included in g_{aRj} and g_{aP} from (19.3) and (19.5), is to boost the pass feasibility in the field positions that match player orientation, whilst decreasing backwards' locations. Three examples are displayed in Figure 19.6, where the difference between the

minimum and maximum values in the image domain decreases when increasing σ_a. Among all possible values in the random-search grid, the optimal one in terms of pass accuracy is $\sigma_a = 0.75$.

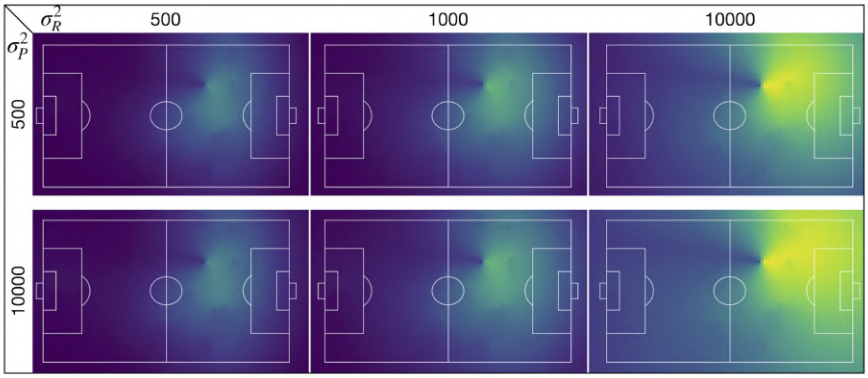

Figure 19.5: Offensive maps M_O are adjusted by tweaking σ_R and σ_P. Notice that the model is quite robust to the choice of these parameters, being $\sigma_R^2 = 10^3$ and $\sigma_P^2 = 10^4$ suitable values in terms of passer / receiver reach.

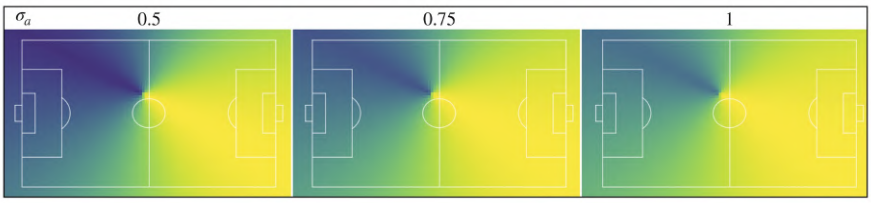

Figure 19.6: Individual defensive maps M_{Di} are adjusted by tweaking σ_a; noticeable differences emerge in the opposite direction of the player's orientation, especially when $\sigma_a = 0.75$.

19.3.3 Defensive Size and Offensive Boost Weight

As detailed in Section 19.2, while modeling defenders, our goal is to encompass the influence area in which the player might steal the ball. Apart from the major role of the own defender's orientation, this area is designed to be proportional to the pairwise distance between the defender and the ball. In particular, given a defender at position $\mathbf{D}_i$ and a passer at $\mathbf{p}$, the parameter σ_{Di} in (19.6) is defined as:

$$\sigma_{Di} = \frac{\|\mathbf{d}_i - \mathbf{p}\|}{\sigma'_D}. \tag{19.8}$$

Thus, the only parameter to be adjusted is σ'_D. Besides, σ'_D has to be optimized while taking into account the offensive boost weight κ in (19.1). Suitable values for both parameters lead to understandable visual pass maps, which can be directly interpreted by coaches or analysts. Some combinations are displayed in Figure 19.7.

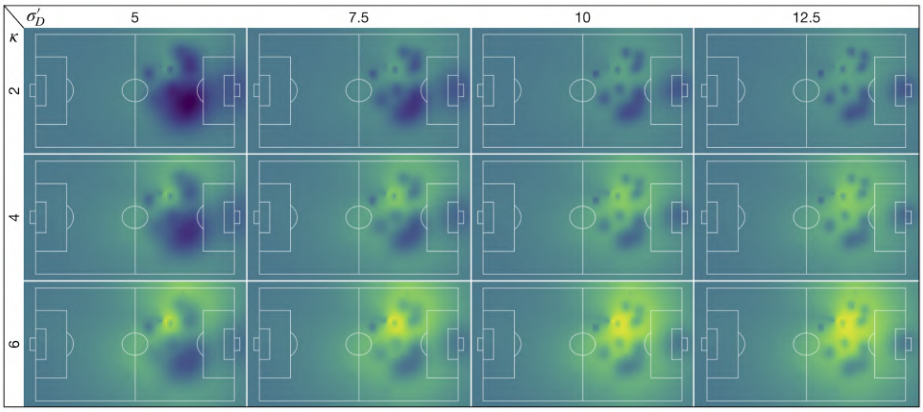

Figure 19.7: Pass feasibility maps M are adjusted by tweaking κ and σ'_D. By using a reasonable trade-off (*e.g.* $\kappa = 4$ and $\sigma'_D = 12.5$), the appropriate relevance is given both to the offensive and the defensive teams.

20 Pass Feasibility Results

In this Chapter, a complete ablation study of both presented methods (discrete states and pass feasibility maps) is given. Note that for all mentioned tests, data from FCB_{DS} have been used.

Since the problem of building pass feasibility has not been fully explored yet, there is still a lack of universal evaluation metrics. To quantitatively assess it, we extract, for a given play, one feasibility value per player, which are later sorted to get a ranking of the most likely receivers. Then, a Top_N ($N \in \{0, 1, \ldots, 10\}$) accuracy measure is implemented, similarly as [106; 51]. The Top_N measure computes the number of times the actual receiver of the pass is within the first N options of the given model. For instance, on the one hand, Top_1 accuracy indicates the percentage of events where the best candidate predicted by the model matches the actual receiver; on the other hand, Top_3 accuracy returns the number of times the actual receiver is among the best 3 predicted candidates. Note that this type of metric has been used for the evaluation of both discrete states and pass feasibility maps.

20.1 Discrete States

In this Section, several experiments will be detailed with one main goal: to study if proper orientation of soccer players is correlated with successful receptions, thus maximizing the probability of creating a potential goal opportunity. Hence, in order to examine the effect of including the orientation, another baseline pass model will be used for testing, which will only use the output of F_p (proximity) and F_d (defense), thus getting rid of F_o (orientation); more concretely, F (defined in Equation (18.2)) will be compared with F_{pd}, defined as:

$$F_{pd}(R_i) = F_p(R_i)F_d(R_i). \tag{20.1}$$

Moreover, histograms will be plotted for each scenario. In all cases, the number of bins is 9, as it corresponds to the number of potential receivers of a play; note the goalkeeper has been excluded because it does not appear in the frame domain in many events. The height of each particular bin B_n (with $n \leq 10$) represents the number of times that the ground truth receiver has been considered the n best candidate according to the feasibility values (for instance, B_1 equals the number of times that the actual receiver was considered as the best option). In these Figures, the histograms of successful (blue) and unsuccessful (orange) passes are plotted together.

20.1.1 Orientation Relevance in Pass Feasibility

The importance of orientation in the computation of the proposed feasibility F will be shown by comparing the results of F with the ones obtained with the baseline feasibility F_{pd}, which does not include orientation. As it can be seen in Table 20.1, in both cases the $\text{Top}_{1/3}$ metric shows that the introduced features in the feasibility computation are directly correlated to the outcome of the play: the difference in Top_1 accuracy between successful and non-successful passes is more than the double, and in Top_3 accuracy, it is more than 0.2. Besides,

orientation makes a difference by complementing distance and defenders. Apart from boosting the difference between successful and non-successful passes by a margin of 0.04 / 0.02, F outperforms F_{pd} Top_1 accuracy by 0.07 and Top_3 by 0.05. Visually, this difference can be spotted in the first bins of the histogram displayed in Fig. 20.1.

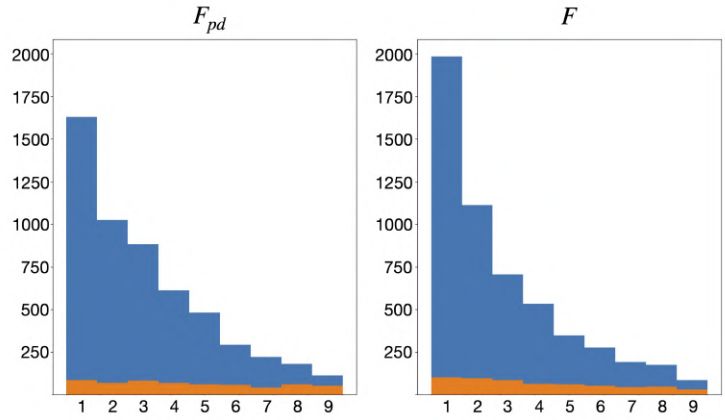

Figure 20.1: Histogram distribution comparison between F_{dp} and F; note that the latter includes the computed orientation feasibility.

	Top_1 (Succ.)	**Top_1 (NSucc.)**	**Top_3 (Succ.)**	**Top_3 (NSucc.)**
F_{pd}	0.299	0.149	0.650	0.411
F	0.367	0.175	0.702	0.487

Table 20.1: $\text{Top}_{1/3}$ accuracy for successful / non-successful passes obtained before (F_{pd}) and after (F) including orientation as a feasibility measure.

Decomposed F_o - F_d - F_p Performance. In order to show how useful the individual estimations are, the performance of the three individual feasibility measures (F_p, F_d, and F_o) is studied together with

their combination. These results are shown in Table 20.2 and Figure 20.2. For the successful passes, the histogram of all three components shares more or less the same shape. However, the top bins of F_p have higher values (0.34, 0.70 for Top_1 and Top_3 accuracy, respectively); as a result, the bottom bins have low values, which means that it is unlikely to pass the ball to players placed far away with respect to the ball. For the unsuccessful passes, F_d and F_p components seem to be the most and less relevant ones, respectively. This means that passing to a player who is far away does not always imply a turnover, but passing to a well-defended player does (0.14 difference in Top_1 accuracy). Generally, F_o resembles F_p, but the former histogram is more distributed (flat shape). Combining all three methods (by computing their product) adds some value due to contextualization. For instance, orientation by itself does not take pairwise distances into account: this means that, in particular scenarios, players placed far away in the field might be the best potential candidates in terms of orientation, but as it has been proved, these passes will hardly ever exist. Besides, our proposed feasibility measure F (defined in (18.2)) combines all three components and keeps the high Top_1 and Top_3 metrics of F_p whilst preserving the difference between the successful / not-successful passes of F_d. The bottom-right histogram shows that this goal has been accomplished.

