Fakultät für Informatik
Technische Universität München

# Automated Semantic Annotation of Football Games from TV Broadcast

Dissertation

*Francisco Siles Canales*

Intelligent Autonomous Systems Group
Fakultät für Informatik
Technische Universität München

# Automated Semantic Annotation of Football Games from TV Broadcast

*Francisco Siles Canales*

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

| | | |
|---|---|---|
| Vorsitzender: | | Univ.-Prof. Dr.-Ing. Darius Burschka |
| Prüfer der Dissertation: | 1. | Univ.-Prof. Michael Beetz, Ph.D.<br>Universität Bremen |
| | 2. | Univ.-Prof. Dr. Daniel Cremers |
| | 3. | Univ.-Prof. Dr. Martin Lames |

Die Dissertation wurde am 16.09.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 07.01.2014 angenommen.

# Abstract

Football is one of the most widespread, richest and complex team sports currently practised, and a growing interest for devising and implementing computational systems for its analysis, arose in the last years. Several applications can be supported by the aforementioned computational systems, for instance, game summarisation using play highlights, strategical, tactical and statistical analyses, development of better training strategies, verification of referee decisions, and many more.

The main objective of this thesis is to investigate mechanisms for the creation of a computational system, for the automated semantic annotation of football games from TV broadcast videos. For the realisation of this system, an abstract model for the representation of football is proposed. The model is used for the storing and retrieval of game-related data and information, that can be used for the answering of queries posed by football-interested people: coaches, journalists, sport scientists, fans, and others. The principal hypothesis of the current work is that the model can be populated, based on the trajectories on the field of play followed by the targets (players, referees, and the ball).

The contributions of this thesis are threefold. In the first place, the development of a *temporal segmentation* algorithm for the selection of a set of image sequences, corresponding to the far-view scenes in the input video. A dissimilarity measure comprising chromatic and structural features of the scenes is used to perform the selection. Secondly, the devise of a *spatial segmentation* algorithm for the detection and localisation of the targets in the selected image sequences. Maximal matching of a bipartite graph, with weights corresponding to shape and texture characteristics of the targets, is used to perform the estimation of their trajectories. Finally, the development of a *semantic segmentation* algorithm for the population of the proposed abstract model. The algorithm classifies the situations, actions, events, episodes, tactics, and strategics of the game, based on the estimated trajectories of the targets.

# Kurzfassung

Fußball ist eine der populärsten, reichsten und komplexesten Mannschaftssportarten, die derzeit praktiziert werden. In den letzten Jahren wuchs das Interesse für die Konzeption und die Umsetzung der computationalen Systeme, zur Analyse dieses Sports. Verschiedene Anwendungen wie Spielzusammenfassungen inkl. Spiel-Highlights, Entwicklung besser Trainingsstrategien, Überprüfungen von Schiedsrichterentscheidungen sowie strategische, taktische und statistische Analysen aber auch vieles mehr, können so durch die genannten computationalen Systeme unterstützt werden.

Das Hauptziel dieser Dissertation liegt in der Erforschung von Mechanismen, zur Erstellung eines computationalen Systems, welches der automatisierten und semantischen Annotation von Fußballspielen im Fernsehen und in Videos dient. Eine Voraussetzung für die Realisierung dieses Systems ist ein abstraktes Model für die Darstellung des Fußballsports. Das Modell wird für die Speicherung und den Abruf von spielbezogenen Daten und Informationen verwendet, welche anschließend für die Beantwortung von Fragen der Fußballinteressierten genutzt werden können. Die wichtigste Hypothese dieser Arbeit ist, dass das Modell durch die auf dem Spielfeld befindlichen Trajektorien (Spieler, Schiedsrichter und Ball), welche das Spielgeschehen verfolgen, mit Daten gefüllt werden kann.

Diese Dissertation gliedert sich dabei in drei Bestandteile. Zuerst ist ein zeitlicher Segmentierungsalgorithmus, welcher die Bildsequenzen entsprechend der Szenen im Eingangsvideo auswählt, zu entwickeln. Um die Auswahl der Szenen durchzuführen, wird ein Differenzierungsmaß mit chromatischen und strukturellen Eigenschaften verwendet. Den zweiten Bestandteil stellt die Entwicklung eines räumlichen Segmentierungsalgorithmus, welcher der Entdeckung und Lokalisierung der Zielobjekte in der ausgewählten Bildsequenz dient, dar. Zur Schätzung der Trajektorien wird das maximale Matching eines bipartier Graphen verwendet, wobei jedoch die Gedichte entsprechend der Form sowie der Texturmerkmale der Zielobjekte berücksichtigt werden müssen. Abschließend ist ein semantischer Segmentierungsalgorithmus zu entwickeln, welcher das abstrakte Modell mit relevanten Daten füllt. Der Algorithmus klassifiziert so die Situationen, Aktionen, Ereignisse, Folgen sowie die Taktiken und Strategien des Spiels, basierend auf den geschätzten Trajektorien der Zielobjekte.

# Resumen

Fútbol es uno de los deportes colectivos más difundidos, ricos y complejos practicado actualmente, y en los últimos años ha habido un creciente interés por el desarrollo e implementación de sistemas computacionales para su análisis. Diversas aplicaciones pueden estar soportadas por los sistemas computacionales mencionados anteriormente, por ejemplo, resumen del juego a partir de jugadas sobresalientes, análisis estratégico, táctico y estadístico, desarrollo de mejores estrategias de entrenamiento, verificación de decisiones arbitrales, y muchas más.

El objetivo principal de esta tesis es el de investigar mecanismos para la creación de un sistema computacional, para la anotación semántica automatizada de juegos de fútbol a partir de señales de televisión. Para la realización de dicho sistema, se propone un modelo abstracto para la representación del fútbol. Este modelo es usado para el almacenamiento y recuperación de datos e información relacionados al juego, los cuales pueden ser utilizados para responder consultas planteadas por personas interesadas en fútbol: entrenadores, periodistas, científicos deportivos, aficionados, y otros. La hipótesis principal del presente trabajo consiste en poder poblar el modelo, utilizando las trayectorias seguidas en el campo de juego por los objetivos (jugadores, árbitros y bola).

Las contribuciones principales de esta tesis son tres. En primer lugar, un algoritmo de *segmentación temporal* es desarrollado, para la selección de un conjunto de secuencias de imágenes correspondientes a escenas de plano general en el video de entrada. Una medida de disimilitud que consta de características cromáticas y estructurales de las escenas es utiliza para realizar la selección. En segundo lugar, la elaboración de un algoritmo de *segmentación espacial* para la detección y localización de los objetivos en las secuencias de imágenes seleccionadas. Apareamiento máximo de un grafo bipartito, con pesos correspondientes a características de forma y textura de los objetivos, se usa para realizar la estimación de sus trayectorias. Finalmente, el desarrollo de un algoritmo de *segmentación semántica* para completar el modelo abstracto propuesto. El algoritmo clasifica las situaciones, acciones, eventos, episodios, tácticas y estrategias del juego, basado en las trayectorias estimadas de los objetivos.

# Dedication

One single grateful thought raised to heaven is the most perfect prayer.

---

*(Gotthold Ephraim Lessing)*

To **Tía Rosa** with love.

# Acknowledgements

> I can't simply type "thanks" so I'll add a pointless sentence.
>
> *(Yotam)*

First of all, I would like to thank my advisor Prof. Michael Beetz, Ph.D., for the opportunity he gave me to work in his amazing research group. The guidance he provided me during this project was invaluable and always pertinent. His passion for science is admirable, and I hope to have learnt at least part of it. I also like to thank Prof. Dr. Bernd Radig for the work together, and for the things he taught me during our joint work, also for his great support during my stay. I am also grateful to Prof. Dr. Martin Lames for the time he spent in helping me from his expertise area, which was fundamental to the conclusion of the present thesis.

To the *Technische Universität München (TUM)* I will be forever in gratitude, since their doors were opened to me, and that signified a great opportunity to grow in the academia. I would like to thank Sabine, Manuela, Doris and Quirin, who were always ready to assist me. Distinguished thanks to the ASPOGAMO-team, for fantastic years of together-work: to Jan for his enormous talent, to Suat for his enthusiastic efforts, to Bernhard for his indefatigable meticulousness, to Nicolai for his astonishing brilliance, to Murat for his overwhelming human quality, and to Malte for his impetuous vocation. Also to my other colleagues, a special thanks: to Alexandra for her unbeatable kindness, to Zahid for his tireless courage, to Alexis for his tolerant disposition, and to Federico for his feisty friendship. To my other great colleagues at *TUM*, I have learnt a lot from you, I expect to keep in touch and to collaborate in the future.

I own to the *Universidad de Costa Rica (UCR)* all the great support during this years, the scholarship that I enjoyed during this period made possible the conclusion of my doctoral studies. A life-long gratitude to Jorge Romero and Eddie Araya, for their disinterested support and diligent aids. I am specially thankful to the *Oficina de Asuntos Internacionales y Cooperación Externa (Office of International Affairs and External Cooperation) (OAICE)*, as well as to the *Escuela de Ingeniería Eléctrica (Department of Electrical Engineering) (EIE)* for their help and trust. Special thanks to Vivian, Yamileth, Fátima, Dra. Ana Sittenfeld, and Dra. Julieta Carranza for their very competent and supportive effort.

Thank you Michele Weisbach for your small but great contribution.

Finally to my friends and family, since they were the engine who kept me going since the very first day I came to Germany. To my positive mother in law, who helped me from afar and sent me good vibrations. To my brave mother, who shaped me to fearlessly face anything and everything in life. To my beautiful and superb wife, who love me even when I deserved it the least, and without who nothing have been done. Last but not least, to my three vigorous lovely kids who are my joy and strength.

Thank you all.

Francisco Siles Canales

*Costa Rica, March 2013*

# Contents

# List of Figures

# List of Tables

# List of Algorithms

# Nomenclature

| | |
|---|---|
| 🟩 | Official FIFA’s pitch lines |
| 🟩 | Imaginary reference line on the pitch |
| 🔴 $h$ | Home player, with t-shirt number $h$ |
| 🔵 $a$ | Away player with t-shirt number $a$ |
| ● | Referee |
| ∘ | Ball |
| ↗ | Pass |
| ↝ | Dribble |
| ↗ | Shot |
| ⇢ | Movement of a player without the ball |
| ⋰ | Displacement of a player in a time period |

# Abbreviations

| | |
|---|---|
| ASPOGAMO | Automated SPOrts Game Analysis MOdel. |
| AFC | Asian Football Confederation. |
| CAF | Confédération Africaine de Football. |
| CCD | Charge-Coupled Device. |
| CONCACAF | Confederation of North, Central American and Caribbean Association Football. |
| CONMEBOL | Confederación Sudamericana de Fútbol. |
| CRF | Conditional Random Field. |
| DBN | Dynamic Bayesian Network. |
| DFG | Deutsche Forschungsgemeinschaft. |
| DFL | Deutsche Fußball Liga GmbH. |
| EC | Euro Cup. |
| EIE | Escuela de Ingeniería Eléctrica (*Department of Electrical Engineering*). |

| | |
|---|---|
| FA | Football Association. |
| FIFA | Fédération Internationale de Football Association. |
| FSM | Finite State Machine. |
| GMM | Gaussian Mixture Model. |
| HD | High Definition. |
| HMM | Hidden Markov Model. |
| HVBP | Hue-Variance Bhattacharyya Product. |
| IFAB | International Football Association Board. |
| MC | Markov Chain. |
| OAICE | Oficina de Asuntos Internacionales y Cooperación Externa (*Office of International Affairs and External Cooperation*). |
| OFC | Oceania Football Confederation. |
| RFID | Radio-frequency Identification. |
| SVM | Support Vector Machine. |
| TUM | Technische Universität München. |

| | |
|---|---|
| UCR | Universidad de Costa Rica. |
| UEFA | Union of European Football Associations. |
| WC | FIFA's World Cup. |

# Glossary

| Term | Description |
|---|---|
| ACTION | Something done. The state of being active. Carry through: put in effect; "carry out a task". |
| AUTO-GOAL | A goal scored by a player against its own team. It is usually accidental, and may be a result of an attempt at defensive play that failed or was spoiled by opponents. |
| BALL LINE | The imaginary line which is parallel to the field middle line and passes through the ball. |
| BUILD-UP RHOMBUSES | The tactical work performed by the players of the offensive team, as to place themselves for a possible pass of a team-mate. |
| CAMERA CALIBRATION | Is the process of estimating the intrinsic and extrinsic parameters of the camera taking the image or video. The extrinsic parameters are the three-dimensional coordinates describing the position of the camera into the world coordinates, as well as the angles at which it is directed: pan $\gamma$, tilt $\alpha$ and roll $\beta$. The intrinsic parameters include the focal length, format size, principal point, and lens distortion $\kappa$. |

| **Term** | **Description** |
|---|---|
| CONE | An imaginary reference triangle formed between the goal base and the player with the ball. |
| CORNER KICK | Is the way of restarting the game after the ball has been sent out of the field through the goal line by a defensive player. A pass is awarded to the attacking team, from the corner of the side of the field where the ball went out. |
| COUNTER-ATTACK REFERENCE | The tactical work performed by the players of the defensive team, such as to serve for a possible counter-attack against the opposing team. |
| CUT LINE | The imaginary perpendicular line between goal line and the ball. |
| DIRECT FREE KICK | Awarded to the opposing team following offences considered to be careless, reckless, or using excessive force. See laws 12 and 13 from the *Laws of the Game*. |
| DOUBLE TEAM | To attack or defend against with twice the usual force. For example, to cover an attacking player by means of two defenders. |
| DROPPED BALL | A dropped ball is a way of restarting play after an interruption called by the referee which is not typified in the *Laws of the Game*. It consists of the referee dropping the ball near the place where the play was temporary stopped. Play restarts when the ball touches the ground. |

| **Term** | **Description** |
|---|---|
| EPISODE | A happening that is distinctive in a series of related events. An incident that forms part of a story and is significantly related to it. Episodes may be either self-contained narratives or events that depend on a larger context for their sense and importance. |
| EVENT | Something that happens at a given place and time. A phenomenon located at a single point in space-time. Consequence: a phenomenon that follows and is caused by some previous phenomenon. An event is a segment of time at a given location that is conceived by an observer to have a beginning and an end. |
| FIELD OF VIEW | The area of the inspection captured on the camera's image. |
| FIELD OF VIEW SIZE | How much of the objects of interest and its surroundings are visible within the camera's *field of view*. |
| FOUL | An unfair act committed by a player against law 12 of the *Laws of the Game*. |
| GAME PHASE | A game phase is one of the two possible football match stages, whether the ball is *in play* or *out of play*. |
| GAME SHEET | Diagram depicting the relative positions of the ball, players and referees over the pitch, as well as a symbolic graphical presentation of the actions performed by the involved actors: passes, dribbles, shots, etc. |

| **Term** | **Description** |
|---|---|
| GOAL | A pair of upright posts linked at the top by a horizontal crossbar, often with a net attached behind it, forming a space into which the ball has to be sent in order to score. An instance of sending the ball into this space as a unit of scoring in a football game. |
| GOAL KICK | A free kick taken by the defending team from within their goal area after the attacking team having sent the ball over the goal line. |
| GOAL LINE | The two shorter boundary lines of the field of play. |
| INDIRECT FREE KICK | Awarded to the opposing team following "non-penal" fouls, certain technical infringements, or when play is stopped to caution or send-off an opponent without a specific foul having occurred. A goal may not be scored directly from an indirect free kick. See laws 12 and 13 from the *Laws of the Game*. |
| INTERRUPTION | Break: an act of delaying or interrupting the continuity, some abrupt occurrence that interrupts an ongoing activity. Pause: a time interval during which there is a temporary cessation of something. |
| KICK-OFF | The start or resumption of a football game, in which a player kicks the ball from the centre of the field. |
| Laws of the Game | the codified rules that help define association football. These laws are published by the sport's governing body *Fédération Internationale de Football Association (FIFA)*, with the approval of the *International Football Association Board (IFAB)*, the body that writes and maintains the laws [FIF11]. |

| **Term** | **Description** |
|---|---|
| MATCH ANALYSIS | Objective recording and examination of behavioural events occurring during competition. |
| MOTION ANALYSIS | Family of techniques of match analysis that records the match to review the game subsequently. |
| NOTATIONAL ANALYSIS | Family of techniques of match analysis that adopts a system of coding for those activities relevant to an assessment of performance, which allows the events to be annotated to later be collated. |
| OFFSIDE | A player is in an offside position if he is nearer to his opponents' goal line than both the ball and the second-last opponent. It is not an offence in itself to be in an offside position, it is only penalised if, at the moment of the player being in offside, the ball touches or is played by a team-mate, and the player is, in the opinion of the referee, involved in active play. |
| OFFSIDE TRAP | Is when the defenders move in a straight line formation to force the opponents to be in an offside position. |
| POSSESSION | A player is in ball possession, or simply in *possession*, when the player has the ball temporary under control. A team is in possession, when the players of the team are in possession consecutively. |

**Term** **Description**

PREVENTIVE COVERING The tactical work performed by the players of the offensive team, such as to protect against a possible counter-attack of the opposing team.

SHOT Consecutive sequence of frames that constitutes a unit of action in a film or video.

STRATEGY Scheme: an elaborate and systematic plan of action. Long-term planning and manoeuvring. The general plan or direction selected to accomplish incident objectives.

TACTIC A plan for attaining a particular goal. In chess, a tactic refers to a short sequence of moves which limits the opponent's options and may result in tangible gain. Tactics are usually contrasted with strategy, in which advantages take longer to be realized, and the opponent is less constrained in responding. A manoeuvre, or action calculated to achieve some end. Tactics are the specific actions, sequences of actions, and schedules you use to fulfil your strategy.

THE BEAUTIFUL GAME Is a synonym of Association Football.

THROW-IN The throw-in is executed when the ball has gone out of play crossing over one of the touch lines on either side of the pitch. The player must throw the ball back into play by delivering the ball from behind and over the head in one continuous motion.

| Term | Description |
|---|---|
| TOUCH LINE | The two longer boundary lines of the field of play. |
| TRACKING | The process of estimating over time the state of the targets using measurements taken from the input. |
| VIDEO TRACKING | The process of estimating over time the location of objects of interest using a video signal as input. |
| WING | External sides of the pitch, with respect to the central length axis. |
| WORK-RATE | Distance covered for a player in a game. |

# Chapter 1

# Introduction

> Not everything that can be counted counts, and not everything that counts can be counted.
>
> *(Albert Einstein)*

Since the, relatively recent, development of more powerful computer systems, the interest in devising and implementing computational systems for analysing human activities has seen an increasing growth, from the research area, as well as from the commercial companies. For surveys covering the available professional literature, see for example [AR11, TCSU08, MHK06, AC97]. This interest has spread also to sports in general and team sports in particular, and it is mainly due to the many possible applications that can be successfully addressed by using such mechanisms. Association football, as one of the diffused, richest and complex team sports currently practised, is not the exception to this ubiquity of computational analysis. See [DL10, WP04, Set03, Nee03, LKH$^+$02] for some related professional literature, and [Pro11, Spo11a, Eli11, Spo11b, Pan11, Cai11, Opt11, Asc11, TRA11, Ora11] for some commercially available products. Several important applications for different football-related groups can be mentioned, that might be dealt with by using computational analysis: game summarisation using play highlights, strategical, tactical, and statistical analyses, helping sports scientists to devise better training strategies, multimedia annotation – 2D and 3D video enhancements for broadcast –, helping in verifying referee decisions – e.g. goal awarding and offside penalisation –, and many more.

To fulfil the different necessities of the diverse football-related groups, such as coaches, journalists, sport scientists, fans, and others, and at the same time, to reduce the ambiguities due to subjectivity while analysing the game, a method for performing an objective analysis of football games is required. The main objective of this thesis is to investigate mechanisms for the creation of a computational system, for the automated semantic annotation of football games from TV broadcast videos. The system have to support mechanisms for the objective analysis of football games, and for that, an enough-informative abstract model – allowing the appropriate extraction and presentation of the relevant information to the football-related groups – has to be generated. With this model, it should be possible to answer questions about scoring chances of a team in determined situations, fail passes and their reasons, statistics about performance of players, team actions and its evaluation, as well as several others.

The principal hypothesis of the present work is that the required abstract model can be constructed based on the positional data of the involved objects of interest, that is, the trajectories of the players, the referees and the ball. To obtain such positional data, the first step is to process the video input in order to generate a useful *temporal segmentation*, which means to find those scenes where the interesting objects can be found and their positions observed. Once those meaningful and useful input video pieces are selected, the detection and localisation of the objects of interest, which correspond to a *spatial segmentation*, is to be effectively and efficiently performed. Then, a tracking of the spatial coordinates of those objects during the game must be carry out, producing the desired trajectories, which represent the sought positional data. Additionally, the game must be partitioned into smaller semantic components – e.g. episodes, events, actions – in order to represent the activities occurring during the match, and with that, to generate a better understanding of it. The above process represents a *semantic segmentation* of the game, where the actions of the players are to be detected and classified, as well as related actions chained together, in order to organise the happenings during the game. For devising such a high-level representation of the game, the development of an abstract model is required. This sought abstract model is a rich representation of the game that can be stored and later retrieved, and should be aimed at answering the diverse queries from the users of the system. All this together constitutes the desired automated semantic annotation system for association football, which is the primal interest of the present work.

## 1.1 Aims and Motivation

Association Football (or simply Football, from now on) has evolved from its origins 150 years ago[1], from a casual game performed by a group of amateur gentlemen to the now highly sophisticated, money-machinery sport known today. Football is the fastest growing and most popular team sport in the world, in both, number of spectators and number of active participants. The *FIFA*, its governing body, estimates the number of footballers that play in official

[1] In 1863, representatives of 12 clubs met in London to form the *Football Association (FA)* and to develop the laws of the game [Rus06].

competitions to be more that 50 million, and more than 240 million play the game on a regular basis [Rus06]. Considering the last 6 *FIFA* World Cups, the cumulative TV viewing audiences has astonishingly reached more than 24 billion spectators! [FIF07, Wikb]. Besides the *FIFA*, other regional confederations exist, each with their own competitions and figures: *Confederation of North, Central American and Caribbean Association Football (CONCACAF)*, *Confederación Sudamericana de Fútbol (CONMEBOL)*, *Union of European Football Associations (UEFA)*, *Confédération Africaine de Football (CAF)*, *Asian Football Confederation (AFC)*, and *Oceania Football Confederation (OFC)*. This enormous football community, comprises not only the players, referees, coaching staff and spectators visiting the stadiums, but also, journalists, sport scientists, physiotherapists, physicians, fans watching the games on TV and Internet, merchandisers, betting companies, insurance companies, software developers, and many more.

Everyone in this huge football community have different interests in diverse aspects of the game. Sports scientists would like to have access to the physiological and psychological aspects, for example, in the aim of prevention of sports injuries, identifying fatigue, differentiating positional differences in the *work-rate* and fitness levels, and development of training schemes. Reporters might be interested in highlights of the game for summarising purposes: goals, fouls, well performed play, shots to goal, and similar. Coaches will be more interested in technical and tactical aspects of the game, for example, in finding strengths and weaknesses of their own teams, and of the opposition, in order to choose new strategies and tactics to implement for current or future games. Also while recruiting new players, the skills such as passing, dribbling and shooting, as well as general performance of the candidates are of the highest importance for them. Of course, each one is eager to give their own view point with respect to the game, but even in the analysis from experts, there are always differences in opinions, mainly due to subjectivity. So it is a highly important matter in this context, to be able of stating an objective view point of the aspects of the game in order to analyse it.

With the aim of performing an objective analysis of football by coaches, match analysis was developed. *Match analysis* is the objective recording and examination of behavioural events occurring during a football match. It acquired widespread attention during the last two decades, and nowadays would be even seen as irresponsible from a coach not to practice some kind of performance analysis [Chr05]. This analysis technique can be used to create a benchmark, with which for example, the performance of the team at different times may be obtained. Two different approaches to perform match analysis are the motion analysis and the notational analysis.

The idea of *motion analysis* is to record the games, so that the trainer can replay them during the debriefing sessions, in order to discuss and analyse the relevant parts. Activities performed by the studied player are classified into different classes, such as walking, jogging, cruising and sprinting. Also, the frequencies of each such activity and the corresponding covered distances are also annotated. In this way, the work-rate of a player can be related to its physiological consequences, for example, in recognising fatigue or differentiating fitness levels. This technique has the advantage that, even if the whole of the interpretation must be performed by the coach, it might help in cancelling negative effects, such as little recall of key events occurring during the game by the trainer – due to limitation of human memory [Bad90] –, set views and prejudices that bias the objectivity of the coach, as well as effects of emotions such as anger or stress. Mechanisms as simple as a video recorder can be used to implement this technique, but of course, newer and better approaches utilises computer software. Several commercial tools are already available that have tried to address the issue (see chapter 2 for a review of the available products). Some of them, only solves the problem partially, while providing for example, only video indexing software, or by allowing the manual annotation of the game.

The other approach for match analysis is the *notational analysis*, which is a means of notating the events occurring during the match. A system of coding is adopted, in which, those activities recognised as relevant to asses both, the performance of the players and that of the team, can be accurately notated. Besides, positional data can be recorded by using a scheme of numbering zones representing the pitch. The notational analysis is basically ball-centred, which means, that the ball contacts determine when to start to notate the relevant information: what kind of action, which players are involved, where on the pitch is the action taking place, when during the game is the action happening, and what is the outcome of the action: successful or unsuccessful. Using this generated information, it is possible for the coach to evaluate the success rates of several actions of the players, and provides significant insights of the patterns of play and about the intensity of the game. For example, it should be possible to analyse the attacking episodes of a team, that is, to evaluate the number of passes, dribbles, and shots executed by the analysed team during its corresponding offensive phase. This will provide a great deal of information to the coach in order to devise a better strategy to increase the chances of scoring.

Before the advent of computer-aided techniques, simply pen-and-paper were used to accomplish the notational analysis, using for example a tally sheet to record the frequencies of the actions. Software products have been created to cope with such needs, but most of them require special hardware to be installed in the stadiums, such as special cameras or transmitter-antenna sets, a fact that, in most of the cases increases the price of the systems, or simply makes them non-portable. Another disadvantage of most of those systems is that too much user interaction is demanded in order to generate the desired information, for example, the actions are required to be manually annotated, player texture samples are required to be delimited, as well as player positions on the pitch. Such extensive manual interactions are, in addition to being highly error prone, tedious tasks at the very least[2]. A review of the related software systems in existence is provided in chapter 2.

In view of the above, an automated system that combines the characteristics of both techniques (motion analysis and notational analysis), and at the same time, facilitates the derived analysis using the gathered information, is very desirable and useful, not only for coaches, but for all the already mentioned members of the football community. Several fundamental questions arise while considering such a proposition, for example: which features are to be used to adequately describe the game, that are worth and meaningful to include in such a system?, how to extract, from the input video, the required features that allow the construction of the model to represent the game?, and how exactly should such models be constructed, in order to permit later queries to fulfil the information needs of the users?. The concerning problem will be detailed next, and the pinpointed questions, and others will be then addressed.

## 1.2 Problem Description

The efforts of the current work are dedicated to the creation of a software system for the automated semantic annotation of football games from TV broadcast video. A diagram of the expected information processing flow is depicted in the figure 1.1. The problem is a complicated one, since it requires the synergy of several parts for its accomplishment. The project is ambitious and of a very broad scope, being the logistics necessary to assemble the pieces together, a big challenge in itself. Several areas of expertise are related to the project, from which varied tools must be brought together and integrated into the software system. Some of those areas are: Pattern Recognition, Knowledge-based Systems, Data Mining, Machine Learning, Tracking, Image Processing and Computer Vision, Sport Science, as well as more fundamental ones such as Probability and Statistics, Databases, Software Engineering, and of course Programming, and Algorithms and Data Structures.

In order to start studying the problem that occupy us, the first aspect to consider is the choice of data input for the system. The choice of using broadcast as video input, is due to

[2] By the experience of the author and his colleagues, it is known that for the generation of the data of a few minutes of the game, an enormous amount of time is demanded, due to the intensive manual interactions required.

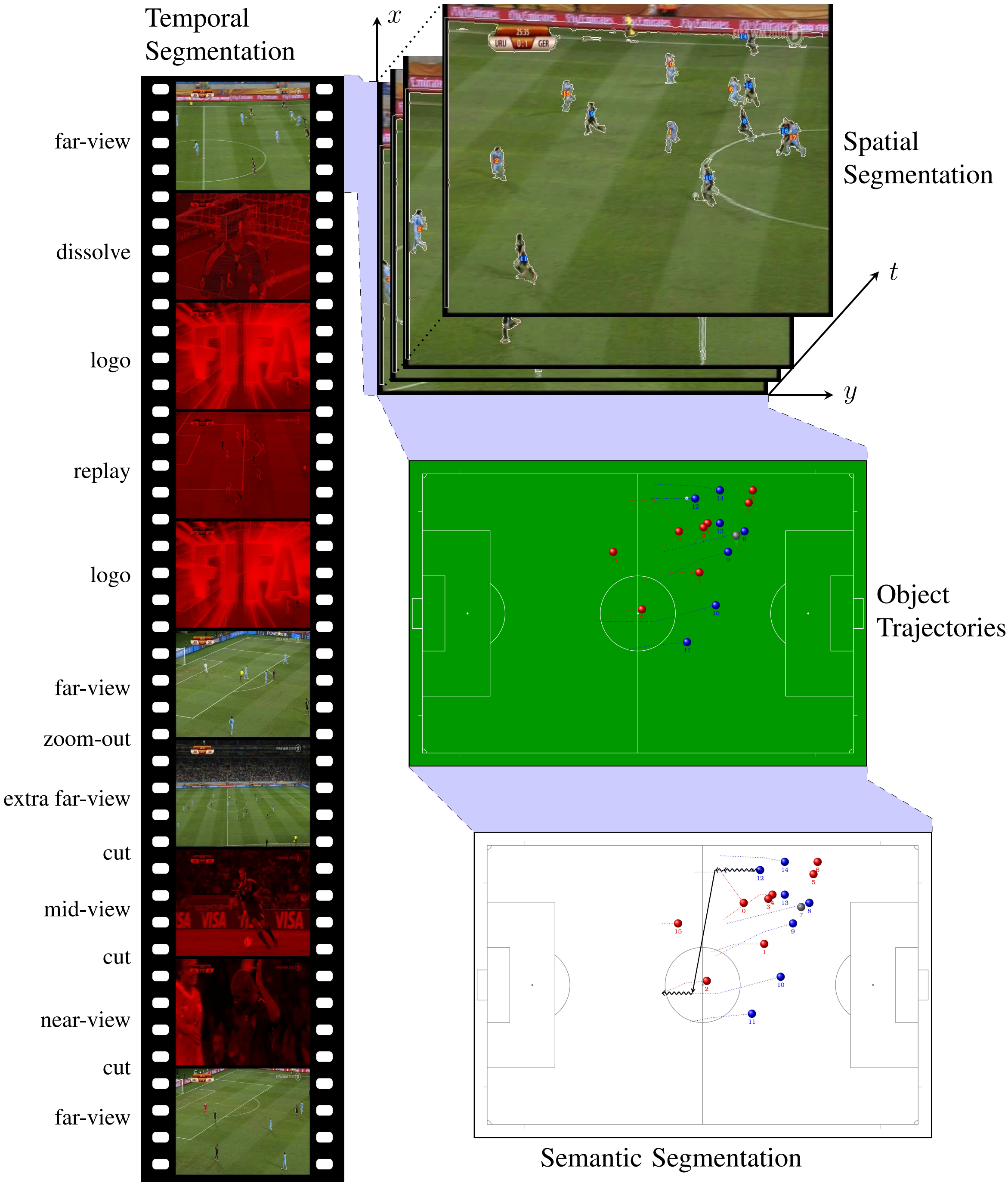


FIGURE 1.1 *Schematic of the desired information processing flow. The input video (left) might be composed of several shot types, including different camera views, replays, effects, and others, therefore, a* ***temporal segmentation*** *to select only the interesting scenes (far-view shots) is required. Once the desired scenes are selected, a* ***spatial segmentation*** *for the detection and localisation of the interesting objects is required (top right), and the object trajectories can be estimated (middle right). With this information appropriately gathered, the actions and events during the game can be recognised in a* ***semantic segmentation*** *stage, from which an abstract hierarchical model can be generated to semantically represent the game. As an example, consider the generation of game sheets, which correspond to attack episodes of a team, and include the movements and actions of the players (bottom right).*

the ampler range of users that can be reached by such a tool in comparison to the smaller coach staff groups enjoying an expensive fixed system installed in their own stadium. Also, the possibility of analysing old important games and events – *Wunder von Bern*, *La Mano de Dios*, *Gol del Siglo* – is of a great impact to the football community. Nonetheless, even if most of the input videos used come from the broadcast cameras, the use of any other available video sources should not be ignored and must be supported and encouraged. For example, positional data from other sources (e.g. manual data, data generated from other systems), or videos from static cameras installed in the stadium.

The second fundamental question to answer is, which features to use that can be extracted from the input video, and that can permit the system to accomplish its tasks. From the discussion in the last section, it is clear that the positive characteristics of both techniques (motion analysis, and notational analysis) shall be combined in the system. In both cases, the features used to accomplish the analysis are based on the positions, velocities and accelerations of the players during the match, as well as on the actions performed by them, the times of the occurrences of the relevant events and the hierarchical classification of such events into semantically meaningful sequences. Therefore, the features selected as part of the envisioned system are those encompassed in the common core of both analysis approaches: positional data and semantic information.

For the sake of simplicity, and to better describe what is to be done, the problem can be fundamentally divided into two main components. The first part deals with the *perception mechanisms* required to obtain the necessary positional data out of the input broadcast video. The second part is responsible for the *cognitive constructs* required to shape the abstract model that will help during the game interpretation. Each part is by itself a proper complex problem, and in the following paragraphs, a more detailed description of them is presented. Also, the main challenges that must be overcame in order to succeed are exposed.

The first part of the problem correspond to the *perception mechanisms*, which have as its input, the TV broadcast video of the game, and as desired output, the positional data of the interesting objects, that is, the position on the field of the players, referees, and the ball. At first, the input video must be processed in order to select only those *shots*, that are going to be useful as input for the remaining system pipeline. This process is termed **temporal segmentation**. For example, zoom-ins of the stands, where no players are visible, are of no help in seeking the players' positions, therefore, they must be removed. Also, digital overlays, such advertisements, logos, scores and the like, must be dealt with to avoid obstructions in the posterior tracking of the players. The replays and slow motion scenes are to be deleted, since they are out of the time-line of interest. This pre-processing of the video generates a sequence of scenes, that are rich in positional information, and are suitable for further processing. Which features to use, in order to do the classification of *shots*, is an interesting problem. No definite solution has been proposed in the professional literature, capable of correctly deal with the detection of cuts and dissolves for the general case. The success of such algorithms heavily depends on the particular contents of the video itself. A more detailed treatment of this will be covered in the chapter 3.

Once the video sequences have been selected, then the interesting objects are to be detected and located in the video sequences, which corresponds to the **spatial segmentation**. Each image of each sequence must be pre-processed in order to treat possible noise, either by filtering or masking it. Also, any colour space conversion that better suits the current needs have to be performed, for example, to enhance a particular characteristic useful to the segmentation algorithm. The segmentation algorithm provides the necessary means for the detection of the candidate regions, which represent the possible regions in the image of the objects of interest. To operate, the segmentation algorithm must use a set of selected features (or combination of them) chosen from the many available: colour of the team uniforms, player contours, local intensity values, and others. The quality of the localisation of the candidate regions procedure depends on the informative value of the selected features, so they must be carefully selected. Another problem is the unknown or changing illumination conditions, since the segmentation algorithm must be robust enough to deal with the possible encountered adverse scenarios. Those may include, but are not limited to: strong shadows on the pitch – produced by the roof of the stadium or by other surrounding objects –, multiple shadowing – when the game is played under artificial illumination and several light sources are available –, and variable weather conditions – cloudy, rainy, sunny –. Clutter in the background, due to advertisement or to partial occlusions among the actors, have a negative impact on the segmentation performance. Fast camera movements produce motion blur, which negatively affects the accuracy of the player detection algorithms. A final practical consideration is that, those algorithms processing massive image data consume the biggest amount of computing resources, so efficient algorithms are to be used.

The tracking algorithm uses the candidate regions found during the object localisation as input and generates the positional data, more specifically the trajectories of the objects of interest during the match. A decision must be made about the space in which the tracking should be performed, for example, it can be done in the 2D image coordinate system, in the 3D field coordinate system, or in a hyperspace of state variables, or in any other combination of the previous ones. An important issue to be faced is how to deal with target occlusion, either partial or total, since the algorithm must decide, once the occlusion finishes, to which of the targets does each of the possible tracks belongs to. This is known as the association problem in the multiple-target tracking literature [Sit64, BSF88]. This is not a simple decision, and several algorithms have been proposed to deal with this matter, with no definite success for the general case [BBK05, XLO04, Str06, SH05, VDP03, FLB$^+$04, GP08]. It is to be noted that a tight relation exists between the object localisation generated using the segmentation algorithm and the tracking algorithm. It is a complex matter to design a tracking algorithm, and satisfy its individual requirements, and at the same time to develop a segmentation algorithm, also with its own limitations and characteristics, in order to find the best combination in benefit of the global outcome. How this is performed will be explained in detail in the chapter 4.

Once the positional data is collected, the second part of the problem is the *cognitive modelling*, which deals with how to build an abstract model that helps in the game interpretation. As already discussed above, several aspects of the game are of interest, since several varied queries from the users must be resolved by using the created model. A multiple-layer structure, in which each new layer is built on top of the information of the previous layers, seems to be appropriate for this task. Such a hierarchical structure produces a logical scheme starting from the simplest and going through the more complex aspects of the game, producing in the way up a very rich representation. What exactly should the layers contain, in order to keep the representation functional, concise and simple, is a matter to be carefully decided, since that would be the ground for the representational power of the model. One problem that might occur is that, since the broadcast camera is following the ball action, it is hence not covering the whole of the pitch, and some players are simply out of the *field of view* of the camera. This generates incomplete positional data, which might or might not be a problem, depending on what exactly is being represented in a specific model. As an example, consider that the interesting aspect to investigate is the shot opportunities of a player with the ball near the opposing goal. In this situation, practically no relevant information is being lost if the attacking-team defenders located in their own half side are not visible. Therefore, the capabilities of the system strongly depend on the amount of input data available and its quality. With respect to practical considerations, it has to be decided which database structure and tables shape are the better ones to store the data that is obtained during the generation of the layers in the model.

For the automatic recognition of the actions performed by the players, such as passes, dribbles, shots, runs, and others, features producing good classification rates must be chosen, in order to correctly generate the desired semantic annotation. The required features are to be obtained based on the gathered positional data. Once the actions are correctly classified, good presentation mechanisms are to be devised, that allow the quick and easy understanding by the user. This methods include graphical representations, in which a lot of information can be conveyed in a very compressed way (a few diagrams for example). An attacking episode of a team is a good example, since in the attacking episode, the occurred passes, dribbles and shots are the most relevant bit of information to be expressed, a good way of representing it would be to show the relative player positions on the pitch, as well as annotations of which actions are being carried on (or were already carried out) by the players. Similar other representations must be chosen for the remaining information available in the model. The relevant aspects of this will be thoroughly discussed in the chapter 5

Another very important concern is how to generate new knowledge based on existing data from several past games, and specially, how to use such knowledge to analyse newer games. During the evaluation of particular situations, like in the case of evaluating the scoring chances of a player in a position to shoot in front of the opposing goal, certain features must be selected to establish a performance measure. Many features can be proposed for that matter, such as distance to goal, angle with respect to the length field axis, number of surrounding opponents, aperture angle to the goal. But which of those features or combination of them is really decisive in the assessing of probabilities for the evaluation of the scoring chances is to be deeply investigated.

A final consideration about the desired system is that, to succeed in this project, a holistic view is required, one that considers the interrelations among the different sub-processes, and not only the optimisation of the individual parts, since the later will not produce the desired expected results. The issue is then, how to keep an integrated perspective of the problem while developing its parts. This might seem to be a more philosophical question, but it rather has serious far-reaching practical repercussions.

## 1.3 Contributions

The main contribution of this thesis is the implementation of a software system capable of generate, in an automated manner – that is, with a minimum of required user interaction –, a semantic annotation of football games. Semantic annotation is understood in this context, as the adding of interpretative information to the game, with the aim of understanding it. For example, attacking episodes are automatically recognised, and information about the structure of this episodes is represented using game sheets, including relevant information about the performed player actions. A *game sheet* is a diagram of the relative positions of the ball, players and referees over the pitch, as well as a symbolic graphical presentation of the actions performed by the involved players. Among the represented actions it can be mentioned: passes, dribbles, shots and runs. More advanced football analysis tools are also provided, such as fail passes analysis, where the unsuccessful passes are collected, and could be used for pinpointing the reasons of the failure.

The required perception mechanisms producing the positional data out of the input video are also implemented. In particular, a new method for *temporal segmentation* is implemented. The framework used for the classification is based on the formal description of the *shot* boundary detection problem presented in [YWX+07]. The novelty of the new method is the dissimilarity measure used to represent each scene. This proposed measure incorporates not only colour spectrum information, but also information about the texture distribution of the scene, making it a very robust feature, capable of reaching high classification rates – the average F1-score obtained was 94,48%. The F1-score is a composed evaluation indicator which comprises the precision and the recall of the classification method under investigation –. Also, a practical implementation for overlay detection is implemented, in order to generate the desired appropriate scene sequences. The method was thoroughly evaluated and validated using several games of the *WC*, and others. As an aside advantage of the developed system, is that the temporal segmentation allows an indexing of the video to bookmark the event occurrences, and later retrieve the desired scenes.

Another important contribution is the devise and implementation of a new *spatial segmentation* method for the detection and localisation of the objects of interest on the sequence of images selected. The method uses colour and texture clues to classify pixels belonging to the background (field) and those corresponding to interesting foreground objects (players, referees and ball). The method has proven to be very robust to the severe illumination changes encountered during the evaluation of the algorithm. The detection-localisation algorithm uses an adaptive modelling of the interesting objects, and by means of a bipartite graph to model one iteration in the image sequences, between time $k-1$ and $k$, is capable of produce better results compared to those obtained by the project in the past. Despite the apparent simplicity of the detection-localisation strategy, it has proven to be very convenient, even when dealing with partial or total occlusion of the targets.

The re-interpretation of the abstract model and the set up of a database with the required structure to support the abstract model for the *semantic segmentation*, was also a contribution of the present work. The design contains the necessary information for answering the queries that the varied users may have. Also, functionality to update and expand such models are also provided, so more features can be incorporated to the main structure at any time. A simple interface using a web protocol was implemented and with it, users can remotely use the program to process their own videos.

The implemented algorithms were thoroughly evaluated against several complete games taken from international competitions, specifically from *WC*2010, *WC*2006, and *WC*2002. Also games from national-level competitions, such as the *1. Bundesliga* in Germany. In particular, games of the *FC Bayern München* during the 2008/2009 season were used for the testing and evaluation of the implemented algorithms. This is in contrast with most papers in the professional literature, where commonly, small video sequences of only few minutes or seconds of a single game were use for the evaluation [DL10].

The current work was done as part of the multidisciplinary project: *Automated SPOrts Game Analysis MOdel (*ASPOGAMO*)*, which jointly with sport scientists, tries to analyse sport games in an automated fashion [BvHHK+09, BKL05].

## 1.4 Thesis Outline

The remaining chapters of the current thesis are organised as follows:

**CHAPTER 2 — SYSTEM DESCRIPTION** A review of the proposed systems in the professional literature, that attempt to solve the problem is presented in this chapter. Also, some existing commercially available products are presented too. Not only football-related products are mentioned, but also more general sport systems are also described, in order to present useful characteristics of them. The drawbacks and advantages of the presented systems are revised in detail. References in the professional literature, Internet and other sources are provided that described the presented systems. Finally, our proposed solution is presented, and the computational framework used to accomplish it will also be described. The layered structure of the modelling for the semantic annotation is presented and justified.