	Top_1 (Succ.)	**Top_1 (NSucc.)**	**Top_3 (Succ.)**	**Top_3 (NSucc.)**
F_o	0.260	0.232	0.566	0.546
F_p	0.340	0.320	0.704	0.665
F_d	0.243	0.107	0.604	0.336

Table 20.2: $\text{Top}_{1/3}$ accuracy for successful / non-successful passes obtained with F_o (orientation), F_p (proximity), and F_d (defensive pressure).

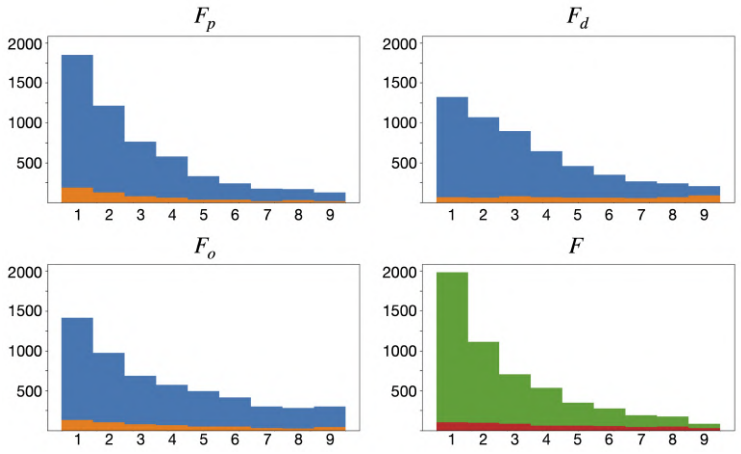

Figure 20.2: Histogram distribution among potential receivers (feasibility components). From left-right, top-bottom: (a) proximity F_p, (b) defensive pressure F_d, (c) orientation F_o and (d) Combination.

20.1.2 Players' Field Position / Game Phase

Once analyzed the impact of orientation as a feasibility measure, in this Subsection, its effect on the different kinds of players and game phases (explained in Section 2.2) are analyzed. By classifying them according to the basic field positions (defenders, midfielders, and forwards), Figure 20.3 and Table 20.3 show the differences, in terms of orientation-based feasibility, among them, which state that midfielders are the ones under bigger F_o influence. When introducing orientation in the feasibility measure, both the Top_1 and the Top_3 accuracy have a boost of 0.10 while preserving a similar difference in successful-unsuccessful differences (first 3 bins of the midfielders histogram). Defenders are not heavily affected by orientation, mostly because of the many security passes that they perform: in this type of pass (usually between defenders), both players have no opponents surrounding them, and they can freely pass to their closest team-mates without having to be strictly oriented towards them. Forwards are also affected by orientation, but they give and receive

fewer passes; besides, in their domain, passes do not only have a high turnover risk, but also a high potential reward.

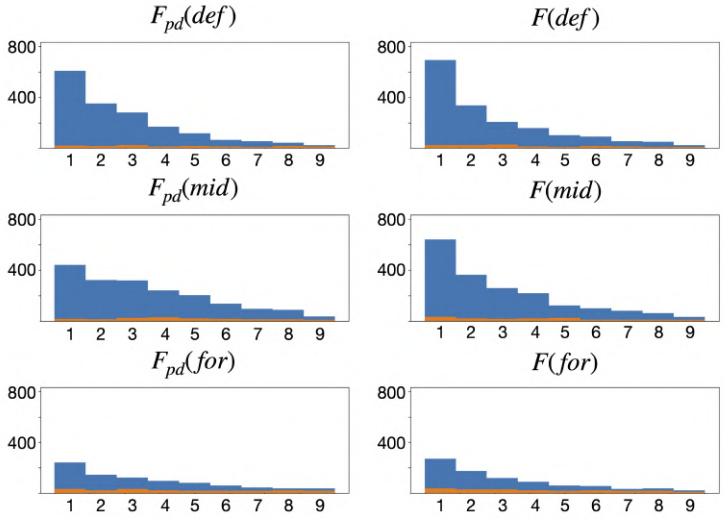

Figure 20.3: Histogram distribution, obtained with (left) F_{dp} and (right) F_{dpo}, for different player positions. From top to bottom: defenders, midfielders, and forwards.

	Top$_1$ (Succ.)	**Top$_1$ (NSucc.)**	**Top$_3$ (Succ.)**	**Top$_3$ (NSucc.)**
F_{pd} (def.)	0.354	0.134	0.724	0.436
F (def.)	0.404	0.162	0.720	0.521
F_{pd} (mid.)	0.235	0.114	0.575	0.341
F (mid.)	0.341	0.196	0.673	0.456
F_{pd} (for.)	0.278	0.158	0.589	0.426
F (for.)	0.315	0.178	0.653	0.459

Table 20.3: Top$_{1/3}$ accuracy for successful / non-successful passes, before / after including orientation, split by player position.

In a similar way, passes can be also classified according to the location of the passer in relation to the defensive team spatial configuration, thus indicating the game phase: (a) build-up, (b) progression, or (c) finalization. Results are displayed in Figure 20.4 and Table 20.4. Once again, the effect of orientation is vital in the half-court, with a notable difference between successful and non-successful passes in the progression phase (around 0.2 difference in both Top_1 and Top_3, and more than 0.7 Top_3 accuracy). As expected, the build-up and finalization game phases are, respectively, the ones with lower and higher risk, but even in these extreme cases, the inclusion of F_o also boosts the pass accuracy metrics.

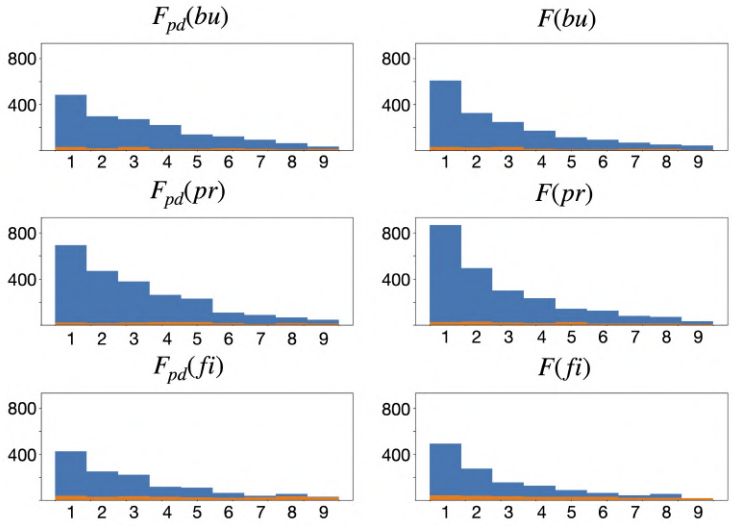

Figure 20.4: Histogram distribution, obtained with (left) F_{dp} and (right) F_{dpo}, for different game phases. From top to bottom: build-up, progression and finalization.

	Top$_1$ (Succ.)	**Top$_1$ (NSucc.)**	**Top$_3$ (Succ.)**	**Top$_3$ (NSucc.)**
F_{pd} (bu.)	0.282	0.143	0.610	0.382
F (bu.)	0.355	0.162	0.688	0.444
F_{pd} (pr.)	0.297	0.128	0.659	0.365
F (pr.)	0.372	0.162	0.712	0.480
F_{pd} (fi.)	0.326	0.185	0.687	0.490
F (fi.)	0.376	0.203	0.710	0.534

Table 20.4: Top$_{1/3}$ accuracy for successful / non-successful passes, before / after including orientation, split by player game phase (*bu* - build up, *pr* - progression, and *fi* - finalization).

20.1.3 Combination with EPV

As mentioned throughout this manuscript, EPV is a recently introduced indicator that tries to boost individual / team performance by assigning value to individual actions, using (among others) a pass probability model. However, the EPV model of [34] does not take the body orientation of players into account, thus producing results that, despite being notably accurate, can be refined. An example is shown in Figure 20.5; for the displayed pass event, the spatial output of the pass probability model (left) and the EPV map (right) can be seen in the middle row. As observed in the original frame, the passer (white circle) is the central midfielder, who is directly facing the right-central defender; for this reason, the passer cannot see in his field of view the left-central defender, hence lowering the latter's receiving chances. However, the output of the pass probability model considers the left-central defender as a notable candidate, but EPV does not penalize this pass as a risky one. Nevertheless, by combining our orientation-based feasibility measure F_o with the output of the (a) original probability model or the (b) output of the EPV model, maps could be adapted accordingly, thus enhancing potentially good receivers in particular regions as it is displayed in the last row of

Figure 20.5.

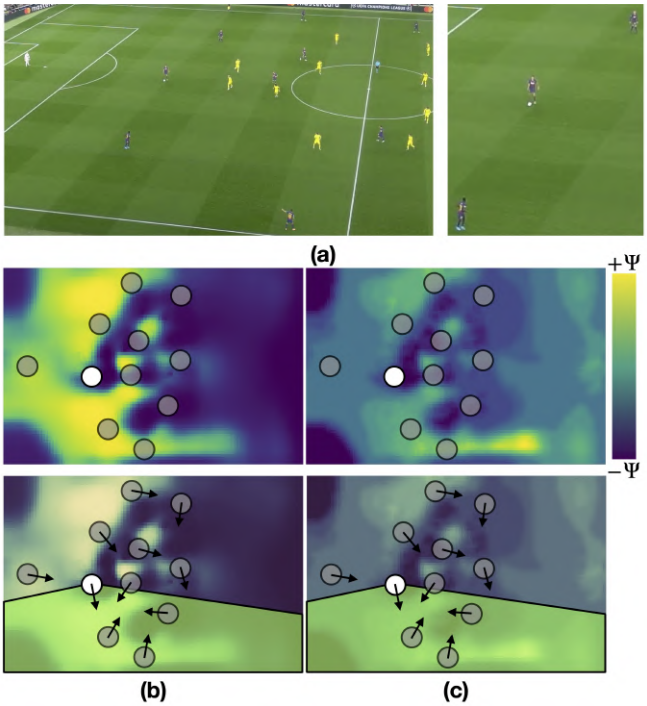

Figure 20.5: (a-left) Pass event and (-right) zoom in the passer region; (b,c-top) output of the pass probability / EPV models respectively of [34], typically Ψ equals 0.015, (b,c-bottom) output example made by hand; the combination of the existing models with body orientation would refine the restricting the area of potential receivers.

The main challenge when combining both methods is the dimension miss-alignment: both the pass probability and EPV models extract an output map with a value for each discretized field position (downscaled to 104×68), whilst the proposed model defines an individual feasibility value for each of the 10 potential receivers. In order to get a single probability / EPV value for each player in the field, and being ρ the output map (defined by the pixels of the downscaled field), a geometrical solution is provided; its approach is based on the idea

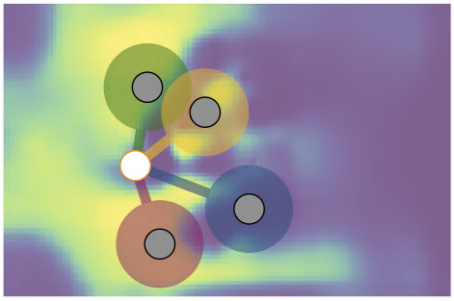

Figure 20.6: Geometrical approach to assign discretized pass probability / EPV field values to particular potential receivers.

that an individual value can be obtained by integrating the probability / EPV values on a meaningful area that extends from the passer to the receiver. In particular, for a given receiver R_i, first, a disc Q_i of radius $q > 0$ is defined around his/her 2D field position, and then, a tubular region S_i of fixed-width $s > 0$ is defined from P (starting position) to R_i (thus, its length is proportional to the distance between the passer and the potential receiver). The final individual value for receiver R_i, denoted here as $V(R_i)$, can be obtained as:

$$V(R_i) = \frac{1}{\text{Area}\,(Q_i \cup S_i)} \int_{Q_i \cup S_i} \rho(x)dx \tag{20.2}$$

where $\text{Area}\,(Q_i \cup S_i)$ denotes the area of the region $Q_i \cup S_i$. In practice, q and s have been set to $\frac{5}{W_\rho}$ and $\frac{2}{W_\rho}$, respectively, being W_ρ the width of the output map ρ (*i.e.* 104). Note that Equation (20.2) can be used for both types of maps, being ρ the output of either the pass probability model (from now on V_P) or the EPV generic model (from now on V_E). Visually, this whole procedure can be seen in Fig. 20.6 for four different receiver candidates.
For comparison purposes, the individual probabilities V_P / expected values V_E are multiplied by our feasibility orientation estimation F_o, (Subsection 18.1); in this way, the effect of orientation itself can be tested for $V_P F_o$ and $V_E F_o$. Note that the other components F_p and F_d

have not been used, as both pass probability and EPV models already include this type of information in their core. Results are displayed in Table 20.5 and Fig. 20.7. As it can be seen, better accuracy is obtained when taking orientation into account in all scenarios, especially in the Top_1 accuracy case, obtaining a boost of almost 0.1 in the output of the current pass probability model. Moreover, orientation also improves the raw performance of V_E (0.07 improvement in Top_1 accuracy), especially by solving miss-leading cases in which players are located out of the field of view of the passer. In conclusion, it has been proved that merging orientation in the SoA implementation of EPV [34] could help to get a more accurate model, which can lead to a better understanding of the decision-making process.

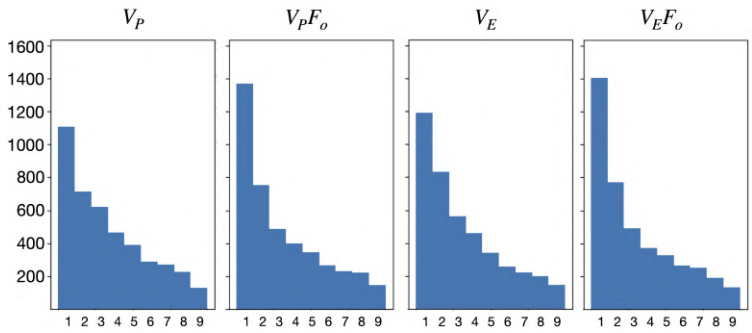

Figure 20.7: Histogram distribution of V_P and V_E, plus the corresponding addition of F_o component.

	Top_1 (Succ.)	**Top_3 (Succ.)**
V_P	0.243	0.567
$V_P + F_o$	0.332	0.612
V_E	0.266	0.606
$V_E + F_o$	0.337	0.637

Table 20.5: $\text{Top}_{1/3}$ Accuracy of the EPV models' output, plus their comparison when merging orientation feasibility.

20.2 Pass Feasibility Maps

In this Section we define the proposed tools to evaluate the pass feasibility map. We show not only the accuracy of the proposed model but also the importance of orientation as a map feature. An ablation study about the role of a star player is also included, in which the addition of prior biases is analyzed. In the upcoming examples, non-successful pass events of FCB_{DS} have been filtered out; while successful passes are well-defined, where a player B receives the ball from a player A, non-successful passes are diverse, including several kinds of failed passes or even dribbling turnovers tagged as such. The filtered dataset contains more than 5000 passes and it is split into 80% train and 20% test. Note that, in this Section, both training and test sets only include passes from the tracking-based FCB_{DS} dataset.

20.2.1 Map Evaluation

Three evaluation suggestions are displayed in Figure 20.8, which consist of:

1. The most simple one would be integrating the feasibility map values around each potential receiver $\mathbf{r}_j$ by placing a disk $D(\mathbf{r}_j, \rho)$ of radius $\rho > 0$ centered at $\mathbf{r}_j$, that is,

$$V_1(\mathbf{R}_j) = \frac{1}{\text{Area}\left(D(\mathbf{r}_j, \rho)\right)} \int_{D(\mathbf{r}_j, \rho)} M(\mathbf{x}) d\mathbf{x}. \tag{20.3}$$

   In practice, $\rho = \frac{\text{field}_{\text{length}}}{20}$ (~ 5), being 20 the number of players that do not play as goalkeepers. From now on, this approach will be named *disks evaluation*.

2. The second one stems from the premise that in *disks evaluation* all surrounding positions have the same weight in the integral. However, defenders receive short passes without moving drastically, whereas forwards usually have to sprint towards their

front to get the ball. To obtain a mask that resembles the receiving area of players, first, the displacements of players inside the training set are analyzed. This is, how does a player move during a pass, from the very moment where a player kicks the ball until the receiver gets the sphere. More concretely, considering pass events, a single displacement map is built for each (a) player position / role λ (goalkeeper / central / full-back / midfielder / forward) and (b) field side γ (left / right). Then, data are fit with kernel density estimation (KDE) [112], hence obtaining a kernel $K^{\lambda,\gamma}$ that is placed on top of every receiving candidate $\mathbf{r}_j$ according to his/her field position λ and side γ:

$$V_2(\mathbf{R}_j|\lambda,\gamma) = \frac{\int_\Omega K^{\lambda,\gamma}(\mathbf{x})M(\mathbf{x})d\mathbf{x}}{\int_\Omega K^{\lambda,\gamma}(\mathbf{x})d\mathbf{x}}, \tag{20.4}$$

where Ω denotes the image domain. This technique will be called *KDE evaluation*.

3. The third proposal to extract results out of M is directly related to the latter, and it is obtained by thresholding $K^{\lambda,\gamma}$ with a given threshold $\tau > 0$. A binary mask is obtained whilst preserving the potential receiving area of players; that is:

$$V_3(\mathbf{R}_j|\lambda,\gamma) = \frac{1}{\text{Area}\left([K^{\lambda,\gamma} > \tau]\right)} \int_{[K^{\lambda,\gamma}>\tau]} M(\mathbf{x})d\mathbf{x}, \tag{20.5}$$

where $[K^{\lambda,\gamma} > \tau] = \{\mathbf{x} : \text{ such that } K^{\lambda,\gamma}(\mathbf{x}) > \tau\}$ denotes the effective support of $K^{\lambda,\gamma}$. This last example will be named *bKDE*, which stands for *binary-KDE*. In practice, τ has been set to 0.75.

Using the training set, the tuning parameters explained in Section 19.3 have been optimized, and the KDE masks have been built; afterward, test accuracy scores are extracted for the best performances on the training set. Table 20.6 shows the best results for all three types of evaluations plus the effect of orientation in the output feasibility maps. In order to assess the importance of incorporating

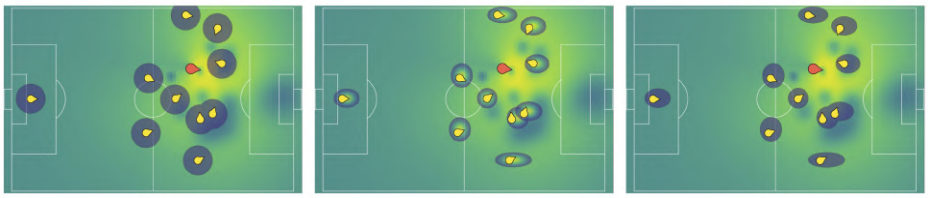

Figure 20.8: Different evaluation proposals: (a) disk-, (b) KDE- and (c) bKDE-evaluation.

orientation information, the same process has been repeated whilst building oriented-less maps; *i.e.* g_{aRj} and g_{aP} in (19.3) and (19.5) have been omitted when defining the feasibility map.
From the obtained results, several facets and choices can be discussed:

- First, it can be spotted that maps without orientation suffer a huge accuracy drop, both in Top_1 and Top_3 metrics (around 0.2 in both cases).
- Since Top_1 and Top_3 metrics are similar in both the training and testing stages, it can be said that the model is not overfitted regardless of the choice of parameters. Besides, the KDE masks created for evaluation also seem to generalize properly.
- Despite using different evaluation approaches, results are akin. This means that, although using a *disks evaluation* might seem a naive approach, it produces similar results (even better in some cases) when comparing with *KDE* or *bKDE evaluation*.
- Computing the pass feasibility for each field spot (including *e.g.* open spaces) outperforms the presented discrete-states model, which only considers a single feasibility value at the receiver's location.

For a more thorough assessment, the upcoming paragraphs analyze particular scenarios.