**CHAPTER 3 — TEMPORAL SEGMENTATION** In this chapter, the method for finding the images sequences of interest selected from the input video is presented. The different effects used by the TV broadcast companies to convey the emotions of the game are explained. The formal framework for detecting shot boundaries, such as cuts, dissolves, and overlays is described. A new proposed dissimilarity measure to represent the scene variations based on the chromatic and structural characteristics of the scenes is explained. The results of the classification of the shots are presented.

**CHAPTER 4 — SPATIAL SEGMENTATION** This chapter basically presents the two separated but strongly inter-related aspects of the positional data generation, that is, the detection and localisation of the objects of interest in the selected image sequences. The algorithms that are used for the detection and localisation of the candidate regions of the interesting objects are presented, as well as all the challenges that have to be circumvented. The detection-localisation is based on an adaptive appearance model of the followed targets, together with a bi-graph matching algorithm to cope with the problem of target association. The spatial segmentation qualitative appraisal using the games of the *WC*2010 is presented, and how this new algorithm supersedes the old method used in the ASPOGAMO project in the past is also discussed.

**CHAPTER 5 — SEMANTIC SEGMENTATION** In this chapter, the main relevant aspects of *The Beautiful Game* are presented and discussed, that consists of the football semantics, which is important to figure it out what and how to model from the game. The actions or events that can occur during a match are described, and it is explained how they amalgamate together to form higher meaningful elements, such as episodes. The abstract hierarchical model used to represent the game is presented, as well as each of its constitutive layers explained. The semantic segmentation, that is, the building of the information corresponding to each of the layers in the model is also addressed. In particular, the classification of the actions and events of the match and the corresponding evaluation is also presented and discussed. Finally, examples of the application of the abstract model for game interpretation are presented in this chapter.

**CHAPTER 6 — CONCLUSIONS AND RECOMMENDATIONS** The final conclusions of the thesis are presented in this chapter, as well as future directions that can be taken for the research in this area.

# Chapter 2

# System Description

"Madness" is continually doing the same thing while expecting a different result.

*(Alcoholics Anonymous)*

In the current chapter, a review of some of the proposed systems in the professional literature, that attempt to solve the problem, as well as some related commercial products will be presented. The focus of the presentation will be football, but of course other important inputs from other sports will be taken into account. The relevance of each system, from the perspective of the current work, will be assessed, and if justified, some instances will be thoroughly discussed.

Finally, our proposed solution will be presented, and the computational framework used to accomplish it, will also be described. The layered structure used for the abstract modelling will be discussed. The corresponding input and output of each block in the proposed architecture of the system will be shown and described in detail.

## 2.1 Related Work

Several approaches attempting to solve the problem of football analysis have been proposed in the professional literature. A comprehensive review about the existing modern developments in the professional literature for football video analysis is presented in [DL10]. From the references cited there, only around a third of them deal with the high level semantic analysis of the games, while the remaining tackle just some of the related sub-problems – such as shot boundary detection, camera tracking, ball tracking, player and referee tracking, etc. –. Another problem found is that most of the algorithms were evaluated only on very small video sequences (see test sequences in tables 3, 4, 5 and 6 in [DL10]), which of course, makes quite difficult to evaluate the accuracy, precision, and generalisation properties of the proposed methods.

In addition to the algorithms mentioned in the professional literature, there are a number of commercial companies offering different products for analysing sports in different ways. For references appearing in the professional literature covering some of the existing companies and their products see [Set03, Nee03, WP04]. The main disadvantage of the company products is that, since the used algorithms are proprietary, no access to their technical characteristics is available. Therefore no objective comparison is possible, nonetheless, good ideas can be obtained while analysing the existing systems.

In the following, a review of the proposed systems in the professional literature, as well as the commercial products will be presented. The relevance of each system, from the perspective of the current work, will be assessed, and if justified, some instances will be thoroughly discussed.

### 2.1.1 Professional Literature

For the presentation of the different proposed algorithms, a classification approach based on the main application areas of soccer video analysis is used in [DL10]. The three types of application domains employed by the authors for the classification are: video summarisation, provision of augmented information, and high level analysis. Richer information is provided to the user while going from video summarisation, to provision of augmented information, and to high level analysis. Consequently, more complex algorithms are required to extract the semantic information supporting the applications. The algorithmic requirements increase as going up in the classification hierarchy, but at some points, similar problems have to be faced by the algorithms of the three classes. For example, in a video summarisation scenario, the idea is to provide the user with a video containing only highlights of the game, for which certain events of interest such as goals, corner kicks and the like have to be recognised. But of course, similar recognition of game events have to occur in a high level analysis application, in order to enable a tactical analysis. Therefore, a better way to categorise[1] the algorithms would be one that does it not only on an application domain base, but on the level of required semantics and on the particular problems that have to be solved.

In the present work, an application-independent and problem-focused categorisation scheme based on the levels of semantic analysis required will be used to present the proposed algorithms. The three categories that will be used are low-level semantics, middle-level semantics,

[1] For an article addressing the difference between categorisation and classification see [Sin06].

and high-level semantics. The classification will be described next, mentioning the more important challenges that have to be faced in each class, as well as some of the proposed solutions to each of them. Also, the relevance of each class with respect to the current work will be assessed.

#### 2.1.1.1 Low-level Semantics

The class of *low-level semantics* algorithms requires the least complex semantic analysis to fulfil the needs of the applications in this domain. That means that this category includes those algorithms that perform a very primitive semantic analysis. The methodology followed in this category is to find correlations between low-level visual and aural features extracted from the video and audio signals respectively, and model their evolution over time, relating them with the searched semantic events.

As an example, consider a semantic indexing task like video summarisation, where, from a long input video, a smaller output video has to be generated, containing the more interesting highlights occurring during the game. Along with reducing the video size, which can be important for transmission applications constrained to reduced bandwidths, it is possible to index the game and allow the users to search for the parts that are relevant to them, which represents a content-based search problem [DL10]. Among the possible interesting events that can be searched for, there are: goals, penalisation cards, player substitutions, and penalties. In the particular case of goals, normal game commentators use louder expressions to announce them, and broadcasters practise camera zoom-ins directed to the celebrating stands, to convey the on-site emotions to the TV spectators. Therefore, an increase in the audio activity, and changes in the camera behaviour (e.g. camera close-ups, and slow motions replays) are features that can be associated to scoring events, hence, can be used to detect goal situations.

The problems faced by the low-level semantics algorithms are video pre-processing, feature extraction, and time modelling of feature evolution. The video pre-processing refers to the need of selecting only those video segments that are relevant to the application at hand. That is to say, shot boundaries have to be found in order to classify the shots depending on its type: far-view, mid-view, near-view, replay, etc. The following references from [DL10] apply some kind of shot boundary detection [CSCZ04, XXC$^+$04, ETM03, AKM03, WMZL05, HSC06, TCP04, YLC04]. The shot boundary detection is relevant to the present work, and therefore a more detailed presentation of the state-of-the-art for shot boundary detection is delayed until the chapter 3.

The visual features used as low-level characteristics include: camera view (close-ups, far-view, mid-view, near-view), camera motion (moving camera, lack of motion, fast zooms, fast pan), pitch characteristics (dominant colour, line orientation), crowd detection (image texture), and text detection (broadcast score boards). And among the aural features, audio loudness is the preferred, aiming at detecting: ambient crowd noise, whistles, clapping, and plain, exiting, and very exiting moments. Finally, this visual-aural information is combined and several models try to extract from this low-level features, the desired representation. Among the techniques used to build the representation models there are: *Markov Chains (MCs)*, *HMMs*, *Support Vector Machines (SVMs)*, *Dynamic Bayesian Networks (DBNs)*, *Finite State Machines (FSMs)*, and *Conditional Random Fields (CRFs)*.

From the point of view of the present work, the main disadvantage of the low-level semantics algorithms is precisely their primitive semantic analysis nature, since the analysis is based only on the extraction and modelling of simple low-level features [DL10]. As reported in [BLM00], individual low-level features such as lack of camera motion, and camera parameters (pan and zoom), are not producing the desired results, and using for example, a visual-aural joint characterisation with a controlled Markov chain model has proven to be insufficient for the task [LMP03]. Because the principal aim of the current work, is to provide the user with the tools that allows for a deeper semantic analysis, low-level semantics algorithms lie outside the scope of this thesis, and therefore will not be considered in the following. Besides, it will be shown that to a certain degree, generation of semantic indexing (video summarisation, for example), is possible using our system (see chapter 5).

#### 2.1.1.2 Middle-level Semantics

Algorithms in the *middle-level semantics* class require a semantic interpretation of the input video sequences. Generally, the required information consists of the camera parameters during the game (known as camera calibration), and the ball, players and referee detection, recognition, and tracking. If accurate, they will provide the data in order to estimate distances on the pitch, for example between players, or between a player and the ball or the opposing goal. Several commercial applications, related to the multimedia area, benefits directly from such algorithms, for example, by allowing the addition of advertisements to the broadcast games on the fly, the supply of statistics about the game and particular players, and the generation of virtual animations of the game, as well as others.

Some of the available middle-level semantics algorithms deal with the problem of camera calibration [Tho07, bHPV04, **?**], which is not directly addressed in the current work, but is a fundamental part of the architecture of our system (see chapter 4). For the ball tracking, algorithms that use TV broadcast videos are [YLXT06, LLHG06, SSMS06, CS05, PMMS08]. The approach followed by them is twofold: first, candidate ball regions must be found in each processed frame, and second, the ball candidates must be evaluated in order to produce valid ball trajectories. The challenges of the detection of the candidates regions are the small size of the ball region in the image, the variation in its shape and texture because of the motion blur due to its fast movement and to the camera operation, the possible occlusions suffered when being driven by a player, and the noise from player's uniforms or lines, or from the stands when the ball is flying. It is clear that since is a very complex problem it has to be supported by a posterior procedure to generate the desired trajectories. This validation procedure is based on a ball motion model, which guides the trajectory selection process. In [LLHG06] the ball is assumed to be white-coloured and therefore, the candidates are found by empirically thresholding the image and searching for the white pixels in the normalised RGB space. Morphological operations are used to cope with noise, and a region growing algorithm is applied to join and smooth the regions. Also some size constraints are used to finally obtain the ball candidates. This ball candidates are added to a weighted graph, and the Viterbi algorithm is used to find the optimal path in the graph, which would correspond to the ball path. In [YLXT06], after the candidates have been found, then candidate ball trajectories are generated for all the found ball candidates, and assuming that the ball is the "most active" object in the video, then some of those candidate trajectories are ruled out. The selection procedure is also supported by heuristics that gives more weight to some of the candidate trajectories, and finally a ball trajectory is generated. Similarly to the above two methods, in [SSMS06] ball candidates are generated based on a colour-shape segmentation procedure, and in order to track the ball even after possible occlusions (with lines, stands or players), then, once a ball is lost, a so called transition graph, representing the possible transitions of the ball to the surrounding objects, is generated. This graph is searched to find ball candidate routes between the lost-ball and the found-again-ball points. The ball final path is selected as the most feasible one, which is based on a normalised measure of the scores divided by the length of the path. In [CS05], the results of a player tracking algorithm were used to generate better hypotheses for the ball position. The ball is tracked in a batch processing mode, that is, the blobs representing ball candidates are accumulated for 20 frames. A particle filter is used to estimate the posterior density based on the observation till now (Markov assumption). The algorithm also uses a "chase" technique to keep track of the objects that are around the ball when it gets lost, and when the ball appears again, then the trajectory can be reconstructed. While the ball is in a visible state, the discontinuities are filled using a first order dynamic model under Gaussian noise.

Other algorithms tackle the problem of player tracking, some of them are focused on the fixed multiple-cameras case [RRH+05, IS07a], while some others are facing the problem using broadcasting videos as input [AWKG05, YYLL08, bHPV04, WXC+08]. The works related to the object tracking will be described in the chapter 4.

#### 2.1.1.3 High-level Semantics

*High-level semantics* algorithms deals with the answering of questions related to the tactical behaviour of the team, the roles of the players and their performance during the game, the scoring opportunities of the teams, their weaknesses and strengths, the interesting events of the game, such as attacking episodes, the particular actions executed by the players: passes, dribbles, shots, goals, etc, and their possessed skills and abilities. The presented algorithms in this area are the most interesting to mention for the present work, since they provide the level of information richness and complexity that we are dedicated to produce. In order to be able to generate the answers to such questions, the systems require the most complete set of accurate information with respect to the position of the moving objects, like the ball, the players, and the referees, as well as a way of representing the interesting occurring events of the game. Once the positional data is gathered, then the semantic information can be generated.

In [ZHX+07, XZZ+08, ZXZ+08, ZXH09a, ZXH+09b] a semantic and tactical analysis for goals is presented. The event detection from broadcast videos is done by using web-cast[2] text to find the goal events, and to generate afterwards tactical features for analysis. Two features for the tactical analysis are extracted from the ball and player trajectories: *attack route* and *attack interaction*. The attack route differentiates between centre and side attacks of a team, using three positional features to identify the play region. The positional features use for identification of play region are: field line location, goal mouth location and central circle location. In the other hand, the attack interaction classifies the attack into several categories in a coarse-to-fine manner. In a first level cooperative and individual patterns are found, and later more detail classifications are performed for each of these classes: e.g. unhindered-attack, interceptive-attack, direct attack, and dribbling-attack, each of which is trying to represent the particular behaviour of the attacks. The system was tested using 168 goals from the *WC*2006 games, obtaining 92% of accuracy for the web-cast text-based event detection, and recall and precision figures between 61.9% and 88.2% for the classification tasks. One disadvantage of the use of web-cast text is that it depends on the proper annotation and web-cast of the respective data for the particular game to be analysed. Another disadvantage is that even the presented tactical analysis features are very interesting and throws some information about the behaviour of the team during the attack, it is an approach that lacks a more structured framework. This topic will be better dealt with in the chapter 5. Also, there are very few algorithms belonging to this category, as stated in [DL10, ZHX+07, ZXH+09b].

[2] Web-cast means broadcast over the Internet.

### 2.1.2 Commercial Products

Each of the companies offering products for sport analysis tackles the problem with a different perspective and with different aims. This is reflected in the way their systems are implemented, and in the properties that are offered to the final user. There are several sport analysis systems available for almost every imaginable sport: football, rugby, cricket, basketball, tennis, car racing, horse racing, gridiron, and a plenty of others. In the following, a selection of sport analysis companies was chosen, their systems described and the worth-mentioning capabilities discussed. Naturally, not all the reviewed products have emphasis in football, as is the main focus of this work, but even if not football-focused, by reviewing some of those systems it is possible to get an overview of interesting technology features, as well as information and capabilities offered to the final users. By the same token, some other systems are presented, even if no automatic tracking is offered. Normally these systems are simple tools to carry out some of the required analysis tasks in a more easier way for the user.

Due to the proprietary nature of the algorithms used by those companies, it is in general very difficult to obtain complete and accurate information about performance and implementation details of the algorithms used. Nevertheless, their products will be described to the possible extent.

#### 2.1.2.1 Prozone Sports

*ProZone Sports* is a company established in the United Kingdom in 1998, which provides several products for match analysis in football. Their main products are *ProZone 3* and *MatchViewer*, with which is possible to semi-automatically track the players every tenth of a second, and to get access to the manually annotated ball-related actions such as passes, dribbles, shots, as well as other events such as fouls, corner kicks, and the like [MRG02, Pro11, Pro, BOWT07].

The architecture of the system consists of 8 to 12 static cameras placed at different locations of the stadium, connected to a central information processing unit in charge of controlling the recording process. Each camera has a different perspective view of the field, and the positioning of the cameras is such, as to jointly cover the whole of the pitch. Each camera covers a particular region of the pitch, and the regions covered by the cameras overlap to some degree, a feature aiming at having good accuracy, resolution and resilience, as well as to better deal with occlusions in the later processing steps.

During the installation of the system in the stadium for the first time, the cameras are calibrated (using a *camera calibration* algorithm), in order to later be able of converting from the 2D image coordinate system of each frame in the video signal to the 3D real world coordinate system of the pitch. This step is necessary, since the interesting player positional data is with respect to the coordinates of the field and not with respect to the cameras' coordinate systems. Due to the static nature of the cameras, that is, the cameras will not be moved after being finally set, this calibration is performed only this single time, and the 2D-3D transformation can be carried on using a homography transformation [HZ03]. Initially, a linear 4-point transformation is used for the initial matching between the image and the 3D model of the pitch, and then this estimation is refined by a proprietary algorithm using 50-points dealing with the $\kappa$-distortion due to the lens curvature [SCBC06].

When a game is to be analysed, an operator at the stadium, controls the beginning and end of the game recording, which, once the match is completely recorded, is transmitted to the remote ProZone's data processing centre, where the semi-automatic tracking of the players, and the manual event annotation is performed. Each video signal (from the different sources) is independently tracked, and then combined to form a single data-set.

Based on the fact that the cameras are static, a background subtraction method is employed to find the foreground pixels (supposedly corresponding to the players). For that, a set of reference images of the background without players under different illumination conditions is generated for each of the cameras. If the weather conditions changes, changing with it the illumination conditions of the scene, the algorithm can select to use a different background image from this previously generated reference image set. A mask is also created per camera, to mask out the unimportant regions not required for tracking, such as the stands.

Once the required reference images and masks are created, they are used during the tracking phase as follows. First, the corresponding reference image is subtracted from the incoming image, the resulting image is then thresholded using a threshold value manually selected, to choose only those pixels with hight brightness, which correspond to the foreground pixels (players). If the threshold value is chosen too low, then too much noise from the pitch will be kept, and if selected too high, many players' parts will be ignored, a compromise must be found by manually adjusting this value. After that, the mask is applied to remove the uninteresting image regions. A player is defined as a continuous block of pixels not deleted in the previous steps. Morphology operations are required to add undesirably removed pixels to keep the assumption of continuity of the players pixels. To each pixel block, an identification box is assigned, where the middle of the bottom line is defined as the position of the player. Kalman filtering is used to predict possible player movement directions based on current speeds of the players, and size constraints are used to remove those objects for which the sizes do not correspond to human beings. When collisions of the bounding rectangles of the found players occur or when no player candidates for the next frame exist, then manually changes by the operator are required to solve such problems. No ball tracking is explicitly performed, but the ball trajectory is obtained from the manually annotated events and the positions of the players [MRG02]. Finally, the positional data is validated by a quality control operator by visually comparing the generated tracks with an outside broadcast signal, and verifying that the trajectories correspond to the actual players. If required, the trajectories can then be re-identified.

The event annotation process can be also performed in parallel by assigning different segments of the video to different operators and allowing them to manually select the corresponding ball-related action or general event by selecting the start and end frames, the involved players, as well as the class of event and pitch location. The annotation requires a team of 7 operators: A *Setup operator*, four *Observers* to whom equally large segments of the game are assigned, and two *Team Leaders* for quality control [BOWT07]. The software used by the observer operators requires to systematically and linearly go trough the corresponding game segment every tenth of a second until an event occurs and manually add the required information about the event to the system. 48 different events are available to select from: pass, dribble, shot, goalkeeper catch, drop ball, foul, offside, clearance, corner pass, corner cross, and others. Some logic incorporated into the program assists the operator by not permitting the concatenation of events that makes no sense, for example having a goalkeeper catch after a ball going out bounds, since a corner kick (for example) must be in between.

In [BOWT07], four different types of errors in the event annotation system were found. Mismatch in the event being annotated, since sometimes it was considered to be constituted by two smaller events instead of a single one; errors arising from the confusion between players of the same team; errors in the event timing, since sometimes the start or end times where differently annotated by different observers; and finally, positional errors, were the annotated position on the pitch were the event is taking place was showing errors in the order of meters (3.6 m as mean absolute error). This errors are *"... an inevitable consequence of using a human operator as part of the measurement instrument."* [BOWT07].

The system provides blending animations, video content and statistics in an interactive format that the user must navigate to seek the required information (See figure 2.1). Examples of the provided statistics are distance covered by each player, passes made, and work intensity. One important drawback that can be mentioned, is that even if the number of provided statistics is large, this can be considered as a negative aspect, since the user, besides requiring to possess the minimum knowledge level to interpret the data, must have been trained into the use of the software, and it has to have the time to navigate through the tool in order to get to the particular important and interesting aspects [Set03]. Other drawbacks that can be accounted have to do with the lot of work required from the operators in order to manually register all the interesting events occurred during the match, as well as to aid in the player tracking. And importantly enough, the installation costs have to be taken into account, since the several cameras must be placed in the stadium, and each game analysis have an extra cost that have to be paid.

#### 2.1.2.2 Sport Universal Process

Established in 1995, *Sport Universal Process* is a French company providing tracking technology and performance analysis solutions for professional football, for football clubs in one hand, and to media providers in the other. *Amisco Pro*[3] is the main product offered by the company [Spo11a, Spoa, Spob, GB00].

The *Amisco Pro* is very similar to *Prozone 3*, the architecture of the system consists of 6 to 10 digital cameras installed around in the stadium. It uses a patented proprietary "tracking" technology to follow the players during the match. Also, the operators manually annotates all events, such as fouls, offsides and cautions, that happen during the game. The information provided by the *Amisco Pro* is basically the same as that provided by *Prozone 3*. 2D representations of the game are available to the final user, as well as summary statistics: how much a player has ran, how fast, and other similar measurements. See figure 2.2 for examples of the offered information.

The disadvantages of the *Amisco System*, are the same as for the *Prozone*: installation costs[4], non-portability (since the system is installed and fix to the particular stadium) and a lot of manual work, not only to generate the data, but to operate the system, since commonly, an operator must be trained and assigned to use the programs.

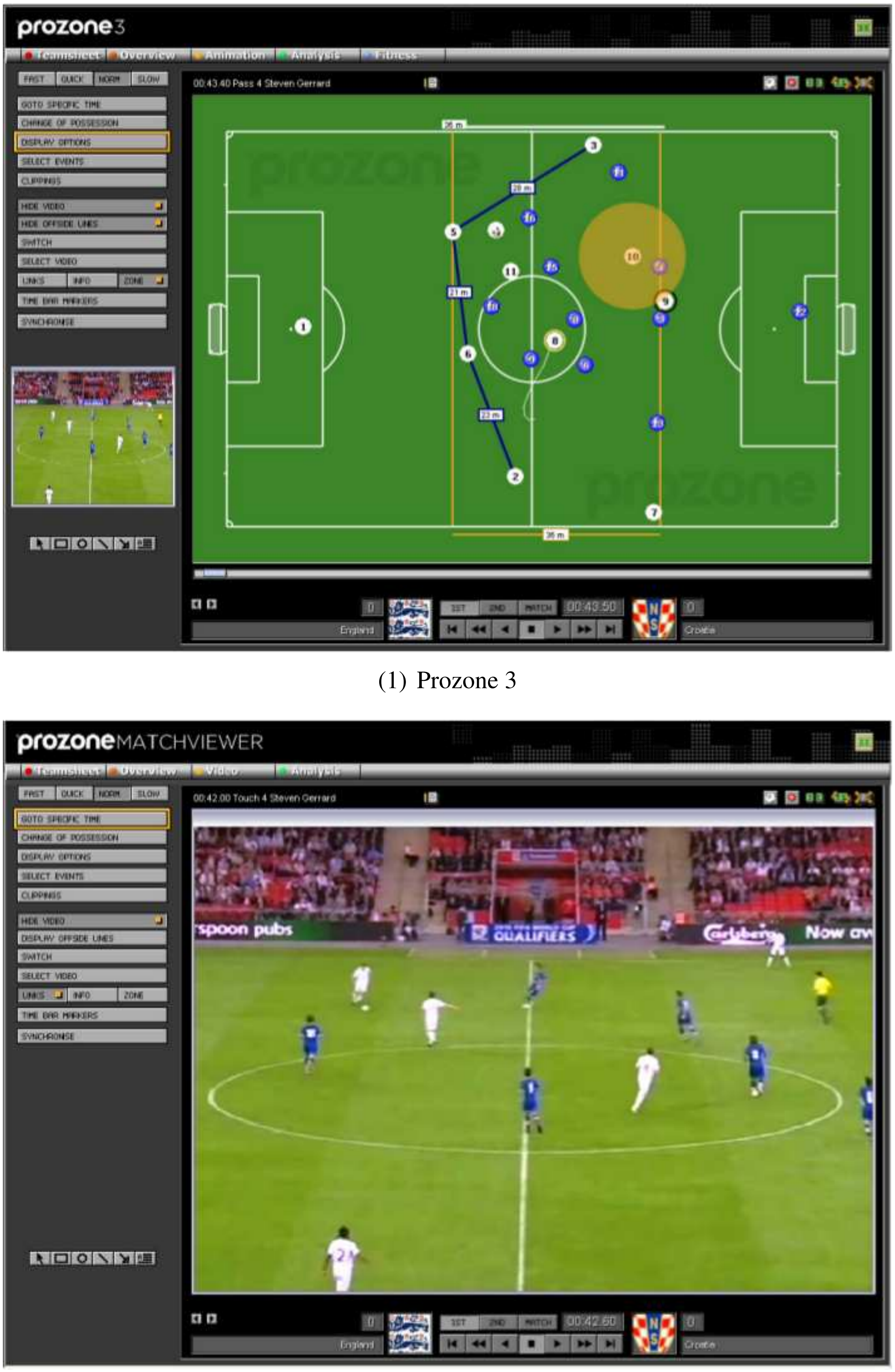
As a interesting side fact, it has been recently announced that *Sport Universal Process* and *Prozone Sports*, the two leading companies of sport analysis, made a strategic agreement in order to become a global industry leader in the area ([Spoc]).

#### 2.1.2.3 Cairos Technologies

*Cairos technologies AG* is a German company providing tools for the localisation of dynamic objects in 3D. Two products are offered, the first of them is for goal-line technology, and it is called *GLT* system. The goal-line technology is a technology that tries to find out if the ball has crossed the goal line, in order to avoid human error when awarding (or not) a possible goal.

[3] The *Amisco* acronym comes from the French: *Analyseur Modéliseur Informatique des Sports Collectifs*, which means *Computerized Analyser and Modeller of Team Sports*.

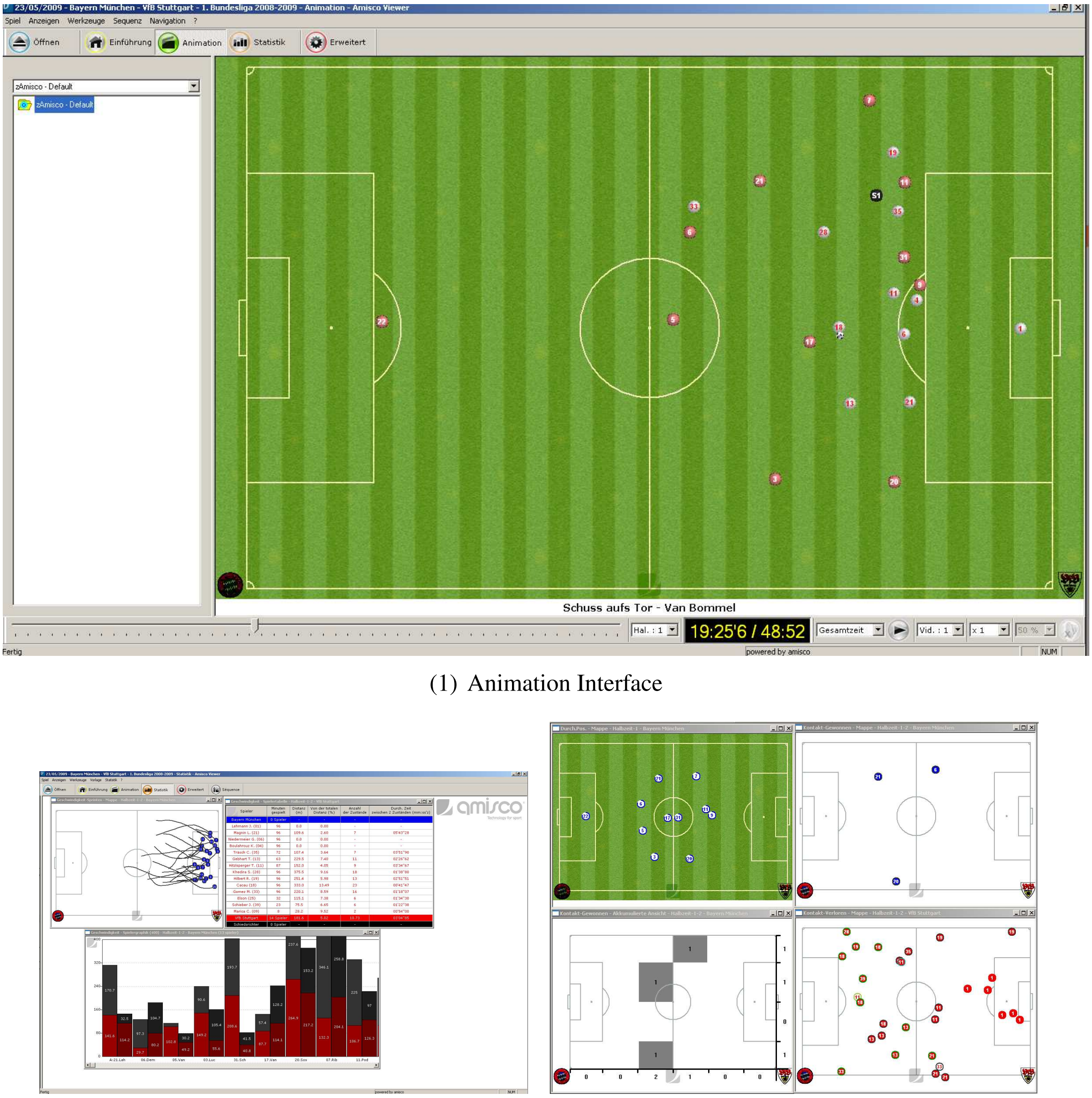
[4] The cost of the installation rounds the £100 000.



(1) Prozone 3

(2) MatchViewer

FIGURE 2.1 ProZone 3 *and* MatchViewer *are the football analysis tools provided by the* ProZone *company. They provide 2D animation and video-enhanced physical, technical and tactical performance insights, as well as access to ball actions and other events of the game.*


23/05/2009 - Bayern München - VfB Stuttgart - 1. Bundesliga 2008-2009 - Animation - Amisco Viewer
Spiel Anzeigen Werkzeuge Sequenz Navigation ?
Öffnen
Einführung
Animation
Statistik
Erweitert
zAmisco - Default
zAmisco - Default
Schuss aufs Tor - Van Bommel
Hal. : 1
19:25'6 / 48:52
Gesamtzeit
Vid. : 1
x 1
50 %
Fertig
powered by amisco
NUM
amisco

(1) Animation Interface

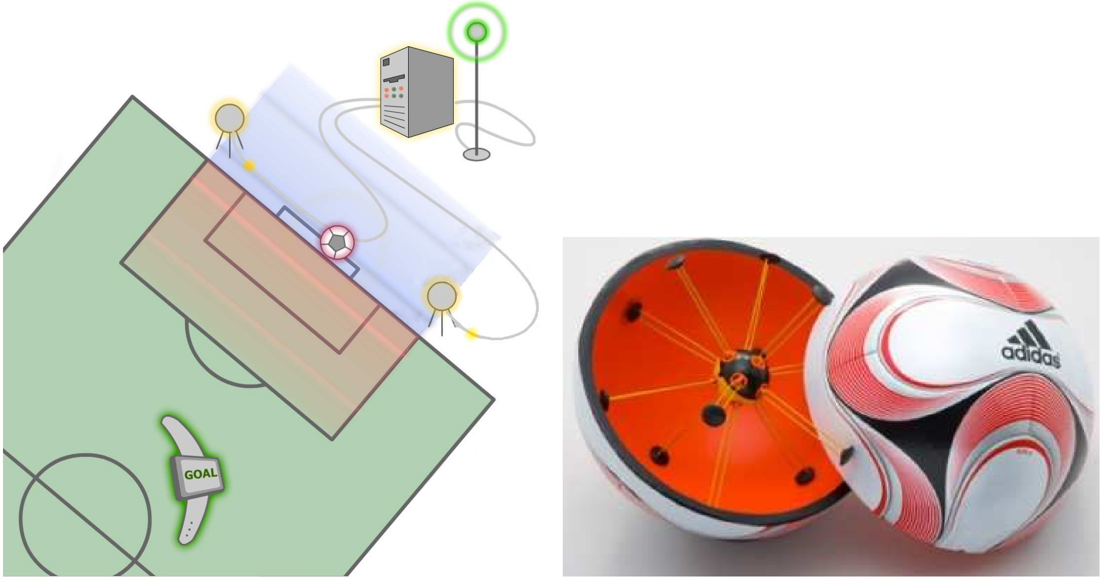
(2) Fitness Statistics

(3) Tactics Statistics

FIGURE 2.2 Amisco Pro *System. The animation interface of the system can be seen in subfigure (1). With it, a 2D animation of the game can be observed. The subfigure (2) shows basic statistics about fitness features of the teams as well as a summary of the action statistics for each half time. Other information presentations, such as the diagrams shown in subfigure (3), are also available. For example tactical line-ups, and ball contacts and possession recovering and corresponding pitch areas are shown in the graphs of subfigure (3). Much more information in different formats (tables, listings, graphs, diagrams, and others) can be obtained too.*

The second offered system is *VIS.Track*, a system for gathering the positional information of the players and the ball during the match [Sü05, Cai11, Impb, Aut].

The *GLT* system uses cables under the penalty area and behind the goal line, which produces a magnetic field detectable by a sensor imbued in the ball ([McK], see figure 2.3(2)), with the aim of detecting, by means of antennas also placed behind the goals, the goal-line-crossings of the ball. All this information is processed by a central computer that decides if the ball has crossed the goal line, or not, sending a signal to the referee. See figure 2.3(1) for a diagram of the system. The technology has been develop since 2005, and it has been tested in several competitions, but the officials of the *FIFA* rejected the incorporation of such technology in the game [McG, Wika]. After the *WC*2010, where several errors of referring were detected, the discussion about the use of the technology was open again [Dug, Mar].

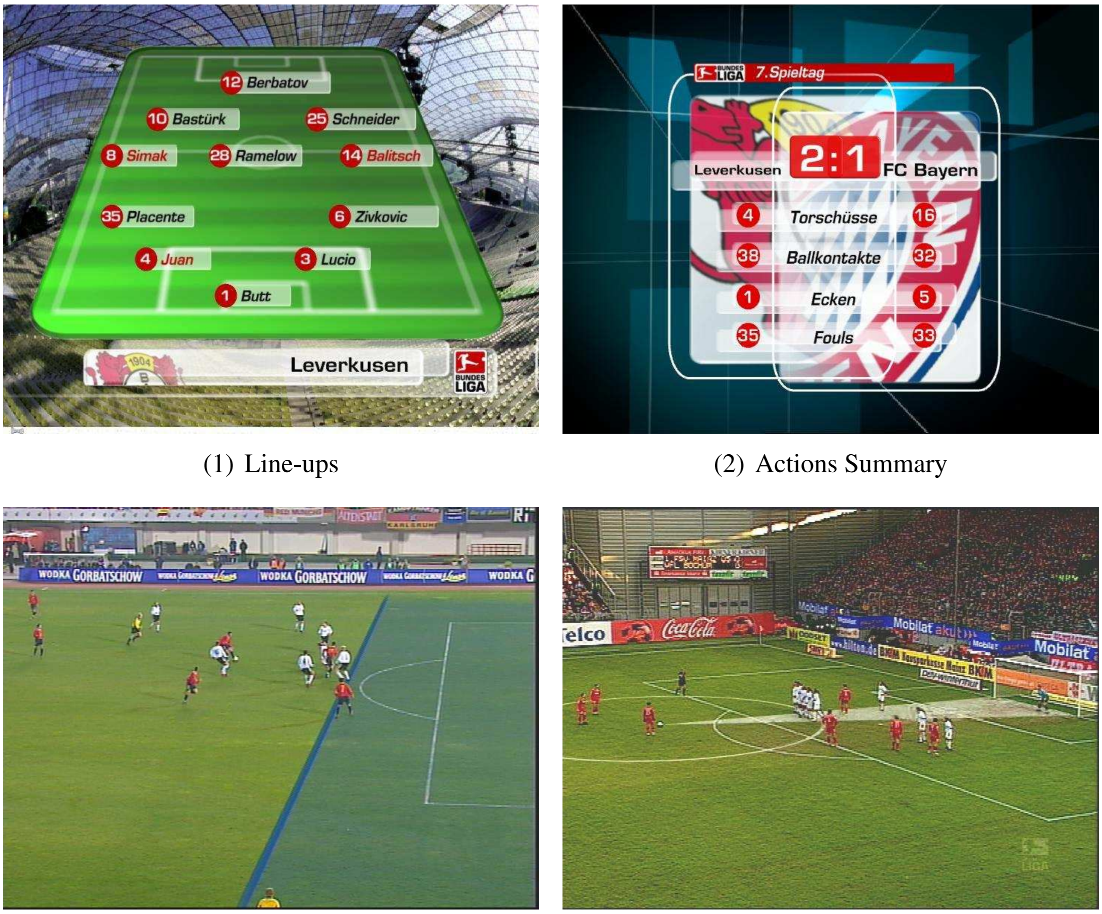


(1) *GLT* System

(2) *Adidas* Smartball

FIGURE 2.3 *The diagram of the* GLT *system is shown in subfigure (1). The red and blue zones in front and behind the penalty area are installed with cables that produces a magnetic filed, which is sensed by the sensor inside the ball, which position is measured by the antennas behind the goal. Finally a computer system has to decide if the ball has crossed or not the goal line. The proposed ball fabricated by* Adidas *is shown in the subfigure (2).*

In 2003, the German company *Impire* was absorbed by *Cairos Technologies*. *Impire* was created in 1988, and it was providing media services in Germany. Information about the game, such as statistics, air graphics and others services, were provided to TV stations, Internet sites, and football clubs (Bundesliga). Currently, they provide the *VIS.Track* system, which it is claimed to be able to track in real-time, the ball and the players during the match by using two *High Definition (HD)* video cameras. Similarly to *Prozone 3*, and *Amisco Pro*, it provides statistics about the game, the players, as well as 2D and 3D animations, but with the difference that it also includes the ball positional data. Examples of the services provided, as well as animations created using the system can be seen in the figure 2.5. The advantage is that the system can be mounted in approximately 30 minutes and it is portable, which allows for installation in away stadiums, as well as in the training places.



(1) Line-ups

(2) Actions Summary

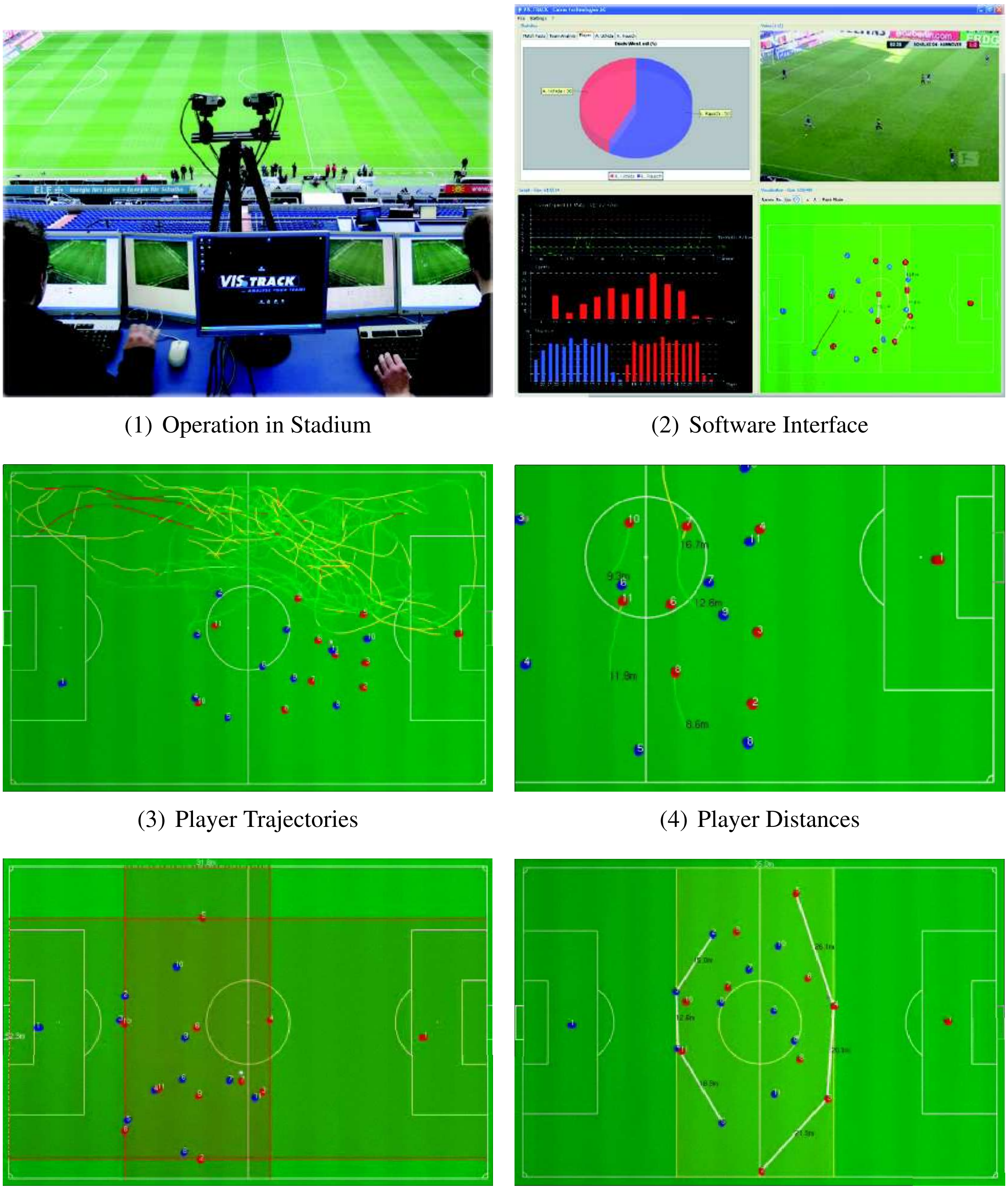
(3) Offside Overlay

(4) Shot Properties

FIGURE 2.4 *Impire offered services. Summaries of the teams, such as line-ups and counts of the actions were provided, as can be seen in the first row. Overlays on the TV signal were also possible, where for example, an offside regions is depicted in subfigure (3), as well as angle of shot and distance for a free-kick in subfigure (4).*

The system is having great success, since it has been recently announced that[5] "the *Deutsche Fußball Liga GmbH (DFL)* commissioned the company, to be the responsible for generating the game data from the first and second leagues (*1.Bundesliga* and *2.Bundesliga*) for the season 2011/2012, as well as to be the exclusive commercialisation company in the area of media and

[5] Taken from [DFL] and translated from the German.



(1) Operation in Stadium

(2) Software Interface

(3) Player Trajectories

(4) Player Distances

(5) Team Blocks

(6) Defence Lines

FIGURE 2.5 VIS.Track *System. The operation of the system showing the two* HD *cameras mounted on the side stands can be seen in subfigure (1). The software interface is shown in subfigure (2). It contains the tactics window, showing tactical behaviour of the team; running performance window, presenting player speeds, covered distances, speed sectors, and similar; statistics/actions window, offering several game actions (goals, corner kicks, shots, and others), linked to the video, as well as general overview statistics able to compare players too. A sample of the possible animations and available informations are shown in subfigure (3) to subfigure (6), where for example, the trajectories followed by the player, interesting distances of the players, compactness of the teams and finally defence lines linking the corresponding team-mates are visually depicted.*

sports betting. The award of the contract is conditioned to a successful technical performance test, currently being accomplished by the *Institut für Spielanalyse*, a *TUM*-founded company[6]" [Impa, DFL, Ins].