Evaluation	σ'_D	κ	Top1 Train	Top3 Train	Top1 Test	Top3 Test
Disk (O)	10	6	0.4591	0.7819	0.4632	0.7852
Disk (O)	12.5	6	0.4523	**0.7888**	0.4690	**0.7984**
Disk	15	6	0.2823	0.6032	0.2614	0.6230
KDE (O)	12.5	6	0.4502	0.7812	0.4563	0.7855
KDE (O)	15	6	0.4387	0.7806	0.4661	0.7898
KDE	15	6	0.2784	0.5871	0.2555	0.5908
bKDE (O)	12.5	6	0.4511	0.7709	**0.4734**	0.7821
bKDE (O)	10	6	**0.4604**	0.7809	0.4567	0.7784
bKDE	15	6	0.2819	0.5982	0.2790	0.6135
Discrete-States	-	-	-	-	0.3710	0.7098

Table 20.6: Evaluation results of pass feasibility maps.

20.2.2 Players' Field Position / Game Phase

Once again, specific scenarios such as the player's role or the ball location are checked to study the effect of orientation under several circumstances. In order to do so, the same testing set, the best tuning configuration (σ'_D,κ) displayed in Table 20.6 and a *disks evaluation* will be used. Starting with the players' field role, four different positions (excluding goalkeepers) are analyzed: defensive (1) center-backs and (2) left / right-backs, (3) midfielders, and (4) forwards. Results are shown in Table 20.7.

As expected, higher values are obtained for those players that attempt less risky passes (centrals), while forwards are the ones with lower Top_1 and Top_3 accuracy. Nonetheless, it is worth mentioning that forwards have the biggest difference in terms of accuracy when including orientation in the pass feasibility map. If the orientation is not taken into account, the pass feasibility in unoccupied free spaces is very low; however, this feasibility increases notably when a player is oriented towards that same space. This scenario is especially critical for forwards, who receive a considerable amount of long passes. Moreover, midfielders are also notably affected by orientation: since they are the ones in charge of organizing the offense, they normally

	Top$_1$ Test	**Top$_3$ Test**	**Diff Top$_3$**
Central Def. (O)	**0.5342**	**0.8174**	0.1689
Central Def.	0.2723	0.6485	
Left/Right Def.(O)	0.4604	0.7723	0.1376
Left/Right Def.	0.2420	0.6347	
Midfielders (O)	0.4340	0.7683	0.1466
Midfielders	0.2903	0.6217	
Forwards (O)	0.4309	0.7724	**0.2114**
Forwards	0.2358	0.5610	

Table 20.7: Evaluation results of pass feasibility maps with respect to player's position.

receive the ball from defenders when oriented towards the defensive field, and then they move the ball forward, after turning around and facing the offensive side.

Similarly, passes can be also classified according to the game phase; *i.e.* location of the passer in relation to the defensive team spatial configuration (preliminaries of Section 2.2), since passes from the same player can be very different depending on his/her situation in the field with respect to the defenders.
Results in Table 20.8 show a similar pattern to the previous analysis. The best results are obtained in the game phase with the lowest defensive pressure (build-up), and the region where orientation makes the biggest difference is the one where riskier passes are given (finalization). In the intermediate phases, differences can be spotted between L1, where midfielders might receive whilst looking backward, and L2, where the pass itself aims to create a goal opportunity, so players generally look forward. Given the fast backward-forward orientation changes, the first part of progression is the most confusing one, thus obtaining, by a small margin, the *lowest* Top$_1$ and Top$_3$ measures.

	Top$_1$ Test	**Top$_3$ Test**	**Diff Top$_3$**
Build-Up L0 (O)	**0.4800**	**0.8010**	0.1398
Build-Up L0	0.3026	0.6612	
Progression L1	0.2313	0.6122	
Progression L2 (O)	0.4910	0.7874	0.1774
Progression L2	0.2800	0.6100	
Finalization L3 (O)	0.4490	0.7880	**0.1992**
Finalization L3	0.2455	0.5958	

Table 20.8: Evaluation results of pass feasibility maps with respect to the game phase.

20.2.3 Weighting Players' Characteristics

In this Subsection, we consider the possibility of incorporating individual players' characteristics, since, *e.g.*, the role of a *star* player might alter the computed pass feasibility map. Take for instance the situation shown in Figure 20.9. In this event, there is a partially-undefended receiver placed close to the right sideline; according to our model, this is the best available receiving candidate. However, the final passing decision might be altered if the star player is inside the passer's field of view. In this ablation, we present an approach to incorporate this type of information through two kinds of weights.

Position Weights Position priors are defined by normalizing the number of given / received passes of each player by the number of faced individual opportunities. In order to avoid overfitting, the prior probabilities of receiving a pass are computed from a position-player perspective, which answers the following question: according to the training set, "what is the receiving probability for a player j, who plays as a position k', if the passer plays in a position k?" (being $k', k \in$ [central defender, left-right defender, midfielder, forward]). From now on, the total of received / given passes by player j will

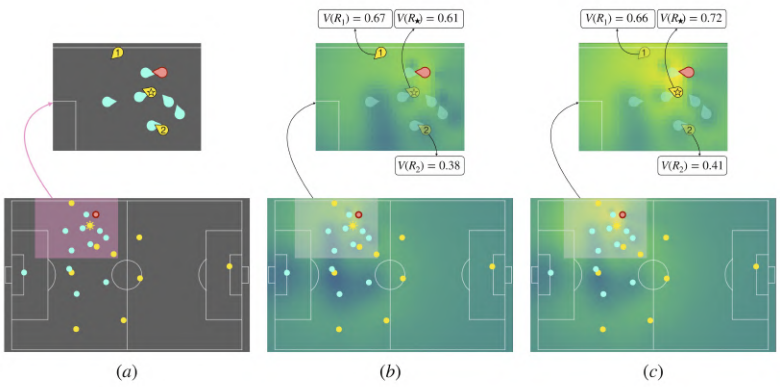

Figure 20.9: Effect of *star* role biases. (a) Initial 2D distribution of a given play. (b) Baseline output, where the model suggests passing the ball to the sideline-player. (c) Output when including weights, with the *star* player emerging as the best receiving candidate.

be named as RP_j and GP_j respectively, and the faced opportunities Op_j. Two cases might emerge in this situation:

- If $k' \neq k$:

$$p_j(\text{receive}|\text{passer } k) = \frac{\text{RP}_j \text{ from } k}{\text{Op}_j \text{ from } k} \tag{20.6}$$

- If $k' = k$:

$$p_j(\text{receive}|\text{passer } k) = \frac{\text{RP}_j \text{ from } k}{(\text{Op}_j \text{ from } k) - (\text{GP}_j)} \tag{20.7}$$

Line Weights As seen in Section 20.2.1, checking the game phase might complement the player position, thus acquiring a better understanding of the passes a player has given / received. In this case, the only probability to be computed is, given all events in the training set, the percentage of passes a player j, placed in the l' line, has

received from a passer who was at the l line of the field. Once again, two similar cases emerge:

- If $l' \neq l$:

$$p_j(\text{receive}|(\text{passer in } l \text{ , } j \text{ in } l')) = \frac{\text{RP}_j \text{ from } l \text{ when } j \text{ in } l'}{\text{Op}_j \text{ from } l \text{ when }_j \text{ in } l'} \tag{20.8}$$

- If $l' = l$:

$$p_j(\text{receive}|(\text{passer in } l,j \text{ in } l')) = \frac{\text{RP}_j \text{ from } l \text{ when } j \text{ in } l'}{(\text{Op}_j \text{ from } l,_j \text{ in } l') - (\text{GP}_j \text{ in } l')} \tag{20.9}$$

Weights Combination: Once position and line weights are computed, the pre-computed priors can be merged into one receiving prior probability / player $p_j(\text{receive})$ for a single event as:

$$p_j(\text{rec.}) = p_j(\text{rec.}|\text{passer } k)p_j(\text{rec.}|(\text{passer in } l,\ j \text{ in } l')) \tag{20.10}$$

To merge this prior with the proposed offensive map (Section 19.1), we suggest:

1. For player j, the receiving prior is converted into a boosting weight by:
$$w_j = \left(\frac{1}{1 - p_j(\text{receive})}\right)^{w_1}, \tag{20.11}$$
where w_1 is a weighting factor that regulates the impact of the computed priors; in practice, this parameter is set to 5.

2. Afterwards, Equation (19.5) can be modified by adding the boosting weight by:
$$M_{wRj}(\mathbf{x}) = w_j \exp\left(-\frac{\|\mathbf{x} - \mathbf{r}_j\|^2}{\sigma_R^2} - \frac{(\angle(\mathbf{x} - \mathbf{r}_j) - \alpha_{Rj})^2}{\sigma_a^2}\right) \tag{20.12}$$

	σ'_D	κ	w1	Top1 Test	Top3 Test
E	12.5	4	-	**0.4651**	0.7634
E_w	12.5	4	5	0.4556	**0.7815**
E_s	12.5	4	-	0.4460	0.7729

Table 20.9: Ablation results: baseline method (E), including weights (E_w) and speed (E_s).

Player	Rec. Passes	Top_1 E_w	Top_3 E_w	Top_1 E	Top_3 E	Top_1 Diff.	Top_3 Diff.
Central	48	0.444	**0.889**	0.500	**0.889**	-0.055	0
Right-Back	204	**0.519**	0.828	**0.544**	0.804	-0.024	0.024
Midfielder	113	0.469	0.823	0.469	0.770	0	**0.053**
Forward	245	0.486	0.830	0.454	0.812	**0.032**	0.018

Table 20.10: Individual performance of different players.

Case Study The effect of weight addition is studied in attacking situations, defined as the events where the passer is over the second (L2 progression) or third (L3 finalization) defensive line. In order to isolate the effect of weights, and using the proposed *disks evaluation*, two experiments (named E and E_w) are performed: the results of E and E_w are obtained using equations (19.5) and (20.12), respectively. The specific parameters and the overall performance of both experiments are displayed in Table 20.9. As it can be spotted, the overall performance when using weights does not abruptly change the final performance result; although there is a slight drop in the Top_1 accuracy, a $+0.02$ boost emerges in Top_3. To dissect the general obtained results, the accuracy has been split into individual players; specifically, four different players are examined in Table 20.10.

As it can be spotted:

- The chosen forward is the one obtaining the most notable Top_1 individual accuracy boost; it also has to be remarked that this forward is clearly an outlier in the *finalization* phase, reaching a

16% receiving probability both from midfielders and forwards, hence justifying the actual bias when introducing a player's roles.

- The displayed midfielder's results do not show an improvement in Top_1 accuracy, but a 5% boost can be seen in Top_3, since midfielders usually have an active role when organizing the team's offense.

- The chosen right-back has significant Top_1 accuracy without weights, but prior data slightly downgrade his performance in favor of team forwards; however, this drop is recovered in Top_3.

- Finally, since central defenders do not usually participate in plays in the *finalization* phase, the presented central's results suffer the biggest drop in Top_1 accuracy and do not improve in Top_3; however, this central defender has only received 48 passes in L2 / L3, which might be a small sample size.

20.2.4 Speed as a Feature

Apart from including roles and existing biases, more features can be added to the models' core. At the moment, 2D location and orientation data have been merged, but the actual conditions / behavior of each player at a time are not taken into account; among all physical capabilities, in terms of receiving (or not) a pass, speed makes the biggest difference. That is, while a static player has few chances of successfully receiving a long pass, since he/she will have to accelerate and then reach the final pass location, a player that is already running towards open spaces is a favorable receiving candidate. For the rest of this ablation study, player speed has been computed directly from 2D tracking data by smoothing player displacements across a temporal window of 11 frames (with respect to the pass timestamp). In order to include speed in the offensive map, the contribution of

each reception map can be adapted by tuning the denominator of the second factor in equation (19.3) as:

$$g'_{aRj}(\mathbf{x}) = \exp\left(\frac{-(\angle(\mathbf{x}-\mathbf{r}_j)-\alpha_{Rj})^2}{\sigma_a^2\nu_r}\right),$$

where $\nu_r = s_j/\bar{s}_{rec}$, being s_j the receiver's speed in meters per second at the given timestamp, and being $\bar{s}_{rec}$ a normalization factor equal to the median speed of receivers in all the set of pass events in the included dataset ($\bar{s}_{rec} = 1.57$ meters / second).
Similarly, the same reasoning can be followed for the defensive team, since defenders that are standing still are the ones creating the largest open spaces in their surroundings given their limited reaction time. Consequently, the numerator in the exponential of the second case in equation (19.6) can be adapted as:

$$-\exp\left(\frac{-(\frac{1}{2\nu_d}(R_{\beta_i}(\mathbf{x}-\mathbf{d}_i))_x^2+(R_{\beta_i}(\mathbf{x}-\mathbf{d}_i))_y^2)}{\sigma_{D_i}^2}\right)$$

where $\nu_d = \max(s_d/\bar{s}_{def}, 1)$; once again, s_d indicates the defender's current speed, and $\bar{s}_{def}$ is the median speed of all defenders in the given dataset, which is 1.86. Note that when dealing with defenders, the back-side of the double-sided Gaussian is not altered with the players' speed, since it corresponds to the opposite defender's orientation.

The effect of speed when modifying the original equations can be seen in Figure 20.10, where both offensive and defensive contributions are displayed. In terms of numerical results, Table 20.9 shows how the effect of speed does not drastically change the obtained results, as the presented tests perform similarly in terms of Top_N accuracy. A further study could be done by taking into account short / mid / long passes, and by assessing the relevance of speed as a feature in each of these. Potentially, long passes will be the ones where speed, together with proper orientation, becomes a crucial factor.

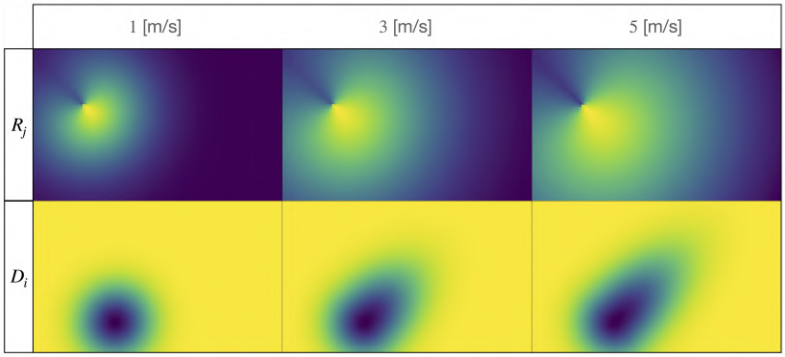

Figure 20.10: Speed can be included as a feature for both the offensive (top) and the defensive (bottom) contributions. The faster the player is moving, the larger space he/she can reach for a proper reception / interception.

20.2.5 Discussion

From the obtained $\text{Top}_1/\text{Top}_3$ accuracy results, it has been proved that not including orientation in the pass feasibility map leads to a notable drop (around 0.2). This fall exemplifies that orientation plays a crucial role in pass events, since offensive players have to exploit open spaces, and also because players outside the passer's field of view will not be able to receive the ball. Overall, the best performance is achieved by central defenders' passes and by passes given in the *build-up* phase, and the worst scenario belongs to forwards or passes given after in the *finalization* phase. However, the latter scenarios are also the ones that benefit the most when adding orientation into the map, meaning that the overall scenario corresponds to a really risky pass where proper orientation makes a difference.
Obtaining 0.79 of Top_3 accuracy means that, by adding orientation, the safest spots in the field are generally detected. Notice that the pass feasibility map indicates how safe it is *a priori* to pass the ball towards each field location in terms of keeping the ball; nevertheless, other concepts such as offensive strategy or match timing, which usu-

ally alter the decision-making process of the passer, are not considered. Although orientation is a crucial skill for soccer players, it is a hard feature to teach by coaches; orientation-based drills are tough to design and the best way to improve this particular skill is through visual examples. By using 2D feasibility maps, results can be filtered individually to detect which players react better / worse in particular field positions or even during different game moments; having the appropriate game video clips plus the right tools to explain them can smooth the communication between analysts-coaches-players, thus favoring a potentially better performance. As in the case of discrete states, the presented pass feasibility maps could also be integrated into existing EPV models [34; 33], which would result in a refined and more realistic output. Besides, apart from coaches benefiting from this type of tool, strength and conditioning staff could also obtain meaningful insights. By checking orientation throughout games and practices, a fine-grained analysis of each players' training load could be performed, thus detecting running types (front / side / back) and becoming a valuable tool for injury prevention.
Moreover, the output of feasibility maps could be adapted to other sports, even though the aim of the presented model could differ from passing events. Some suggestions to adapt feasibility maps to other domains are given in Section 21.1.

21 Conclusions

In the last Part of this manuscript, the feasibility of soccer pass events has been analyzed with two different approaches: discrete states and pass feasibility maps. In both cases, the main contribution is the inclusion of orientation data, estimated directly from video frames, into a passing model, which has proved to be strictly correlated to the play outcome and a key feature to characterize the decision-making process of players.

On the one hand, when dealing with discrete states, orientation feasibility is computed with a geometrical approach among offensive players, and it is combined with two other estimations, based on the faced defensive pressure and pairwise distances between the passer and all potential receivers. Moreover, the combination of the model's output with existing pass probability / EPV models has been studied, obtaining confident results which indicate that SoA methods can be refined by including orientation data. Note that this model is really light in terms of computational complexity and it could be used in real-time.

On the other hand, pass feasibility maps provide a plausibility measure that indicates how safe it is to pass the ball towards any 2D position of the whole field. The proposed feasibility map is modeled through Gaussian functions that depend both on orientation and lo-

cation. For the offensive team, the passer's field-of-view (or reach) is modeled and later combined with the aggregation of receivers. The defensive contribution is also computed by estimating the individual *influence area*, which also relies on player orientation. By merging offense and defense into one same function, the 2D pass feasibility map is obtained.
Three suggestions to evaluate this type of map are given. In this matter, 0.46 / 0.79 $\text{Top}_1/\text{Top}_3$ accuracy are obtained, respectively, with a -0.2 drop in both cases when not taking orientation into account, thus showing that orientation is indeed a vital skill for soccer players. In the presented ablation study, the first steps to build an accurate model including player characteristics (speed) and biases have been detailed, which results in $\text{Top}_1/\text{Top}_3$ accuracy boosts for individual *star* players. The visual information of the proposed feasibility map can be directly used by analysts or coaches, who might detect strengths and weaknesses in the passing spatial patterns of a given team.

21.1 Future Work

Although the obtained pass feasibility results generalize well to the vast majority of situations, other soccer-based facets could be included in the model's core to provide even more realistic outputs. For instance, the following factors have not been taken into account:

- Intrinsic player skills that define the passing reach. Since each player has his/her own strengths and weaknesses and given that passing might be one of them, several scenarios come into play. A midfielder that excels in passing to open spaces should have a larger associated spatial reach, but a central defender whose main goal is to recover the ball and to ensure safety passes to midfielders should have a restricted one. Anyway, large samples of data are required to include all these specifics without

overfitting the model.

- Similarly, game context also matters. For example, a team that is trailing by 1 goal is willing to take a lot of risky moves in the last minutes of the game (*e.g.* goalkeeper in the offensive area in a corner play). All these contextual cues definitely affect the outcome of passes and could be considered as game features in the presented model.

Besides, the model could also benefit from orientation-based Voronoi tessellation. In fact, many of the presented passing tools and models reviewed in Chapter 17 include these kind of cells in order to assign field regions to individual receivers; by including orientation, these regions could be refined according to the player's actual spatial reach. Apart from improving the consistency of the presented computational soccer models, the generalization to other sports could be studied. However, in other scenarios, the aim of the presented model could differ from passing events. In the case of basketball, body-orientation definitely affects the passing outcome, but several other aspects could provide much more meaningful insights, such as:

- The orientation of *off-ball* defenders in the *weak-side* of the play is vital. These types of defenders need to be in a perfect position and orientation in case: (a) the player carrying the ball drives to the basket, (b) their match-up cuts to the basket (back-door), or (c) they have to perform a close-out defense after a skip pass. Likewise, other *off-ball* defensive clues could be obtained, such as the perfect combination of spacing plus orientation in order to build a defensive pressure set (or even zone defense).

- The orientation of *on-ball* defenders is also important, especially in the current NBA context, where the defensive strategy of several teams is based on switches. In these situations, two miss-matches occur: first, there is a big player guarding the small ball-handler, and then, there is a small player covering a

big one. With proper orientation, lateral quickness is favored, thus mitigating an unfavorable outcome of the potential mismatch.

- On the offensive end, other practical applications can be built. An interesting one could be the effect of screening angles according to body-orientation in *pick and roll* or *off-screen* plays. Since the body-orientation of the screener is fundamental to trap the ball defender and generate a scoring opportunity, feasibility offensive maps could be built. Note that in this scenario, the raw computation of orientation might differ from the original one, since the lower part of the body is also needed.