#### 2.1.2.4 LucentVision

*LucentVision* [PJC98, PJC99, PJOC00, POJ00, POJC01] is a system for analysing tennis. Animations are included as part of the system, presence maps (where a player was situated during the match), virtual replays, numerical statistics. Two cameras are used to cover each half of the court each, and with it track the players positions during the match. To accurately track the position of the ball, 6 cameras placed around the court are used. It is a real-time system that has been successfully used in international tennis tournaments, it has been also used in Formula One racing [Nat]. An example of its interface and output statistics is shown in figure 2.6.

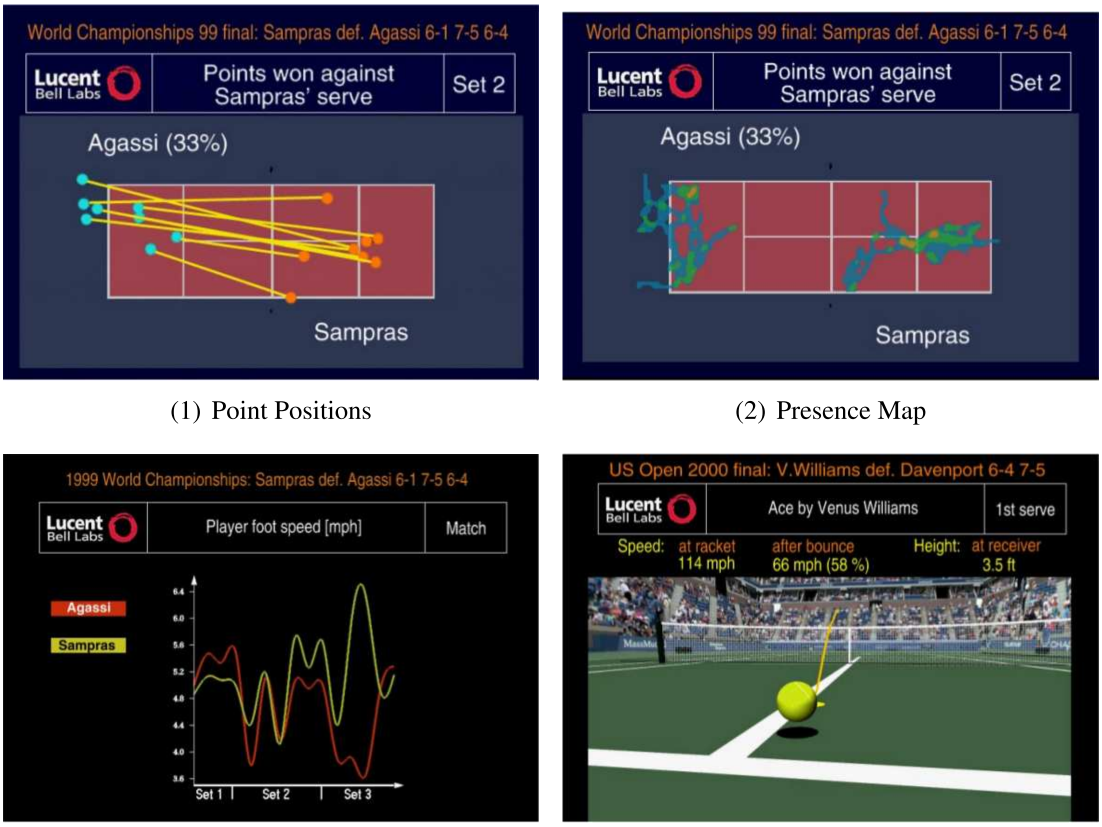
Tennis and football are very different sports, a fact that, besides being obvious, it is crucial for the design of a suitable tracking technology to track the interesting objects. For example, in tennis, only two players moving in restricted areas of the court each, are to be tracked. This is in contrast to the football situation, in which at least 20 players (not counting the goalkeepers nor the referees), in more complex situations such as partial or total occlusions, must be tracked. Also, since the rules are different, the events that can happen vary a lot. For the case of tennis, a particular play pattern is followed, e.g. sets, services, and the like. But for football, the game is more dynamic, and correspondingly, the play options open to its participants is larger.

#### 2.1.2.5 Trakus

The system developed by *Trakus* is called *Digital Sports Information*. The company is based in Medford, MA, USA, and was established in 1997 [Tra, iCo, BI06].

The system uses active transmitters (called tags) in order to send signals to be detected by antennas placed around the sport area. In this way, the position of the tags can be tracked for the duration of the sport. The tags must be carried by the athletes, or in the case of objects, such a puck or ball, they have to be embedded on them. The system has been tested in hockey, car racing, gridiron football, and horse racing. But lately, the company has concentrated its efforts to only cover the horse racing area, where the company have seen more economic opportunities. Similarly to other systems, it provides statistics of the athletes, as well as the opportunities

[6] Prof. Dr. Martin Lames being one of its mentors.


World Championships 99 final: Sampras def. Agassi 6-1 7-5 6-4
Lucent Bell Labs
Points won against Sampras' serve
Set 2
Agassi (33%)
Sampras
World Championships 99 final: Sampras def. Agassi 6-1 7-5 6-4
Lucent Bell Labs
Points won against Sampras' serve
Set 2
Agassi (33%)
Sampras
1999 World Championships: Sampras def. Agassi 6-1 7-5 6-4
Lucent Bell Labs
Player foot speed [mph]
Match
Agassi
Sampras
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
Set 1
Set 2
Set 3
US Open 2000 final: V.Williams def. Davenport 6-4 7-5
Lucent Bell Labs
Ace by Venus Williams
1st serve
Speed: at racket 114 mph
after bounce 66 mph (58 %)
Height: at receiver 3.5 ft

(1) Point Positions

(2) Presence Map

(3) Players' Speed

(4) Virtual Ball Animation

FIGURE 2.6 *Lucent Vision System. In the subfigure (1) and subfigure (2), the end positions of the players for points won by Agassi against Samprass in the World Championship 1999, and the corresponding presence maps are respectively shown. Fitness statistics such as players velocities can be seen in subfigure (3). Also it is possible to generate virtual replays as shown in subfigure (4).*

for virtual animations, such replays, or change of perspective, where game engines are used to generate the required graphics.

The system uses a proprietary wireless communication protocol to track the transmitters placed on the saddle-cloth of each horse. The used proprietary processing technique is called multi-lateration (U.S. Patent No. 6,204,813). In contrast to timing systems (like *Radio-frequency Identification (RFID)*), that can be used to determine the passing of a horse through a point, the *Digital Sports Information* system tracks the position of all the horses at a frequency of 30 measurements per second, during the whole event. The devices are very small, light and portable with sizes approximating that of a credit card and weight about 80 g. The system used between 6 and 10 antennas, that are located at convenient positions around the track surface, for example in the camera turrets, light poles, or stands.

The main disadvantage of this system, and similar ones, is their use of active devices that must be carried during the sport, because some of the official sport regulation associations does not allow the use of such devices. Particularly, in the case of football, the *FIFA* does not allow the use of transmitters of this kind. Another disadvantage is the required installation of the antennas, which makes the system not so portable. Also, a daily fee must be paid, to cover for equipment cost and labour. Finally, no available information exist about the accuracy and precision of the system, so no objective comparisons can be made.

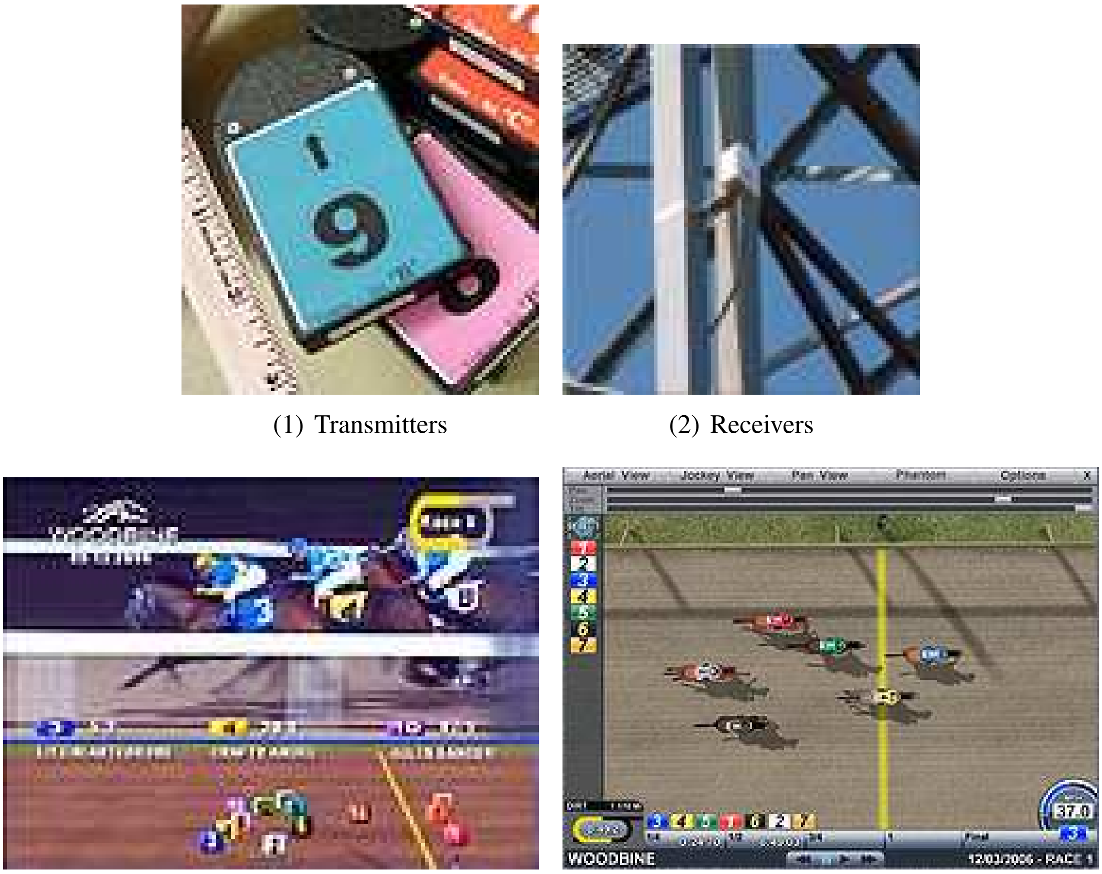
#### 2.1.2.6 Elite Sports Analysis

*Elite Sports Analysis* was established in 1996 in Fife, Scotland, it offers two main products called *Focus X2*, and *Focus X3*. *Focus X2* is a tool for notational based performance analysis. With this tool, it is possible to create a user-specified video index of those interesting events, and later quickly search for the desired events, in order to allow the user to run a more detailed examination of the selected material. The interface of the the *Focus X2* is shown in figure 2.8. *Focus X3* is another tool to show split-screens and multiple-views with support for drawing and measuring tools for those events indexed using *Focus X2* [Eli11, ana].

No tracking of the interesting objects is provided by this system, only tools to ease the manual annotation of interesting events during the game are possible.

#### 2.1.2.7 SportsCode

*SportsCode* is a video editing program (they called it video analysis) which allows the user to annotate the interesting video sequences, as well as storing them for later retrieval. Thus, the actual analysis is entirely left to the user. It was developed by *SportsTec* in Warriewood,



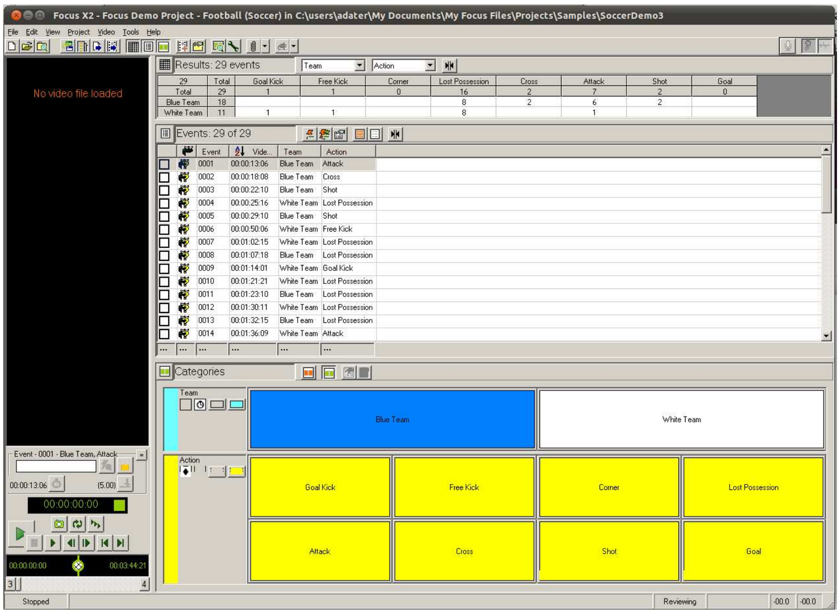
(1) Transmitters

(2) Receivers

(3) Virtual Animation 1

(4) Virtual Animation 2

FIGURE 2.7 *Trakus' System. The first row show the communication hardware used in the Trakus' System, the transmitters (tags) carried by the horses and the receivers (antennas) placed at a near-the-track tower. Finally the last row present a sample of the possible virtual animations that the system is capable of.*



Focus X2 - Focus Demo Project - Football (Soccer) in C:\users\adater\My Documents\My Focus Files\Projects\Samples\SoccerDemo3

Results: 29 events — Team — Action

| 29 | Total | Goal Kick | Free Kick | Corner | Lost Possession | Cross | Attack | Shot | Goal |
|---|---|---|---|---|---|---|---|---|---|
| Total | 29 | 1 | 1 | 0 | 16 | 2 | 7 | 2 | 0 |
| Blue Team | 18 | | | | 8 | 2 | 6 | 2 | |
| White Team | 11 | 1 | 1 | | 8 | | 1 | | |

Events: 29 of 29

| | | Event | Vide... | Team | Action |
|---|---|---|---|---|---|
| □ | | 0001 | 00:00:13:06 | Blue Team | Attack |
| □ | | 0002 | 00:00:18:08 | Blue Team | Cross |
| □ | | 0003 | 00:00:22:10 | Blue Team | Shot |
| □ | | 0004 | 00:00:25:16 | White Team | Lost Possession |
| □ | | 0005 | 00:00:29:10 | Blue Team | Shot |
| □ | | 0006 | 00:00:50:06 | White Team | Free Kick |
| □ | | 0007 | 00:01:02:15 | White Team | Lost Possession |
| □ | | 0008 | 00:01:07:18 | Blue Team | Lost Possession |
| □ | | 0009 | 00:01:14:01 | White Team | Goal Kick |
| □ | | 0010 | 00:01:21:21 | White Team | Lost Possession |
| □ | | 0011 | 00:01:23:10 | Blue Team | Lost Possession |
| □ | | 0012 | 00:01:30:11 | White Team | Lost Possession |
| □ | | 0013 | 00:01:32:15 | Blue Team | Lost Possession |
| □ | | 0014 | 00:01:36:09 | White Team | Attack |
| ... | ... | ... | ... | ... | ... |

FIGURE 2.8 Focus X2 *User Interface. The system allows for manual annotation and indexing of interesting events of a football match. No input video was selected yet.*

Australia. Disadvantages are that the system required still a lot of manual work from the user part, and that the system only works in Apple computers. The advantages are that the system is portable, since only requires the video input for doing the annotations; do not required the "invasive" analysis by external operators, as in *Prozone 3*, and *Amisco*; and finally the cost of the system, which is cheaper [Spo11b, Spod, Spoe]. For some images of the provided interface see figure 2.9.

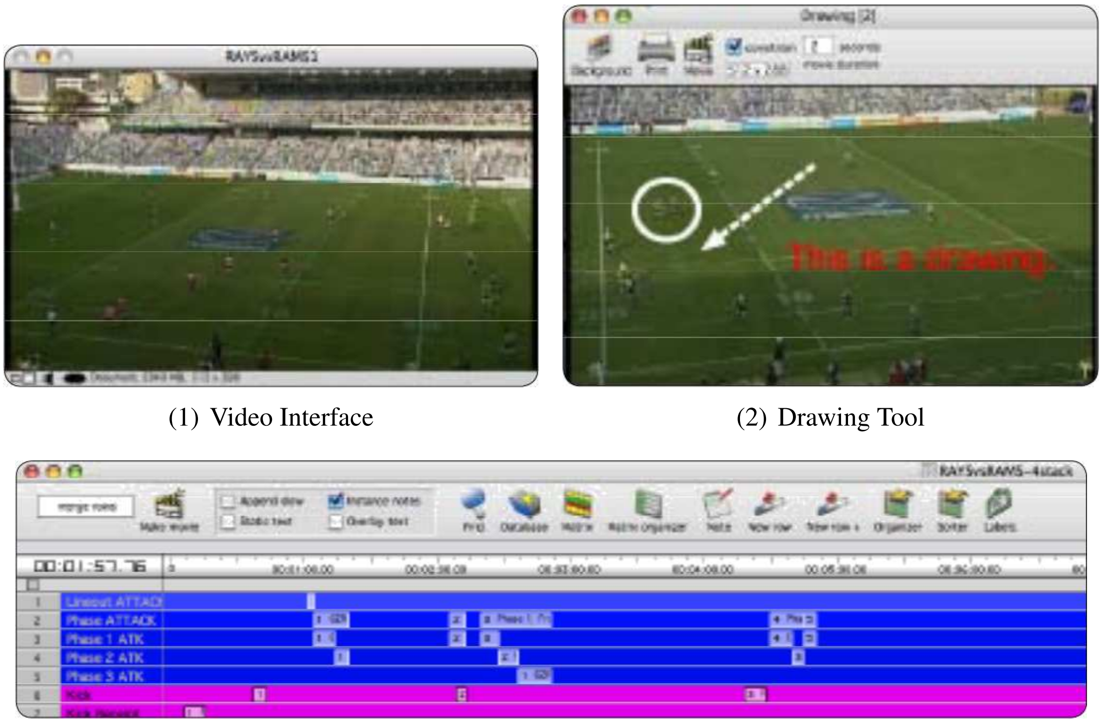
This system provides tools for manual video editing and indexing of interesting events occurring during the match. The advantage of the simpler video editing tools, such as *SportsCode* and the *Focus* tools is that are inexpensive and portable. The main problem is that all the analysis is done by the final user (coaches for example), since the tools only help in the manage of the video information.

### 2.1.2.8 Panini Digital

Formerly known as Digital Soccer Project (started in Modena, Italy) is an Italian company part of the *Panini Group*[7].

Technically is not an analysis system, since all the data about the game is collected manually by a trained company operator, but the statistics provided are very similar to those of the already mentioned systems. For example, tactical line ups, passes flows, ball possession

[7] The *Panini Group* is a very known sticker and trading card collections company.



(1) Video Interface

(2) Drawing Tool

(3) Time Line

FIGURE 2.9 SportsCode *Interface. The subfigure (1) shows the video interface of the* SportsCode *system, and the corresponding time line is shown in the subfigure (3). Finally, a drawing tool for creating video overlays is provided, as can be seen in subfigure (2).*

statistics, players performances, pitch coverage, and others, are provided [Pan11, Pan]. In the figure 2.10, examples of the pass flow and player statistics for the match between Brazil against France in the World Cup 2006 are shown.

#### 2.1.2.9 Opta

*Opta* is a sports data company founded in 1996 in London, England. The company generates data after performing an analysis of the sport, begin afterwards distributed to the TV broadcasting stations, Internet services, betting agencies, newspapers, brands, and others. The variety of sport is large, not only football, but to more of 30 different sports in more than 70 countries [Opt11, Thec].

To generate the data, the company assign 2 to 3 operators for each game. The operators select the corresponding event, where in the pitch it happened and which players were involved. All this generated information is stored in a main database, where some triggers generate new information for the customers consumption. A small example of the usage given to the provided data by some of the clients can be seen in figure 2.11 [Osh, Thea, Theb].

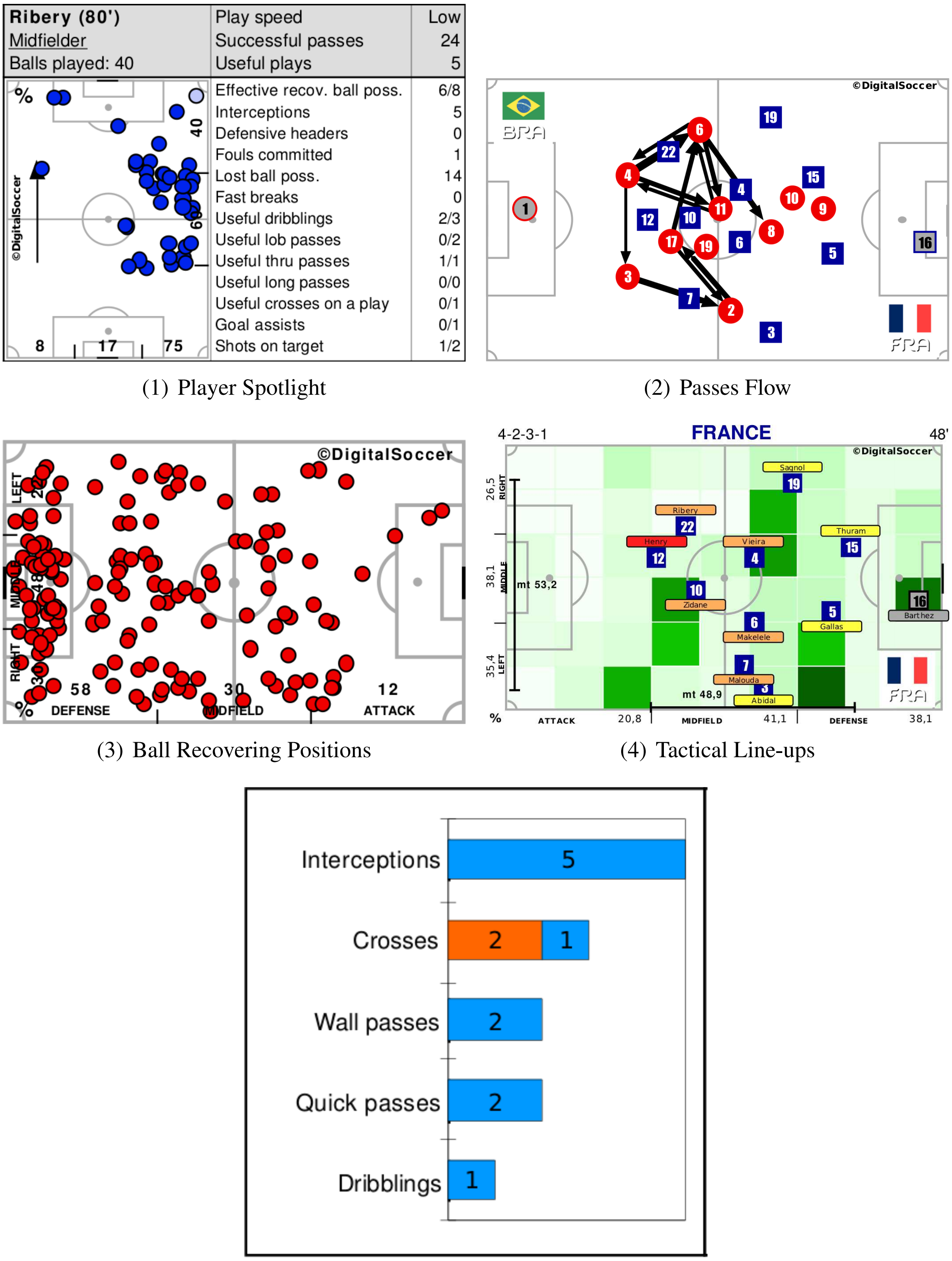
#### 2.1.2.10 Ascensio Systems

Ascensio Systems is a company founded in London, England. The main offered tool is the *Match Expert*, which is a tool for visualisation of the data gathered for a game, which can be analysed and operated by the user (instead as a service like other companies). Bi-dimensional and three-dimensional animations are possible, also statistics of the actions (passes, goals, average positions, speeds, accelerations, trajectories), and tables, and charts are available. The generation of the data is done semi-automatically by means of at least four operators that use four static cameras, in order to cover the pitch. Example of the interface can be seen in figure 2.12 [Asc11, Las].

#### 2.1.2.11 Tracab

*Tracab* started in 2003 in Stockholm, Sweden. Their principal aim is to deliver the positions of the players in a sports arena in real time. They claim to be "the only company in the world providing true 3D tracking in real-time" [TRA11, Heg, Saa].

The $x, y, z$ coordinates of each moving object, including the ball, is tracked by the system, with a latency of no more than 2 frames during the whole game. The architecture of the system consist of two multiple-camera units (8 single cameras), a desktop computer, and a laptop. The



| Ribery (80')<br>Midfielder<br>Balls played: 40 | Play speed<br>Successful passes<br>Useful plays | Low<br>24<br>5 |
|---|---|---|
| | Effective recov. ball poss. | 6/8 |
| | Interceptions | 5 |
| | Defensive headers | 0 |
| | Fouls committed | 1 |
| | Lost ball poss. | 14 |
| | Fast breaks | 0 |
| | Useful dribblings | 2/3 |
| | Useful lob passes | 0/2 |
| | Useful thru passes | 1/1 |
| | Useful long passes | 0/0 |
| | Useful crosses on a play | 0/1 |
| | Goal assists | 0/1 |
| | Shots on target | 1/2 |

(1) Player Spotlight

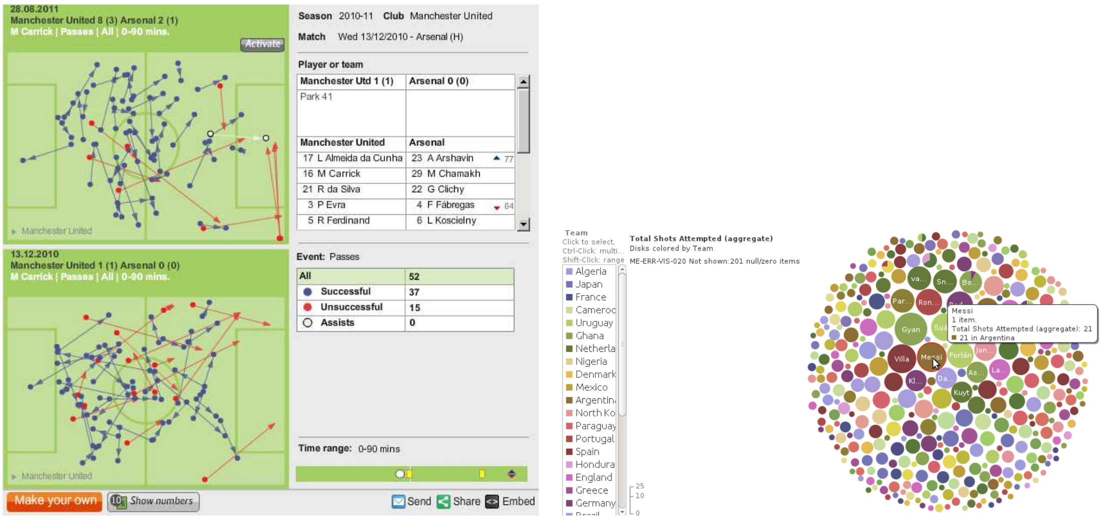
(2) Passes Flow

(3) Ball Recovering Positions

(4) Tactical Line-ups

(5) Shot Actions

FIGURE 2.10 Panini Digital *Example Output. In subfigure (1) the player performance statistics are shown: passes, shots, useful plays. Passing channels are described by the graph shown in subfigure (2), where the thickness of the arrows is proportional to the number of passes between the corresponding players. The positions on the pitch where the ball was recovered is shown in subfigure (3). Also tactical line ups are provided as depicted in subfigure (4) for the French team. Examples of the shots, and from which actions they originated are presented in subfigure (5).*



28.08.2011
Manchester United 8 (3) Arsenal 2 (1)
M Carrick | Passes | All | 0-90 mins.

13.12.2010
Manchester United 1 (1) Arsenal 0 (0)
M Carrick | Passes | All | 0-90 mins.

Season 2010-11 Club Manchester United

Match Wed 13/12/2010 - Arsenal (H)

Player or team

| Manchester Utd 1 (1) | Arsenal 0 (0) |
|---|---|
| Park 41 | |
| **Manchester United** | **Arsenal** |
| 17 L Almeida da Cunha | 23 A Arshavin ▲ 77 |
| 16 M Carrick | 29 M Chamakh |
| 21 R da Silva | 22 G Clichy |
| 3 P Evra | 4 F Fábregas ▼ 84 |
| 5 R Ferdinand | 6 L Koscielny |

Event: Passes

| All | 52 |
|---|---|
| ● Successful | 37 |
| ● Unsuccessful | 15 |
| ○ Assists | 0 |

Time range: 0-90 mins

Team
Click to select,
Ctrl-Click: multi...
Shift-Click: range

Total Shots Attempted (aggregate)
Disks colored by Team
ME-ERR-VIS-020 Not shown:201 null/zero items

Messi
1 item.
Total Shots Attempted (aggregate): 21
21 in Argentina

(1) User-created chalkboard

(2) Goals per player *WC*2010

FIGURE 2.11 Optasports *Data Visualization. Diagrams created by the users at* guardian.co.uk, *showing statistics of actions of the game are shown in the subfigure* (1). *In subfigure* (2) *statistics about the goals-per-player during* WC*2010 is shown.*

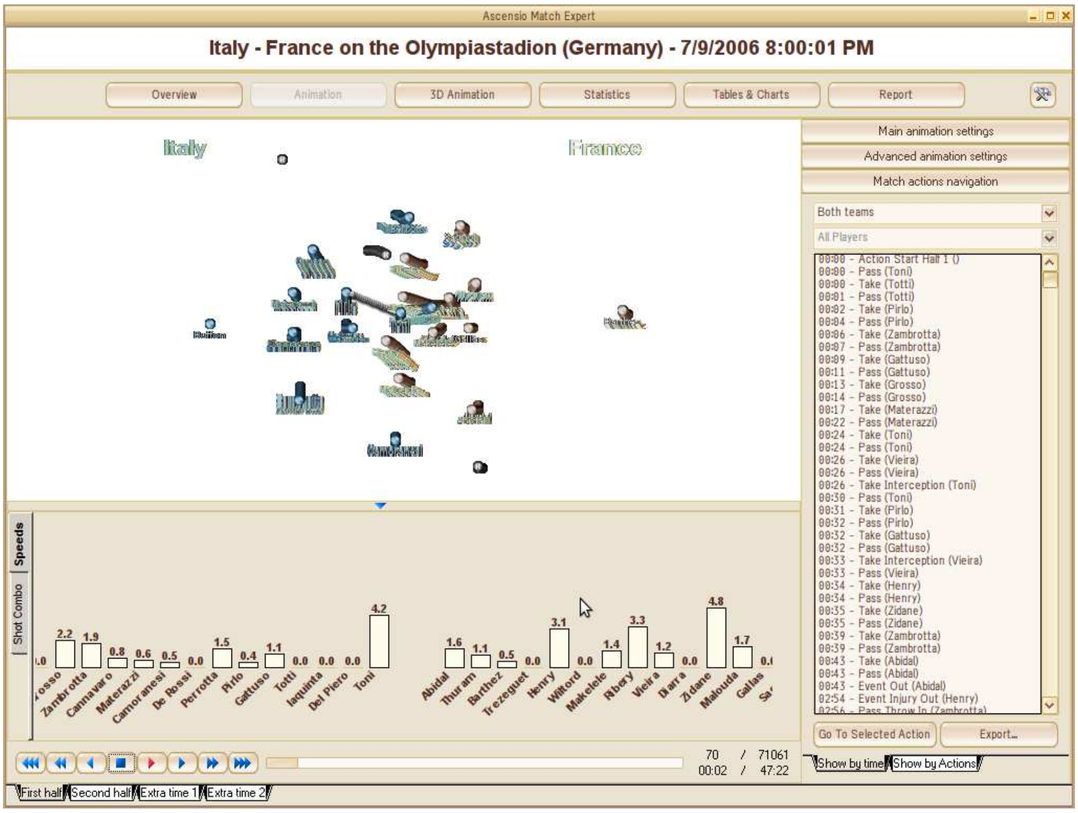
way they reach 3D tracking in real-time is using stereo tracking. The core algorithms were developed by SAAB (an international defence and security company). Animations are possible, the tracked data is fed into graphics engines that allows the creation of live 3D virtual game. Images showing the architecture, hardware and generated graphs is shown in figure 2.13.

### 2.1.2.12 Orad

*Orad* was founded in 1993, it provides 3D graphics for broadcasting, as well as access to a video database using a video server. Their system is called *TrackVision*, which is a tracking system that allows the placement of digital overlays on the field in the input video.

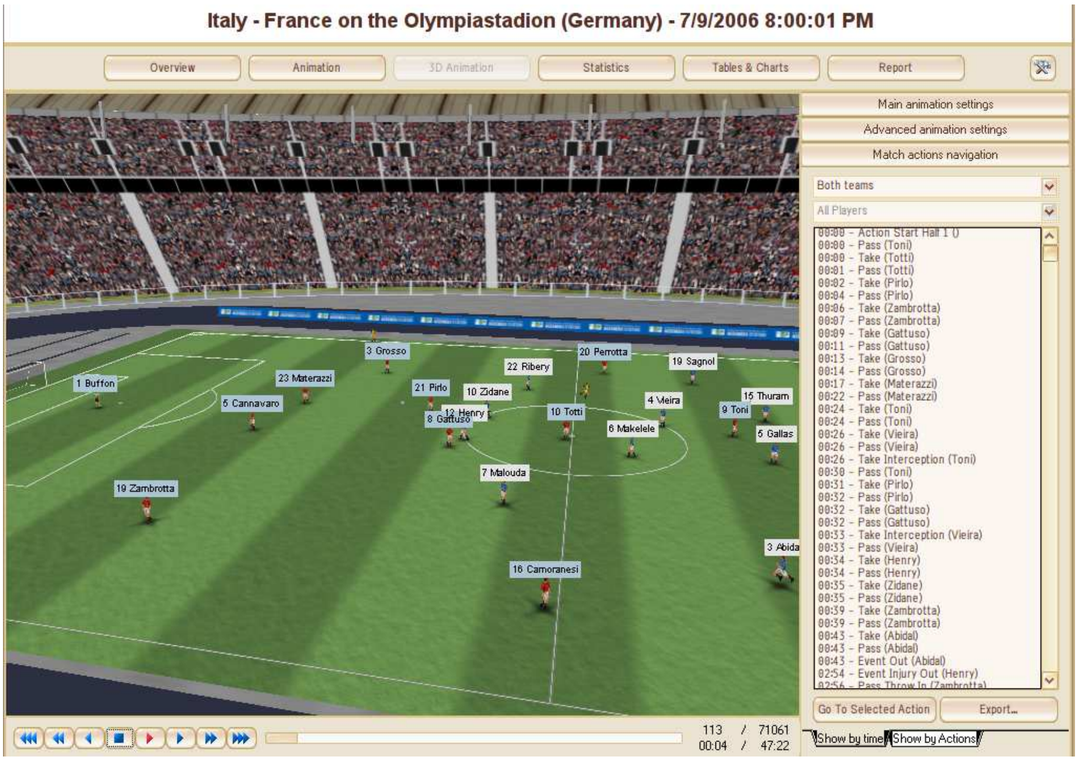
In live mode, only overlaid of distance to goal, 9m free-kick distance circle are available, while in replay mode, the offline is signalised as well as the speed of the ball. Also advertisement can be placed on the input video on both modes. They claim that their sensor-free tracking technology, "*TrackVision* provides perfect results regardless of venue, camera positions, and angles, and weather conditions". See figure 2.14.

Some of the existing products are very invasive, by requiring tags to be carried, a requisite sometimes forbidden by the corresponding official boards. There are products that also require the installation of fixed special equipment in the stadium, such as antennas, or special cameras, making the system stationary and therefore non-portable. Also, some products are very prohibitive with respect to cost (especially for not very wealthy clubs), due to the



```
00:00 - Action Start Half 1 ()
00:00 - Pass (Toni)
00:00 - Take (Totti)
00:01 - Pass (Totti)
00:02 - Take (Pirlo)
00:04 - Pass (Pirlo)
00:06 - Take (Zambrotta)
00:07 - Pass (Zambrotta)
00:09 - Take (Gattuso)
00:11 - Pass (Gattuso)
00:13 - Take (Grosso)
00:14 - Pass (Grosso)
00:17 - Take (Materazzi)
00:22 - Pass (Materazzi)
00:24 - Take (Toni)
00:24 - Pass (Toni)
00:26 - Take (Vieira)
00:26 - Pass (Vieira)
00:26 - Take Interception (Toni)
00:30 - Pass (Toni)
00:31 - Take (Pirlo)
00:32 - Pass (Pirlo)
00:32 - Take (Gattuso)
00:32 - Pass (Gattuso)
00:33 - Take Interception (Vieira)
00:33 - Pass (Vieira)
00:34 - Take (Henry)
00:34 - Pass (Henry)
00:35 - Take (Zidane)
00:35 - Pass (Zidane)
00:39 - Take (Zambrotta)
00:39 - Pass (Zambrotta)
00:43 - Take (Abidal)
00:43 - Pass (Abidal)
00:43 - Event Out (Abidal)
02:54 - Event Injury Out (Henry)
02:56 - Pass Throw In (Zambrotta)
```

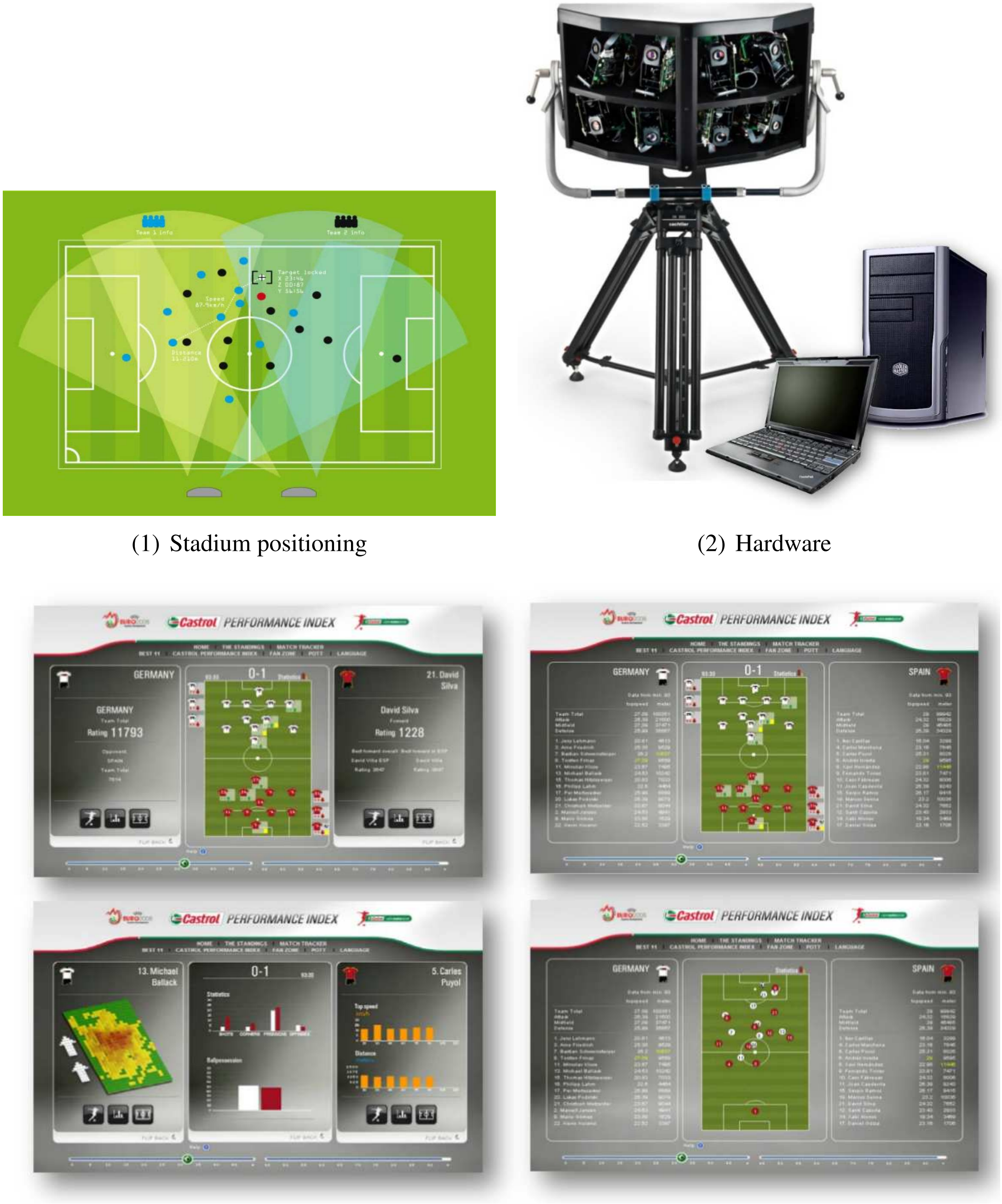
(1) 2D Animation



```
00:00 - Action Start Half 1 ()
00:00 - Pass (Toni)
00:00 - Take (Totti)
00:01 - Pass (Totti)
00:02 - Take (Pirlo)
00:04 - Pass (Pirlo)
00:06 - Take (Zambrotta)
00:07 - Pass (Zambrotta)
00:09 - Take (Gattuso)
00:11 - Pass (Gattuso)
00:13 - Take (Grosso)
00:14 - Pass (Grosso)
00:17 - Take (Materazzi)
00:22 - Pass (Materazzi)
00:24 - Take (Toni)
00:24 - Pass (Toni)
00:26 - Take (Vieira)
00:26 - Pass (Vieira)
00:26 - Take Interception (Toni)
00:30 - Pass (Toni)
00:31 - Take (Pirlo)
00:32 - Pass (Pirlo)
00:32 - Take (Gattuso)
00:32 - Pass (Gattuso)
00:33 - Take Interception (Vieira)
00:33 - Pass (Vieira)
00:34 - Take (Henry)
00:34 - Pass (Henry)
00:35 - Take (Zidane)
00:35 - Pass (Zidane)
00:39 - Take (Zambrotta)
00:39 - Pass (Zambrotta)
00:43 - Take (Abidal)
00:43 - Pass (Abidal)
00:43 - Event Out (Abidal)
02:54 - Event Injury Out (Henry)
02:56 - Pass Throw In (Zambrotta)
```

(2) 3D Animation

FIGURE 2.12 Ascensio Match Expert *interface. 2D and 3D animations of the final game of the* WC*2006.*

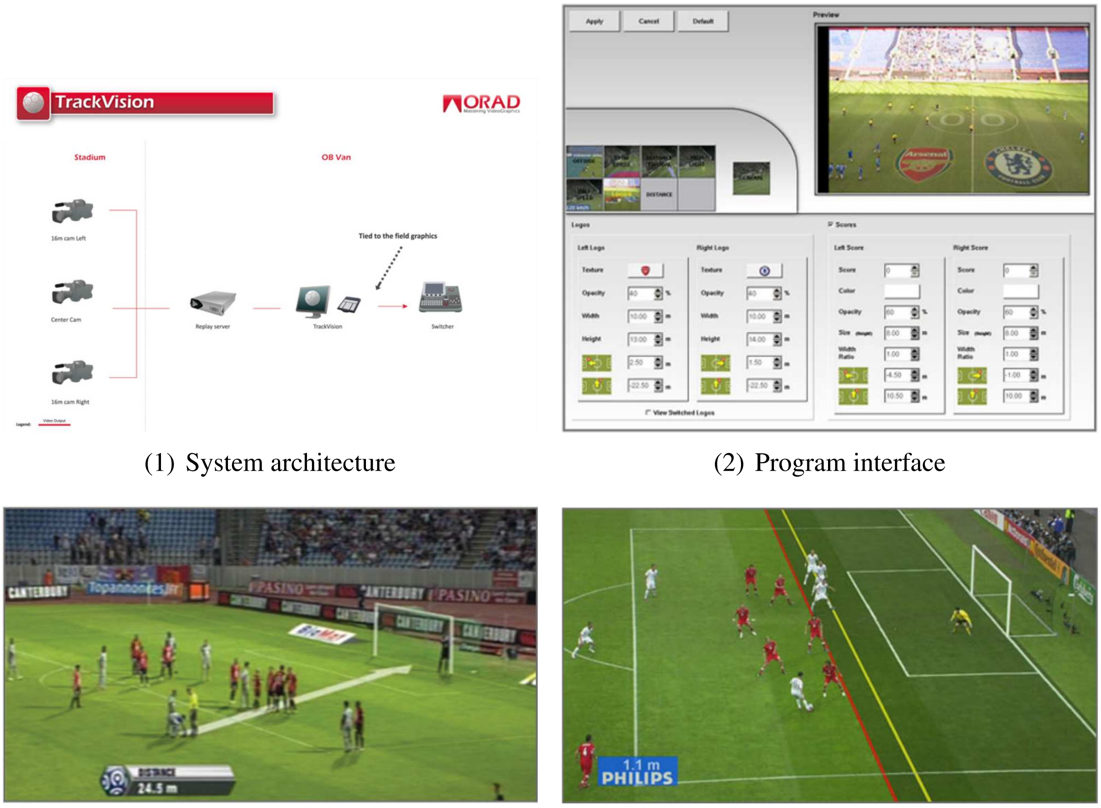


(1) Stadium positioning

(2) Hardware

(3) Generated diagrams, tables and charts

FIGURE 2.13 Tracab *System. In the first row, the required hardware and the disposition of the cameras around the stadium is shown. The subfigure (3) shows a sample of the generated diagrams for the brand* Castrol *for the* EC *2008.*



(1) System architecture

(2) Program interface

(3) Distance to goal

(4) Offside line

FIGURE 2.14 TrackVision *System. In the subfigure (1), a possible set up for the* TrackVision *System is depicted, where the system interacts with the replay server to generate offline overlays. The interface of the program to add the team logos and score overlays can be seen in subfigure (2). In the last row, examples of the possible digital overlays, such as distance to goal, offside lines, respectively.*

requirement of installing specialised cameras around in the stadium, and in general, due to the extra fees that have to be paid per game analysed. Even after paying this game fee, it is required to wait until the next day to get the results of the analysis. On the other hand, other systems provide only partial solutions, since for example, only video analysis tools are provided. Other systems require too much manual interaction from the user, making them very error-prone. Even, there are systems providing a bunch of first order statistics, such as distance ran by a player, velocities, accelerations, number of passes and number of shots, that might not be necessarily very informative from the tactical point of view. Also some systems are not capable of gathering information about the opposing team, feature which might be decisive to create counter-measurements to dealing with them. Finally, those systems that claim to have complete operational and automatic systems use algorithms that are proprietary, and therefore are not available for objective comparisons and evaluations.