Therefore, feasibility maps could be a valid resource for coaches to extract insights, and their power is not only limited to passing events in soccer.

Closure

You can unfasten your seat belt, we have reached the end of this particular journey of CV in sports. During this adventure, several Parts and Chapters have guided us to understand where Computer Vision falls inside this brand new field of sports analytics and which are the data sources and the potential applications that clubs or coaches are looking for.

In Chapter 1, the importance of tracking in sports has been contextualized, hence proving that the overall situation of data science in elite competitions has shifted completely; in this new decision-making paradigm, data-driven automated processes attempt to complement the existing *know-how* of coaches and general managers. Within this context, the inclusion of tracking data has proved to be crucial, since it encompasses other types of data sources, and moreover, several models can be trained on top of this kind of data.

However, since tracking data are still not an exploited resource in a large set of competitions, where teams have unbalanced economical resources, the viability of single-camera basketball multi-tracking algorithms has been studied in Part I. More concretely, through a CV-based pipeline, involving court filtering and player detection, the effect of different feature extraction processes has been analyzed, with the obtention of color- and deep-learning-based features. Once matched all the given instances in the scene across frames, the obtained results have shown almost a 0.7 multi-object tracking accuracy,

which is a notable performance since (up to) 10 targets have to be identified and tracked.

In Part II we have seen how tracking data may fall short in given scenarios since the 2D location of players normally lacks the general context. Under these circumstances, CV is an endless resource of techniques that can be used to extract metrics and to enrich pure tracking data. In this manuscript, a complete research about body orientation has been provided, thus resulting in a new dimension that complements raw tracking (2D location + body orientation). Once defined body orientation as the 2D projection of the 3D normal vector extracted right in the middle of the player's upper-torso, two approaches have been described. The first one, model-based, stems from pose models and combines classical CV and 3D-Vision techniques to map the position of body parts in a 2D template, hence obtaining the desired normal vector with ease. This estimation is later refined with contextual information, *i.e.* ball location. The second one, learning-based, tackles this challenge as a classification problem by fine-tuning a VGG-19 network; by leveraging on a proper compensation of angles with respect to the camera pose, and by introducing a cyclic loss function based on soft labels, the network is able to classify bounding boxes into orientation bins. Results have been validated through EPTS-held devices, which provide ground-truth orientation data; while the model-based method achieves decent performance with less than 28 degrees of median absolute error, the learning-based one improves its performance by a large margin, resulting in an error less than 13 degrees. What is more, some data visualizations are given in order to make raw orientation data understandable at first sight; in particular, OrientSonars-maps, Reaction-maps, and On-Field-maps are presented.

Even though the inclusion of orientation is supposed to be beneficial for any type of post-processing modeling, its real profit has to be validated. Consequently, Part III aims to prove the vital impor-

tance of body-orientation in the most common soccer event: passes. More specifically, the notion of pass feasibility is introduced, which expresses who is the most likely receiver to get the ball, or which are the safest field spots. In the first case, a discrete computational model is presented, which combines (a) the location, (b) the orientation, and (c) the faced defensive pressure of every potential receiver, and ends up ranking the safest receivers. In order to exploit open spaces, the second computational model extends the latter by producing pass feasibility maps. In this sort of tool, pass safety is displayed on every field spot by merging together the offensive and the defensive team's contributions, both of them including location plus orientation. The performance of feasibility-tools has been studied with a Top_N fashion, which indicates if the output of the model matches the real scenario, and promising results have been obtained: 0.36 / 0.70 (discrete), and 0.46 / 0.79 (maps) in Top1 and Top3 respectively. Moreover, the same experiments have been performed without including orientation in the computational model; in all cases, a drop around -0.2 accuracy appeared, thus proving that the included variable definitely had a relevant weight in the proposed models.

Having listed all the presented contributions, I would like to give some personal closure to this manuscript with future guidelines and take-home messages. First of all, apart from the concrete potential improvements described in Sections 7.1 and 21.1, other lines of research that could benefit from the enclosed manuscript's contributions are listed:

- Individual action recognition is probably one of the most interesting ways to enrich player tracking data. By training models from compensated bounding boxes, networks could classify each player into a given set of action-related categories. For instance, basketball-wise, it would be interesting to know, apart from each player's 2D location and his/her orientation, what basic actions players are performing: running, screening, shooting, passing, dribbling... With this new generation of analytics,

several meaningful improvements could be made on existing basketball layers:

  - Playtype data could be automatically collected for each player. At a coarse level, it could be known *a priori* if a player, *e.g.*, takes more shots off the dribble or in spot-up situations. At a fine level, *e.g.* novel defensive metrics could quantify the defensive effort of a player in specific situations (such as *weak-side help* or *pick and rolls*).
  - Advanced tracking statistics could also be redefined. The most obvious example are the defensive shot labels: roughly, if a player shoots without a defender close to him/her, the shot is considered as *open*; otherwise, the shot is considered as *contested*. At the given moment, this definition is only based on pairwise distances: if the shooter has a player nearby in the 2D space (within a radius of 1.25 meters), that shot will be considered as contested. However, without action recognition, it cannot be known if the defensive player is truly contesting the shot (jumping, stretching his/her arm) or he/she has a passive attitude that does not bother the shooter at all. By adding action recognition into the equation, a better definition of this type of statistics would be obtained.

- Collective action recognition could also be really valuable. Although individual players perform independent actions by themselves, the success of team sports' offensive motions depends on the overall synergy among players. For instance, when two basketball players are executing a *ball-screen* play, the ball handler is the one generating a scoring opportunity through the screen, the screener decides whether to roll or to pop, but what is more, there are three other players who have to be in the right place at the right time, thus producing the desired spacing to end up with the better shot for the team. Capturing these collective

patterns could help coaches identifying the playing style of a team in different game phases.

- If all actions are being labeled with individual / collective action recognition, the next step is to generate automatic highlights through action spotting. With this type of tool, not only competitions or broadcasting companies could have accurate and automatic summaries of the game, but also coaches could save a lot of film time while preparing scouting reports. Note that in this case, for the sake of completion, data from other domains, such as eventing or audio cues, could even generate more customized and precise reports.

- During the whole manuscript, we talked about basketball tracking and soccer body-orientation applications, but there is a large set of sports-related scenarios to be explored. Obviously, the most competitive leagues, with respect to their economical resources, will be the ones gathering sport-dependent tracking data first, but all these tools should be multi-sport soon. For instance, in Part 13.2 we have shown how easy-to-adapt models could be tested without requiring lots of input data (unbalanced sets, at least), thus boosting the potential generalization across sports.

- At the given moment, the vast majority of trackers are gathering information only in the 2D space (X, Y coordinates). In the case of basketball, the main reasoning is that tracking cameras consist of an overhead setup that considers each player as a dot in the court; by merging data from the broadcasting camera, the Z dimension could be also inferred. Note that this dimension could also be obtained by using ground-truth data coming from EPTS-held devices, and its output could also be helpful for action recognition models. In this particular domain, the 3D pose models detailed in Chapter 10 could be exploited.

- Pose models have been widely used during the whole thesis for

tracking and orientation purposes, but more applications could be built from the obtained skeletons. For instance, player similarity could be brought to the next level. At the moment, similarity is just computed in terms of numerical statistics or based on human perception through video reports, but by including pose data, inner patterns could be detected as well: *e.g.* how does a player pass in terms of a purely technical biometrical perspective with his/her off-hand? How do these players resemble from the point of view of their shooting mechanics? This tool could be also useful for talent identification (or recruitment) in youth sports.

- Until this point, we have mentioned several open datasets that have been used to train the presented models. Actually, all these models share one feature: included data belong to real games and to real players. One source of data that has still not been exploited are video games, which could be an excellent resource given that: (a) the perception of reality in video games is almost perfect at the moment, (b) the amount of data that can be generated with this data source is boundless – in terms of camera pose and also with respect to the number of games –, and (c) metadata could be easily obtained without requiring the need of sensor-data – *e.g.* player meshes and normal vectors –. I truly believe that training models from video-game data could really improve the generalization capability of existing models.

- Last but not least, visualization tools of all presented methods could be built from a computer graphics / UX-design perspective. Even though the vast majority of post-processing analysis is not done during games, with the appropriate graphic tools, coaches could have real-time information regarding several orientation-based metrics, such as printed pass feasibility maps on top of game footage. These data could be displayed on a tablet being carried by an assistant coach, and automatic per-

formance reports could facilitate the decision-making process of the coaching staff *e.g.* during half-time.

Finally, I would like to encourage all researchers working with sports analytics in academia to keep sharing their work in order to create a solid state-of-the-art that could encompass many game facets across sports. When submitting our work to computer science journals, a common flaw that some reviewers detected was that, since our work could not be compared with previous work, it was unclear if the obtained results could be validated. One decade ago, it was hard to imagine that someone could pursue a professional career in sports analytics, but once that phase has been surpassed, it is now time to prove that sports analytics' research can fit in the scope of reputed journals. Fortunately, a huge effort has been made by researchers / organizations / clubs to create a research-sharing culture in this field, namely:

- In academia, apart from the well-known *Journal of Sports Sciences* and the *MIT Sloan Sports Analytics Conference* (hybrid between companies and academia), there has been a substantial rise in the number of sports-related conferences in the last decade, for instance:
  - International Workshop on Computer Vision in Sports at the Computer Vision Pattern Recognition conference (CVPR) - 7 editions.
  - Workshop on Machine Learning and Data Mining for Sports Analytics at the ACM International Conference on Multimedia - 7 editions
  - International Workshop on Multimedia Content Analysis in Sports at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases - 3 editions.
  - AI for Sports Analytics Workshop at the International Joint Conference on Artificial Intelligence - 1 edition.

- Besides, topnotch clubs / companies / organizations are also hosting enriching events where professional analysts and researchers discuss their research topics (mainly soccer), such as:
  - OptaPro Analytics Forum - 5 editions.
  - Futbol Club Barcelona's Sports Tomorrow Congress - 4 editions.
  - StatsBoom Innovation in Football Conference (first hosted in 2019 and now scheduled for 2021).
  - Despite being virtual sessions, the community of *Friends of Tracking* is also sharing powerful resources weekly, including open code or detailed tutorials.
- Upcoming literature will benefit from open datasets, like the ones described in Section 2.3. For instance, the authors from SoccerNet [40; 26] are currently hosting action-spotting challenges performed on top of their open dataset.

With all these limitless resources, there are no possible excuses: the timing is perfect to research every possible greedy detail of this emerging field. Possibly, the main open question regarding the pursuit of a sports analytics career is how to get started, and as far as I can tell, since sports-analytics-based undergraduate programs still do not exist, there is not a default-valid answer. It is crystal clear that professional analysts must be comfortable coding, handling large structures of numbers, so any computer science / data science career will provide the student with all required technical tools. However, what truly makes the difference has little to do with computers: analysts must be sports enthusiasts, and having some *court / field experience* is definitely a bonus. One of the main challenges for analysts while establishing data science departments in clubs is to create a solid communication environment. Since insights can be easily lost in the communication transfer between analysts and coaches, both of them need to speak the same language, or even better, the analyst

needs to speak the same language as the coach, and he/she must be able to translate technical sports-based argot into numbers. Bearing in mind that each coach is unique, one of the key aspects for this smooth transfer of knowledge is to ask the appropriate questions without getting too technical, thus getting to know their true needs. Therefore, I would say that professional analysts must be hybrid profiles with the appropriate tools to create, analyze, and communicate.

Bibliography

[1] I. Akhter and M. J. Black. Pose-conditioned joint angle limits for 3d human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1446–1455, 2015.

[2] A. Arbués Sangüesa, R. Benítez Iglesias, T. B. Moeslund, and C. H. Bahnsen. Identifying Basketball Plays from Sensor Data; towards a Low-Cost Automatic Extraction of Advanced Statistics. In *IEEE Conference on Data Mining; Data Analysis for the Performance of Succes (DAPS) Workshop*, 2017.

[3] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[4] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *Journal on Image and Video Processing*, 2008:1, 2008.

[5] M. Beuoy. NBA Player Shooting Motions. `https://www.inpredictable.com/2021/01/nba-player-shooting-motions-data-dump.html`, 2021. Accessed: 23-04-21.

[6] J.-Y. Bouguet. Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm. *Intel Corporation*, 5(1-10):4, 2001.

[7] L. Bransen and J. V. Haaren. Player chemistry: Striving for a perfectly balanced soccer team, 2020.

[8] G. Brasó and L. Leal-Taixé. Learning a neural solver for multiple object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6247–6257, 2020.

[9] L. Bridgeman, M. Volino, J.-Y. Guillemaut, and A. Hilton. Multi-person 3d pose estimation and tracking in sports. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[10] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *CoRR*, abs/1611.0, 2016.

[11] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1302–1310, 2017.

[12] F. Cardinale. Isr. `https://github.com/idealo/image-super-resolution`, 2018.

[13] Catapult. CatapultSports. `https://www.catapultsports.com/es/`, 2006. Accessed: 23-04-21.

[14] D. Cervone, A. D. Amour, L. Bornn, and K. Goldsberry. POINTWISE : Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. In *MIT Sloan, Sports Analytics Conference*, 2014.

[15] J. Chen and J. J. Little. Sports camera calibration via synthetic data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[16] J. Chen, J. Wu, K. Richter, J. Konrad, and P. Ishwar. Estimating head pose orientation using extremely low resolution images. In *2016 IEEE Southwest symposium on image analysis and interpretation (SSIAI)*, pages 65–68. IEEE, 2016.

[17] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[18] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381:61–88, 2020.

[19] A. Cioppa, A. Deliege, M. Istasse, C. De Vleeschouwer, and M. Van Droogenbroeck. Arthus: Adaptive real-time human segmentation in sports through online distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[20] L. Citraro, P. Márquez-Neila, S. Savarè, V. Jayaram, C. Dubout, F. Renaut, A. Hasfura, H. B. Shitrit, and P. Fua. Real-time camera pose estimation for sports fields. *Machine Vision and Applications*, 31(3):1–13, 2020.

[21] CMU. OpenPose Repository. `https://github.com/CMU-Perceptual-Computing-Lab/openpose`, 2010. Accessed: 23-04-21.

[22] D. Crundall, G. Underwood, and P. Chapman. Driving experience and the functional field of view. *Perception*, 28(9):1075–1087, 1999.

[23] C. De Vleeschouwer, F. Chen, D. Delannay, C. Parisot, C. Chaudy, E. Martrou, A. Cavallaro, et al. Distributed

video acquisition and annotation for sport-event summarization. *NEM summit*, 8, 2008.

[24] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1851–1861, 2019.

[25] M. F. Deering. The limits of human vision. In *2nd International Immersive Projection Technology Workshop*, volume 2, page 1, 1998.

[26] A. Deliège, A. Cioppa, S. Giancola, M. J. Seikavandi, J. V. Dueholm, K. Nasrollahi, B. Ghanem, T. B. Moeslund, and M. Van Droogenbroeck. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. *arXiv preprint arXiv:2011.13367*, 2020.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[28] R. Diaz and A. Marathe. Soft labels for ordinal regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4738–4747, 2019.

[29] A. Doering, U. Iqbal, and J. Gall. Joint flow: Temporal flow fields for multi person tracking. *arXiv preprint arXiv:1805.04596*, 2018.

[30] F. Er, B. Dežman, G. Vu, J. Perš, M. Perše, and M. Kristan. An Analysis of Basketball Players' Movements in the Slovenian Basketball League Play-Offs Using the SAGIT Tracking System. *Physical Education and Sport*, 6(1):75–84, 2008.

[31] M. Fastovets, J.-Y. Guillemaut, and A. Hilton. Athlete pose estimation from monocular tv sports footage. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1048–1054, 2013.

[32] P. Felsen and P. Lucey. Body shots: Analyzing shooting styles in the nba using body pose. In *Proceedings of the MIT Sloan Sports Analytics Conference, Possession Sketches: Mapping NBA Strategies, Boston, MA, USA*, pages 3–4, 2017.

[33] J. Fernández and L. Bornn. Soccermap: A deep learning architecture for visually-interpretable analysis in soccer. In *ECML-PKDD 2020*, 2020.

[34] J. Fernández, L. Bornn, and D. Cervone. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13 th Annual MIT Sloan Sports Analytics Conference*, 2019.

[35] T. Fischer, H. Jin Chang, and Y. Demiris. Rt-gene: Real-time eye gaze estimation in natural environments. In *European Conference on Computer Vision (ECCV)*, pages 334–352, 2018.

[36] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):267–282, Feb 2008.

[37] H. Folgado, J. Bravo, P. Pereira, and J. Sampaio. Towards the use of multidimensional performance indicators in football small-sided games: the effects of pitch orientation. *Journal of Sports Sciences*, 37(9):1064–1071, 2019.

[38] A. Franks, A. Miller, L. Bornn, and K. Goldsberry. Counterpoints : Advanced Defensive Metrics for NBA Basketball. In *MIT Sloan, Sports Analytics Conference*, 2015.

[39] X. Fu, K. Zhang, C. Wang, and C. Fan. Multiple player tracking in basketball court videos. *Journal of Real-Time Image Processing*, 17(6):1811–1828, 2020.

[40] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem. Soccernet: A scalable dataset for action spotting in soccer videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1711–1721, 2018.

[41] R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, and D. Tran. Detect-and-track: Efficient pose estimation in videos. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 350–359, 2018.

[42] K. Goldsberry. CourtVision : New Visual and Spatial Analytics for the NBA. In *MIT Sloan, Sports Analytics Conference*, 2012.

[43] K. Goldsberry and E. Weiss. The Dwight Effect : A New Ensemble of Interior Defense Analytics for the NBA. In *MIT Sloan, Sports Analytics Conference*, 2013.

[44] L. Gyarmati and X. Anguera. Automatic extraction of the passing strategies of soccer teams. *arXiv preprint arXiv:1508.02171*, 2015.

[45] L. Gyarmati and R. Stanojevic. Qpass: a merit-based evaluation of soccer passes. *arXiv preprint arXiv:1608.03532*, 2016.

[46] M. Hayashi, K. Oshima, M. Tanabiki, and Y. Aoki. Upper body pose estimation for team sports videos using a poselet-regressor of spine pose and body orientation classifiers conditioned by the spine angle prior. *Information and Media Technologies*, 10(4):531–547, 2015.

[47] M. Hayashi, T. Yamamoto, Y. Aoki, K. Ohshima, and M. Tanabiki. Head and upper body pose estimation in team sport

videos. In *2013 2nd IAPR Asian Conference on Pattern Recognition*, pages 754–759. IEEE, 2013.

[48] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[49] R. Henschel, L. Leal-Taixé, D. Cremers, and B. Rosenhahn. Fusion of head and full-body detectors for multi-object tracking. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pages 1509–150909, 2018.

[50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[51] O. Hubáček, G. Šourek, and F. Železnỳ. Deep learning from spatial relations for soccer pass prediction. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*, pages 159–166. Springer, 2018.

[52] S. Hurault, C. Ballester, and G. Haro. Self-supervised small soccer player detection and tracking. In *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports*, MMSports '20, page 9–18, New York, NY, USA, 2020. Association for Computing Machinery.

[53] InStat. InStat. `https://instatsport.com/`, 2007. Accessed: 23-04-21.

[54] N. Johnson. Extracting player tracking data from video using non-stationary cameras and a combination of computer vision techniques. In *Proceedings of the 14th MIT Sloan Sports Analytics Conference, Boston, MA, USA*, 2020.

[55] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[56] S. Kalman and J. Bosch. Nba lineup analysis on clustered player tendencies: A new approach to the positions of basketball & modeling lineup efficiency of soft lineup aggregates. 42 analytics (2020), 2020.

[57] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[58] W. Kim, S.-W. Moon, J. Lee, D.-W. Nam, and C. Jung. Multiple player tracking in soccer videos: an adaptive multiscale sampling approach. *Multimedia Systems*, 24(6):611–623, 2018.

[59] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[60] M. Lewis. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.

[61] Q. Liang, W. Wu, Y. Yang, R. Zhang, Y. Peng, and M. Xu. Multi-player tracking for multi-view sports videos with improved k-shortest path algorithm. *Applied Sciences*, 10(3):864, 2020.

[62] D. Link, S. Lang, and P. Seidenschwarz. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PloS one*, 11(12):e0168768, 2016.

[63] H. Liu and B. Bhanu. Pose-guided r-cnn for jersey number recognition in sports. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[64] LongoMatch. LongoMatch. `https://longomatch.com/`, 2008. Accessed: 23-04-21.