## 2.2 Proposed Solution

The current work is part of a broader project called ASPOGAMO[8] (see figure 2.15), which involves the creation of a comprehensive model for the automated analysis of sport games. The project has been carried on as a collaborative project between computer science and sport science, as well as with cooperation with sport teams and trainers. The principal aim of the project is the generation of descriptive models for the games, that serves as the framework for an automatic analysis of different aspects of the games [BvHHK$^{+}$09]. The three main objectives of the project are, the investigation of the required computational mechanisms to recognise intentional activities, the development of a computational system for the interpretation and analysis of the games, and the demonstration of the impact of the project in several application areas: sport science, football training and coaching, and sports entertainment.

As stated in the introduction, the main objective of the current work is the implementation of a software system capable of generate, in an automated way, a semantic annotation of football games. This semantic annotation is based on positional data of the objects of interest (ball, players, and referees), and it allows the user to generate semantic analyses for a particular input game. In order to describe our proposed solution, a diagram showing the logical components of the system is presented in figure 2.16. As can be seen in the figure, the *input* module consist of the sensor(s) generating the video frames required for the further processing. The signal can be provided by a broadcast camera, or from other sources, such as static cameras, dynamic cameras, and several other camera configurations.

[8] Partly funded by the *Deutsche Forschungsgemeinschaft (DFG)* [DFG].

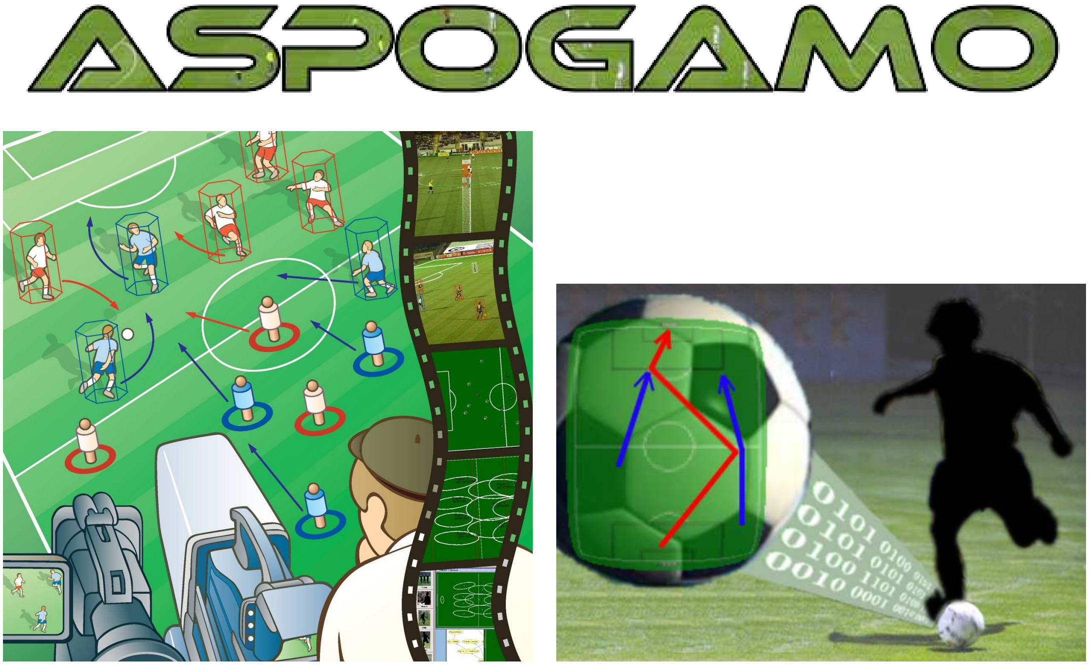


FIGURE 2.15 *Illustration of the* ASPOGAMO *project.*

The video frame sequences are to be processed in order to select only those scenes from the input video, that correspond to the shots where the positional data gathering can be conducted. The process of selecting the appropriate scenes is performed by the *temporal segmentation* module, and it consists of two steps. The first step is the detection of the *shot* boundaries (*shot boundary detection* sub-module), which finds where in the video stream each different shot have occurred. The found shots are fed into the *shot classification* sub-module, which simply selects those shots corresponding to far-view scenes. Other shot types, such as near-view or middle-view, does not show the players in a way that their positions can be accurately estimated, and therefore must be discarded. This module will be treated in the chapter 3.

The third module, the *spatial segmentation* module, is in charge of the generation of the detection and localisation of the objects of interest (or targets), that is the ball, the players, and the referees. A set of candidate regions in each image must be generated for the targets by the corresponding *target detection* sub-module, which are then input to the *target location* sub-module, for the localisation of individual targets inside the detected candidate regions. The output of the *spatial segmentation* module is the location of the targets in the current image. The chapter 4 expounds the challenges that have to be faced, as well as the details of the implementation of our proposed solution.

The locations of the targets, that is the observations of the measurement system, are fed to the fourth module: the *tracking module*. The *target tracking* sub-module is in charge of

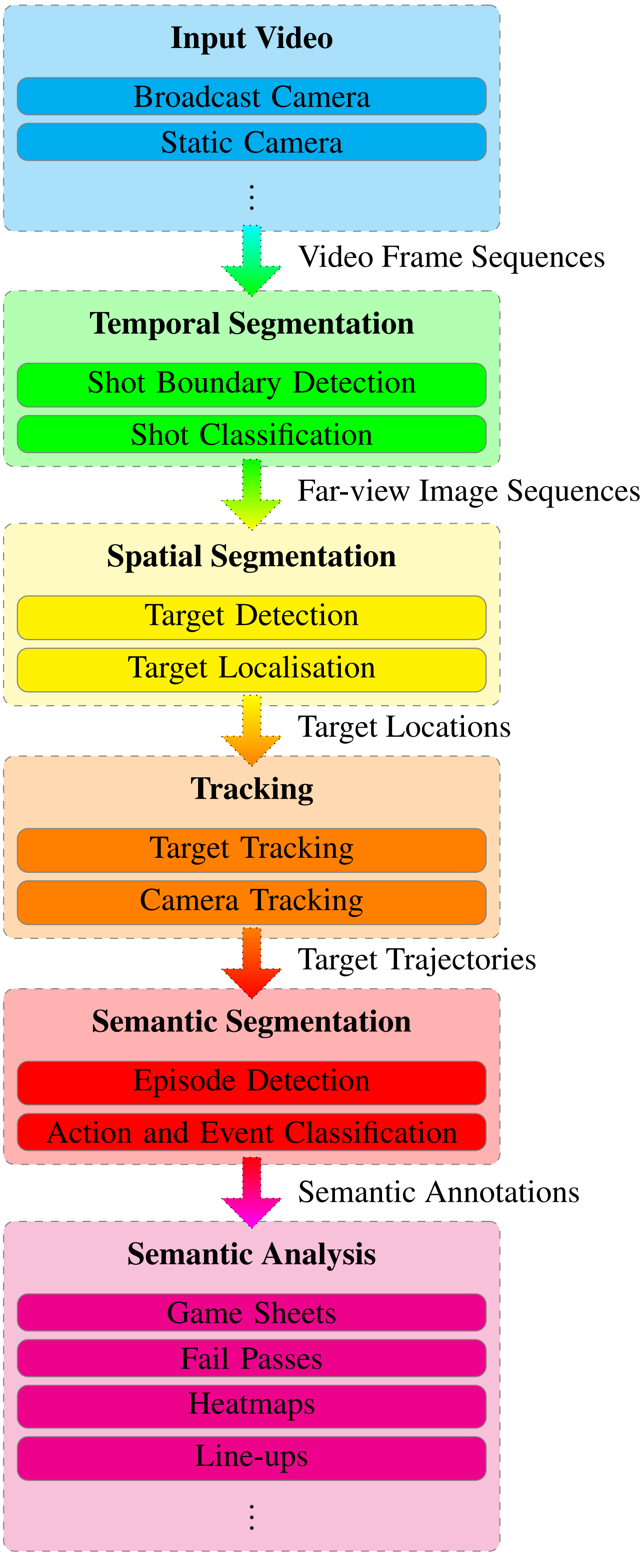


FIGURE 2.16 *Logical components of the semantic annotation system proposed in this thesis. Each block represents a particular processing unit, for which, internal sub-modules tackle smaller problems with the aim of solving the main problem of the block. The expected output of each block is given in the right side of the transitions between the blocks.*

keeping the state vector of the targets, which correspond to the coordinates on the field of play of the targets. The *camera tracking* sub-module is in charge of the continuous estimation of the camera parameters, including the intrinsic and extrinsic parameters, that will allow the estimation of the positions of the targets with respect to the field coordinate system, and therefore permitting the later semantic analysis of this positional data. The *tracking* module strongly depends on the *spatial segmentation* module, since the quality of the localisation results will directly affect the performance of the resulting tracking.

From a high-level point of view, the first four modules deal with the perception in our system, while letting the last two modules to deal with the cognitive part of the system. The object trajectories are used to generate a first semantic categorisation of the events occurred during the game into more explicit player actions, and ball events, such as goals, passes, shots, ball out-of-bounds, and others. This step, of course, is vital for a latter semantic interpretation of the game, as well as for the forthcoming tactical analysis, since it allows to semantically index the game, such that if required, a specific episode can be easily found. An example of such search could be the list of all attacking episodes of a team, that have started in the middle field, and are a result of a 'stealing the ball' event. The two steps involved in the semantic segmentation are the *episode detection* and the *action and event classification*. An *episode* is a sequence of related *events*, that in the case of football, can take the form of a group of consecutive team intentional activities, for example, the sequence: pass, dribble, pass, dribble, and shot. The *episode detection* sub-module, based on ball contacts, find the boundaries of the episodes, which are latter passed to the *event classification* sub-module in order to classify the individual happenings occurring during the given episode, into player actions or game interruptions. This module will be presented in full detail in the chapter 5.

Finally, the last module: *semantic analysis*, based on the game interpretation abstract model, generates informative pieces of information, that can latter be used to interpret, and analyse a game. Examples of such mechanisms are the automatic generation of *games sheets*, with which, the highlights of the game can be easily selected, and that gives in a glimpse a quick overview of a whole game in very few diagrams. Also basic statistical information can be also provided, such as line-ups of the players on the field, heat-maps, representing the paths and relative frequency followed by the players on the pitch. The passing channels used by the players of a team, which describes associations between players, and between player and particular pitch regions. Another interesting example is the analysis of fail passes, which are defined as the failed attempts of players to pass the ball to a team-mate, resulting in a possession gave away. Of particular interest are the *disaster passes*, which are fail passes that lead the opposing team to score a goal. The reasons for the failures can also be delimited by the tools provided

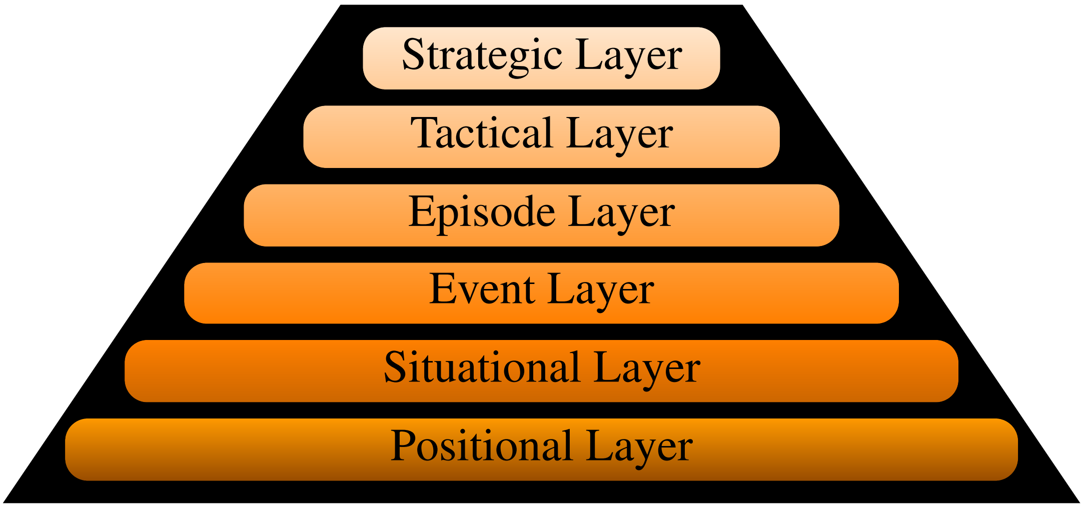


FIGURE 2.17 ASPOGAMO *Abstract Model. Each layer of the abstract model codifies a specific part of the game, allowing to build a hierarchical richer representation of a game, for a later interpretation and analysis of it.*

by this module. The *semantic analysis* module will be covered also in the chapter 5.

The abstract model used to represent the games is depicted in figure 5.8, from this figure it can be seen that there exist a hierarchy of layers, each encoding and describing a particular dimension of the game. The lowest layer is the **Positional Layer**, which constitutes a representation of the kinematic information of the objects of interest during the game. In this layer, the positions, velocities and accelerations of each of the players, ball and referees are stored, such that it is always possible to query the system to obtain quickly such information, and perform more complex calculations in a simpler manner. Also it is possible to obtain positional interpolations for the objects of interest for times between the start and end of a particular scene of interest. The use of positional information as the base building brick of our system respond to a need for acquiring a game as easy and fast as possible, without sacrificing the expressive power reachable by the representation. This is the fundamental layer, since on top of it, the rest of the model is build, therefore greater effort is dedicated to extract from the digital video, the information concerning the positions of the objects of interest.

The next level is the **Situational Layer**, which includes information relative to specific situations of interest during the game. This new layer is the result of the combination between information from the level below it and from a higher-level modelling of the game situations, obtained from the sport science point of view. The situations comprise information about technical as well as tactical actions, which aid in the understanding of the game. For example, the *ball line*, dividing the filed into two parts: behind and ahead of the ball position, the players in *preventive covering* and *counter-attack reference* during offensive and defensive episodes, *build-up rhombuses* for preparing passes, distances to goal, angles to goal, distances to ball, and others. With this information encoded, it is easy to provide information to the upper levels in order to generate more complex representations of the analysed situations.

The next level is the **Event Layer**, which incorporates the logical actions available to the players, such as passes, shots, clearances, tackles, dribbles, and others. Also the logical events of the game are included in this layer: kick-off, throw-in, free kicks, among others. For example, a pass in this layer is defined as a player kicking the ball with the intention to give it to a team-mate[9].

A logical continuation of the event layer is the **Episode Layer**, which combines actions sequences in logical order as to express the becoming of a game in a time lap. For example, a counter-attack of a team can be described as a sequence of actions like: interception-pass-dribble-pass-pass-dribble-centre-shot, which might culminate into a goal. For reaching such a description, not only the actions are to be recognised individually, but the boundaries of the logical sequences are to be detected. Basically two types of episodes were created to describe the game, which are the offensive episodes, including build-up play, final touches, shooting; as well as the defensive episodes, including forcing and intercepting. Another kind of episode are the transition episodes, which are instantiated by defensive restarts, offensive restarts or the transition between attacks and defends stages.

The first of the higher levels of description of the model is the **Tactical Layer**, which aims to describe the tactical work performed by the players, and comprise the individual actions of the actors. For example, during passing, the wanted action can be a depth pass, a meeting type of pass, an encompassing movement, or the like, which deals more with the global tactical work, that to specific isolated actions. Another example is during defending, where anticipation, cone, or T-shape covering are desired.

Finally, the **Strategic Layer** spans over styles of play of the teams: more defensive or offensive, possession football or direct football. Based on this layer, different squads can be classified, and recognised.

[9] Since sometimes, the intention of a player is not easily extracted from the performed curse of action of the player, then it is assumed that if a ball reaches a team-mate coming from another member of the team, the intention was that of a pass.

# Chapter 3

# Temporal Segmentation

> In theory, there is no difference between theory and practice. But in practice, there is.
>
> *(Lawrence Peter Berra)*

The principal aim of this chapter is to show how to obtain a set of frame sequences out of the original input video on which useful semantic annotations would be possible. The original input video comes from a typical TV broadcast source, and as such, it is full of editing elements to try to enhance the spectators experience. Therefore, the analysis of the different editing elements present in the input video, and how they affect our system is of primal importance. By detecting such editing elements, it is possible to split the original video into coherent semantic units of action. Later on, such video units are to be classified in order to select only those frame sequences that will allow the semantic annotations desired.

In the following, the important TV broadcast editing elements will be further described and its relevance to our study will be pointed out. The related work found in the professional literature to deal with such effects will be presented. Finally, our proposed method to accomplish with the frames' selection will be presented and explained in detail.

## 3.1 Broadcast Editing Elements

In an attempt of delivering to the viewers, a better insight into the game, and a more comprehensive convey of the feelings of its actors, and with it, to enhance the viewers' experience, certain video editing elements are used in the football TV broadcasting. Among such elements it is possible to find concatenation of shots from cameras located at different positions and angles in the stadium, effects during shot transitions, and overlays of digital objects showing some game information. In general, a *shot*[1] is a consecutive sequence of frames that constitutes a unit of action, generally taken with a single camera [YWX+07]. From a film-making perspective a shot means a video segment taken in a single rolling of the camera without interruption, and from a film-editing perspective, a shot means the video segment between two consecutive editing effects. Since our current interest is in video editing, then the latter definition will be used.

Most of the editing elements are not only unimportant for the semantic annotation that concern us, but also present some difficulties in its pursuance. Therefore, a proper dealing with such effects, either by ignoring the involved frames, or by removing the effects from the affected frames, is a required first step to reach our goal. A useful taxonomy for the treatment of the editing effects is depicted in the figure 3.1.

*Transitional effects* are used to divide shots in a video, and depending on the time span used by the transitional effect, it can be further classified as abrupt or gradual. An *abrupt transitional effect* (or simply cut) is basically the concatenation of two different shots, where normally, visual continuity is broken. From the filming perspective, the cuts are used to change the perspective of the portrayed shot in a way that makes the viewer suddenly see the shot from a different perspective (place and angle). In live transmissions, the cuts can be generated by suddenly switching to another camera, or more generally, to another video input source. Illustrations of such effect can be seen in figure 3.2.

A *gradual transitional effect* has a larger time span compared to an abrupt transition, and can be subdivided into dissolves and wipes, depending on the spatial intermixing of the shots. A *dissolve effect* gradually overlaps two shots for the duration of the effect. It it used to soften up the abrupt cuts that may surprise the viewer. In live video production, the effect is obtained by interpolating the voltages of the two video signals. See figure 3.3 and figure 3.4 for examples of dissolves. A further classification into *short dissolves*, such as that depicted in figure 3.3, and *large dissolves*, such as those in figure 3.4 can be made.

In a *wipe effect*, a new shot is gradually occupying the space of the old shot, while this old shot is correspondingly being erased. The adjacent shots are not temporally separated but spatially well separated at any time. Sometimes also, an object is used in the transitional space of a wipe effect, such as a logo. For examples see figure 3.5 and figure 3.6.

In the case of *overlays*, they can be classified as static and dynamic overlays, depending on how many frames are affected by the effect. The *static overlays* stayed in the video longer (generally during the length of the whole video) than their dynamic counterparts, and they

[1] The term derives from the analogy between the old hand-cranked cameras and machine guns, both from the early days of film production: one would *shot* film the same way one would *shoot* bullets.

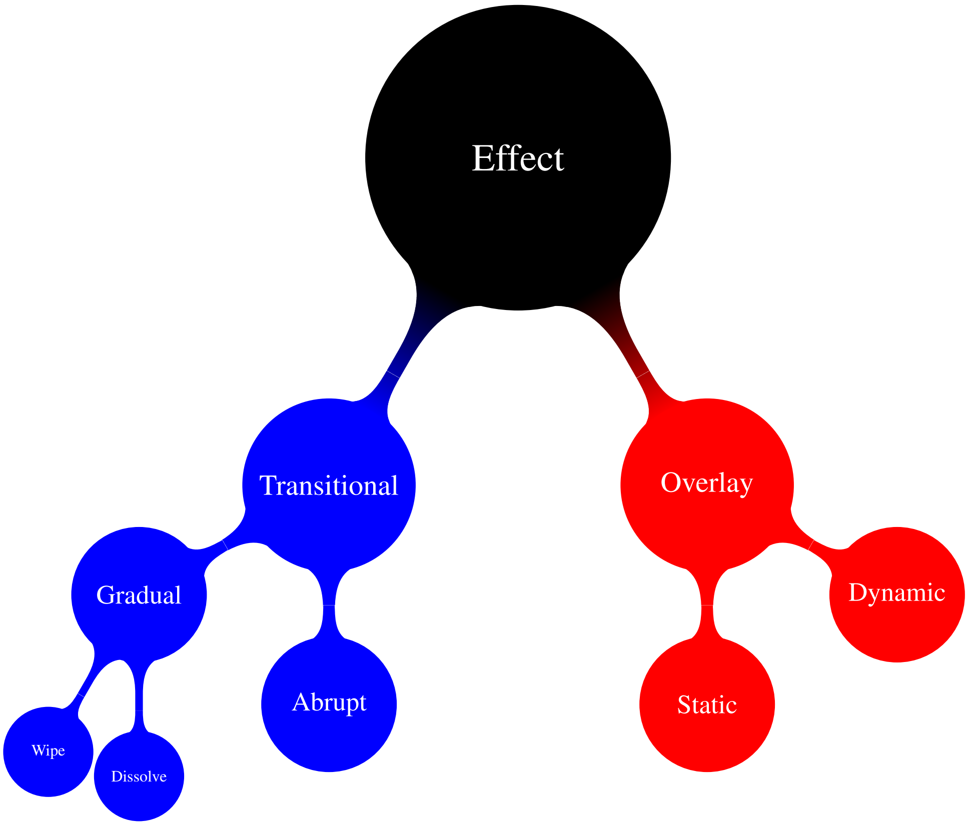
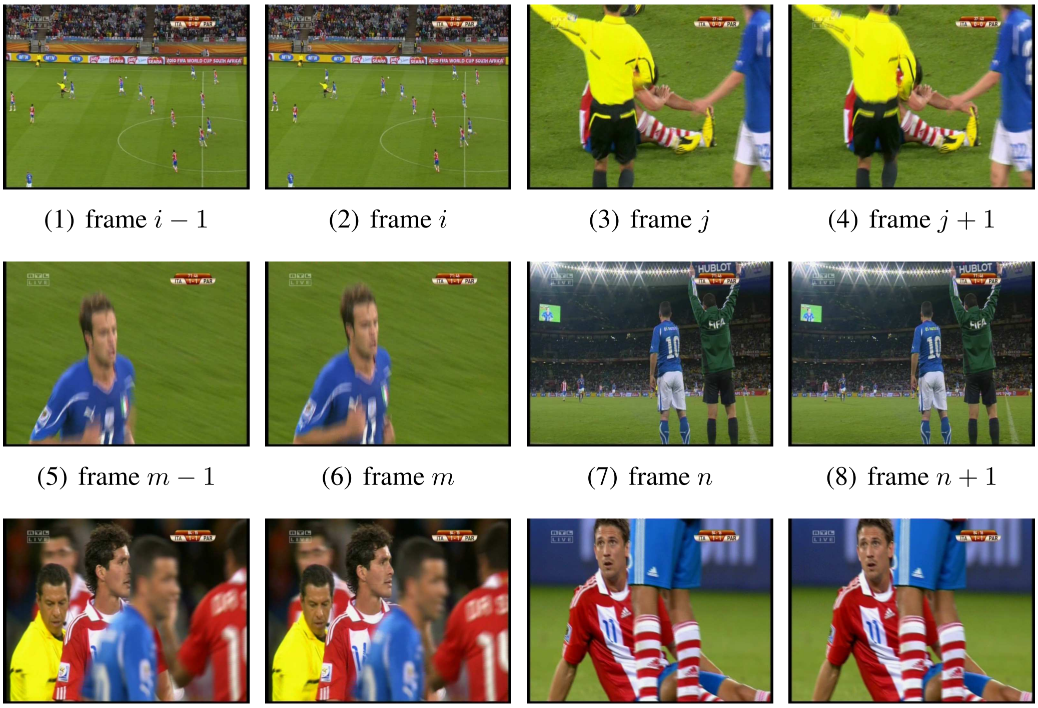


FIGURE 3.1 *Taxonomy of the TV broadcasting editing effects used in football. The transitional effects are used to split shots, and the overlay effects to provide information about the game.*

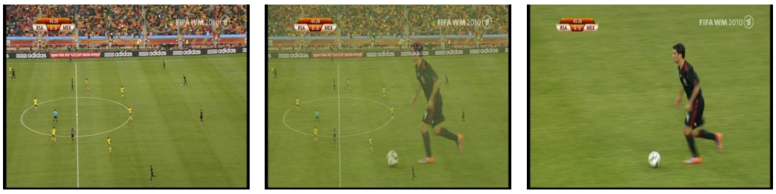


(1) frame $i-1$ (2) frame $i$ (3) frame $j$ (4) frame $j+1$

(5) frame $m-1$ (6) frame $m$ (7) frame $n$ (8) frame $n+1$

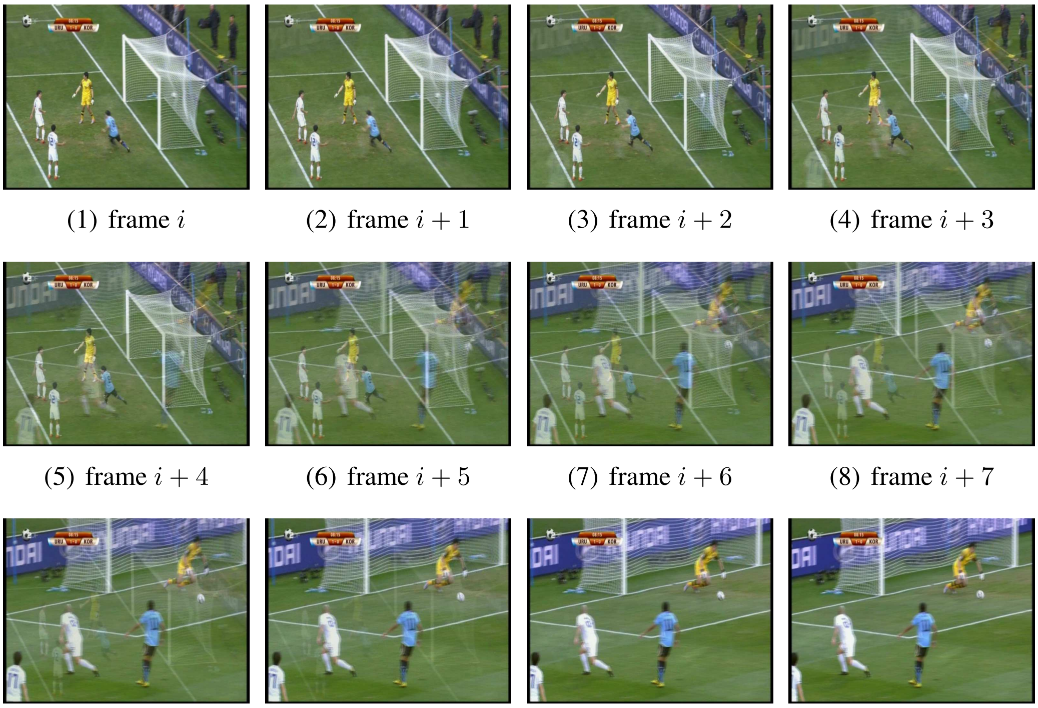
(9) frame $p-1$ (10) frame $p$ (11) frame $q$ (12) frame $q+1$

FIGURE 3.2 *Cuts examples from game #11 of WC 2010: Italy vs Paraguay. Each row corresponds to a different sequence where a cut effect is shown. The exact positions of the cuts can be observed in the transitions from frames $i \rightarrow j$, $m \rightarrow n$ and $p \rightarrow q$ respectively. The two shots involved in a cut are spatially and temporally well separated.*



(1) frame $i-1$ (2) frame $i$ (3) frame $i+1$

FIGURE 3.3 *Dissolve example from game # 1 of WC 2010: Southafrica vs Mexico. The frame sequence shows two shots combined in a dissolve effect. Both shots are spatially and temporally intermixed. This short dissolve uses only 1 frame (frame $i$) to switch between shots, it is much shorter than the one showed in figure 3.4. Also on the top corners, static overlays can be noticed.*

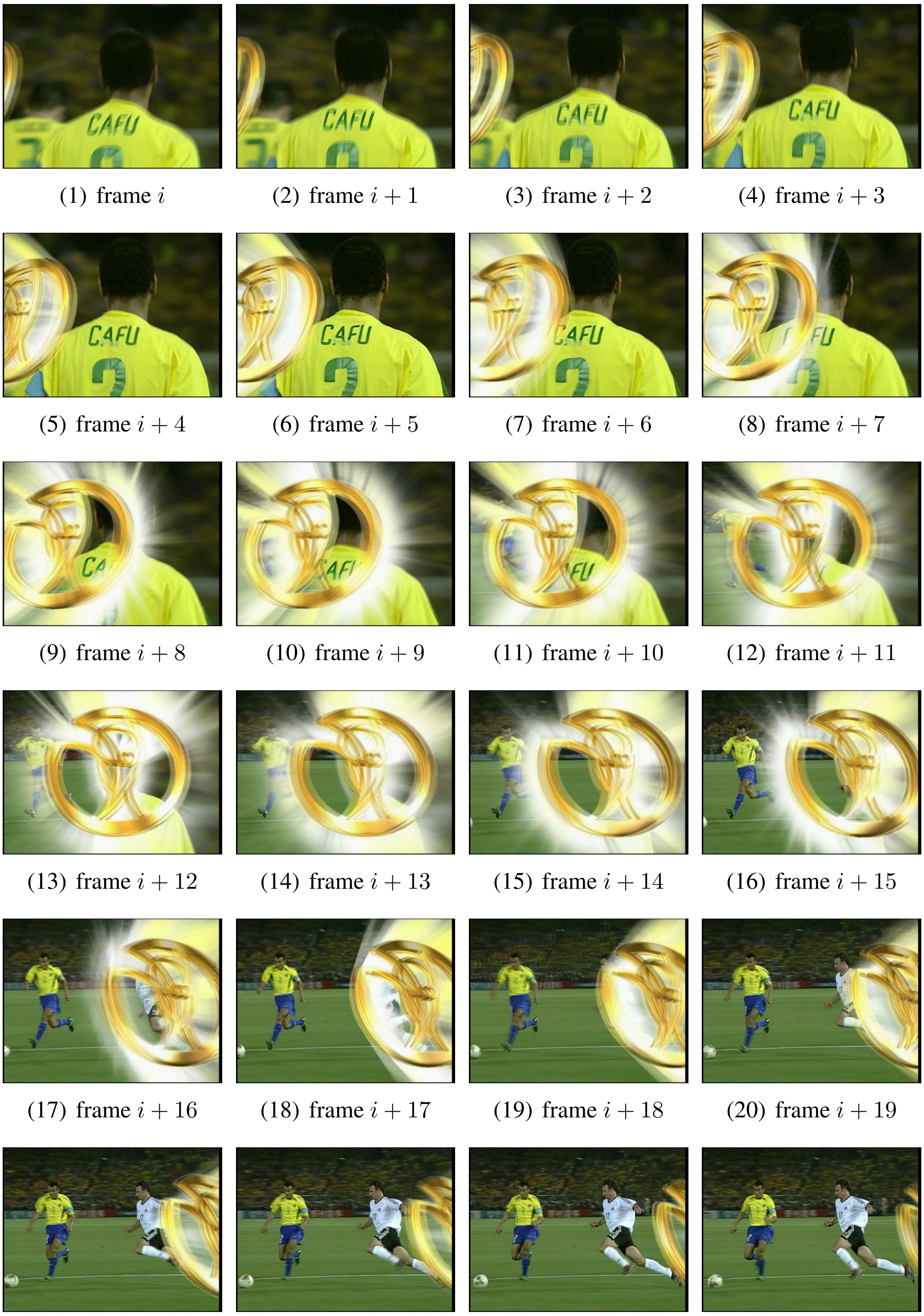


(1) frame $i$ (2) frame $i+1$ (3) frame $i+2$ (4) frame $i+3$

(5) frame $i+4$ (6) frame $i+5$ (7) frame $i+6$ (8) frame $i+7$

(9) frame $i+8$ (10) frame $i+9$ (11) frame $i+10$ (12) frame $i+11$

FIGURE 3.4 *Dissolve example from game #49 of* WC *2010: Uruguay vs Korea Republic. The frame sequence shows two shots combined in a dissolve effect. Both shots are spatially and temporally intermixed. The length of this long dissolve is about 12 frames, it is much larger than the one showed in figure* 3.3.



(1) frame $i$ (2) frame $i+1$ (3) frame $i+2$ (4) frame $i+3$

(5) frame $i+4$ (6) frame $i+5$ (7) frame $i+6$ (8) frame $i+7$

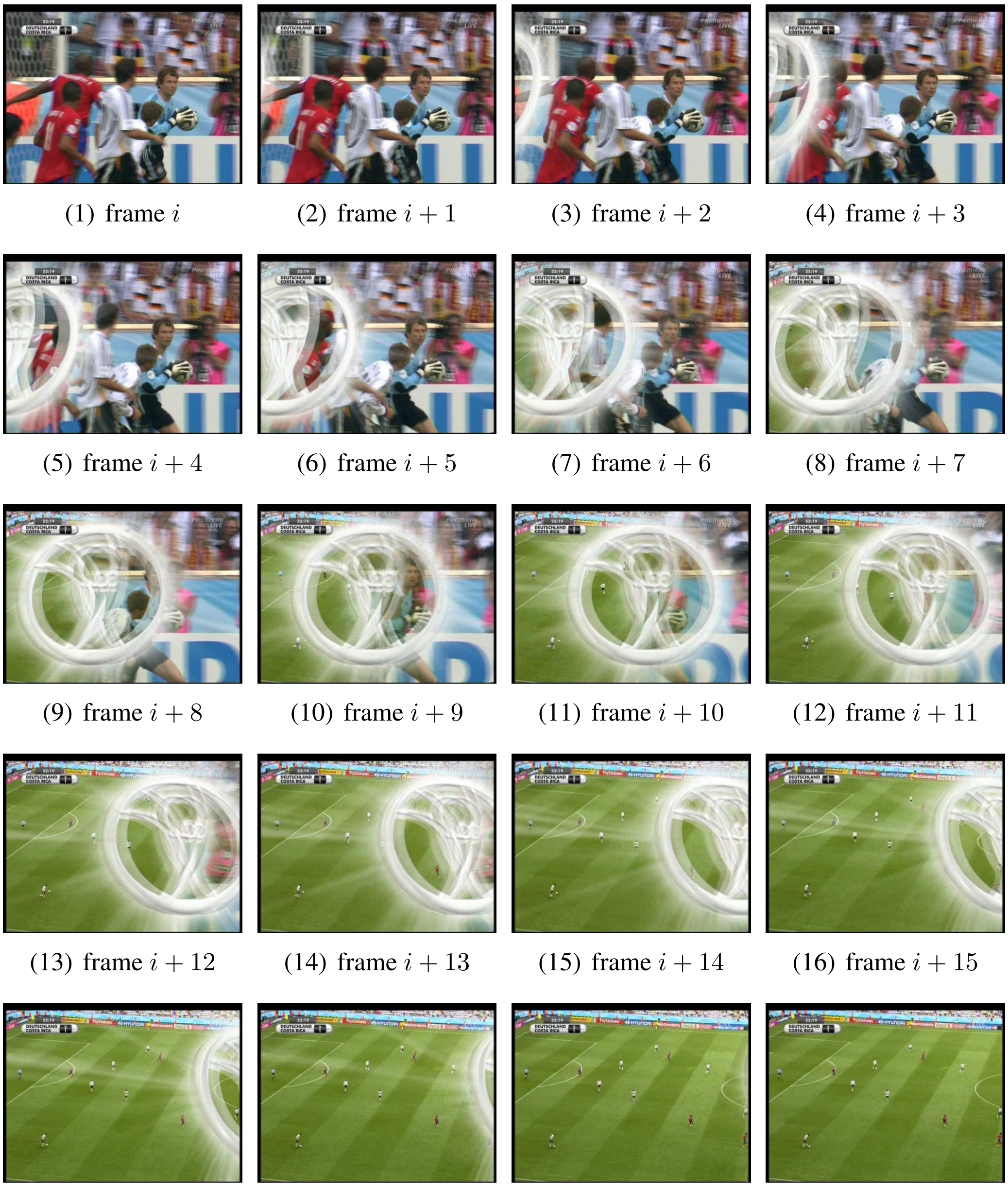
(9) frame $i+8$ (10) frame $i+9$ (11) frame $i+10$ (12) frame $i+11$

(13) frame $i+12$ (14) frame $i+13$ (15) frame $i+14$ (16) frame $i+15$

(17) frame $i+16$ (18) frame $i+17$ (19) frame $i+18$ (20) frame $i+19$

(21) frame $i+20$ (22) frame $i+21$ (23) frame $i+22$ (24) frame $i+23$

FIGURE 3.5 *Wipe example from the game #64 of* WC *2002: Brazil - Germany. A gold-coloured object is being used as borderline between the two spatially separated shots. The duration of the effect is about 24 frames.*



(1) frame $i$ (2) frame $i + 1$ (3) frame $i + 2$ (4) frame $i + 3$

(5) frame $i + 4$ (6) frame $i + 5$ (7) frame $i + 6$ (8) frame $i + 7$

(9) frame $i + 8$ (10) frame $i + 9$ (11) frame $i + 10$ (12) frame $i + 11$

(13) frame $i + 12$ (14) frame $i + 13$ (15) frame $i + 14$ (16) frame $i + 15$

(17) frame $i + 16$ (18) frame $i + 17$ (19) frame $i + 18$ (20) frame $i + 19$

FIGURE 3.6 *Wipe example from the game #1 of* WC *2006: Germany - Costa Rica. A silver-coloured object is being used as borderline between the two spatially separated shots. The duration of the effect is about 20 frames.*

can be logos of the broadcasting channels, game time, game scores, or others. For example, in the left top corner of the figure 3.3(3), game times, score and teams' abbreviations can be observed, as well as the channel logo on the right top corner of the same image. For the *dynamic overlays*, the duration is normally of about some few to tens of frames, and they generally show temporary information about player changes, score change, or advertisement (see figure 3.7). Also game information can be coded into such overlays, such as the one showing the offside area in figure 3.8.

All the aforementioned effects appear combined amongst themselves, for example, a dissolve combined with a dynamic overlay is shown in figure 3.9. Also some static overlays are shown in the same image, such as the channel logo, and the game time, in the left top corner. Other effects will not be considered, such as fade ins, fade outs, or other more specific types of cuts, since its use is not common in modern football broadcasting.