[65] P. Lucey, A. Bialkowski, P. Carr, Y. Yue, and I. Matthews. " How to Get an Open Shot ": Analyzing Team Movement in Basketball using Tracking Data. In *MIT Sloan, Sports Analytics Conference*, 2014.

[66] R. Maheswaran, Y.-h. Chang, A. Henehan, and S. Danesis. Deconstructing the Rebound with Optical Tracking Data. In *MIT Sloan, Sports Analytics Conference*, 2012.

[67] R. Maheswaran, Y.-h. Chang, J. Su, S. Kwok, T. Levy, A. Wexler, and N. Hollingsworth. The Three Dimensions of Rebounding. In *MIT Sloan, Sports Analytics Conference*, 2014.

[68] A. Maksai, X. Wang, and P. Fua. What players do with the ball: A physically constrained interaction modeling. *CoRR*, abs/1511.06181, 2015.

[69] M. Manafifard, H. Ebadi, and H. A. Moghaddam. A survey on player tracking in soccer videos. *Computer Vision and Image Understanding*, 159:19–46, 2017.

[70] R. Marty. High resolution shot capture reveals systematic biases and an improved method for shooter evaluation. In *MIT Sloan, Sports Analytics Conference*, pages 1–10, 2018.

[71] A. Mcintyre, J. Brooks, J. Guttag, J. Wiens, and A. Arbor. Recognizing and Analyzing Ball Screen Defense in the NBA Learning to Classify Defensive Schemes. In *MIT Sloan, Sports Analytics Conference*, 2016.

[72] N. Mehrasa, Y. Zhong, F. Tung, L. Bornn, and G. Mori. Deep Learning of Player Trajectory Representations for Team Activity Analysis. In *MIT Sloan, Sports Analytics Conference*, pages 1–8, 2018.

[73] MetricaSports. Metrica Sports - Football Open Data. https://metrica-sports.com/football-open-data/, 2020. Accessed: 23-04-21.

[74] A. Milan, L. Leal-Taixé, K. Schindler, and I. Reid. Joint tracking and segmentation of multiple targets. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 5397–5406, 2015.

[75] A. Miller, L. Bornn, R. Adams, and K. Goldsberry. Factorized Point Process Intensities: a Spatial Analysis of Professional Basketball. In *International Conference on Machine Learning*, pages 235–243, 2014.

[76] A. C. Miller and L. Bornn. Possession Sketches : Mapping NBA Strategies. In *MIT Sloan, Sports Analytics Conference*, 2017.

[77] N. MJ. 2016 NBA Raw SportVU. https://github.com/tcash21/BasketballData/tree/master/2016.NBA.Raw.SportVU.Game.Logs, 2016. Accessed: 23-04-21.

[78] G. Myklebust, A. Skjølberg, and R. Bahr. Acl injury incidence in female handball 10 years after the norwegian acl prevention study: important lessons learned. *British Journal of Sports Medicine*, 47(8):476–479, 2013.

[79] NacSport. NacSport. https://www.nacsport.com/index.php?lc=es-es, 2004. Accessed: 23-04-21.

[80] NBN23. Nothing But Net 23. https://www.nbn23.com/, 2016. Accessed: 23-04-21.

[81] G. Ning, J. Pei, and H. Huang. Lighttrack: A generic framework for online top-down human pose tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1034–1035, 2020.

[82] D. Oliver. *Basketball on paper: rules and tools for performance analysis*. Potomac Books, Inc., 2004.

[83] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, and F. Giannotti. Playerank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5):1–27, 2019.

[84] J. L. Pena and H. Touchette. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012.

[85] M. Perse, M. Kristan, J. Perš, and S. Kovacic. A Template-Based Multi-Player Action Recognition of the Basketball Game. In *CVBASE '06 - Proceedings of ECCV Workshop on Computer Vision*, pages 71–82, 2006.

[86] P. Power, H. Ruiz, X. Wei, and P. Lucey. Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1605–1613, 2017.

[87] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M.-H. Yang. Hedged deep tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4303–4311, 2016.

[88] J. Quiroga, H. Carrillo, E. Maldonado, J. Ruiz, and L. M. Zapata. As seen on tv: Automatic basketball video production using gaussian-based actionness and game states recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 894–895, 2020.

[89] V. Ramakrishna, D. Munoz, M. Hebert, J. Andrew Bagnell, and Y. Sheikh. Pose Machines: Articulated Pose Estimation

via Inference Machines. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 33–47, Cham, 2014. Springer International Publishing.

[90] V. Ramanathan, J. Huang, S. Abu-El-Haija, A. Gorban, K. Murphy, and L. Fei-Fei. Detecting events and key actors in multi-person videos. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3043–3053, 2016.

[91] RealTrack. RealTrack Systems. `http://www.realtracksystems.com/`, 2010. Accessed: 23-04-21.

[92] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[93] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[94] J. Sánchez. Comparison of motion smoothing strategies for video stabilization using parametric models. *Image Processing On Line*, 7:309–346, 2017.

[95] N. Sandholtz and L. Bornn. Replaying the NBA. In *MIT Sloan, Sports Analytics Conference*, pages 1–13, 2018.

[96] R. Sanford, S. Gorji, L. G. Hafemann, B. Pourbabaee, and M. Javan. Group activity detection from trajectory and video data in soccer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 898–899, 2020.

[97] Second Spectrum. Our Work, 2013.

[98] T. Seidl, A. Cherukumudi, A. Hartnett, P. Carr, and P. Lucey. Bhostgusters : Realtime Interactive Play Sketching with Synthesized NBA Defenses. In *MIT Sloan, Sports Analytics Conference*, pages 1–13, 2018.

[99] L. Sha, J. Hobbs, P. Felsen, X. Wei, P. Lucey, and S. Ganguly. End-to-end camera calibration for broadcast videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13627–13636, 2020.

[100] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[101] SkillCorner. SkillCorner - Open Data. `https://github.com/SkillCorner/opendata`, 2020. Accessed: 23-04-21.

[102] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.

[103] Stats Perform. STATS Perform, 1981. Accessed: 23-04-21.

[104] StatsBomb. StatsBomb - Open Data. `https://github.com/statsbomb/open-data`, 2020. Accessed: 23-04-21.

[105] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 246–252. IEEE, 1999.

[106] M. Stein, H. Janetzko, T. Breitkreutz, D. Seebacher, T. Schreck, M. Grossniklaus, I. D. Couzin, and D. A. Keim. Director's cut: Analysis and annotation of soccer matches. *IEEE computer graphics and applications*, 36(5):50–60, 2016.

[107] M. Stockl, T. Seidl, D. Marley, and P. Power. Making offensive play predictable - using a graph convolutional network to understand defensive performance in soccer. In *Proceedings of the 15th MIT Sloan Sports Analytics Conference*, 2021.

[108] O. Sumer, T. Dencker, and B. Ommer. Self-supervised learning of pose embeddings from spatiotemporal relations in videos. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[109] SynergySports. Synergy. `https://synergysports.com/`, 2004. Accessed: 23-04-21.

[110] M. Sypetkowski, G. Kurzejamski, and G. Sarwas. Football players pose estimation. In *International Conference on Image Processing and Communications*, pages 63–70. Springer, 2018.

[111] Ł. Szczepański and I. McHale. Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2):513–533, 2016.

[112] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on image processing*, 16(2):349–366, 2007.

[113] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton. Computer vision for sports: current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18, 2017.

[114] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159:3–18, 2017.

[115] V. Vercruyssen, L. De Raedt, and J. Davis. Qualitative spatial reasoning for soccer pass prediction. In *CEUR Workshop Proceedings*, volume 1842, 2016.

[116] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32(4):722–732, 2010.

[117] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 24–25, 2020.

[118] J. Wang, I. Fox, J. Skaza, N. Linck, S. Singh, J. Wiens, and A. Arbor. The Advantage of Doubling: A Deep Reinforcement Learning Approach to Studying the Double Team in the NBA. In *MIT Sloan, Sports Analytics Conference*, pages 1–12, 2018.

[119] K.-c. Wang and R. Zemel. Classifying NBA Offensive Plays Using Neural Networks. In *MIT Sloan, Sports Analytics Conference*, pages 1–9, 2016.

[120] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li. Unsupervised deep tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1308–1317, 2019.

[121] Q. Wang, J. Gao, J. Xing, M. Zhang, and W. Hu. Dcfnet: Discriminant correlation filters network for visual tracking. *arXiv preprint arXiv:1704.04057*, 2017.

[122] X. Wang, A. Jabri, and A. A. Efros. Learning correspondence from the cycle-consistency of time. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2566–2576, 2019.

[123] G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean. Characterizing concept drift. *Data Mining and Knowledge Discovery*, 30(4):964–994, 2016.

[124] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional Pose Machines. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4732, 2016.

[125] X. Weng and K. Kitani. A baseline for 3d multi-object tracking. *arXiv preprint arXiv:1907.03961*, 2(5), 2019.

[126] J. Wiens, G. Balakrishnan, J. Brooks, and J. Guttag. To Crash or Not To Crash : A quantitative look at the relationship between offensive rebounding and transition defense in the NBA. In *MIT Sloan, Sports Analytics Conference*, 2013.

[127] L. J. Williams. Cognitive load and the functional field of view. *Human Factors*, 24(6):683–692, 1982.

[128] K.-H. Wu, W.-L. Tsai, T.-Y. Pan, and M.-C. Hu. Robust basketball player tracking based on a hybrid detection grouping framework for overlapping cameras. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 5094–5100. IEEE, 2019.

[129] D. Zecha, M. Einfalt, and R. Lienhart. Refining joint locations for human pose tracking in sports videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[130] F. Zhang, X. Zhu, and M. Ye. Fast human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[131] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu. Multi-camera multi-player tracking with deep player identification in sports video. *Pattern Recognition*, 102:107260, 2020.

[132] W. Zhang, Z. Liu, L. Zhou, H. Leung, and A. B. Chan. Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation. *Image and Vision Computing*, 61:22–39, 2017.

[133] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.

[134] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. Conditional Random Fields as Recurrent Neural Networks. *arXiv.org*, cs.CV:2015, 2015.

[135] C. Zhi-chao and L. Zhang. Key pose recognition toward sports scene using deeply-learned model. *Journal of Visual Communication and Image Representation*, 63:102571, 2019.