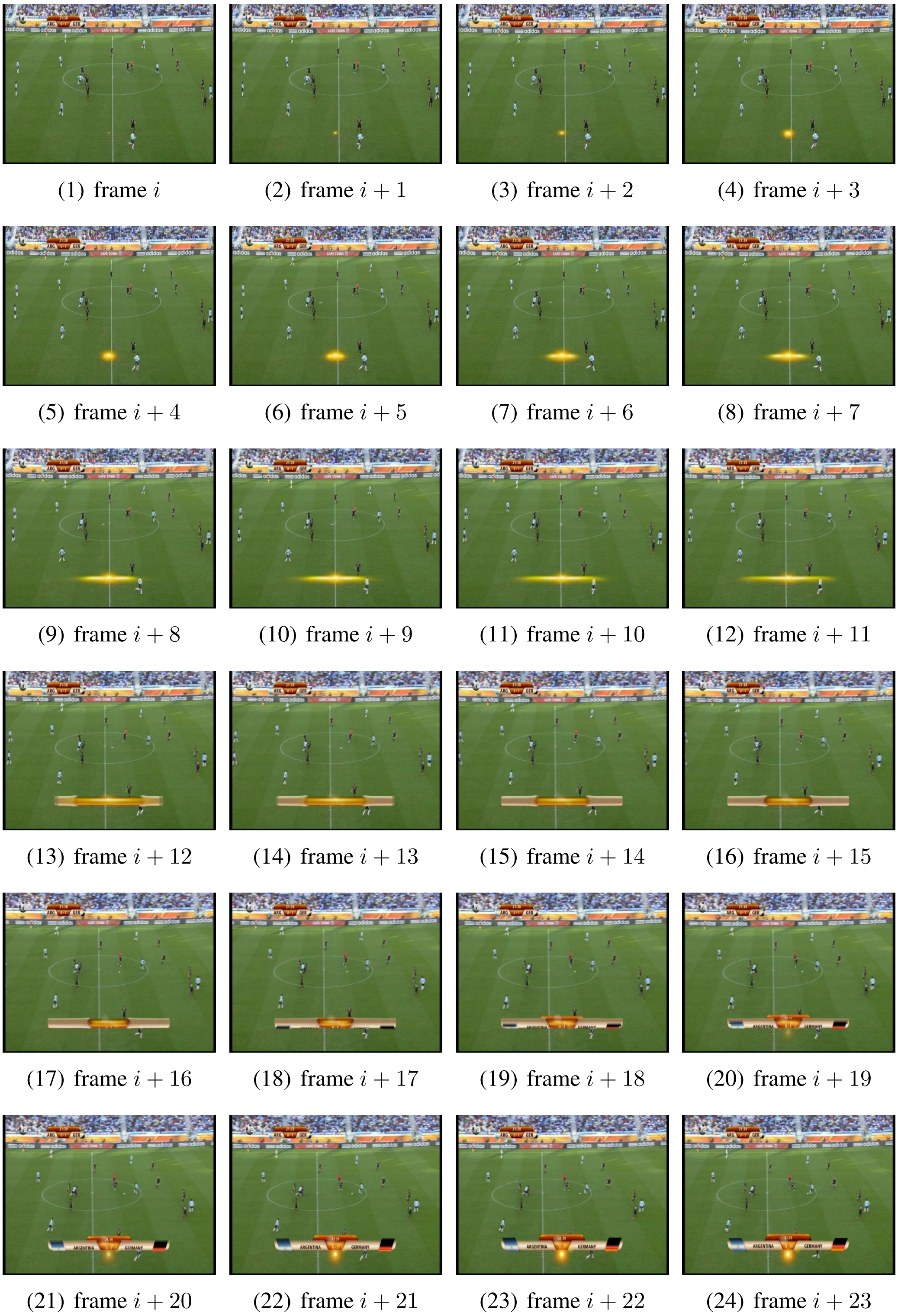
It is very important for the current work to be able to find, in the frame sequences, the positions on the field of the visible players, because their spatial situation is the basis for the construction of the semantic annotation system that interest us. To reach that objective, the shots containing cut-ins to the players, or zooms to the stands, or to the trainer areas are to be removed, since they do not provide any information of the relative positions of the players on the field. The replays have to be removed too, since they are out of the normal running time line, and therefore, the positional information of the players in this segments would not be useful. A video temporal segmentation and shot classification is required, to select the shots containing far-view scenes, since the far-view scenes contains the interesting objects in a convenient relation, as to extract the most appropriate and informative positional data.

The frame selection method must be capable of detecting the shot boundaries, locating the replays, and classifying the shots depending on the camera of interest and camera view. In order to represent this situation, the input video $V$ is partitioned into a set of video segments: $V = \{v_1, v_2, \cdots, v_N\}$, such that each video segment $v_i$ corresponds to a desired video segment, a shot from a camera of no interest, a shot from a camera of interest but with severely altered intrinsic parameters (e.g. zooming factor), a transitional effect (cuts, dissolves or wipes), a replay, and others. Therefore, our interest is to find a mapping $f$ such that:

$$f : V \to S \tag{3.1}$$

$$\{v_i\} \to \{s_i\} \tag{3.2}$$

from the video partition $V$ to a subset of it $S = s_i, s_2, \cdots, s_n$, such that the elements of $S$ are appropriate and informative enough for further processing, that is, $S$ is a set of far-view shots.



(1) frame $i$ (2) frame $i+1$ (3) frame $i+2$ (4) frame $i+3$

(5) frame $i+4$ (6) frame $i+5$ (7) frame $i+6$ (8) frame $i+7$

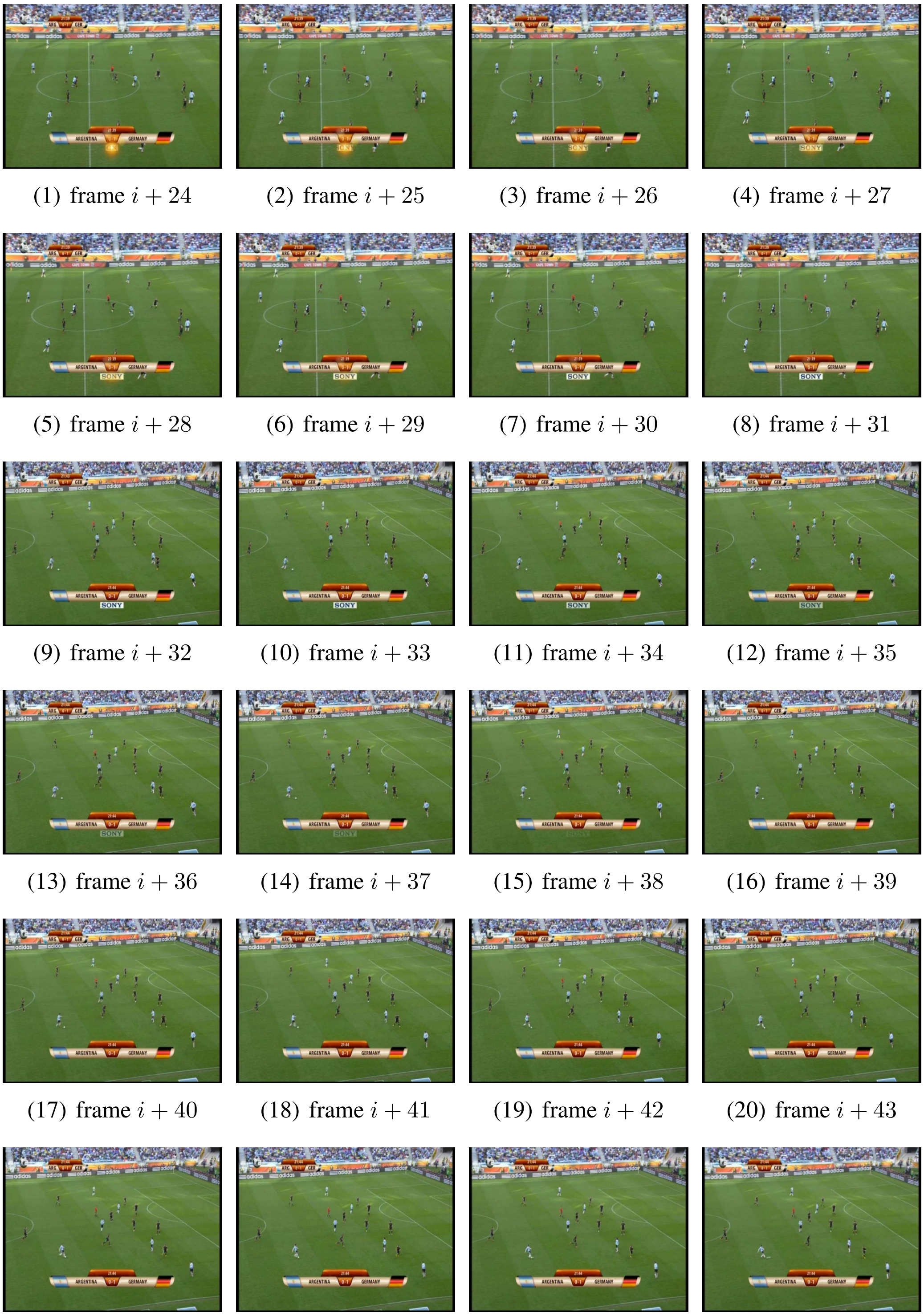
(9) frame $i+8$ (10) frame $i+9$ (11) frame $i+10$ (12) frame $i+11$

(13) frame $i+12$ (14) frame $i+13$ (15) frame $i+14$ (16) frame $i+15$

(17) frame $i+16$ (18) frame $i+17$ (19) frame $i+18$ (20) frame $i+19$

(21) frame $i+20$ (22) frame $i+21$ (23) frame $i+22$ (24) frame $i+23$

FIGURE 3.7 *Dynamic overlay example from the game #59 of* WC *2010: Argentina - Germany. The game time and the current score are being shown. The duration of the overlay is of 56 frames. The effect is showing the teams, the score, the game time as well as advertisement.*



(1) frame $i + 24$ (2) frame $i + 25$ (3) frame $i + 26$ (4) frame $i + 27$

(5) frame $i + 28$ (6) frame $i + 29$ (7) frame $i + 30$ (8) frame $i + 31$

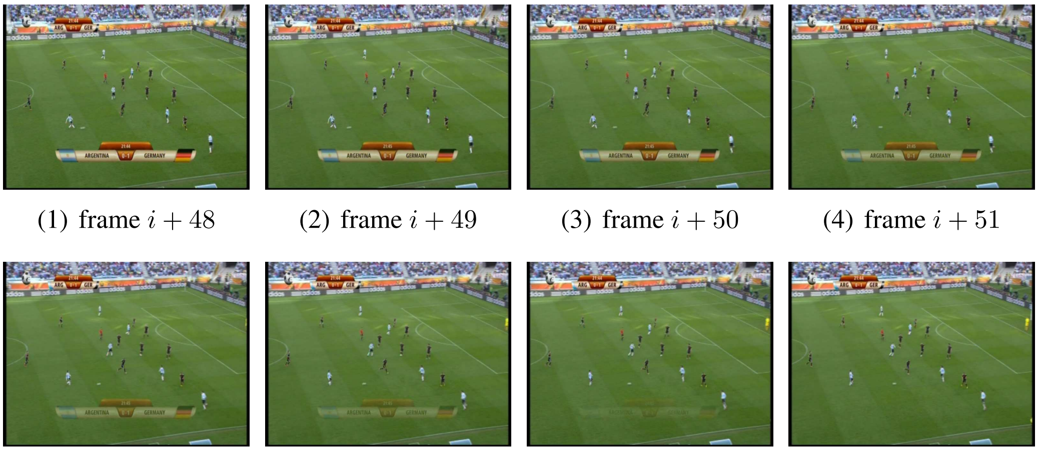
(9) frame $i + 32$ (10) frame $i + 33$ (11) frame $i + 34$ (12) frame $i + 35$

(13) frame $i + 36$ (14) frame $i + 37$ (15) frame $i + 38$ (16) frame $i + 39$

(17) frame $i + 40$ (18) frame $i + 41$ (19) frame $i + 42$ (20) frame $i + 43$

(21) frame $i + 44$ (22) frame $i + 45$ (23) frame $i + 46$ (24) frame $i + 47$

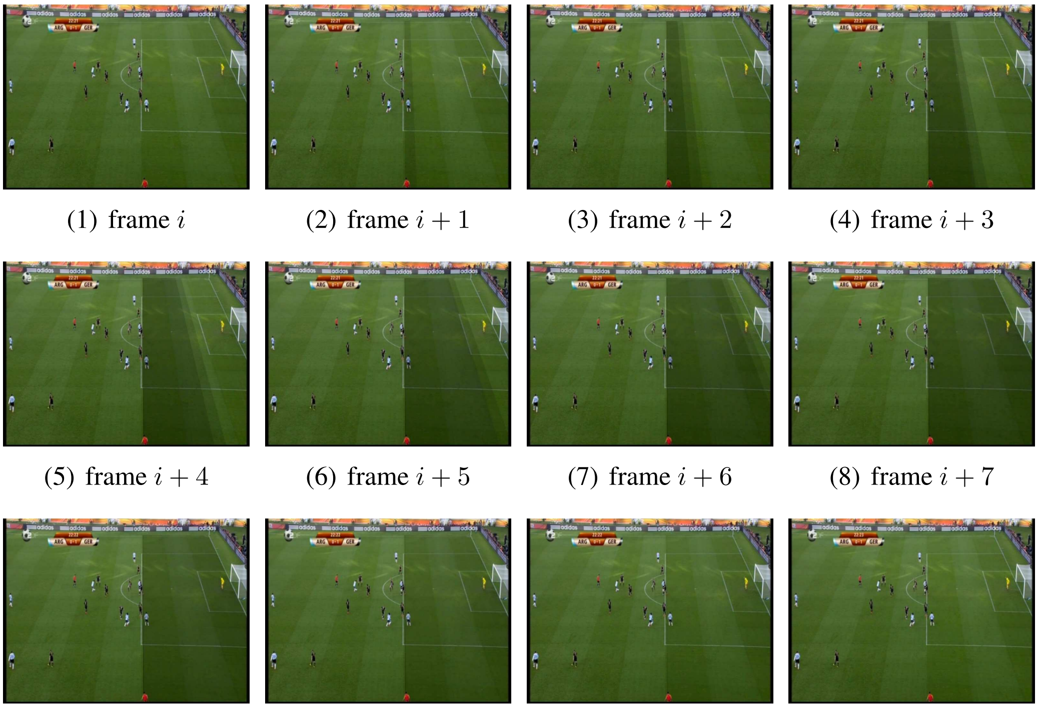
FIGURE 3.7 *(cont. 1). Dynamic overlay example from the game #59 of* WC *2010: Argentina - Germany. The game time and the current score are being. The duration of the overlay is of 56 frames. The effect is showing the teams, the score, the game time as well as advertisement.*



(1) frame $i + 48$ (2) frame $i + 49$ (3) frame $i + 50$ (4) frame $i + 51$

(5) frame $i + 52$ (6) frame $i + 53$ (7) frame $i + 54$ (8) frame $i + 55$

FIGURE 3.7 *(cont. 2). Dynamic overlay example from the game #59 of* WC *2010: Argentina - Germany. The game time and the current score are being. The duration of the overlay is of 56 frames. The effect is showing the teams, the score, the game time as well as advertisement.*

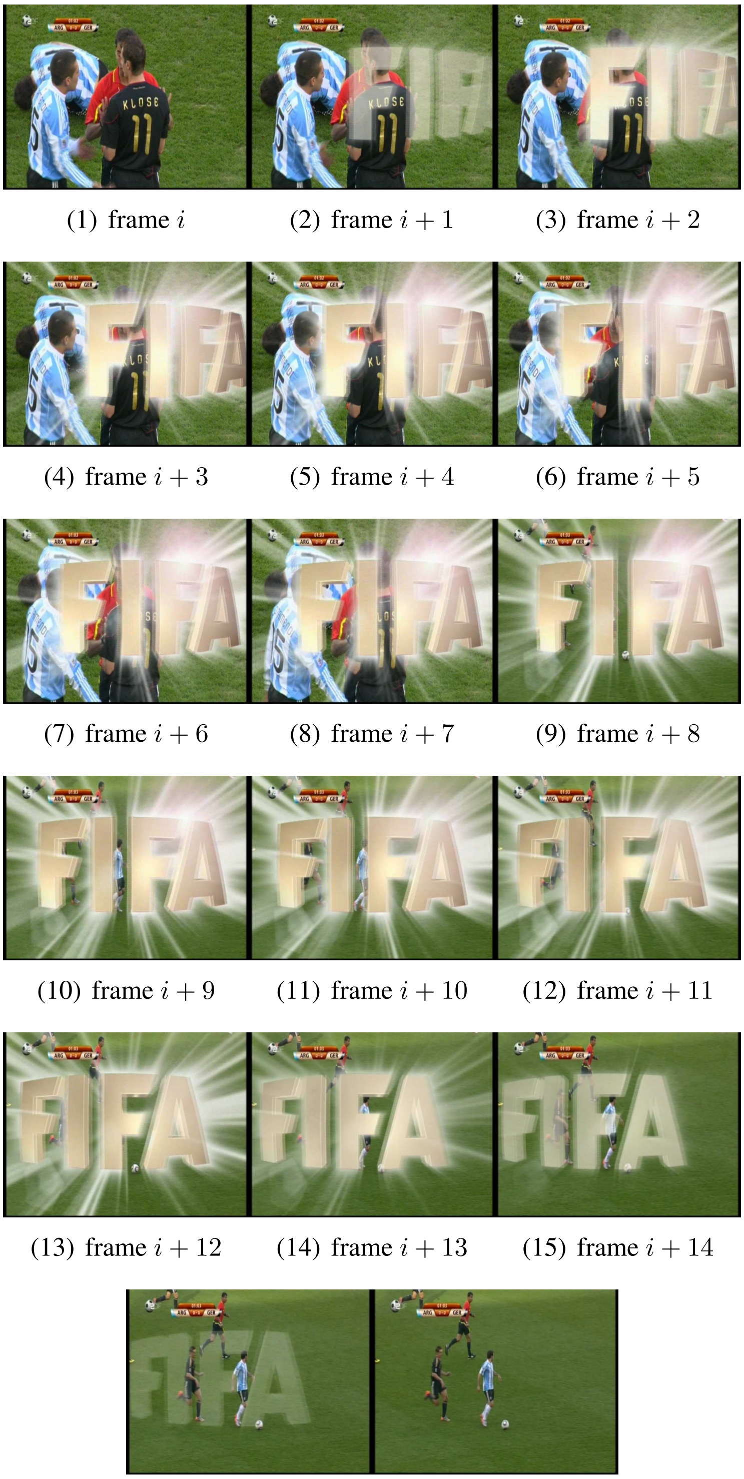


(1) frame $i$ (2) frame $i + 1$ (3) frame $i + 2$ (4) frame $i + 3$

(5) frame $i + 4$ (6) frame $i + 5$ (7) frame $i + 6$ (8) frame $i + 7$

(9) frame $i + 8$ (10) frame $i + 9$ (11) frame $i + 10$ (12) frame $i + 11$

FIGURE 3.8 *Dynamic overlay example from the game #59 of* WC *2010: Argentina - Germany. The offside area is being shown. The duration of the overlay is of 10 frames.*



(1) frame $i$ (2) frame $i+1$ (3) frame $i+2$

(4) frame $i+3$ (5) frame $i+4$ (6) frame $i+5$

(7) frame $i+6$ (8) frame $i+7$ (9) frame $i+8$

(10) frame $i+9$ (11) frame $i+10$ (12) frame $i+11$

(13) frame $i+12$ (14) frame $i+13$ (15) frame $i+14$

(16) frame $i+15$ (17) frame $i+16$

FIGURE 3.9 *Combined dissolve and dynamic overlay example from game #59 of* WC *2010: Argentina vs Germany. The frame sequence shows two shots combined in a dissolve effect and a dynamic overlay effect. The shots are spatially and temporally intermixed. The length of this dissolve effect is about 15 frames.*

Examples of the desired frame sequences are depicted in figure 3.10. The description of the algorithm to obtain the set of frame sequences out of the input TV broadcast video is depicted in algorithm 3.1. This algorithm will be detailed and explained in the following sections.

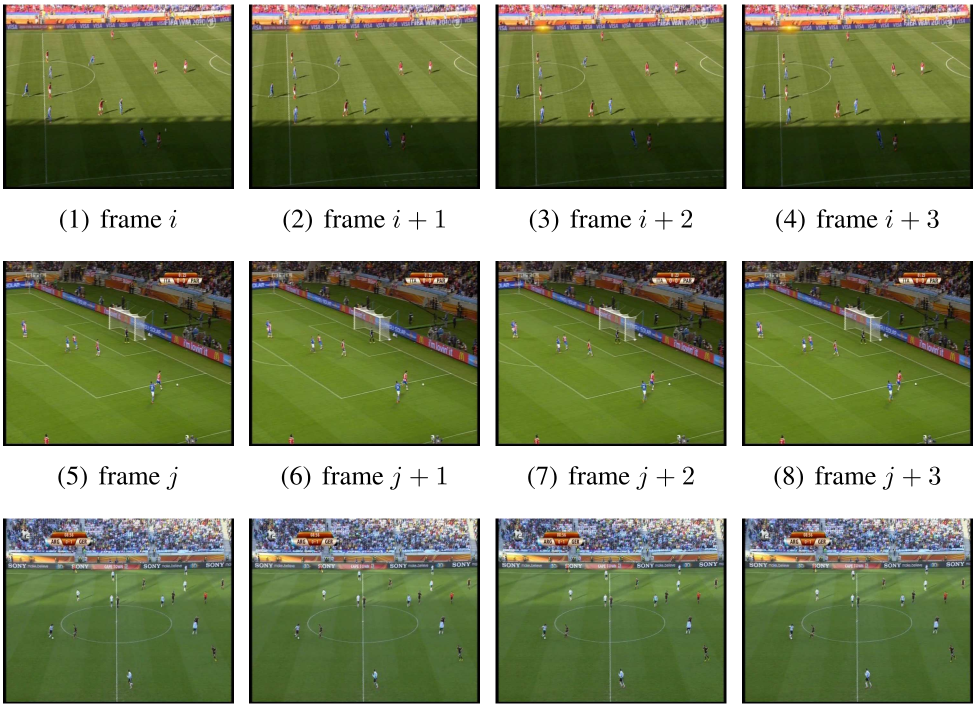
**Algorithm 3.1** Frame Sequences Selection Algorithm $f : V \to S$. The input is the input video taken from a TV broadcast source, the output is a set of frame sequences corresponding to the far-view shots.

```
Detect transitional overlays using the $cd$ and $mcd$: equation 3.6 and equation 3.5
Remove replays using set differences
Detect transitional effects (cuts and dissolves) using $HVBP$: equation 3.23
Classify transitional effects using threshold $t_d$ and $t_c$: equation 3.24 and equation 3.25
Remove transitional effects using set differences
Generate features for remaining shots $GPR$ and $PPR$: equation 3.28 and : equation 3.29
Classify shots using threshold $t_{GPR}$ and $t_{PPR}$: equation 3.30 and equation 3.31
```

## 3.2 Replays

There exist several types of replays, for instance, slow motion replays, normal-speed replays, as well as still images. In the case of slow motion replays, the videos are taken either with a high-speed camera, or with a normal camera. The replays recorded using a high-speed camera are later replayed at a slower frame rate. The replays recorded using a normal camera repeat some of the frames to achieve a similar visual effect as with the high-speed camera method. Finally, still images are sometimes displayed for the analysis of possible offside situations, or similar.

To try to detect replays, some methods have been presented in the professional literature, for example in [PBS01] a *HMM* with 5 states: slow motion, still images, normal replay, editing effect and normal play was used and tested for different sports. The reported success rate was 100%, but most of the clips used for the evaluation are no longer than 1 minute, which makes it difficult to asses the performance that the algorithm will have for complete games of about 90 minutes each. Besides, difficulties are mentioned about the localization of the boundaries of the slow motion replays, due to cuts inside slow motion replays as well as normal motion replays. In [KDD99] a method is presented that exploits some characteristics of the MPEG video coding scheme, but the results apart from being only qualitative, apply only to the detection of slow motion replays without addressing the other subclasses. Finally, in [Ged09] a method is presented using the absolute gradient of the zero crossing measure $g_t = |d_{t-1} - d_t|$, where $d_t$ is the standard deviation of the difference between two consecutive

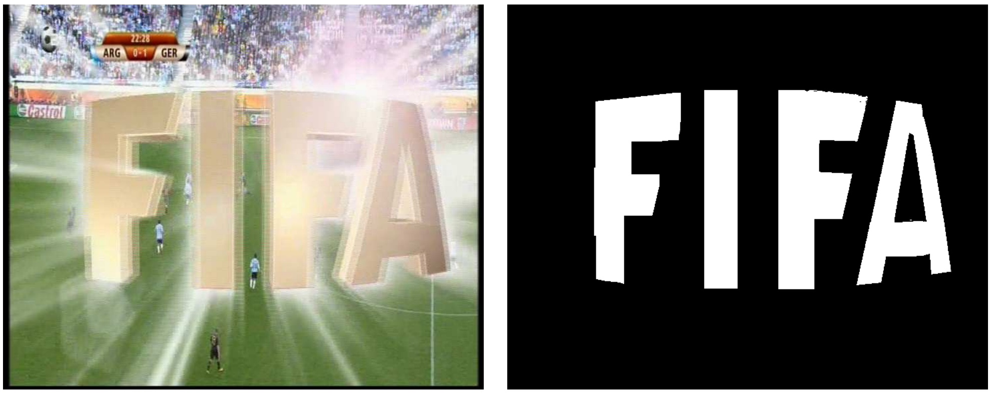


(1) frame $i$ (2) frame $i+1$ (3) frame $i+2$ (4) frame $i+3$

(5) frame $j$ (6) frame $j+1$ (7) frame $j+2$ (8) frame $j+3$

(9) frame $m$ (10) frame $m+1$ (11) frame $m+2$ (12) frame $m+3$

FIGURE 3.10 *Examples of desired frame sequences: taken from the high-angle or tactical cameras. First row: game #4 of* WC *2010: Korea Republic vs Greece. Second row: game #11 of* WC *2010: Italy vs Paraguay. Third row: game #59 of* WC *2010: Argentina vs Germany.*



(1) FIFA's logo used as delimiter for the replays

(2) Mask $M$ used to search for the cumulative difference in equation 3.6

FIGURE 3.11 *Delimiter for replays used in the* WC *2010.*

images, as an indicator of the slow motion sequence. The reported results are very good, with a precision of 100% and a recall of 99,74%, but these results are relevant to slow motion detection only, since no results for other kinds of replays were given. It is also to be mention, that the replay boundaries are not found by these method, and that the algorithm was evaluated using only one game with 42 slow motions.

The first two methods are very general and do not take into account specifics about the sport type, and about the transitional effects used, and therefore the accuracy and robustness of such methods for the football case is not optimal as desired. The last method is a better one, since it takes into account the sport type, but it ignores the effect of the overlays as replay delimiters. Based on the observation that for the last *WCs*, the replays, and slow motions have been delimited by overlays such as those shown in figure 3.5, figure 3.6, and figure 3.9, and that these transitional effects are kept constant during each of the championships, a simpler yet powerful method to locate the replays is to find the specific effects being used in the transitions, and used them as delimiters for the replay segments. For example, during the *WC* 2010, the *FIFA*'s logo was used as delimiter of the replays, see figure 3.11(1). The idea of the proposed algorithm is to find a mapping $r$ from the video $V$, to the replays $R$ such that:

$$r : V \to R \tag{3.3}$$

$$\{v_i\} \to \{r_i\} \tag{3.4}$$

The main idea is to find the replays' boundaries and then simply, by set differences, obtain the video segments $Q$ without the replays. The description of the algorithm used to remove the replays from the video input is described in algorithm 3.2.

**Algorithm 3.2** Algorithm $q : V \rightarrow Q$ to obtain the video sections $Q$ removing the replays $R$ from the original video $V$

```
Obtain the $mcd$ for each frame in the video $V$: equation 3.5
Threshold the $mcd$ values using $t_0$
Generate the set $R$ as delimited by consecutive replays boundaries
Obtain the set difference $Q = V \setminus R$
```

The proposed method uses the mean cumulative difference $mcd$ (equation 3.5) as a measure of dissimilitude between the current frame $I$ and the reference frame $E$ containing the effect sample (see figure 3.11(1)), which is defined in terms of the cumulative difference $cd$ (equation 3.6).

$$mcd = \frac{\sum_{\forall i \in [r,g,b]} \overline{cd(r,g,b)}}{N} \tag{3.5}$$

$$\overline{cd(r,g,b)} = \sum_{\forall (x,y) \in M} I(x,y) - E(x,y) \tag{3.6}$$

where $M$ (see figure 3.11(2)) is the set of spatial positions where the required measure is desired, in the simpler case it corresponds to the total image area span by $I$.

The measure $mcd$ is obtained for all the frames in the input video, and the locations where the low values are present, represent the positions of the overlays. Examples of the calculated dissimilitude measure are shown in figure 3.12. To determine the positions of the overlays a threshold $t_0$ is selected, such that the number of false negatives was reduced to 0, and the number of the false positives was the achievable optimum without increasing the number of false negatives, since missing an overlay has more negative consequences to our interests than finding some fake ones.

For the evaluation and comparison of the method, three metrics are used: precision $p$, recall $r$, and $F1$-score $F1$:

$$p = \frac{tp}{tp + fp} \tag{3.7}$$

$$r = \frac{tp}{tp + fn} \tag{3.8}$$

$$F1 = 2\frac{p \cdot r}{p + r} \tag{3.9}$$

where $tp$, $fp$ and $fn$ are the number of true positives, false positives, and false negatives respectively.

The threshold was set to $t_0 = 25$, producing the results shown in table 3.2. From the table it can be seen that $5189$ transitional overlays were analysed, from which only $83$ false positives, and 0 false negatives were obtained. This $5189$ overlays correspond to the games of the *WC* 2010. Also, it can be seen that the metrics for the evaluation: precision, recall and $F1$-score are respectively $98,43\%$, $100,00\%$, and $99,21\%$, which means that the proposed method performs a lot better than the ones presented in the professional literature. Also it is important to note, that a more thoroughly evaluation was performed in the present work, since more than five thousand overlays were evaluated along several complete games, compared to only few frames, as found in the professional literature.

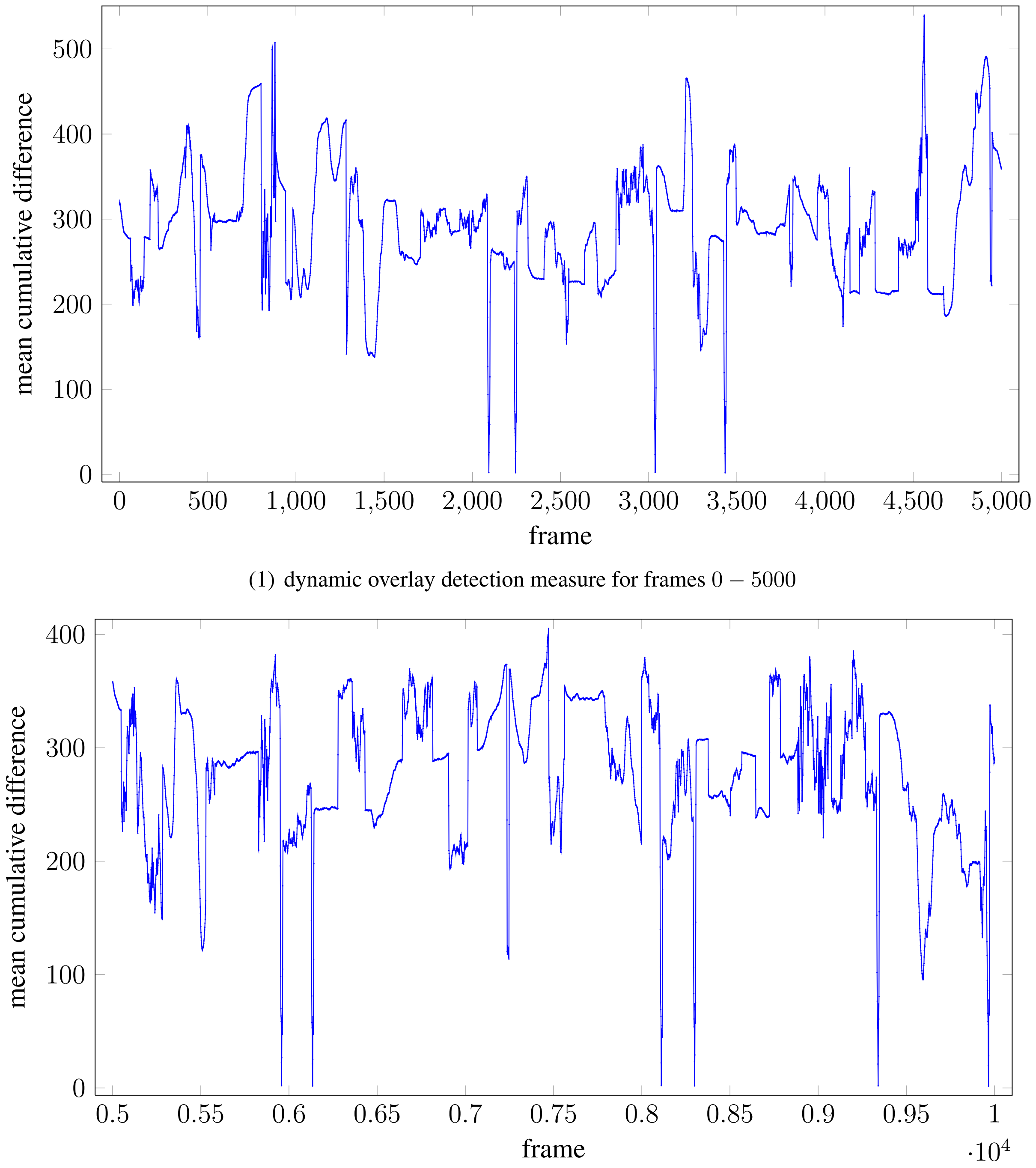
## 3.3 Transitional Effects

The next step is to detect the shot boundaries, particularly cuts and dissolves. For the detection of such transitions, in the professional literature, several methods have been proposed, as well as summarising surveys [BBR96, YCK98, Lie99, KC01, YWX+07, GN08]. The main idea followed by these methods is to detect the discontinuities in the visual content of the video. This approach can be described as composed of three main elements: **visual content representation**, **construction of continuity signal** and **classification of the discontinuity values** [YWX+07].

Formally, to represent the visual content, a mapping $\rho$ (equation 3.10) from the image space $I$ to the feature space $F$, extracting features from the image is required, such that the two somehow opposed requirements *invariance* and *sensitivity* reach a useful trade-off. That is, the features must be sensitive enough to be able to detect the transitions, but invariant enough to account for variations in the visual content, such as illumination changes, and players or camera movement. Of course, the selection of the appropriate features is vital for the correct performance of the detector. The subscript $t$ represents the frame number.

$$\begin{aligned} \rho : I &\rightarrow F \\ I_t &\rightarrow f_t \end{aligned} \tag{3.10}$$

To construct the continuity signal, some similarity (or dissimilarity) measure between adjacent features is defined. That is, a map $\delta$ (equation 3.11) between the Cartesian product of the



(1) dynamic overlay detection measure for frames $0 - 5000$

(2) dynamic overlay detection measure for frames $5001 - 10000$

FIGURE 3.12 *Dynamic Overlay Detection Measure for game #4 of the WC 2010: Korea Republic vs Greece. The frames where the mean cumulative difference (equation 3.5) is below the threshold $t_0$, represent the locations of the dynamic overlays:* 2094, 2246, 3037, 3434, 5958, 6134, 8111, 8300, 9340, *and* 9966. *In the subfigure (1) the measure for the first* 5 *thousand frames of the video is shown, and the next* 5 *thousand frames on subfigure (2). The total number of frames in this particular video is* 182543*, and only a small percentage (around* 11%*) of it is shown.*

| | | | |
|---|---|---|---|
| 1 | Southafrica vs Mexico | 30 | Portugal vs Korea DPR |
| 4 | Korea Republic vs Greece | 31 | Chile vs Switzerland |
| 5 | England vs USA | 32 | Spain vs honduras |
| 9 | Netherlands vs Denmark | 33 | Mexico vs Uruguay |
| 10 | Japan vs Cameroon | 37 | Slovenia vs England |
| 11 | Italy vs Paraguay | 39 | Ghana vs Germany |
| 12 | Newzealand vs Slovakia | 41 | Slovakia vs Italy |
| 13 | Coete d'Ivoire vs Portugal | 43 | Denmark vs Japan |
| 14 | Brazil vs Korea DPR | 45 | Portugal vs Brazil |
| 15 | Honduras vs Chile | 47 | Chile vs Spain |
| 16 | Spain vs Switzerland | 49 | Uruguay vs Korea Republic |
| 17 | Southafrica vs Uruguay | 50 | USA vs Ghana |
| 18 | France vs Mexico | 51 | Germany vs England |
| 19 | Greece vs Nigeria | 52 | Argentina vs Mexico |
| 20 | Argentina vs Korea Republic | 53 | Netherlands vs Slovakia |
| 21 | Germany vs Serbia | 54 | Brazil vs Chile |
| 22 | Slovenia vs USA | 55 | Paraguay vs Japan |
| 23 | England vs Algeria | 56 | Spain vs Portugal |
| 24 | Ghana vs Australia | 57 | Netherlands vs Brazil |
| 25 | Netherlands vs Japan | 58 | Uruguay vs Ghana |
| 26 | Cameroon vs Denmark | 59 | Argentina vs Germany |
| 27 | Slovakia vs Paraguay | 60 | Paraguay vs Spain |
| 28 | Italy vs Newzealand | 63 | Germany vs Uruguay |
| 29 | Brazil vs Coete d'Ivoire | 64 | Netherlands vs Spain |

TABLE 3.1 *Analysed Games from the WC 2010. Not all of the games played during the tournament were analysed, this happen due to a variety of reasons, for example, some games were broadcast at the same time, and only one video capture card was available for the recording.*

feature space $F$ to the continuity space $C$ must exist:

$$\begin{aligned} \delta : F^{2\times n} &\to C \\ (f_{t-n+1}, \cdots, f_t, f_{t+1}, \cdots, f_{t+n}) &\to d_t \end{aligned} \tag{3.11}$$

where $n$ is the size of the neighbourhood used to obtain the dissimilarity measure $d_t$ between frames $I_t$ and $I_{t+1}$. Commonly, the size used is $n = 1$, but other values are also possible.

The last component is the classification of the dissimilarity measures, so that the transitions are found. This is a mapping $\kappa$ (equation 3.12) between the Cartesian product of the continuity space $C$ to the decision space $W$.

$$\begin{aligned} \kappa : C^{2\times m+1} &\to W \\ (d_{t-m}, \cdots, d_t, d_{t+1}, \cdots, d_{t+m}) &\to w_t \end{aligned} \tag{3.12}$$

where $m$ is the neighbourhood size used by the classifier and $W$ is the decision space: transition, no-transition; or types of transitions. In the most of the cases, $m = 0$ is used, but again, other values are also possible.

As pointed out in [Lie99, YWX$^+$07], the detection of cuts has been successfully tackled, while the dissolves are still an open problem and a difficult one indeed. This is due to the fact that the cut normally produces continuity values which are easier to detect than for dissolves, especially because the time-space intermixing of the shots in a dissolve. Also because the patterns in the continuity signal produced by the dissolves are similar to those produced due to camera or player movement. Another problem is the illumination changes, since they might produce significant discontinuities in the continuity signal, which affects those methods where the colour is used to represent the visual content. Finally, large movements of the players, and of the camera may generate values as big as those for cuts or similar patterns to the dissolves. It is therefore difficult to detect the boundaries using only features based on colour.

Different ways have been proposed to perform the visual content representation, broadly classified into **pixel-wise**, **block-based** and **histogram methods**[2]. Pixel-wise techniques use differences between corresponding pixels in consecutive images as the features retrieved by $\rho$. Commonly, the difference used for a colour image of size $X \times Y$, and with $Z$ channels is

[2] Other methods based on a particular codec compression coefficients (for example MPEG coefficients) of the video are not considered into this thesis, since the objective is to be able to deal with the most general video input, disregarding the compression codec.

given by equation 3.13.

$$\rho = \frac{\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}\sum_{z=0}^{Z-1}|I_t(x,y,z) - I_{t+1}(x,y,z)|}{(X-1)\cdot(Y-1)} \tag{3.13}$$

This methods have been shown to be very sensitive to movements of the camera and the objects [BBR96, KC01], which in the current case of faster player, and camera movements is not adequate.

In the block-based techniques, the image is spatially divided into non-overlapping sub-areas which are then paired with their corresponding sub-areas in the image at the next time step, and a difference measure is obtained per block pair. The dissimilarity measure to use can be some statistical quantity, for example a function of the likelihood ratio $\lambda_k$ of the $k$-th block.

$$\rho = \sum_{k=0}^{B-1} c_k \lambda_k \tag{3.14}$$

$$\lambda_k = \frac{\left[\frac{\sigma_{k,t}+\sigma_{k,t+1}}{2} + \left(\frac{\mu_{k,t}+\mu_{k,t+1}}{2}\right)^2\right]^2}{\sigma_{k,t}\cdot\sigma_{k,t+1}} \tag{3.15}$$

where $\sigma_{k,t}$ and $\mu_{k,t}$ are the standard deviation and mean of the $k$-th block at frame $t$ respectively, and $c_k$ is a predetermined weight of the respective block[3]. This method is more robust against slow and small objects' motions compared to the pixel-based methods, but it is slower due to the complexity of the dissimilarity measure used, and it is reported to produce many false positives [BBR96].

The algorithms based on histogram assume that the grey-level or colour does not change inside shots, but across them. So the idea basically consist on finding the peaks on the continuity signal, using consecutive differences, or more generally $k$-apart differences of the histograms. Since only the grey-intensity or colour is being considered, it might occur that two frames with same colour distributions have totally different content. The absolute sum of the histogram differences is used as a dissimilarity measure, see equation 3.16, or the histogram intersection,

[3] There seems to be an error with the equation 4 as edited by [KC01] pg. 480, because otherwise, the second term in the numerator of $\lambda_k$ would be zero. Here, this apparent error is corrected.

see equation 3.17, can be used too. Other histogram comparison methods are also available.

$$\rho = \sum_{n=0}^{N-1} |H_t(n) - H_{t+1}(n)| \tag{3.16}$$

$$\rho = 1 - \frac{\sum_{n=0}^{N-1} \min\left(H_t(n) - H_{t+1}(n)\right)}{\sum_{n=0}^{N-1} \max\left(H_t(n) - H_{t+1}(n)\right)} \tag{3.17}$$

where $H_t$ is the histogram (colour or grey-scale) of the image $I_t$ at time step $t$.

Another variation of the histogram technique is the so called local histogram, where the block-based and histogram paradigms are mixed. Basically, the image is divided into non-overlapping sub-areas, and then, a histogram continuity measure between corresponding blocks is obtained, see equation 3.18.

$$\rho = \sum_{k=0}^{B-1} \lambda_k \tag{3.18}$$

$$\lambda_k = \sum_{n=0}^{N-1} |H_t(n,k) - H_{t+1}(n,k)| \tag{3.19}$$

In [LI94] they claim to perform better than using global histogram comparison, but the results are more qualitative than quantitative, so a more objective performance evaluation is therefore difficult at least. See [LI94] or [KC01] for the formulation of the selective HSV measure. Mainly colour has been used as feature for both techniques, also other structural representation of the visual contents have been proposed, but with no better results, such as edges change ratio, where for example, the false positives for dissolves reach sometimes values as high as $37100\%$! [Lie99]. Also, the videos used to evaluate the results are few and are in general short ($<$ 30 minutes, when specified). Another drawback of the presented methods in the professional literature is that they are not specialised to the football case, where another problems arise, for instance: green dominance in the colour space, fast camera movement, fast player movements, as well as others.

A way to improve the detection of these effects is to use complementary features, describing orthogonal characteristics of the frame content. The main idea of our proposed method is to use a histogram approach combining colour information with structural information so as to form a more reliable feature to detect the effects. The colour information is important, since the dominant colour in football is green, and this naturally leads to a rejection of frames without green as the main colour. And since colour is not enough, as mentioned above, then a representation for the structural information is required. As a colour feature, the histogram of the hue component of the image in the HSV colour space is selected, and for encoding the structural information, the histogram of the local spatial variance in the intensity component of the image is used. Therefore, another map from the video segments without replays $Q$ to the shot candidates $C$ is seek:

$$c : Q \rightarrow C \tag{3.20}$$

$$\{q_i\} \rightarrow \{c_i\} \tag{3.21}$$

The method used for this task is shown in the algorithm 3.3.

**Algorithm 3.3** Algorithm $c : Q \rightarrow C$ to obtain the shot candidates $C$

```
  Obtain the dissimilarity measure $d$ for each frame in the video $Q$: equation 3.23
  Threshold the $d$ values using $t_d$ and $t_c$: equation 3.24 and equation 3.25
  Generate the set $C$ of shot candidates as delimited by the found transitional effects
```

The two particular features used were selected amongst a set of several candidates, by comparing their performance in the classification to the manually-annotated ground-truth. For the comparison, the recall and precision were used, as well as their product, taken as a way to take into account both evaluation mechanisms into a single measurement. As can be seen in the figure 3.13, the selected combination hue-var product, produces the best classification results.

In order to create the specific dissimilarity measure $d$, with metric properties, the natural logarithm of the product of the Bhattacharyya distance ($B$) between the histograms of hue and local spatial variance is used, and the size of the neighbourhood is $n = 1$. The proposed new dissimilarity measure is named *HVBP*. The Bhattacharyya distance $B$ between two histograms $h1$, and $h2$ is given by:

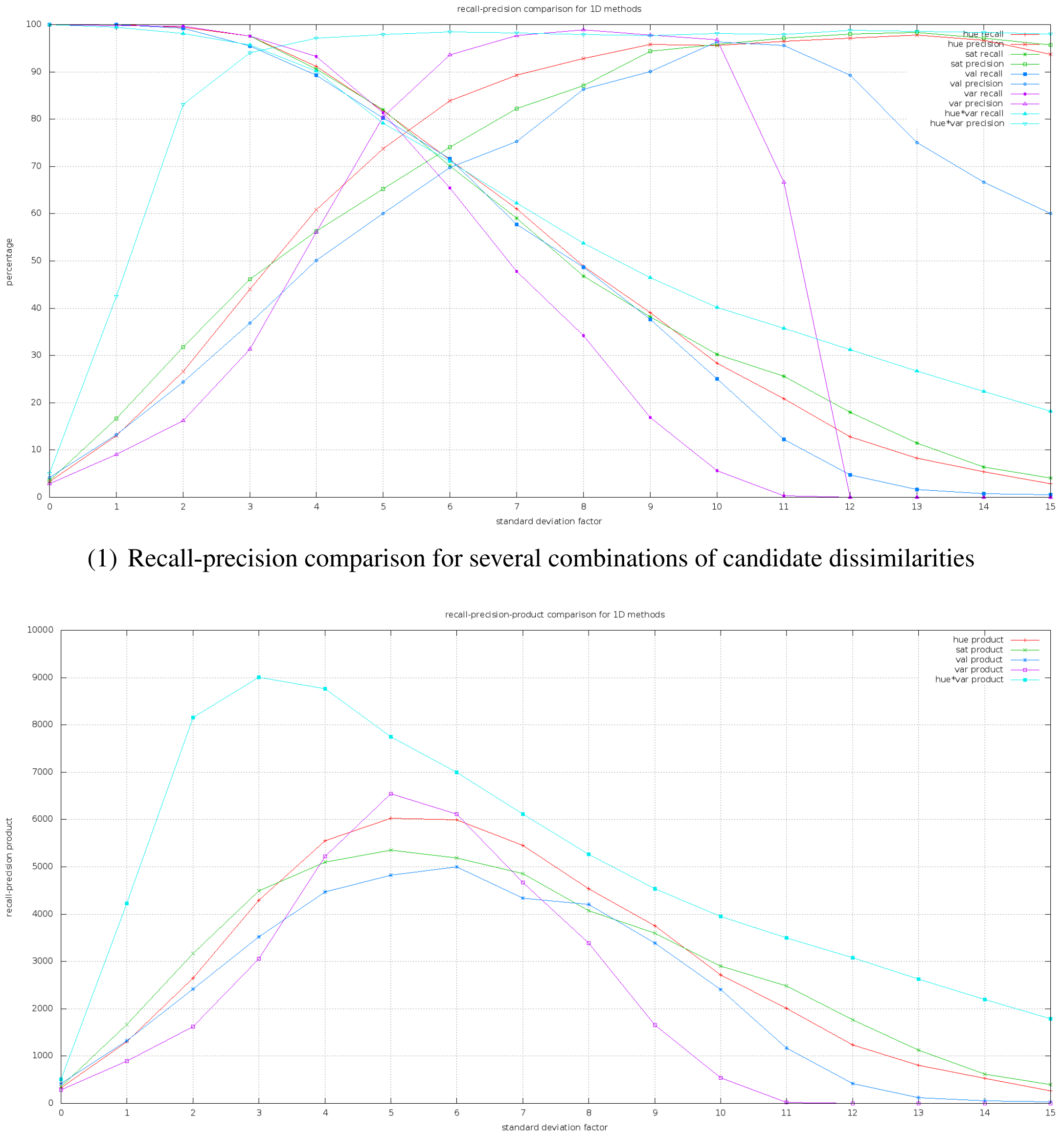
$$B(h1, h2) = \sqrt{1 - \frac{1}{\sqrt{\bar{h1} \cdot \bar{h2} N^2}} \sum_i h_1[i] h_2[i]} \tag{3.22}$$

therefore the dissimilarity measure $d$ is given by:

$$d = \log(B(h_H(I_{t-1}), h_H(I_t)) \cdot B(h_V(I_{t-1}), h_V(I_t))) \tag{3.23}$$

where $h_H$ is the histogram of the hue component of the HSV colour space representation of the image $I$, and $h_V$ is the histogram of the local spatial variance image of $I$.

In the figure 3.14, the histogram of the logarithms of the features used to detect the transitional effects is shown. The dissimilarity probability density is modelled as a uni-variate



(1) Recall-precision comparison for several combinations of candidate dissimilarities

(2) Recall-precision product comparison for several combinations of candidate dissimilarities

FIGURE 3.13 *Candidate Dissimilarity Factor Analysis. In the upper sub-figure the curves of recall and precision are shown for dissimilarity candidates such as: hue, saturation, value, local spatial variance, and hue-variance product. In the lower sub-figure, the product of recall-precision is shown as a combined measure of evaluation for the dissimilarity candidates. The hue-var produces the best results.*

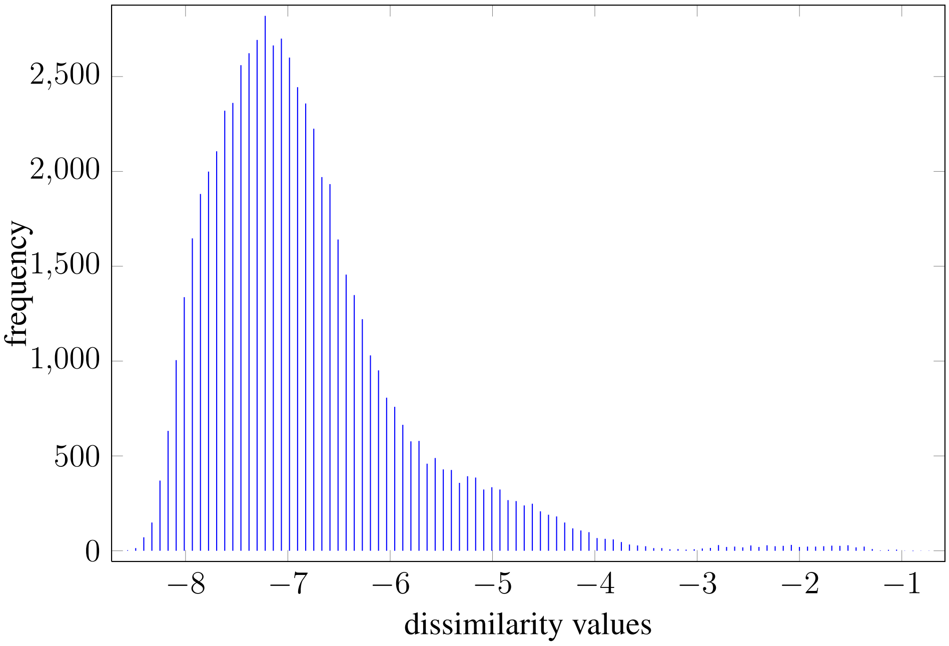


FIGURE 3.14 *Histogram of the dissimilarity values from the* HVBP *metric used to detect transitional effects*

Gaussian, since it represents the normal values of the metric for the commonly not-changing frames. Therefore any departure from this model will enclose the behaviour of especial cases such as the searched effects: cuts and dissolves. The selection of the required thresholds necessarily yields a trade-off between recall and precision, which is to be guided by the application at hand. A threshold $t_d$ is used for the detection of dissolves, and a threshold $t_c$ for the detection of the cuts, see equation 3.24 and equation 3.25. Since the cut generally produces a sharper peak in the feature sequence, due to the abrupt change of scene, then a higher threshold guarantees the correct detection of the cuts, while a lower value for the dissolve threshold allows for the detection of the softer transitions produced by such effect. Another aspect for selecting this dissimilarity measure is because is easy and fast to compute.

$$t_d = \frac{\sum_i d[i]}{N} + \sqrt{\frac{\sum_i (d[i] - \bar{d})^2}{N-1}} \tag{3.24}$$

$$t_c = 2 \cdot t_d \tag{3.25}$$

As discussed above, the cut detection was a simple problem, since they produce a high peak in the feature sequence, easy to detect using a simple threshold. In the figure 3.15(1), the dissimilarity measurements $d$ are shown for the game #11 of the *WC* 2010, for a span of 2170 frames. A zoom of a particular peak in that sequence can be seen in figure 3.15(2), where a cut can be easily distinguished, showing the good performance of the *HVBP* as a dissimilarity measure. Also for dissolves (see figure 3.16), the *HVBP* correctly represents the existence of a given effect. The combined results of the transitional effects detection and the shot classification are presented in the table 3.3, and will be discussed in the section 3.4.

For the case of long dissolves (see figure 3.17), the *HVBP* produces high peaks in the feature sequence, but the content of the video in such cases also confuses the detector, since similar high peaks are produced. This is still an open problem and more attention is to be given to it. For the present study, no many long dissolves were present in the video material, and therefore no much attention was given to that particular problem. In the conclusions section, possible directions will be pinpointed.

## 3.4 Shot Classification

At the end of the frames' selection process, there is the classification and selection of the shots that are relevant for the rest of the processing. The classification is performed based on the *field of view size*[4], which defines how much of the objects of interest and their surrounding elements are visible in the camera's *field of view*. Two factors determine the field of view size: the distance between the objects and the camera, and the focal length of the used lens[5]. This parameter has a very important effect in the narrative power of a shot, therefore the TV broadcasting staff, make use of it to convey different emotions of the game to the viewers. Based on the classification of the shots with respect to the field of view size, it is possible to name three main standardised type of shots (*together with the desired emotional effect*):

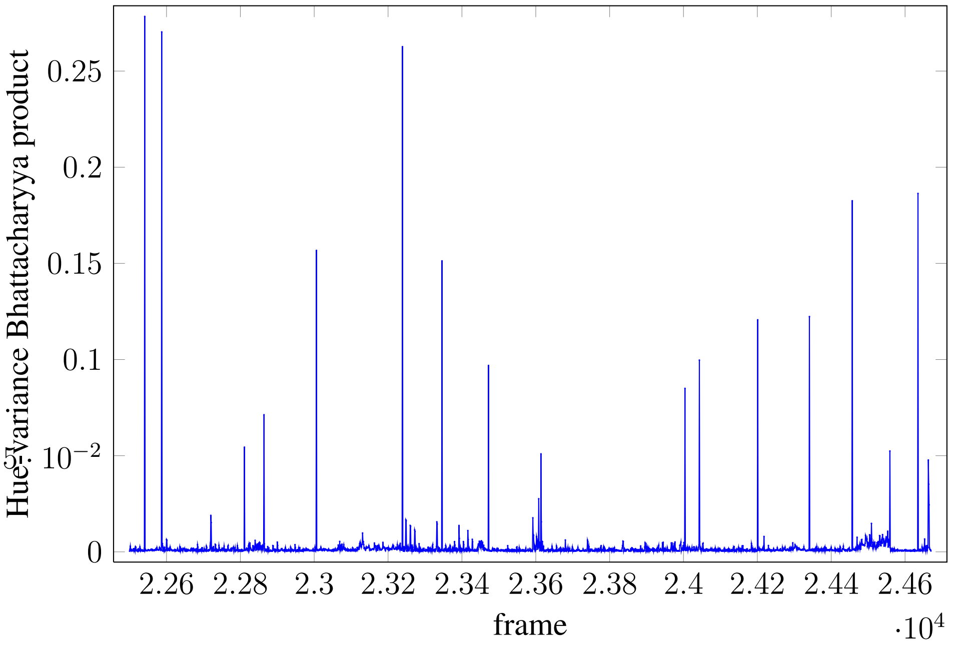
**Long shot** Small human figures are dominated by the surroundings. *Characters seem vulnerable to larger forces beyond their control.*

**Medium shot** A full-length view of a human subject. *They are good in showing facial expressions and body language.*

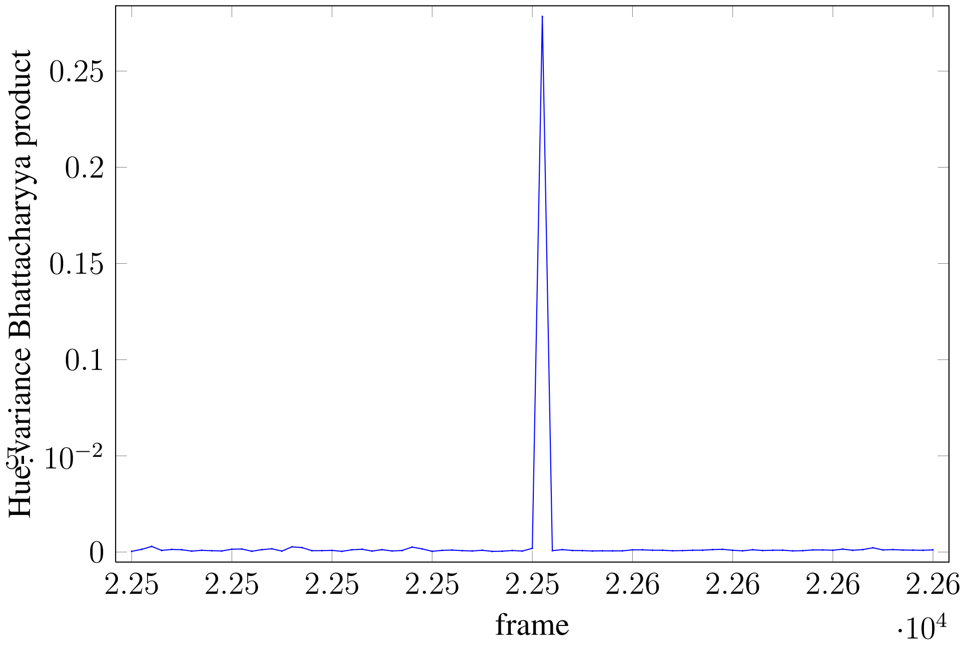
**Close-up** There are various degrees of close-up depending on how tight (zoomed in) the shot is: Medium-close-up (for example: head and shoulders), Close-up (for example head), and Extreme Close-up (for example just an eye). *Close-ups show characters' emotions.*

[4] Commonly, the term used is *field size*, but in the current context, it might be confused with the pitch size, therefore the less ambiguous term field of view size is used.

[5] The shorter the lens's focal length, the wider the angle of view.



(1) dissimilarity values for frames 22500 − 24670



(2) zoom to the cut at 22540-22541

FIGURE 3.15 *Dissimilarity values of cuts examples for game #11 of the WC 2010: Italy vs Paraguay. In (1), the frames corresponding to cuts: 22540-22541, 22586-22587, 22810-22811, 22863-22864, 23005-23006, 23238-23239, 23345-23346, 23471-23472, 23613-23614, 24003-24004, 24042-24043, 24200-24201, 24340-24341, 24456-24457, 24558-24559 and 24634-24635, present high discontinuity values, which makes them simple to detect. In (2) a zoom showing the single peak corresponding to the cut at 22540-22541.*

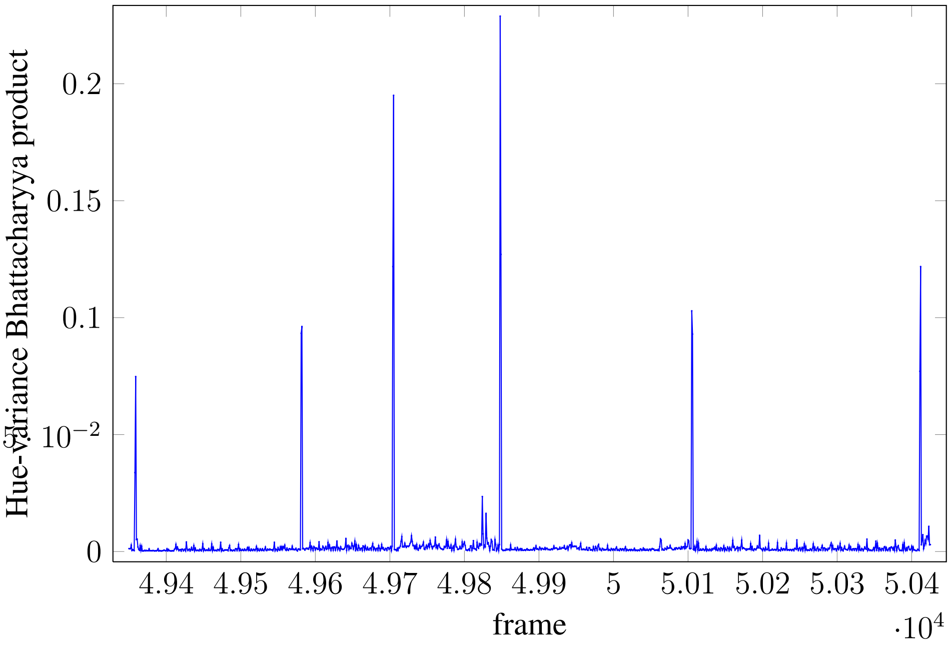


FIGURE 3.16 *Dissimilarity values of short dissolves examples for game #18 of the WC 2010: France vs Mexico.The frames corresponding to dissolves: 49358, 49581, 49704, 49848, 50105, 50411 present a relative high amplitude, easy to detect using a simple threshold.*

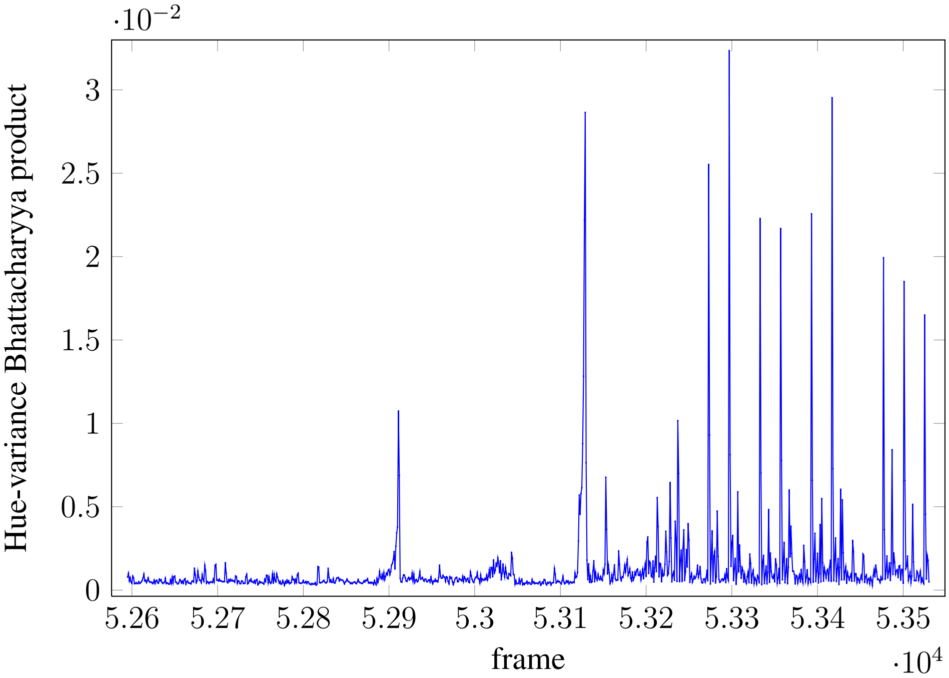


FIGURE 3.17 *Dissimilarity values of long dissolves examples for game #04 of the WC 2010: Korea Republic vs Greece.The frames corresponding to dissolves: 52280-52811, 53016-53044, and 53121-53129 present similar amplitudes with respect to the remaining frames, where normal content variation is present, which complicates the detection of the dissolves with a simple threshold.*

Another classification is based on the camera placement and angle: for example the high-angle camera (in the middle line of the pitch), or the crane camera (behind the goals). Depending on each of the cameras used and their focus adjustments, the size of the players vary in a range of $[10 \times 30, 60 \times 300]$ pixels$^2$. Examples of the view by each of those cameras is shown in figure 3.18(1) and figure 3.18(2)). But the proposed system for player detection and localisation of the players can deal with such variations, therefore, the classification of the camera type by its placement is immaterial. Each of the different field of view size can be associated with a specific camera view, for example, the long shot corresponds to a far-view shot, a medium shot is a mid-view shot, and finally, a close-up is a near-view respectively. For examples of how this kind of camera view shots are in the case of football, see figure 3.18. Since our interest is to estimate the position of the players on the pitch, then the shots of interest are those taken with the far-view perspective (see figure 3.18(3), and figure 3.18(4)). A further distinction between the two different far-views shown in figure 3.18(3), and figure 3.18(4) is not made, since the proposed system is capable of dealing with the player size in any of those views, as mentioned above for the camera placement too.

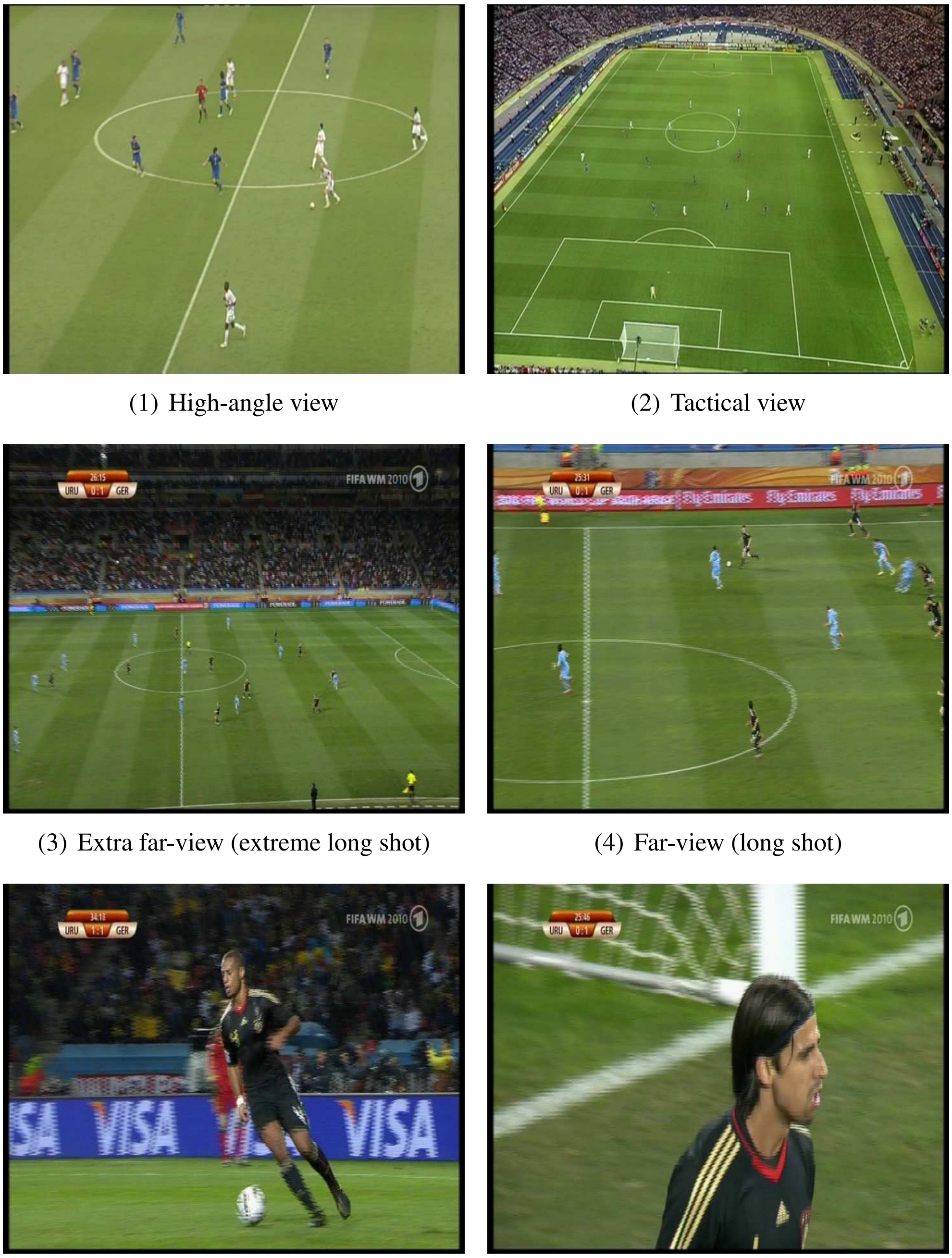
For the classification of the shots, features describing the dominant pitch colour are used, as well as features to describe the structural content of the video, since it has been seen in the professional literature, that the use of only colour information, or only structural features is not informative enough to correctly perform the classification [Lie99]. Therefore, it is necessary to use combined features to represent this both orthogonal characteristics of the signal. The features used in the proposed algorithm for classification are the *grass pixel ratio* and the *players pixel ratio*. The procedure to classify the shot candidates $C$ into the interesting far-view shots $S$, that it the mapping $s$ (see equation 3.27), is described in the algorithm 3.4.

$$s : C \to S \tag{3.26}$$

$$\{c_i\} \to \{s_i\} \tag{3.27}$$

**Algorithm 3.4** Algorithm to find the mapping $s : C \to S$ to obtain, out of the shot candidates, the far-view shots

```
Calculate the thresholds $t_{GPR}$ and $t_{PPR}$ from $C$: equation 3.30 and equation 3.31
Uniformly sample each candidate shot $c_i$ to select some representative frames $\{f_i\}$
Calculate the features $GPR$ and $PPR$ for each $f_i$: equation 3.28 and equation 3.29
Obtain a mean value for the features $\overline{GPR}$ and $\overline{PPR}$ for the current shot $c_i$
Classify each candidate shot $c_i$ based on the criteria from equation 3.32
```



(1) High-angle view

(2) Tactical view

(3) Extra far-view (extreme long shot)

(4) Far-view (long shot)

(5) mid-view (medium shot)

(6) near-view (close-up)

FIGURE 3.18 *Examples of different types of camera views (types of shots), depending on the field of view size*

The *grass pixel ratio* $GPR$ represents the ratio between the number of pixels with green colour with respect to the number of pixels in the image inside the valid mask $M$, which eliminates from the original image $I$, the possible static overlays, such as logos or brands; or other undesired image areas not for consideration. The $GPR$ is a simple, quick and easy to obtain characteristic, but its decision power is high as it will be shown below. The $GPR$ is obtained as shown in equation 3.28.

$$GPR = \frac{\sum_{Gr \neq 0} 1}{\sum_{M \neq 0} 1} \tag{3.28}$$

where $Gr$ is the image corresponding to the grass regions, which is obtained as described in the algorithm 3.5.

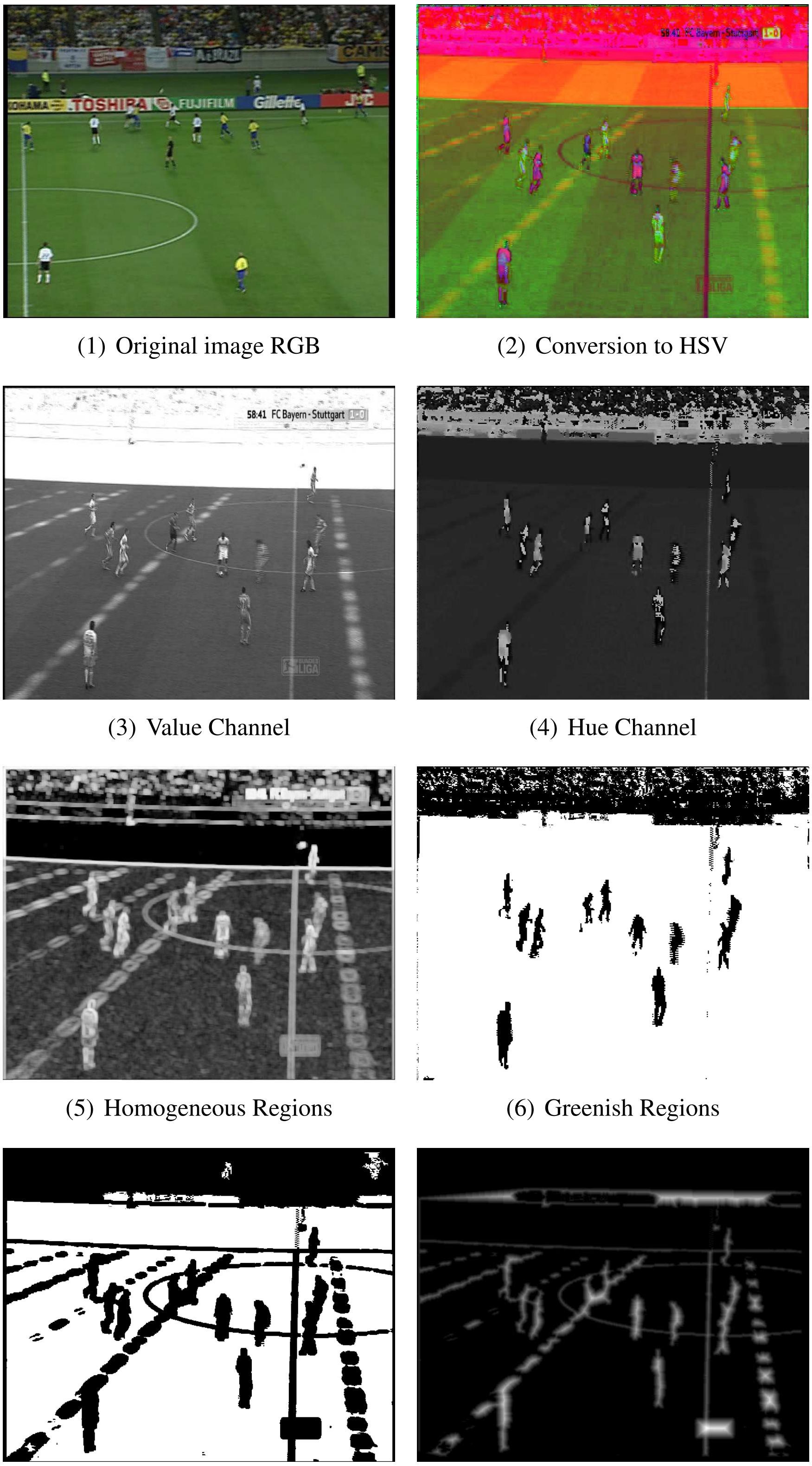
In the other hand, the $PPR$ feature represents the intrinsic geometrical relation in the images, that allows the incorporation of structural information into the classification process. The $PPR$ measure corresponds to a mean distance with respect to the number of elements that correspond to players in the image. The distance $D$ that is averaged is obtained as an approximation to the Voronoi diagram for the binary image corresponding to the objects on the pitch. The *players pixel ratio* $PPR$ is obtained as shown in equation 3.29.

$$PPR = \frac{\sum D}{\sum_{P \neq 0} 1} \tag{3.29}$$

where $D$ is the distance transform of the objects image $O$ on the pitch, including players, referees, ball, lines and others, and $P$ is the image corresponding to the players' regions inside the grass regions $Gr$. The algorithm 3.5 describes how the required images are obtained, and the figure 3.19 shows some intermediary resulting images from that process.

**Algorithm 3.5** Algorithm to find the required images for the calculation of the shot classification features $GPR$ and $PPR$

```
Convert $RGB \rightarrow HSV$
Threshold $H$ to the range $[60, 180]$ to obtain simple greenish regions $G$
Generate the local spatial log variance $Var$ for the value image $V$
Threshold (Otsu's method) the bi-Gaussian distribution of $Var$ to obtain the homogeneous regions $Hom$
Generate the grass regions as $Gr = Hom \wedge G$
Obtain the pitch region $Pi$ as the convex hull of $Gr$
Get players image $P = \{\neg g | \forall g \in Gr\} \wedge Pi$
Get objects image $O = \neg Gr \wedge Pi$
Calculate the distance transform $D$ of $O$
```



(1) Original image RGB

(2) Conversion to HSV

(3) Value Channel

(4) Hue Channel

(5) Homogeneous Regions

(6) Greenish Regions

(7) Grass Regions

(8) Distance Transform

FIGURE 3.19 *Steps in the generation of the shot classification features*

For each candidate shot $c_i$, a set of frames are sampled to calculate the $GPR$ and $PPR$ features only for that selected frames, in order to accelerate the classification. Empirically, a value of $10\%$ of the images per candidate shot $c_i$ are selected as descriptors for the whole shot. The classification is performed by means of two thresholding processes, applied iteratively to each of the mentioned features. The values of the thresholds $t_{GPR}$ and $t_{PPR}$ are obtained as follows:

$$t_{GPR} = med_{GPR} - std_{GPR} \tag{3.30}$$

$$t_{PPR} = med_{PPR} \tag{3.31}$$

where $med_X$ and $std_X$ are the median and the standard deviation of the corresponding empirical density distribution of the feature $X$. The models for the density distribution of the features used are uni-Gaussians. Then, the far-view shots $s_i$ are those that satisfied the following relation:

$$t_{GPR} < \overline{GPR}_{c_i} \wedge \overline{PPR}_{c_i} < t_{PPR} \tag{3.32}$$

In the table 3.3, the *true positives* are the number of shots that are far-view, and the *true negatives* are the shots that are non-far-view. For example, for the game #1 in the table, there were $201 + 196 = 397$ candidate shots, from which $201$ were far-view shots, and $196$ were non-far-views. The table also presents the shot classification results, with an account of the false positive classified shots (those non-far-views miss-classified as far-views), and the false negative classified shots (those far-views miss-classified as non-far-views). Also, the performance of the classifier is evaluated by means of the standard metrics: precision, recall and F1-score, from equations: 3.7, 3.8, and 3.9 on page 64.

As can be seen from the table totals, 39269 candidate shots were classified, from which, the $43,96\%$ $(17261)$ were true positive, and the remaining $56,04\%$ $(22008)$ were true negatives. This produces a classification with an overall precision of $93,61\%$, a recall of $95,36\%$, and a F1-score of $94,48\%$. This results show that the performance of the classifier is excellent, and also that is reliable, since it was tested in many games of the *WC* 2010.

| Game id | Transitional Overlays | False Positives | False Negatives | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|---|---|
| 1 | 52 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 4 | 76 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 5 | 146 | 5 | 0 | 96,69 | 100,00 | 98,32 |
| 9 | 105 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 10 | 116 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 11 | 110 | 1 | 0 | 99,10 | 100,00 | 99,55 |
| 12 | 64 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 13 | 80 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 14 | 88 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 15 | 100 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 16 | 114 | 1 | 0 | 99,13 | 100,00 | 99,56 |
| 17 | 90 | 1 | 0 | 98,90 | 100,00 | 99,45 |
| 18 | 100 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 19 | 109 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 20 | 112 | 8 | 0 | 93,33 | 100,00 | 96,55 |
| 21 | 85 | 6 | 0 | 93,41 | 100,00 | 96,59 |
| 22 | 118 | 2 | 0 | 98,33 | 100,00 | 99,16 |
| 23 | 98 | 5 | 0 | 95,15 | 100,00 | 97,51 |
| 24 | 134 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 25 | 112 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 26 | 100 | 1 | 0 | 99,01 | 100,00 | 99,50 |
| 27 | 118 | 1 | 0 | 99,16 | 100,00 | 99,58 |
| 28 | 96 | 2 | 0 | 97,96 | 100,00 | 98,97 |
| 29 | 122 | 0 | 0 | 100,00 | 100,00 | 100,00 |

TABLE 3.2 *Evaluation of the proposed method based on the mean cumulative difference for the detection of the transitional dynamic overlays for games of the* WC *2010*

| **Game id** | **Transitional Overlays** | **False Positives** | **False Negatives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 30 | 128 | 17 | 0 | 88,28 | 100,00 | 93,77 |
| 31 | 96 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 32 | 110 | 1 | 0 | 99,10 | 100,00 | 99,55 |
| 33 | 116 | 6 | 0 | 95,08 | 100,00 | 97,48 |
| 37 | 101 | 4 | 0 | 96,19 | 100,00 | 98,06 |
| 39 | 116 | 5 | 0 | 95,87 | 100,00 | 97,89 |
| 41 | 117 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 43 | 116 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 45 | 116 | 1 | 0 | 99,15 | 100,00 | 99,57 |
| 47 | 87 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 49 | 86 | 6 | 0 | 93,48 | 100,00 | 96,63 |
| 50 | 118 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 51 | 134 | 2 | 0 | 98,53 | 100,00 | 99,26 |
| 52 | 86 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 53 | 120 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 54 | 92 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 55 | 98 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 56 | 120 | 5 | 0 | 96,00 | 100,00 | 97,96 |
| 57 | 102 | 3 | 0 | 97,14 | 100,00 | 98,55 |
| 58 | 186 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 59 | 115 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 60 | 114 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 63 | 90 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 64 | 180 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| **Total** | **5189** | **83** | **0** | **98,43** | **100,00** | **99,21** |

TABLE 3.2 *(cont.) Evaluation of the proposed method based on the mean cumulative difference for the detection of the transitional dynamic overlays for games of the* WC *2010*

| Game id | True Positives | True Negatives | False Positives | False Negatives | Preci-sion (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|---|---|---|
| 1 | 201 | 196 | 20 | 22 | 90,95 | 90,13 | 90,54 |
| 4 | 560 | 534 | 27 | 30 | 95,40 | 94,92 | 95,16 |
| 5 | 315 | 545 | 47 | 41 | 87,02 | 88,48 | 87,74 |
| 9 | 583 | 424 | 21 | 18 | 96,52 | 97,00 | 96,76 |
| 10 | 400 | 544 | 14 | 2 | 96,62 | 99,50 | 98,04 |
| 11 | 418 | 689 | 24 | 21 | 94,57 | 95,22 | 94,89 |
| 12 | 141 | 242 | 22 | 15 | 86,50 | 90,38 | 88,40 |
| 13 | 187 | 335 | 16 | 14 | 92,12 | 93,03 | 92,57 |
| 14 | 150 | 32 | 13 | 4 | 92,02 | 97,40 | 94,64 |
| 15 | 754 | 425 | 73 | 81 | 91,17 | 90,30 | 90,73 |
| 16 | 767 | 427 | 54 | 50 | 93,42 | 93,88 | 93,65 |
| 17 | 246 | 501 | 10 | 3 | 96,09 | 98,80 | 97,43 |
| 18 | 260 | 512 | 23 | 8 | 91,87 | 97,01 | 94,37 |
| 19 | 213 | 343 | 25 | 10 | 89,50 | 95,52 | 92,41 |
| 20 | 249 | 428 | 39 | 32 | 86,46 | 88,61 | 87,52 |
| 21 | 286 | 475 | 31 | 18 | 90,22 | 94,08 | 92,11 |
| 22 | 198 | 472 | 11 | 7 | 94,74 | 96,59 | 95,65 |
| 23 | 301 | 385 | 16 | 6 | 94,95 | 98,05 | 96,47 |
| 24 | 429 | 539 | 42 | 33 | 91,08 | 92,86 | 91,96 |
| 25 | 935 | 386 | 40 | 38 | 95,90 | 96,09 | 96,00 |
| 26 | 238 | 384 | 13 | 7 | 94,82 | 97,14 | 95,97 |
| 27 | 995 | 376 | 128 | 62 | 88,60 | 94,13 | 91,28 |
| 28 | 407 | 488 | 37 | 15 | 91,67 | 96,45 | 94,00 |
| 29 | 504 | 589 | 38 | 32 | 92,99 | 94,03 | 93,51 |
| 30 | 304 | 540 | 9 | 7 | 97,12 | 97,75 | 97,44 |

TABLE 3.3 *Evaluation of the proposed method for shot classification for games of the* WC *2010 [SC13]*

| **Game id** | **True Positives** | **True Negatives** | **False Positives** | **False Negatives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|---|
| 31 | 376 | 630 | 24 | 28 | 94,00 | 93,07 | 93,53 |
| 32 | 329 | 468 | 17 | 14 | 95,09 | 95,92 | 95,50 |
| 33 | 280 | 463 | 31 | 2 | 90,03 | 99,29 | 94,44 |
| 37 | 125 | 230 | 17 | 17 | 88,03 | 88,03 | 88,03 |
| 39 | 788 | 461 | 90 | 113 | 89,75 | 87,46 | 88,59 |
| 41 | 128 | 333 | 18 | 7 | 87,67 | 94,81 | 91,10 |
| 43 | 279 | 508 | 54 | 45 | 83,78 | 86,11 | 84,93 |
| 45 | 390 | 399 | 37 | 34 | 91,33 | 91,98 | 91,66 |
| 47 | 174 | 410 | 14 | 6 | 92,55 | 96,67 | 94,57 |
| 49 | 351 | 575 | 35 | 18 | 90,93 | 95,12 | 92,98 |
| 50 | 382 | 644 | 53 | 42 | 87,82 | 90,09 | 88,94 |
| 51 | 305 | 503 | 16 | 6 | 95,02 | 98,07 | 96,52 |
| 52 | 314 | 480 | 15 | 12 | 95,44 | 96,32 | 95,88 |
| 53 | 361 | 403 | 45 | 34 | 88,92 | 91,39 | 90,14 |
| 54 | 182 | 434 | 14 | 8 | 92,86 | 95,79 | 94,30 |
| 55 | 229 | 465 | 27 | 9 | 89,45 | 96,22 | 92,71 |
| 56 | 326 | 526 | 21 | 11 | 93,95 | 96,74 | 95,32 |
| 57 | 260 | 448 | 22 | 11 | 92,20 | 95,94 | 94,03 |
| 58 | 678 | 757 | 57 | 18 | 92,24 | 97,41 | 94,76 |
| 59 | 266 | 537 | 20 | 12 | 93,01 | 95,68 | 94,33 |
| 60 | 174 | 409 | 15 | 1 | 92,06 | 99,43 | 95,60 |
| 63 | 250 | 258 | 40 | 29 | 86,21 | 89,61 | 87,87 |
| 64 | 273 | 561 | 27 | 18 | 91,00 | 93,81 | 92,39 |
| **Total** | **17261** | **22008** | **1502** | **1071** | **93,61** | **95,36** | **94,48** |

TABLE 3.3 *(cont.) Evaluation of the proposed method for shot classification for games of the* WC *2010 [SC13]*

# Chapter 4

# Spatial Segmentation

> I find that the harder I work, the more luck I seem to have.
>
> *(Thomas Jefferson)*

In the chapter 3 the selection mechanism of the informative shots required by our semantic annotation system was explained. The selected shots correspond to the far-view scenes, scenes where our proposed semantic annotations are possible. The next step is the extraction of useful information from those selected shots. By useful information, it is meant the positional data of the objects of interest (or targets[1]): players, referee, and the ball, during the course of the game, or more specifically during each of the previously selected shots. The positional data of the objects of interest is represented by the trajectories followed by each object over the field of play. Later, in chapter 5 it will be seen how an abstract model, that helps in the interpretation of the game, is built using the trajectories of the objects of interest.

In this chapter, the challenges that have to be surpassed in order to generate the desired trajectories will be presented and explained. The proposed solutions found in the professional literature to deal with the aforementioned challenges will be discussed. Finally our proposed method for generating the required trajectories will be described and explained in detail.

[1] Target is the name given to the object of interest in the tracking specialized literature, therefore both terms will be used interchangeably.

## 4.1 Tracking Challenges

In general terms, *tracking* is the process of estimating over time the state of the targets[2] using measurements taken from the input [BSF88]. The state of the targets is defined depending on the particular application, and may be comprised of kinematic components such as position, velocity or acceleration, or other features extracted from the input image sequence, such as colour, or shape. In order to obtain the trajectories of the targets out of the selected frame sequences, a method of video tracking must be employed. *Video tracking* is the process of estimation over time of the location of the objects of interest using a video signal as input [Emi11]. Tracking the positions of targets from a video signal is a difficult task, due to the complex relation between the objects, and the corresponding projected images into the camera plane. Many factors influence this complex relation, such as the relative positions of the objects with respect to the camera and among themselves, varying illumination conditions, local deformations of the objects, partial or total occlusions, as well as many more. A classification of the different challenges faced by a tracking system is depicted in the figure 4.1. *Scene change challenges* refers to those challenges due to reasons attributable to the external environment. *Clutter* for example, is the existence of other objects besides the targets with similar appearance characteristics, which might generate false negatives and confuse the tracker. Examples of cluttering can be seen in figure 4.2.

Another scene challenge is related to the *Illumination Conditions*. This includes effects of shadows: like those produced by the roof of stadiums, or by a flying camera, as well as the shadows produced by the players themselves. Also the ambient illumination, which changes with the weather conditions is considered in this subcategory, for example, if clouds temporary occludes the sun rays or if it is raining, then the appearance of the targets will change. During the 90 minutes of a game, the Earth rotation will change the angles of the sun rays with respect to the surface normals of the targets, and therefore the radiance[3] that reach the camera sensors will change. Illustrations of the above mentioned challenges can be seen in figure 4.3.

The last subcategory: *Camera Challenges*, refers to events occurring in the camera, such as noise, that depends on the quality of the sensors[4], or errors during the transmission of the video signal. Also, when there is a fast camera movement, motion blurring occurs that affects

[2] In the professional literature a clear distinction is established between the single-target scenario and the multiple-target scenario, but in the present study were several objects are of interest, it is simply immaterial. The distinction will be made only when necessary to describe others' algorithms.

[3] The radiance $L$ defines the amount of radiant flux per unit solid angle per unit projected area of the emitting source, and it is given by the following expression: $L = \dfrac{d^2\Phi}{d\Omega dScos\theta}$ [JH00].

[4] In the case of digital video, the sensors are commonly *Charge-Coupled Devices (CCDs)*.

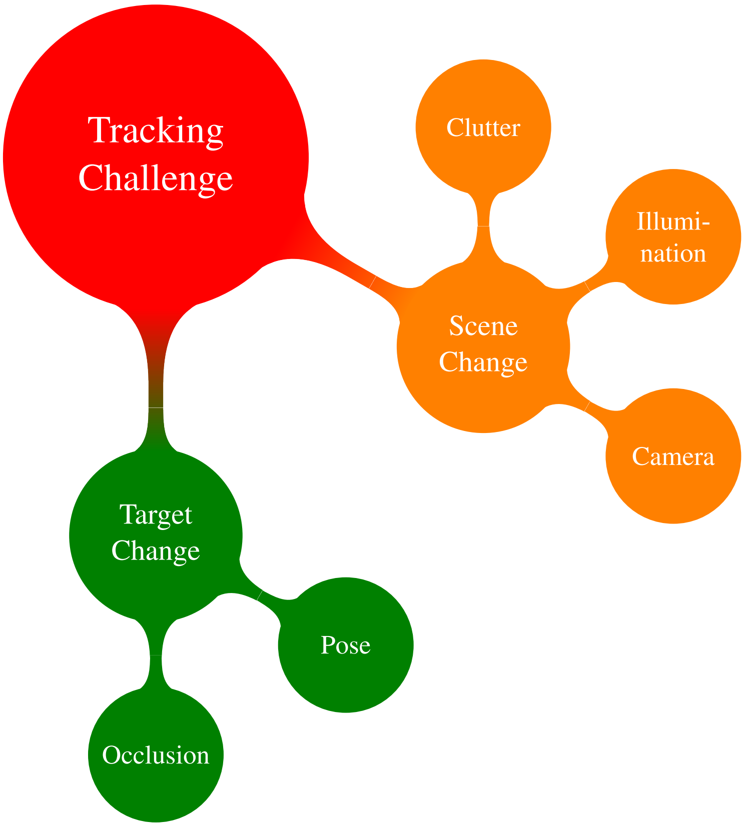
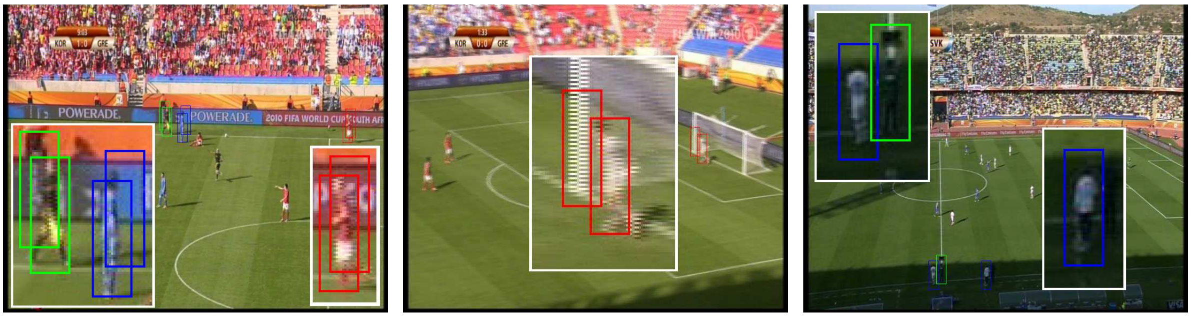
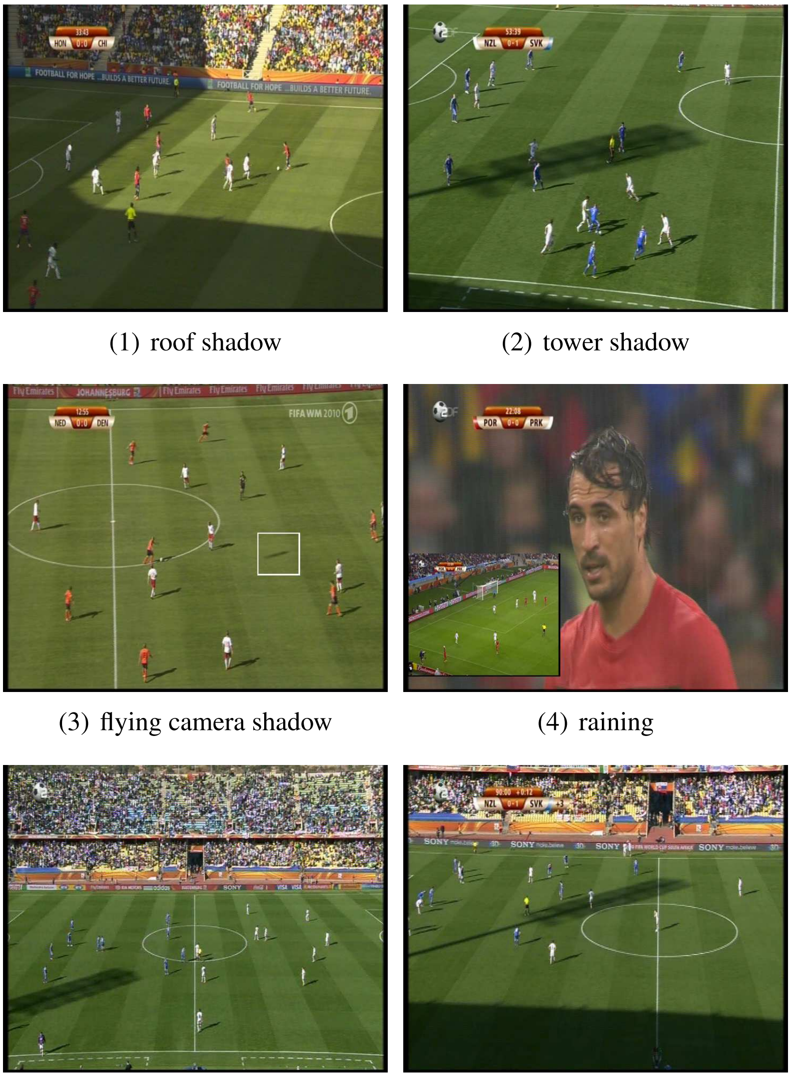


FIGURE 4.1 *Tracking Challenges Taxonomy. The target changes refer to variations in the target appearance, or to interactions between the targets that make it difficult to follow their trajectories. The scene changes refer to alterations in the contextual situation of the tracking, due to external agents that are not directly related to the targets.*



(1) advertisement clutter (2) poles clutter (3) staff clutter

FIGURE 4.2 *Clutter examples. In the subfigure (1) the background region contains colours and shapes that can be misinterpreted as targets. In the subfigure (2), the poles at the goal might be confused as the goalie and generate false negatives. Finally in subfigure (3), members of the team staff, referees and other players share similitudes with the real targets. Zoom-ins of the interesting regions are inscribed in white rectangles.*



(1) roof shadow

(2) tower shadow

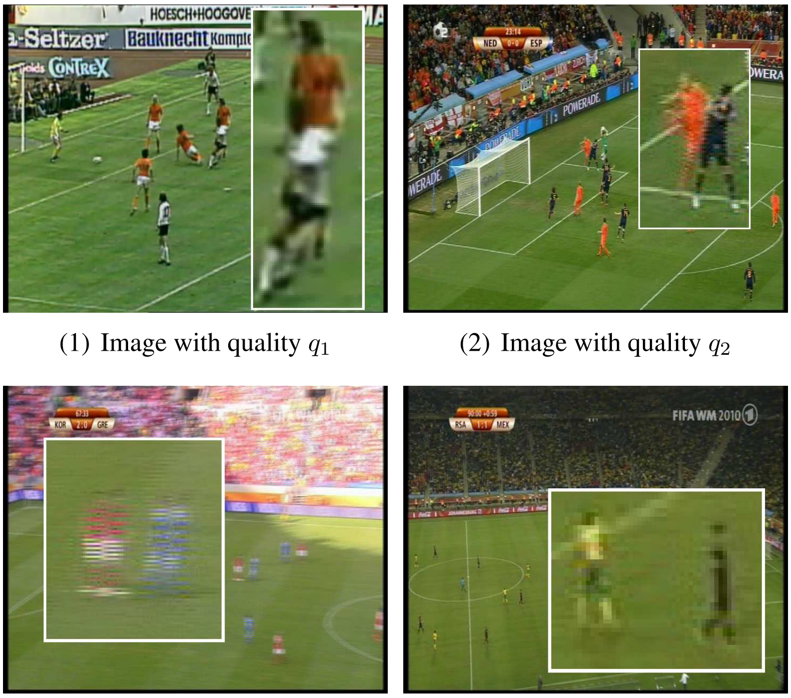
(3) flying camera shadow

(4) raining

(5) sun rays at time $t$

(6) sun rays at time $t + \delta_t$

FIGURE 4.3 *Examples of illumination conditions that can affect the performance of the tracker. In subfigure (1) the shadow of the roof is laying on the field, partially changing the appearance of some targets. Subfigure (2) and (3) show the shadows of other objects on the field: a tower, and a flying camera (inscribed in a white rectangle). The subfigure (4) shows a game where it was intermittently raining, causing certain change in appearance of the targets. Subfigure (5) and (6) show the effect of the Earth rotation during the course of the game, and how this changes the radiation angles from the sun rays on each target's surface. The sub-figures also show the shadows of the players due to different illumination arrangements.*



(1) Image with quality $q_1$

(2) Image with quality $q_2$

(3) Motion blur

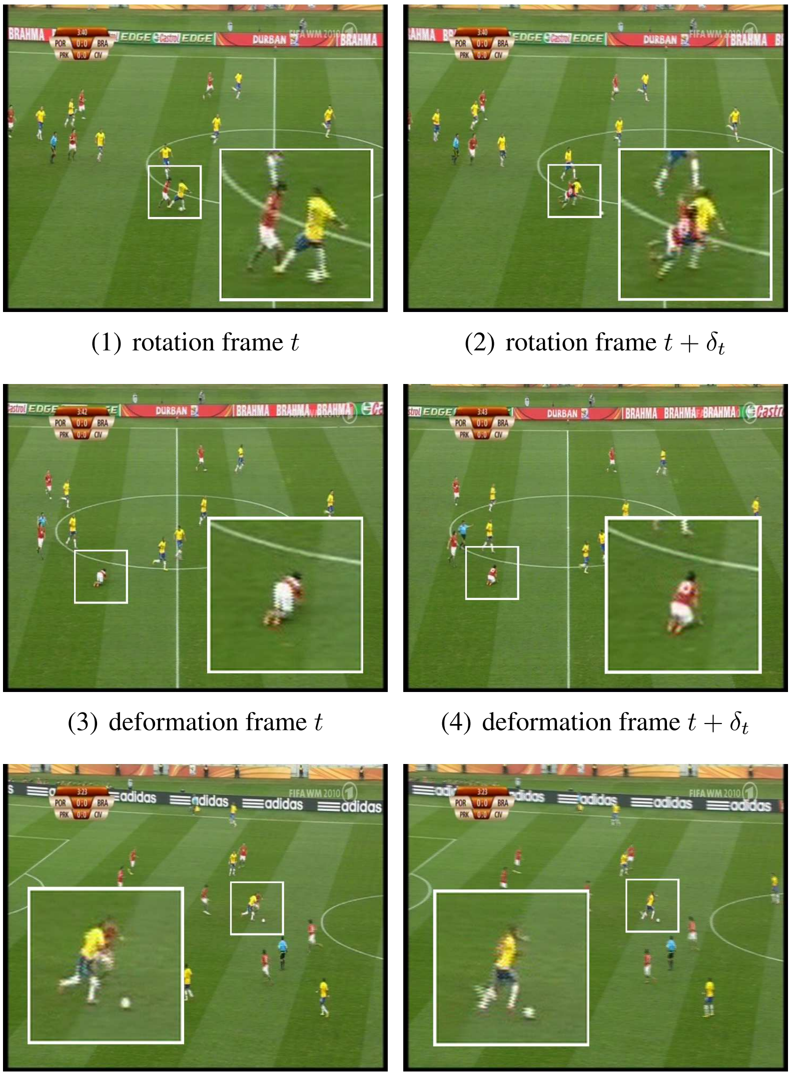
(4) Dominant colour effect

FIGURE 4.4 *Camera related challenges. In the first row, images with different qualities are shown. In the subfigure (3) the effect of a rapid camera movement producing motion blur can be observed. In the same image, the interlacing is visible due to an increasing in the shift between the captured lines in the interleaving (known also as combing). The subfigure (4) shows how even if the players dressed coloured uniforms, the pixels surrounding the players borders are always mixed up with the dominant colour (green in this case), affecting the discrimination by colour of the players. Zoom-ins of the interesting regions are inscribed in white rectangles.*

the look of the targets. In outdoor sports in general, and particularly in football, the play field is covered by natural grass or some artificial turf, that has a particular dominant colour. This dominant colour, which is in general the scene dominant colour too, influences the way how the colours of the targets are seen by the camera, by mixing it with the targets' colours. Examples of such effects are depicted in figure 4.4.

With respect to the *Target Change Challenges* category, it includes the changes in the targets that may generate troubles for the tracker. Target challenges include the target pose changes and target occlusions. The *Pose Changes* include rotations and deformations of the targets[5], and the *Occlusions* can be of total or partial nature, depending if the involved target is completely covered by another object or only if it is partially hidden. Illustrations of this challenges category are shown in figure 4.5.

[5] Technically pose changes should include translations too, but since the aim of the tracker is to follow moving targets, then it makes no much sense to consider translations as as special problem.



(1) rotation frame $t$

(2) rotation frame $t + \delta_t$

(3) deformation frame $t$

(4) deformation frame $t + \delta_t$

(5) partial occlusion

(6) total occlusion

FIGURE 4.5 *Target changes challenges. In the first row a player falling down is shown. The player is rotating while keeping his body straight. This rotation is changing the relative positions of the colours of the target with respect to the camera view, as well as changing the bounding box of the target, which might be problematic too. The second row displays a deformation of the player since he is bending his body until being sit. Certain assumptions made about the shape of a player must be managed with care, since they may be unsatisfied. A partial occlusion and a total occlusion are shown in the subfigure (5) and (6) respectively. The interesting regions are inscribed in white rectangles.*

## 4.2 Related Work

In the following paragraphs, the related professional literature will be discussed. The ideas and problems present in the methods will be presented. It is important to consider too, that there is also a lot of industrial research taking place in this field, driven by the commercial potential of such systems, as already discussed in the chapter 2.

When the videos are taken from static cameras, it is possible to use background subtraction to detect the targets in the images, as is done in [FLB06a] and [NSC06]. Background subtraction has been used in the analysis of sports (see [KCM03, JdOA04, IS07b, LMS$^+$08, GP08]). The technique exploits the fact that since the camera is not moving, or more precisely, not changing its intrinsic and extrinsic parameters, then a model of the background is created previously to the acquisition of a new image for analysis. This model is then used as a reference that is compared to the analysed image, and using thresholding techniques, the regions corresponding to the targets are extracted. The models range from a simple reference image, a Gaussian model for each pixel ([WADP97], [SG99]), and a model extracted using principal component analysis: eigenbackgrounds [ORP00, RRG$^+$04, HWTH11]. In the simpler case, an image is taken before the targets enter the scene, and this image is the reference with which the new images will be compared to. Pixel-wise subtraction and thresholding will produce the regions that do not match with the reference, and since the camera was not moving, those areas must correspond to the sought targets. The main problem is that the background image is not fixed, since it has to adapt to illumination changes (e.g. shadows due to Earth rotation), to motion changes in the scene (changes in the stands), as well as others. For surveys covering the research in those areas see [Pic04, PF08].

The background subtraction technique is of course not suitable for the more general case of dynamically moving cameras, as in our case, since the TV cameras are typically following the actions happening around the ball. For moving cameras, it is possible to gradually construct a mosaic image that contains all the visible background during the game, and then use image differences with respect to this mosaic image to find the players, as in [BBK05], and [KH00]. This mosaic image consist of creating a bigger image containing the field area and its surroundings, such that, when a new image is to be analysed, its position in the mosaic is searched and when found, image differences – as in the static camera case mentioned above – will produce the desired target regions. Since the camera have a limited view area, this mosaic image have to be built by stitching images one after another, taking new pieces of information, each time the camera moves to regions not visited before. For some more references see [BDH04, LTSH07, YlJBkY02, RCKL06, JdOA04, FLB06b]. A problem of using such technique is that the reference image is to be gradually built as more areas get discovered by the movement of the camera, so the mosaic is not ready for the first images of the input video. Besides, since the camera is changing its intrinsic parameters, the focus with which the shot is being filmed is changing, and then mosaics of different sizes are to be constructed. Also the shadows (of players, roofs, or other structures in the stadium) might affect the creation of the reference image, dynamic overlays (see chapter 3) will negatively affect the creation of the mosaic, as well as the variation of the appearance in the stand regions. In order to deal with the shadows problem, in [ROTJ04] the colour of the shadows is learnt using an unsupervised learning approach, and then are excluded from the targets regions.

Since the football players are dressed with colours to differentiate the team affiliations, and since the colour of the field is designed to be distinguishable by the referees and the audience, segmentation by colour should be well suited for the task of foreground segmentation (see [JC03, VMP03, NB01, BGB+07, GBvHH+07, BBG+06a, SMND07]). In [VMP03] an adaptive colour space segmentation is used to find the regions in the image containing targets. It is also possible to learn colour classes using machine learning techniques in an unsupervised fashion, as in [SMND07]. A popular method for the colour modelling of targets is the use of *Gaussian Mixture Models (GMMs)* to represent the classes of colours corresponding to different geometrical regions of the player, so as to model the t-shirt, the shorts, and the socks (see [NB01, BGB+07, GBvHH+07, BBG+06a]). In fact, the *GMM* was used in the ASPOGAMO project until our new method, that will be presented below, superseded it. The main problem was the selection of the colour samples to build the models, since they were selected manually, and it was a time-consuming task. Some other methods, by means of a trained colour histogram of the green colour, which is the dominant colour in the scene, find the connected regions representing the field area, and the holes that were left correspond to the sought players. In this method, the colours for the players are not modelled directly, but only the dominant green. For more on this, see [AG06, HLB07, LTL+07].

Finally, segmentation by motion is used to detect the target regions [EBMM03, GP08, KZW+08, HWL+09]. The motion of the background is modelled under the assumption of a consistent motion, which make it suitable for the static and dynamic camera cases. The motion is represented by motion vectors, obtained by means of the optical flow (see for example [LK81]), which is based on the colour consistency between the frames. To reduce outliers and improve the robustness of the motion model, RANSAC [FB81] is commonly applied. The idea is similar to that of the background subtraction, in the sense that, all pixels whose motion vectors do not match the expected background motion, are selected as target regions. But this method is not accurate enough due to limitations of the optical flow.

An associated problem to the detection of the target regions, is the dealing with the field lines, since a subtle interaction with the targets can be present. For example, for the static camera case, a method for learning the colour of the lines based on a Gaussian mixture learning approach is used in [vHH10]. Also, a method of projecting virtual field lines onto the real image to delete them, based on the camera parameters estimation proposed in [Ged09], together with morphological operations for keeping the lines at the official length of 13cm ([FIF11]) is used in [vHH10] for dynamic cameras, in order to detect players that went on crossing any field line. The problem with this approach is twofold: first, circles and arcs are approximated by line segments, and therefore sometimes the representation is not precise enough. Second, and more importantly, when the estimation of the camera parameters is not accurate enough, then the lines are not appropriately detected.

Another problem found is that for the evaluation of some of this systems, only short videos were used [NSC06], and it is important to use whole games, to test the robustness of the system at facing environmental and illumination condition changes. Additionally, most of the proposed methods work offline with the whole sequence as input, and not online as our proposed method.

In the next two sections, our proposed method to generate the spatial segmentation will be explained. Basically, the spatial segmentation can be divided into two main parts: the *Target Detection* sub-module, which consists in the detection of the target region candidates, and the *Target Localisation* sub-module, which is in charge of finding each individual target inside the candidate regions.

## 4.3 Target Detection

Our new proposed method assumes that the field area is a green coloured region mostly homogeneous in texture, and that objects not satisfying these criteria must correspond to target areas. The main idea is to combine colour information using a particular colour space representation for the image, and homogeneity descriptors to delimit the field area. Then, the found field region is inverted, and the remaining regions not belonging to the field are considered as candidate regions corresponding to the sough targets. The colour space representation used is the $HSV$: hue, saturation, and value; that codifies the colour in a convenient manner to fulfil our purposes. As a measure of the homogeneity of a region, we propose the use of the local spatial variance [SC04, Fre00].

The local spatial variance consist of the calculation of the variance $\sigma^2$ in local neighbourhoods in the image of value $V$, obtained from the $HSV$ representation. This local neighbourhood is defined by systematically moving a square window of predefined size $(2W + 1) \times (2W + 1)$ around the image, where $W \in \mathbb{N}$. In order to obtain the local spatial variance, the local spatial mean $m$ on the image $V$ is first required. For a particular pixel position, denoted by $(i, j)$ (row, and column in the image coordinate system), the particular value of the local spatial mean is obtained as follows:

$$m(i,j) = \frac{1}{(2W+1)^2} \sum_{a=-W}^{W} \sum_{b=-W}^{W} V(i+a, j+b) \tag{4.1}$$

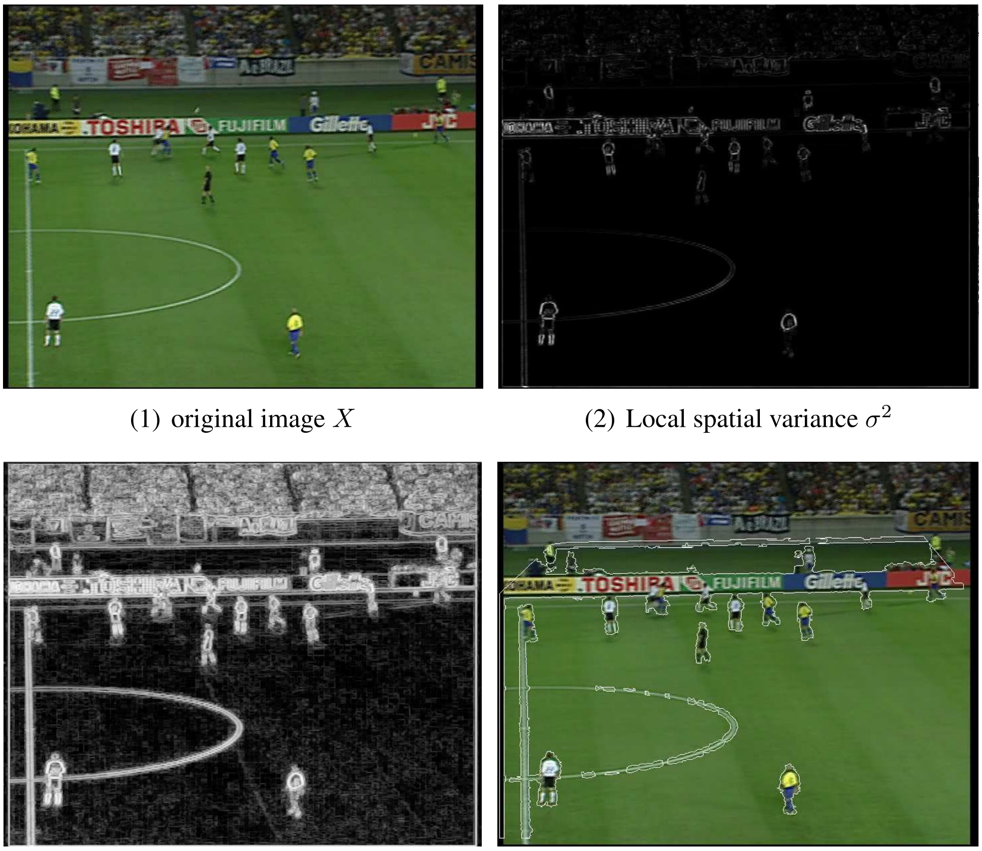
Similarly, for the local spatial variance (see figure 4.6):

$$\sigma^2(i,j) = \frac{1}{(2W+1)^2+1} \sum_{a=-W}^{W} \sum_{b=-W}^{W} \{V(i+a, j+b) - m(i,j)\}^2 \tag{4.2}$$

The selection of $W$ depends on the homogeneity characteristics of the input image, and this is a matter of the quality of the input video. The worst the quality of the video, the larger the window size. This prevents that the noise negatively affects the expected shape of the probability distribution (this will be explained later). Most of the time a general neighbourhood of $3 \times 3$, that is a value of $W = 1$, will perform adequately, but some particular cases require the increase of this parameter. To increase the speed of the calculations, the concept of integral images is used.

Once the local spatial variance is calculated, the empirical probability density function is obtained from the normalised histogram, as can be seen in figure 4.7. This empirical density function is comprised of two components that correspond to the regions of low variance, and high variance in the image, respectively. But, as can be seen from equation 4.2, the local spatial variance will never produce negative values due to the square of the differences, therefore the distributions representing the low and high variances will necessarily be skewed. This characteristic leads to the argument that the joint probability density function is not well represented by a bi-normal distribution. We have considered that the joint probability density function fits a bi-log-normal distribution, [LSA01]. This fact can be visually checked by taking the logarithm of the local spatial variance, and plotting the resulting histogram to see a bi-normal distribution, [LSA01], as can be seen from figure 4.7; or by using a probability plot [Jan10].

Once the empirical probability density function has been transformed, it can be observed that it is bi-modal, with each peak corresponding to the low-variance and high-variance components in the image. This two peaks basically corresponds to the homogeneous and the heterogeneous regions on the image, respectively. Based on a Maximum Likelihood approach



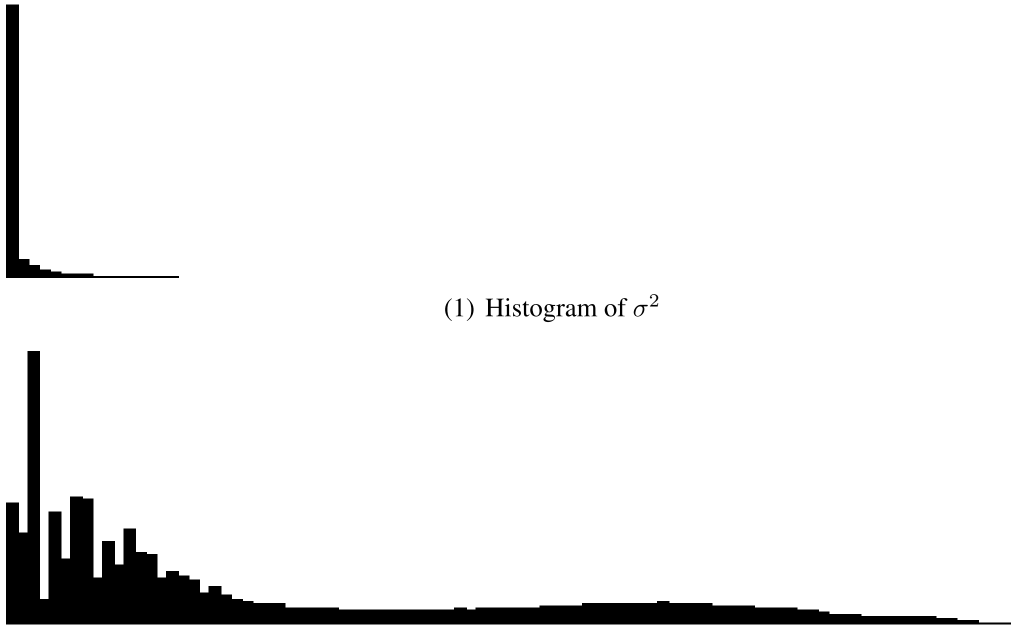
(1) original image $X$

(2) Local spatial variance $\sigma^2$

(3) Logarithm of the local spatial variance $\log(\sigma^2 + 1)$

(4) Blobs obtained

FIGURE 4.6 *Example of the local spatial variance. In the subfigure (1), the original RGB image is shown. The subfigure (2) shows the result of applying the equation 4.2 to the original image in subfigure (1). The subfigure (3) shows the result of applying the logarithm to the variance image in subfigure (2). Finally, the subfigure (4) shows the generated player blobs as the detection of the targets.*



(1) Histogram of $\sigma^2$

(2) Histogram of $\log(\sigma^2 + 1)$

FIGURE 4.7 *Histograms of the local spatial variance. The log-normal distributions in subfigure (1) are represented by normal distributions, when applying the logarithm function.*

(Kittler e Illingworth, Kurita, or Otsu methods), the optimum threshold value for the segmentation of the low-variance and high-variance regions is found.

The result of the detection of the homogeneous regions is combined with the chromatic information generated using the dominant colour of the image to describe the field. A simple range thresholding in the hue component $H$ of the image is enough to get the greenish regions. Depending on the quality of the video input signal, sometimes the compression have moved the expected green peak to other colours like blueish, so the hue range is to be empirically selected to match the characteristic of the dominant colour of the image[6]. Morphology was used in order to remove undesired spurious regions. The algorithm describing the strategy is depicted in algorithm 4.1.

Since the inside of shadows projected on the field are greenish and homogeneous, no especial treatments are required to deal with them, which is an advantage, when compare with [ROTJ04]. The borders of the shadows produce high values in the variance, but since they are very small, these regions get deleted by the morphology operations mentioned above. With respect to the field lines issue, it will be explained in section 4.4 how our proposed method deals with the problem.

[6] This adjustment mechanism is also useful when playing football in the snow, as have occurred in other games.

**Algorithm 4.1** Algorithm to obtain the targets candidate regions.

```
Convert $RGB \rightarrow HSV$
Threshold $H$ to the range $[60, 180]$ to obtain simple greenish regions $G$
Generate the logarithm of the local spatial variance $Var$ for the value image $V$
Threshold (Otsu's method) the bi-Gaussian distribution of $Var$ to obtain the homogeneous regions $Hom$
Generate the grass regions as $Gr = Hom \wedge G$
Obtain the pitch region $Pi$ as the convex hull of $Gr$
Get targets image $T = \{\neg g | \forall g \in Gr\} \wedge Pi$
Get the contour $c_t$ for each target candidate region found: $t \in T$
```

## 4.4 Target Localisation

Now that the candidate target regions were found, the task at hand is to find the target silhouettes inside the candidate regions (or blobs, for easiness in the following discussion). In the past, in the ASPOGAMO project ([BGB$^+$07, GBvHH$^+$07, BBG$^+$06b]), the approach for detecting and localising the players was to use a colour template matching technique. The idea was sustained by the fact that the players are dressed in a way as to be visually distinctive for the easy to the spectators in the recognising of the team associations[7]. Also, other contrasting colours are selected, such as the green field and the white ball and lines, making the colour a very informative feature. In order to use the colour as a clue for the target segmentation, then samples of the different colours present in the input video were to be taken. The samples were manually selected from images sparsely chosen from the input video, which of course required the complete sequence to be available, or at least a long enough video segment, as to have objects with all the expected colours of the football objects: goalies, field players, referees, and the ball. This colour samples were used to build colour classes, each associated to a specific colour. The colour classes were modelled by Gaussian distributions in the $RGB$ colour space, generated semi-automatically using k-means clustering in the $HSV$ colour space using the colour samples. Based on this colour classes, colour templates were generated, each for each player type. The colour template consisted of three parts to model the colours of the t-shirt, the shorts, and the socks, as can be seen in figure 4.8. Each section of the template might comprise several colour classes.

Since the size of the players as seen by the camera changes during the game, due to their relative motion, several colour templates were created to match each situation. So, for each image position, a colour template for each player type was created. Also, since the players are,

[7] To avoid the distinction between players and referees, the referees will be considered to be part of another team, with its own distinctive uniform, and therefore will be also called players.

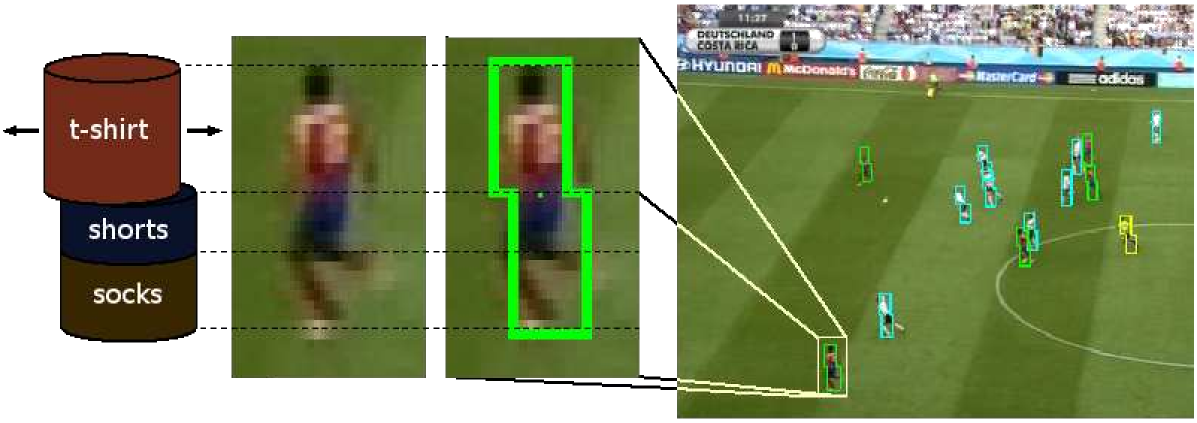


FIGURE 4.8 *Example of a Player Colour Template, showing the three components of the model: the t-shirt, the shorts, and the socks.*

of course, not always in a standing position, then five different configurations of the template were considered, each one shifting the t-shirt part some pixels to the left or to the right by a fixed predetermined amount. Other pose changes of the targets, such as rotations or deformations were not considered by the method. Then the position of the target was assumed to correspond to the place where the colour template similarity measure was higher, this considering the five geometrical configurations discussed. Two additional constraints were used in order to deal with situations when the teams shared some colours, or when the same team uses only one colour in its dressing, or when the size of the template did not match the size of the players, or when occlusions were present. The constraints were the size and the compactness constraints, which tried to keep the target locations limited in dimension and position. Still, too many false negatives were produced by the method, and the template matching approach was shown to be insufficient.

In order to deal with the template matching issues, our already described target detection method was proposed, see section 4.3. The principal idea was to limit the search space of the algorithm, and try to reduce the false negatives, and produce more accurate results. Together with the target detection method, a better localisation technique was devised. Since the players are being subtle to pose changes, then the location mechanism has to be flexible to be able to detect such cases, and therefore the rigid template matching is not appropriate. The proposed model used an adaptive model for each player, that is updated in every new acquired frame. The model is composed of the silhouette contour of the player, as well as of the texture and colour of its inner region.

A block diagram of the target detection and localisation algorithm is shown in figure 4.9. The input block models the image acquisition process, where basically the targets in the scenes are being sensed by the video camera. The noise represents the uncertainty existent in the data acquisition process. The input block generates a sequence of images, and for a particular iteration $k$, the corresponding image is $I_k$. The image $I_k$ is processed by the *Target Detection Algorithm*, which using the chromatic-homogeneity model described in section 4.3, finds a set of blobs $Z_k$ containing possible target regions. Each blob $z_{j_k} \in Z_k$ might contain a single target or multiple targets, depending on the current state of the system.

The state of the system at time $k$ corresponds to the location of each target $t_{i_k}$ in the set of tracked targets $T_k$. To represent the state of the system, the stacked state vector $\mathbb{X}_k = \{X_{i_k} | i \in \{T_k\}\}$ is used. Each target state vector $X_{i_k}$ consists of kinematic information relative to the current position of the target, as well as appearance information, including shape of the silhouette and colour-texture descriptors for the target. This information is referred to as *target descriptors*.

The *Target Localisation Algorithm* generates the estimated locations of the targets for the $k$-th iteration, based on the minimum weight matching of the generally unbalanced and incomplete weighted bipartite graph $B = (\mathbb{X}_{k-1}, \mathbb{X}_k, E, W)$. The edges $E$, and corresponding weights $W$ are given by the expressions in equation 4.3, and equation 4.4.

$$E = \{(X_{1_{k-1}}, X_{1_k}), \cdots, (X_{1_{k-1}}, X_{n_k}), \cdots, (X_{m_{k-1}}, X_{n_k})\} \tag{4.3}$$

$$W = \{w_{1,1}, \cdots, w_{1,n}, \cdots, w_{m,n}\} \tag{4.4}$$

where $m = |\mathbb{X}_{k-1}|$, and $n = |\mathbb{X}_k|$, and the weights $w_{i,j}$ quantifies the match of the target descriptors between the two vertices.

The graph $B$ represents the association of the targets at time $k-1$ and at time $k$. $B$ is said to be generally unbalanced because at some points, a target can get out of sight in the time period $(k-1, k)$, or a new target can appear in the scene in that lapse, so leaving the cardinality of the two disjoint sets of vertices uneven. The characteristic of $B$ of being generally incomplete, is a desired feature, since this allows the minimum matching optimisation to be obtained separately in each partition of the graph, with the consequent increase of speed in finding a solution. Therefore, to make the graph $B$ a generally incomplete graph, a kinematic model of the targets representing basic fundamental physics laws is applied to the motion of the targets, and a consequent prediction of the targets locations can be made, and this will discard the connectivity between some of the vertices of the graph, reducing consequently the search space. Some results of applying the detection-localisation algorithm to artificial and real football games are depicted in figure 4.11 and figure 4.12. The artificial game was picked in order to test the chromatic, and homogeneity features on images with prepared colours,

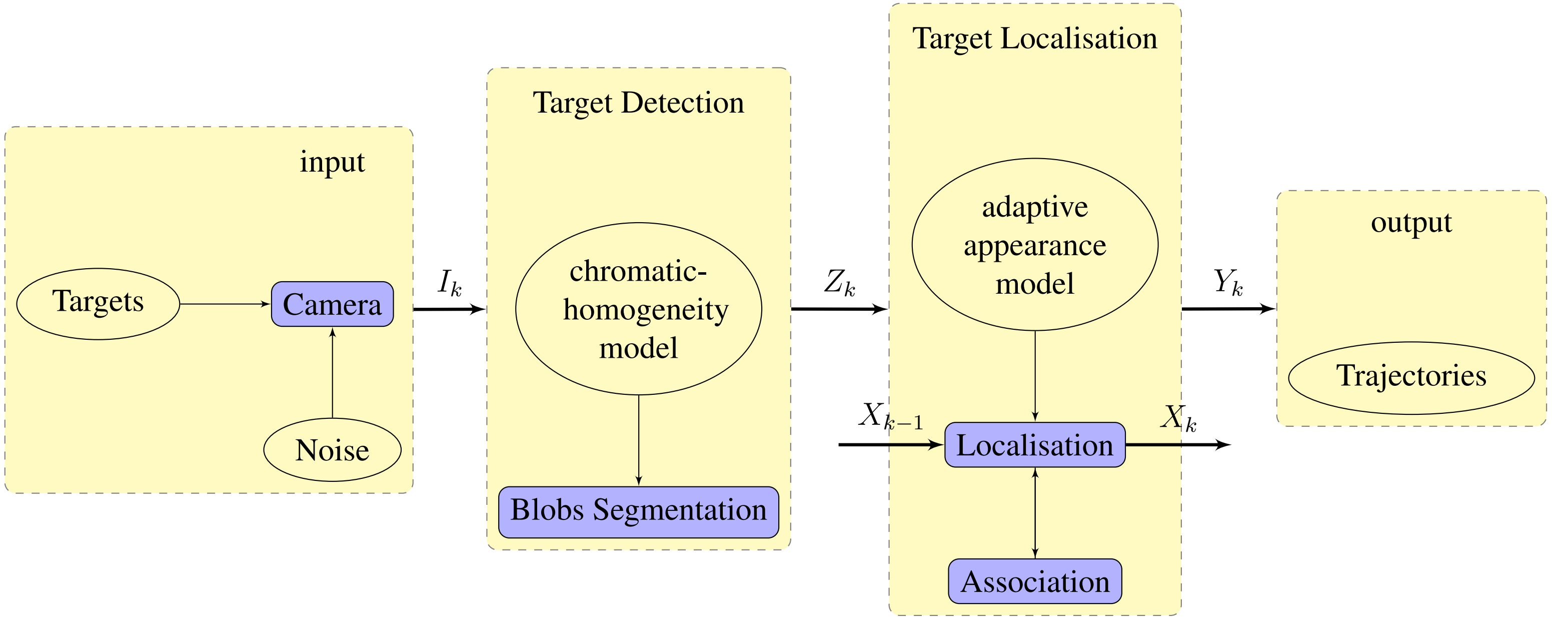
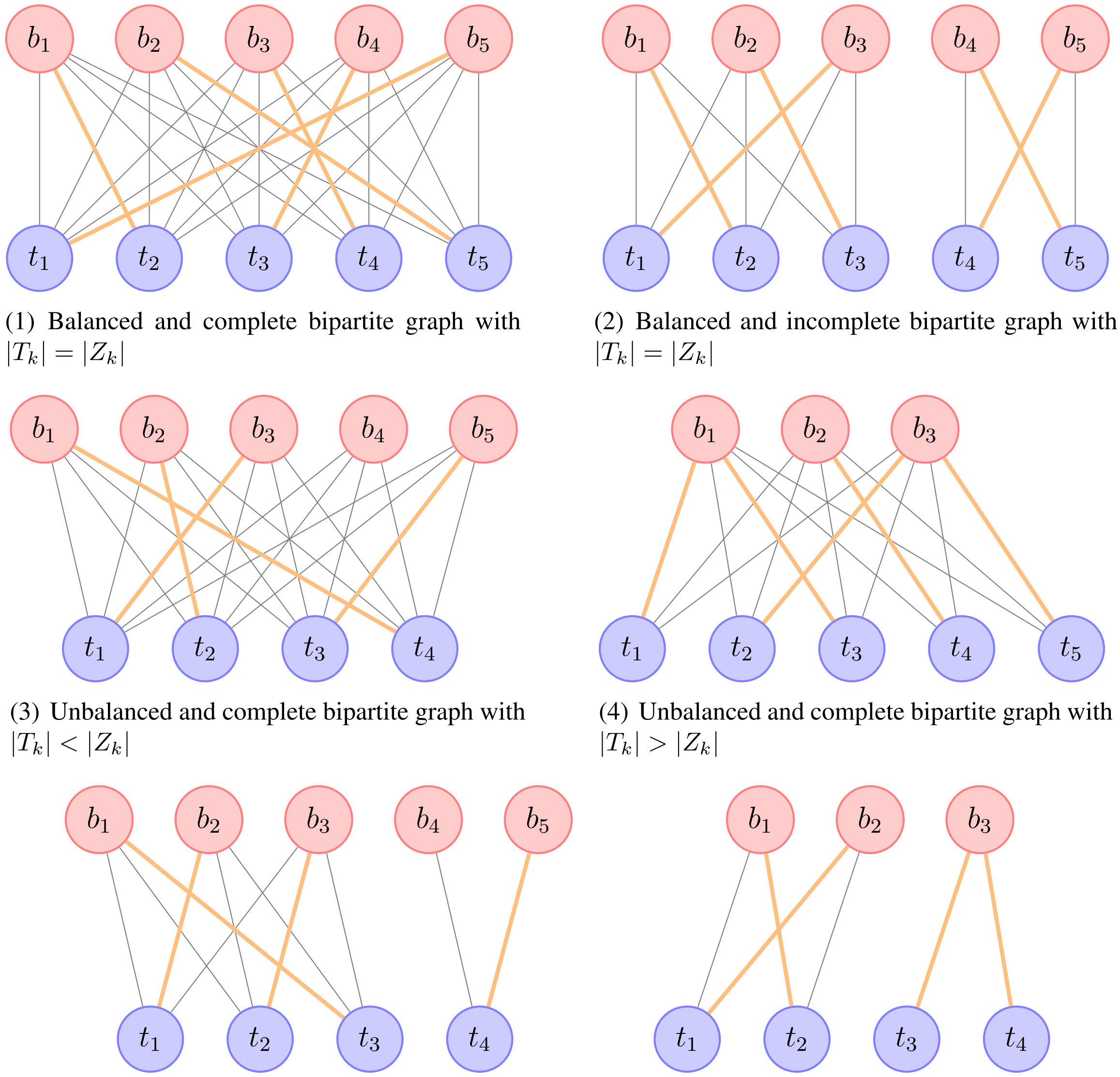


FIGURE 4.9 *Block diagram of the target detection-localisation algorithm. The image $I_k$ is segmented to generate the target candidate regions (blobs) using the chromatic-homogeneity model described in section 4.3. Based on the last state vector $X_{k-1}$, a prediction of the new state over the set of blobs $Z_k$ is generated, using a dynamic model of the targets. By means of an adaptive appearance model, the predictions are validated, and the positions of the targets $Y_k$ are obtained. Also, the new state vector is updated $X_k$ for the next iteration.*



(1) Balanced and complete bipartite graph with $|T_k| = |Z_k|$

(2) Balanced and incomplete bipartite graph with $|T_k| = |Z_k|$

(3) Unbalanced and complete bipartite graph with $|T_k| < |Z_k|$

(4) Unbalanced and complete bipartite graph with $|T_k| > |Z_k|$

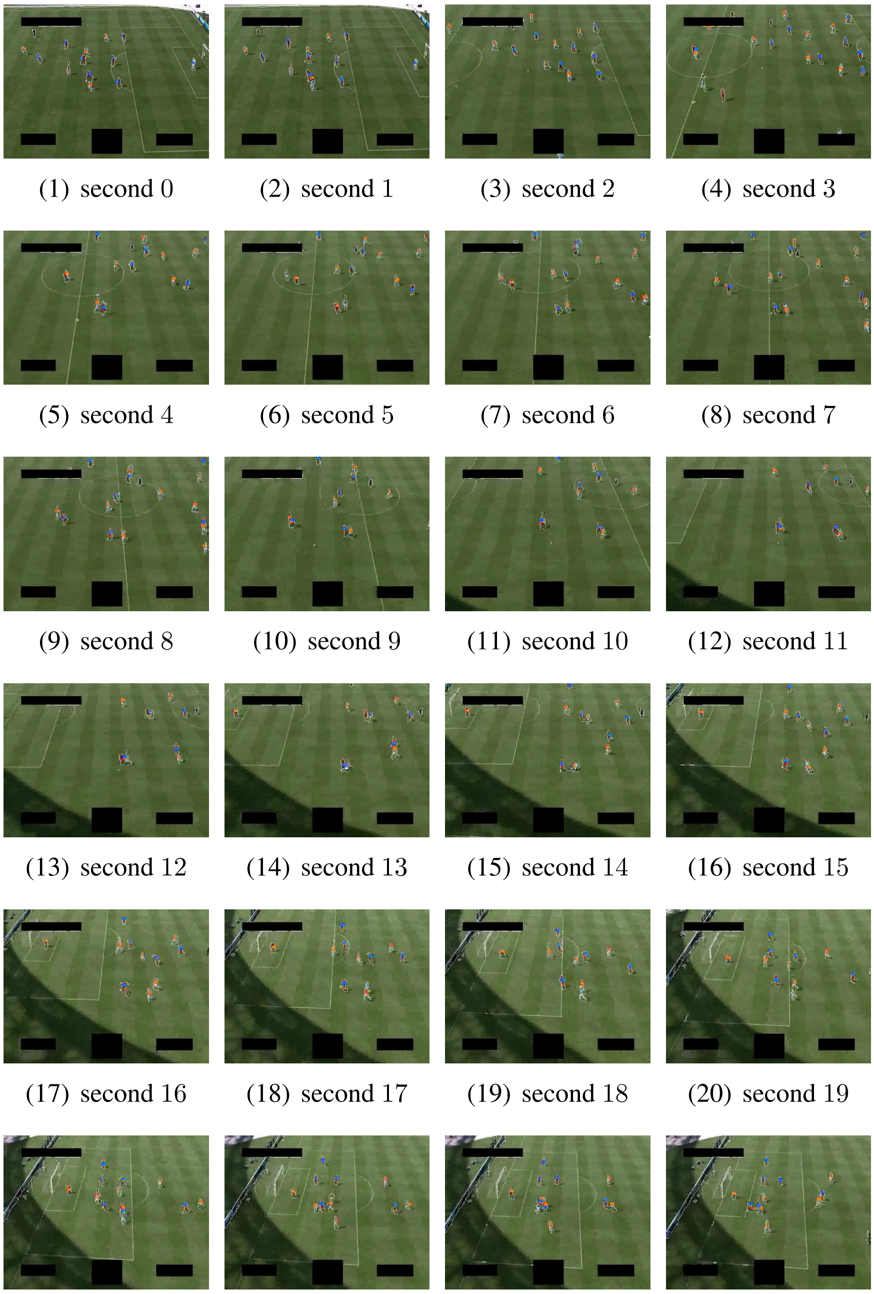
(5) Unbalanced and incomplete bipartite graph with $|T_k| < |Z_k|$

(6) Unbalanced and incomplete bipartite graph with $|T_k| > |Z_k|$

FIGURE 4.10 *Examples of bi-graphs relating the targets with the blobs, with a maximum matching (yellow) solution. As examples of the situations modelled by the bipartite graphs, specific situations from the figure 4.12 will be pinpointed. The subfigure (2) models the situation at time $0s - 1s$, where there are same number of targets and blobs, and due to the dynamic model restriction, not all associations are considered, therefore the incompleteness of the graph. The subfigure (5) models the situation at time $1s - 2s$, where a new blob appears (the one corresponding to the referee), and then a new track is created (*birth*). The subfigure (6) models the situation at time $16s - 17s$, where a join have occurred, that is, due to partial occlusion, two targets correspond to a single blob. A similar join occurred for the targets $t_2$ and $t_5$ in the subfigure (4).*

and without illumination changes, and as can be seen the results are adequate. The real robustness of the algorithm can be observed on the real game shown in figure 4.12, where the shadow of the stadium roof abruptly changes the appearances of the targets. The illumination changes were very strong, as to make two players of the same team placed under different sides of the shadow, look more different than two players of opposing teams under any same side of the shadow. For example, compare the appearances of the red players 61 and 66, and player blue 63 in the figure 4.12(22). The corresponding longer videos can be seen in http://pris.eie.ucr.ac.cr/pris/index.php/Videos.

Once the target locations on the images are found, then the tracking module, consisting of sub-modules: camera tracking, and target tracking (see figure 2.16 on page 44), generate the projection of the targets' locations on the 2D coordinate system of the image, to the 3D positions of the world coordinate system, which are then use to populate the abstract model, that will be described in chapter 5. The description of the tracking modules can be found in [vHH10], and [Ged09], and will not be covered in the present work.



(1) second 0 (2) second 1 (3) second 2 (4) second 3

(5) second 4 (6) second 5 (7) second 6 (8) second 7

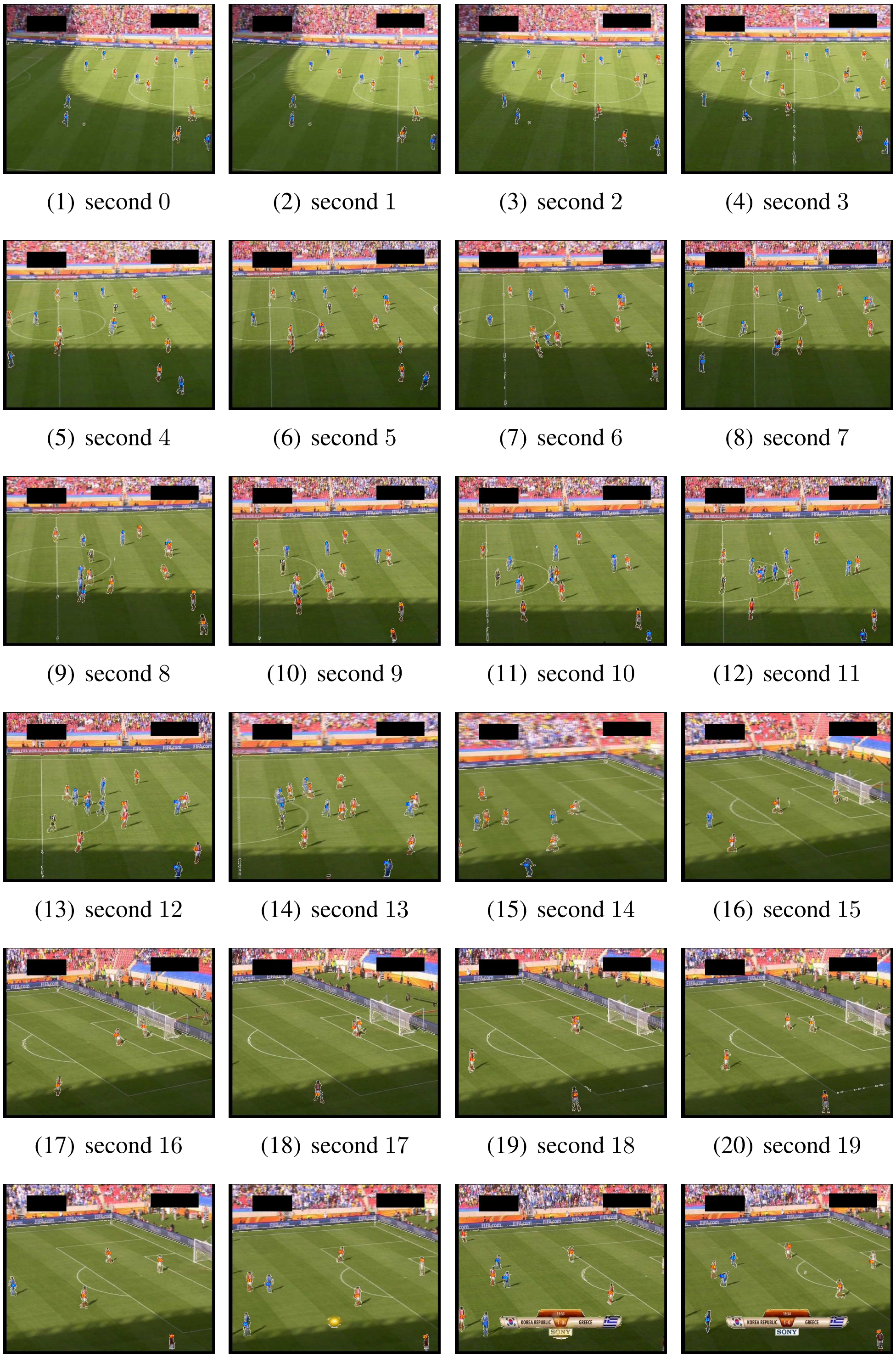
(9) second 8 (10) second 9 (11) second 10 (12) second 11

(13) second 12 (14) second 13 (15) second 14 (16) second 15

(17) second 16 (18) second 17 (19) second 18 (20) second 19

(21) second 20 (22) second 21 (23) second 22 (24) second 23

FIGURE 4.11 *Artificial game for evaluating the detection-localisation algorithm. Check for example how the system can recover from the occlusions, specially at seconds 1-2, 4-5, 6-7, 13-14, and 22-23. The corresponding video can be check at* `http://pris.eie.ucr.ac.cr/pris/index.php/Videos.`



(1) second 0 (2) second 1 (3) second 2 (4) second 3

(5) second 4 (6) second 5 (7) second 6 (8) second 7

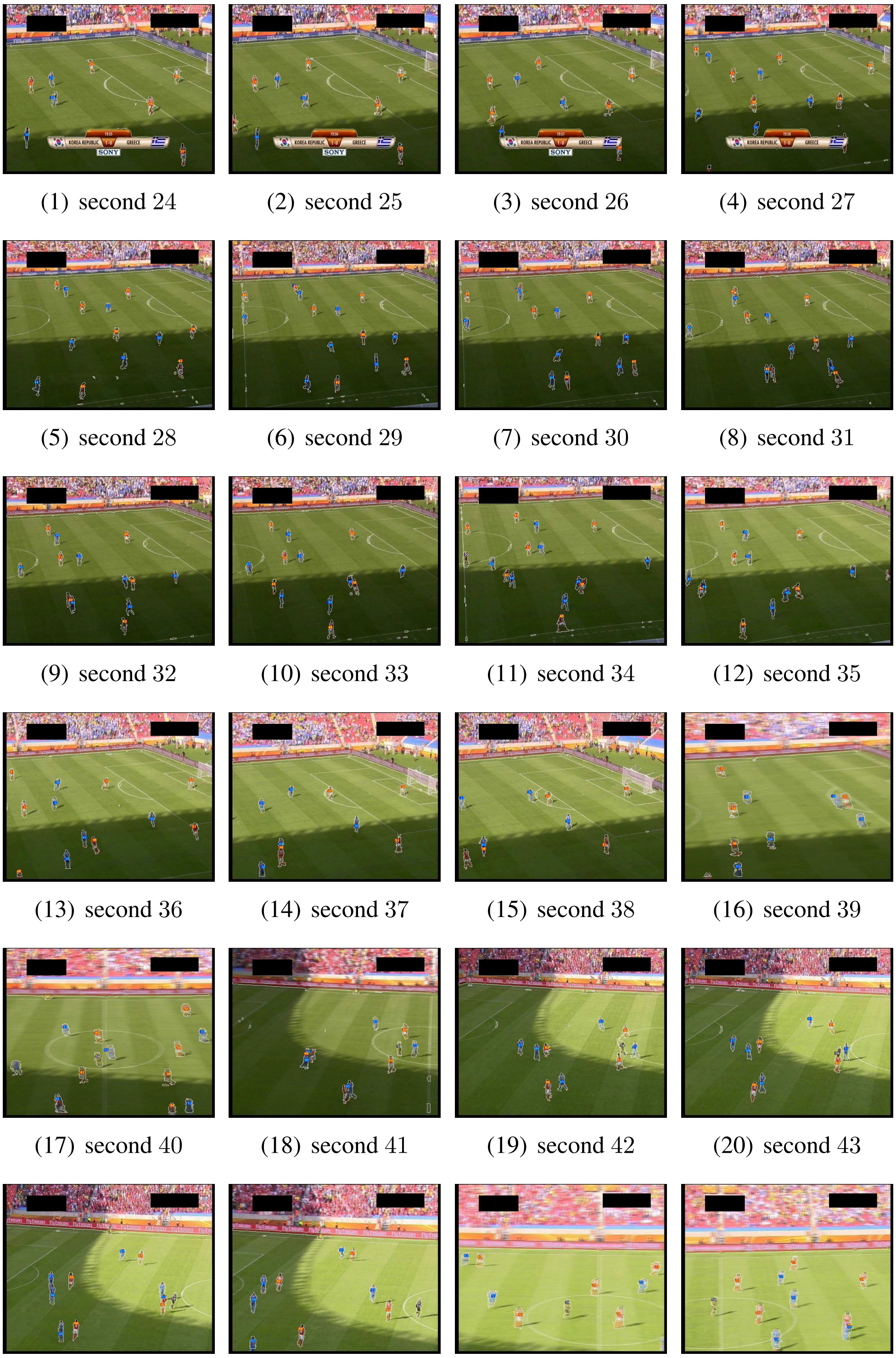
(9) second 8 (10) second 9 (11) second 10 (12) second 11

(13) second 12 (14) second 13 (15) second 14 (16) second 15

(17) second 16 (18) second 17 (19) second 18 (20) second 19

(21) second 20 (22) second 21 (23) second 22 (24) second 23

FIGURE 4.12 *Real game from the WC2010: Greece vs South Korea, for evaluating the detection-localisation algorithm. Check for example how the system can deal with the strong changes in illumination condition, see for example the contrast between second 41 and 47. From second 24-28 the system correctly deals with a dynamic overlay that hides a player. The corresponding video can be check at* `http://pris.eie.ucr.ac.cr/pris/index.php/Videos`.



(1) second 24 (2) second 25 (3) second 26 (4) second 27

(5) second 28 (6) second 29 (7) second 30 (8) second 31

(9) second 32 (10) second 33 (11) second 34 (12) second 35

(13) second 36 (14) second 37 (15) second 38 (16) second 39

(17) second 40 (18) second 41 (19) second 42 (20) second 43

(21) second 44 (22) second 45 (23) second 46 (24) second 47

FIGURE 4.12 *(cont.) Real game from the WC2010: Greece vs South Korea, for evaluating the detection-localisation algorithm. Check for example how the system can deal with the strong changes in illumination condition, see for example the contrast between second 41 and 47. From second 24-28 the system correctly deals with a dynamic overlay that hides a player. The corresponding video can be check at* `http://pris.eie.ucr.ac.cr/pris/index.php/Videos`.

# Chapter 5

# Semantic Segmentation

> The great fallacy is that the game is first and last about winning. It is nothing of the kind. The game is about glory, it is about doing things in style and with a flourish, about going out and beating the other lot, not waiting for them to die of boredom.
>
> *(Robert Dennis Blanchflower)*

The main objective of the current work is the devise and implementation of a software system capable of generate in an automated way, a semantic annotation of football games, which aids in the interpretation of it. In order to fulfil this objective, the game is to be represented using an abstract model powerful enough to extract from it the required information for this interpretation. The main purpose of this chapter is to explain and describe this abstract model, and for that, the objectives of this chapter are threefold. In the first place, in order to model *The Beautiful Game*, its logical structure is presented, with a thorough description of its constitutive parts. This description will be pointed towards to be helpful in identifying how to model the game based on measurable observations. Therefore, in the second place, the abstract model proposed for the representation of the game for its interpretation will be described and explained. The chosen model, as described already in the chapter 2, is a hierarchical model, in which several layers codifies different aspects of the game and give rise to a semantic segmentation of the game. How to build this model, and what specific information it codifies will be extensively exposed. Finally, in the third place, examples of the representational capabilities of the model are described in the last part of the chapter, which shows how the model can be queried to extract the interesting high-level information.

## 5.1 Football Semantics

A review of the semantics of football will be presented here with the aim of stating the language required to model it subsequently. The game can be view as being played in what we called *game phases*, which describe the state of the match with respect to the ball being *alive* or not. Depending on the state of the game and on the state of the players, different events and actions are open to be chosen from, in a particular situation during the game. For example, when a player has the ball, three main courses of action can be taken: passing the ball, dribbling with it, or shooting to the goal. The logical structure of the football game will be described in this section, always with the aim in mind of using such structure as a guide to build an abstract model to represent the game.

### 5.1.1 Game Phases

Football is a game of phases, in which the game is broken down into two main stages: static stage and dynamic stage, with several further divisions each. The *static stage* refers to the moments when the ball is *out of play*, that is, when the whole body of the ball has crossed the *goal line* or *touch line*, or when the referee has stopped the game by any reason. The *dynamic stage* refers to the times when the ball is *in play*, which simply means not to be *out of play*. A classification of the different phases of the game, and their sub-classes is depicted in figure 5.1.

The dynamic stage consists of three phases: the offensive phase, the defensive phase and the transition phase [Mas00, Mas01]. The *offensive phase* of a team is defined as the time periods when the team has the ball, that is, when the team is in *possession*. A team is in possession, when the players of the team are in possession for consecutive time periods. The main objective of the offensive phase is to score a *goal*, of course, and in the pursuing of scoring, other parallel tasks must be performed. For example, the team must refrain itself from losing the ball, by avoiding risking actions. At the same time, the team should move the ball forwards or, when advisable, wait until the team has gotten a better position. Also, the ball has to be passed to an unmarked player in a good position to shoot, which finally, has to shoot effectively.

The offensive phase (or *attacking phase*) can be further divided into three sub-phases: build-up play, final touches, and shooting. The objective of the *build-up play sub-phase* is to get the ball near the opponent's goal without loosing the ball possession. While moving forward, the player with the ball must be supported by the other team-mates, by being offered with pass opportunities. The ideal position of the player in *possession* is in the top or bottom of an imaginary rhombus, formed using supporting team-mates as its vertexes. This rhombus becomes a triangle when the action is carried on near the sides of the field. Several different types of passes can be used for building play, such as long forward passes, diagonal passes or back passes. During the *final touches sub-phase*, the objective is to make a player capable to shoot. For the player having the ball this would mean to dribble to defeat the defence, or

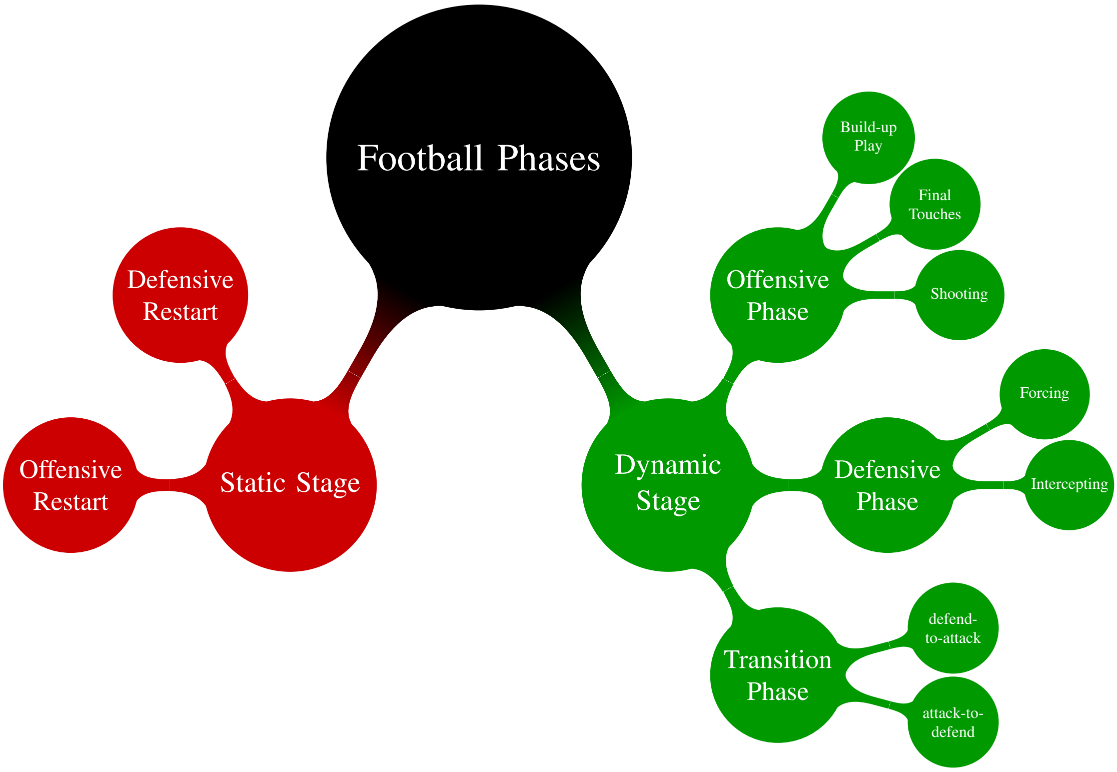


FIGURE 5.1 *Football Phases. In order to analyse, and better understand the game, it can be divided into several phases, depending on the ball situation. In the first level of the classification, there are the dynamic stage, which is the time when the ball is 'alive', and the static stage, which corresponds to the time when the game is stopped. Further divisions are also shown, which tries to describe the main objectives followed by the teams at different game times.*

to pass the ball to another player in a better situation: better shooting angle or less marked, for example. The task for the players without the ball is to assist the team-mate in *possession*, by positioning themselves in order to receive a pass. In the *shooting sub-phase* the player with the ball has to shoot trying to score, as its name obviously indicates. To shoot just after the *final touches sub-phase* is not the only possibility open for a player, the player can shoot after winning a loose ball or after stealing it, or even another possibility is to shoot from a far distance.

The *defensive phase* of a team refers to the moments when the team does not have the ball. Naturally, when a team is in its offensive phase, the opponent team must be in the complementary defensive phase, and vice-versa. The objective of the defensive phase (or *defending phase*) is to gain control of the ball. The sub-phases of the defensive phase are: forcing and intercepting. During the *forcing sub-phase*, the defending team must either push the opponents to play in certain zones of the field, or drive them to pass the ball only when the defence is ready to intercept it. The objective of this actions is to apply pressure to the team in possession to increase the chances of the defending team to gain control of the ball. The defence should get closer to the player with the ball as soon as possible and try to slow down the attacking efforts of the offensive team. In the *intercepting sub-phase* the defensive team wins possession of the ball. There are several ways to accomplish this task, for example, taking advantage after a mistake of an attacking player; intercepting the ball in the middle of a pass; anticipating or tackling the player with the ball in order to steal it; or winning the ball after a loose ball, which may come from an unsuccessful steal attempt. Finally, it is important to note that, offensive and defensive phases are tightly related, that is, the better the recovering of the ball control by a team, then the more will be the attacking opportunities it has.

The last phase of the dynamic stage is the *transitional phase*, which corresponds to the short time period when the ball possession is transferred from one team to another, that is, when the ball is lost or won by any of the teams, e.g. after a stolen ball or after a failed attempt to goal. The principal objective of the transitional phase is to adapt quickly to the new situation. The transition phase can be split into two different transitions: defend-to-attack and attack-to-defend. A *defend-to-attack transition* occurs when the ball is won during defending, for example after a successful interception. For the gaining-possession team it is very important to keep the ball under its control, therefore a decision must be made on whether to start a counter-attack or to turn away from pressure in order to try a more organised attack. On the other hand, an *attack-to-defend transition* happens when during the attack, the ball is lost, for example after a successful build-up play but no goal scored. When the opponent team is preparing the attack, the new defence must be capable of organise itself to reduce the free spaces conceded to the offensive team, and to keep trying to recover the ball again.

The other main football stage is the *static stage*, which consists of the two possible restart types after the game has been stopped: *defensive restarts* and *offensive restarts*. Restarts occur after events such as half-time starts, *throw-ins*, *fouls*, *corner kicks*, *goal kicks* and in general after any other game interruption called by the referee. In a restart, the ball possession might pass from one team to the other one depending on the circumstance that produced the stopping of the game. If for example, a player of the home team (in ball possession) looses the ball after a tackling performed by an opponent, and sends it out through the *touch line*, then the away team wins the ball possession, and begins an offensive restart with a *throw-in*, while the home team conversely begins a defensive restart. In the opposite case, where the away team player is the one who sent the ball out, then an offensive restart is to be performed by the home team, and the defensive restart by the away team. A special case occurs when both teams have to perform a defensive restart, for example in the case of a *dropped ball*, resulting from the interruption by an outside agent.

It is important to note that the restarts normally occur in specific positions on the pitch, for example, the *throw-ins* are executed from the *touch line*; the *corner kicks* from the corners of the pitch; the *direct free kicks* and *indirect free kicks* from near the place where the *foul* was committed; and so on. This will prove very useful when populating the model, since it will allow the classification of the restarting event based on the position of the ball during the restart.

### 5.1.2 Player Actions

It has been mentioned that football is a game of phases, and such phases are changing very rapidly for each team during a match. Depending on the current phase of the game in which a team finds itself, so will be the chosen objectives of the team for that moment. Each of this objectives requires different actions to be executed, and consequently different actions can and must be performed by the players. To describe the possible decisions that a player shall face, a classification depending on the game phases is adequate. This classification is shown in figure 5.2.

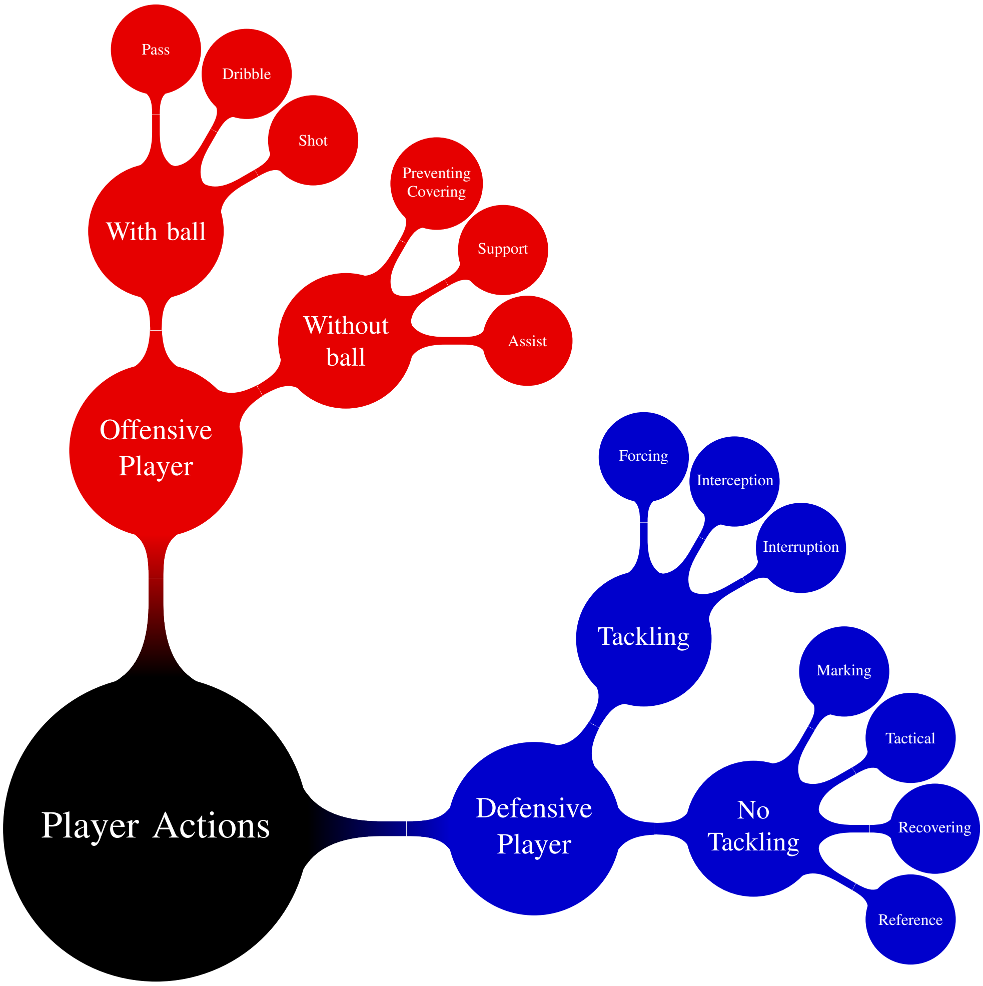


FIGURE 5.2 *Player Actions. The actions available to each player during the game depend on the current phase of the game in which the player's team is playing, therefore two main divisions: offensive player and defensive player are made. Going a step further, the offensive player actions can be subdivided into actions when having the ball and without it. For the defensive player, the actions are aim to tackle the ball or actions without trying to tackle the ball. Further finer divisions showing the individual actions that a single player can performed are also shown.*

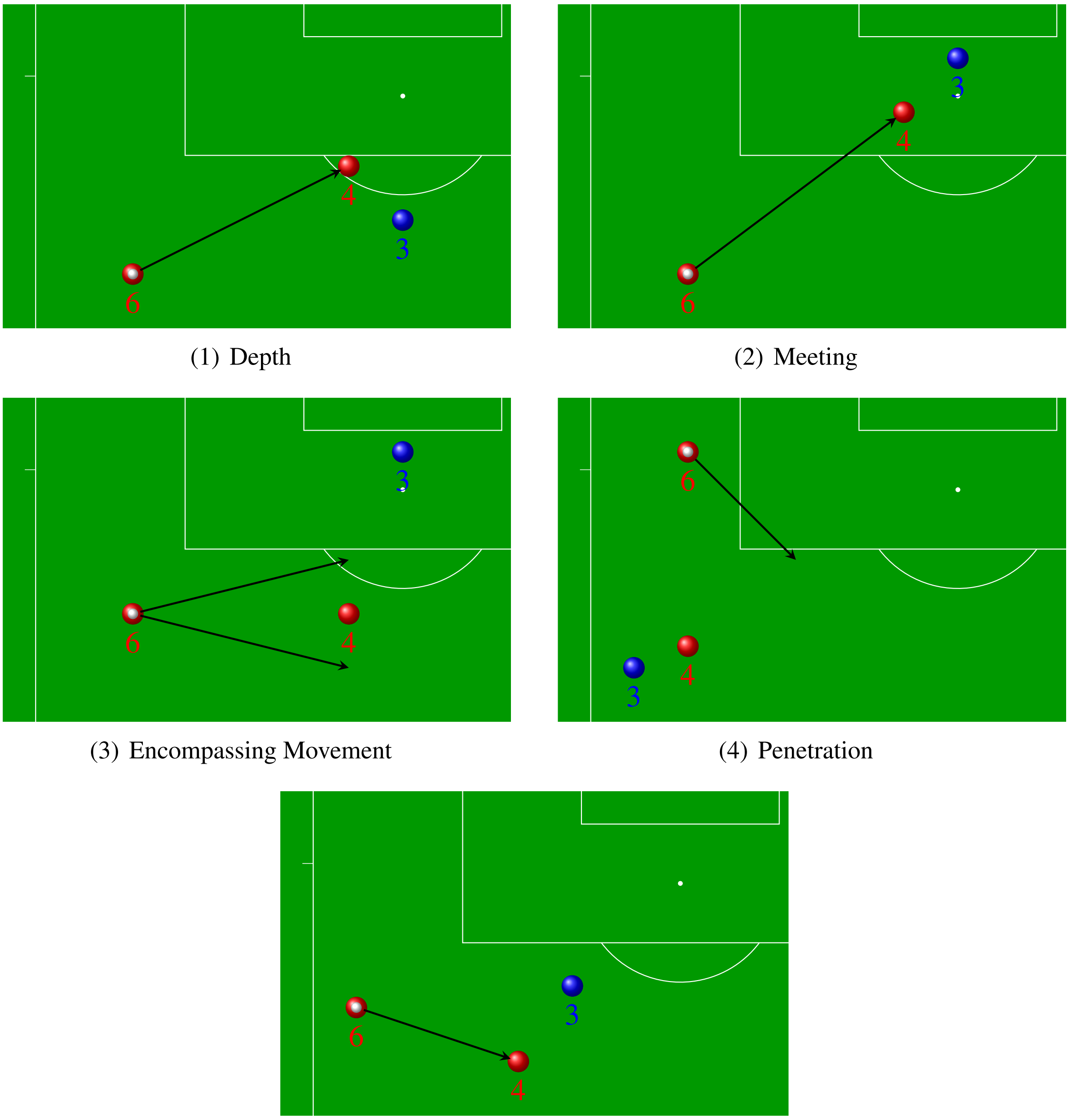
#### 5.1.2.1 Offensive Player

*Passes* are the first option open to an attacker while building-up play, and the most used during a game[1], representing 50% of the actions performed by the players with the ball. Passes are fundamental to move the ball forwards, and therefore passing the ball effectively is a crucial matter. Of course, the successfulness of a pass is not only responsibility of the sender, it is a task that requires also the receiver to make himself available to get the ball: either by getting rid of the marking player, or by reaching the meeting point with good timing. There are several types of passes, which can be useful in different situations. Depending on the relative positions of the sender, the receiver and the defence, the passes can be divided into five categories as shown in figure 5.3. Other classifications of the passes are possible, for example, by considering the length of the path followed by the ball, they can be distinguish between short, medium and long passes [SM10]. An advantage of using long passes is in keeping the position of the team, without giving space to the opponents. It also avoids taking unnecessary risks near the own goal. A disadvantage of long passes is that it makes easier for the opposing team to regain *possession*. Several classifications of passes will be used in our semantic annotation system, the chosen one at a particular point will depend on the appropriateness for the current interpretation objectives.

In a *depth pass*, the ball is moving forwards, with respect to the attacking direction, and the receiving player is getting the ball over the defence line. This kind of pass can lead to the *offside trap* if not carefully performed [Rus06, Rob99], but it also might be a great goal scoring opportunity. The *meeting pass* moves the ball forwards, but the receiver is under the defence. The *encompassing movement pass* is an horizontal pass, where the sender and receiver are located at the same level with respect to the middle line of the field, the defenders are located over this line. The objective of this type of pass is to open the game and to provide an expanded width of action for the attacking team. The ball is going backwards for the *penetration pass* and the *back pass*, but in the former, the receiver is over the defence, and the opposite situation occurs for the latter. The meeting passes, encompassing movements and back passes are the building blocks of the build-up play sub-phase. Penetration and depth passes are important in the final touches sub-phase, because a player can be found to be in a good position to shoot.

The next possible action open to a player with the ball is the dribble. *Dribbling* has the main objective of creating opportunities to score, since if successful, one or more defenders will be passed, which may generate a superiority in numbers with advantage for the attacking team. To dribble is an action more useful in the last third of the field, where more attacks occur,

[1] This was not always the case, since during the earlier days of modern football (England, 1800's), the dribble was the preferred action [Jon09].



(1) Depth

(2) Meeting

(3) Encompassing Movement

(4) Penetration

(5) Back Pass

FIGURE 5.3 *Pass Classes. The subfigure (1) shows a depth pass, in which the ball is moving forwards, and the receiver is over the defence. The meeting pass shown in subfigure (2), is also a forward moving ball type of pass, but the receiver is under the defence. The encompassing movement (or horizontal pass), is a pass in which the sender and receiver are formed in a line approximately parallel to the middle line. The subfigure (4) shows a penetration pass, where the ball moves backwards and the receiver is over the defence. Finally, the back pass depicted in subfigure (5) show a pass, where the ball is moving backwards and the receiver is located under the defence. The complete situation for each of this passes is not totally depicted, that is, the positions of the other defenders and attackers involved are not shown, therefore no inferences about why the respective type of pass was chosen by the player with the ball can be made.*

but also it is more difficult to perform there, since less space is available to manoeuvre. If the dribble is not achieved, it may lead to a loss of ball possession. On the other hand, it might also culminate in an awarded free-kick if a defender commits a foul. In summary, to dribble requires high skills from the side of the player, therefore it should be executed only when is worth or when no team-mates are available for a pass, and finally when losing the ball does not represent an utmost risk.

The last action considered for the player with the ball is the shot. Once the ball is possessed and the location is not inappropriate to shoot, two options are open to the player. The player may shoot immediately or a dribble might be tried in order to get to an even better shoot spot, and with this, increase the chances of scoring, for example, by changing the shoot angle. Other factors such as the height of the trajectory imposed to the ball, the angle and the speed should be considered by the player, with the aim of beating the goalkeeper.

The normal duration of a football match is about 90 minutes, which means that in average, each player (including the goalkeepers) will have the ball for approximately 4 minutes. That means, that the player will be playing without the ball for roughly 86 minutes. This is a very simplistic model of the real situation, but it highlights the tactical importance of what the player should do during the time when not having the ball. As figure 5.2 shows, the actions of the offensive players without the ball are threefold: preventive covering, support and assist.

The players of the offensive team that are located behind the *ball line* are in *preventing covering*, as depicted in figure 5.4. Their task is to defend against a possible counter-attack if their own team looses the ball. To *support* means that the player must make himself available for a back pass or an encompassing movement, so the player with the ball have the opportunity to pass the ball to open the attack width, or to release pressure generated by defenders. To *assist* means to help a team-mate to score a goal. Depth and meeting passes are used to assist, for example. Several forms of assisting exist: cut-ins, combinations, walls, and overlappings [Mas00]. In the *cut-in*, the player without the ball unmarks himself, and tries to occupy a free space, where an assistance pass from the player with the ball can be received. If the player without the ball is moving towards to or away from the *cut line*, then the cut will be a *converging cut* or a *diverging cut* respectively. The *combination* can be a *pass an go*, or a *pass an follow*, and others, in which the attackers try to get rid of the marking. The *wall pass* consist of passing the ball to a team-mate which is penetrating the opposing area. The objective of the wall pass (and overlapping) is to put the defender in an awkward situation since it is not possible to control both: the ball movement and the movement of the receiver. In the overlapping, a team-mate is going forwards alongside the player with the ball. The difference between the wall pass and overlapping is simply the orientation of the player doing the pass, in

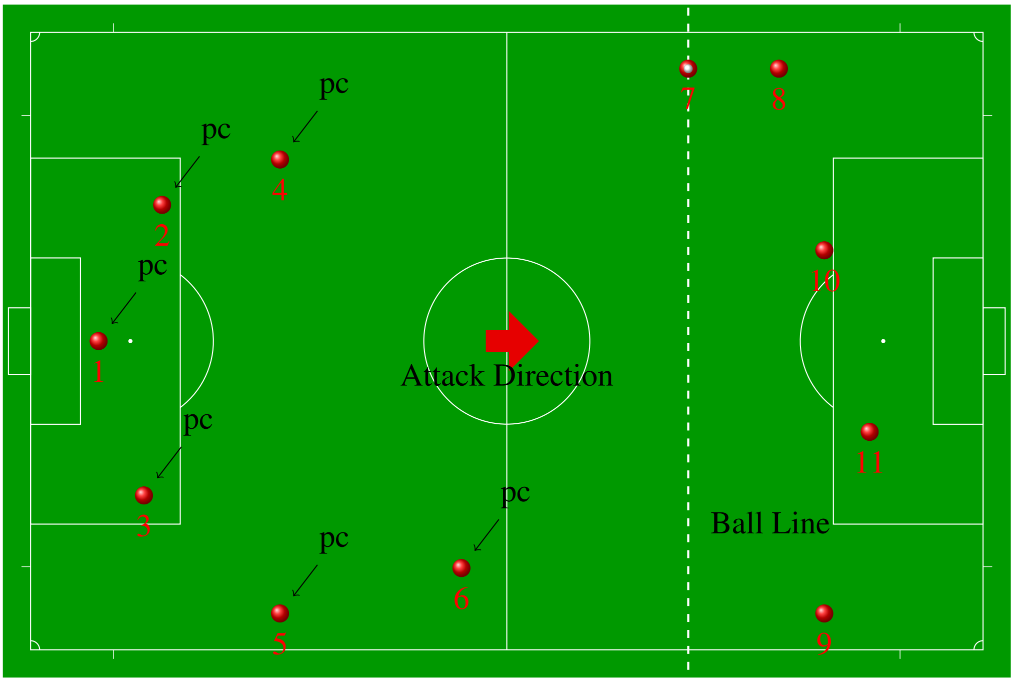


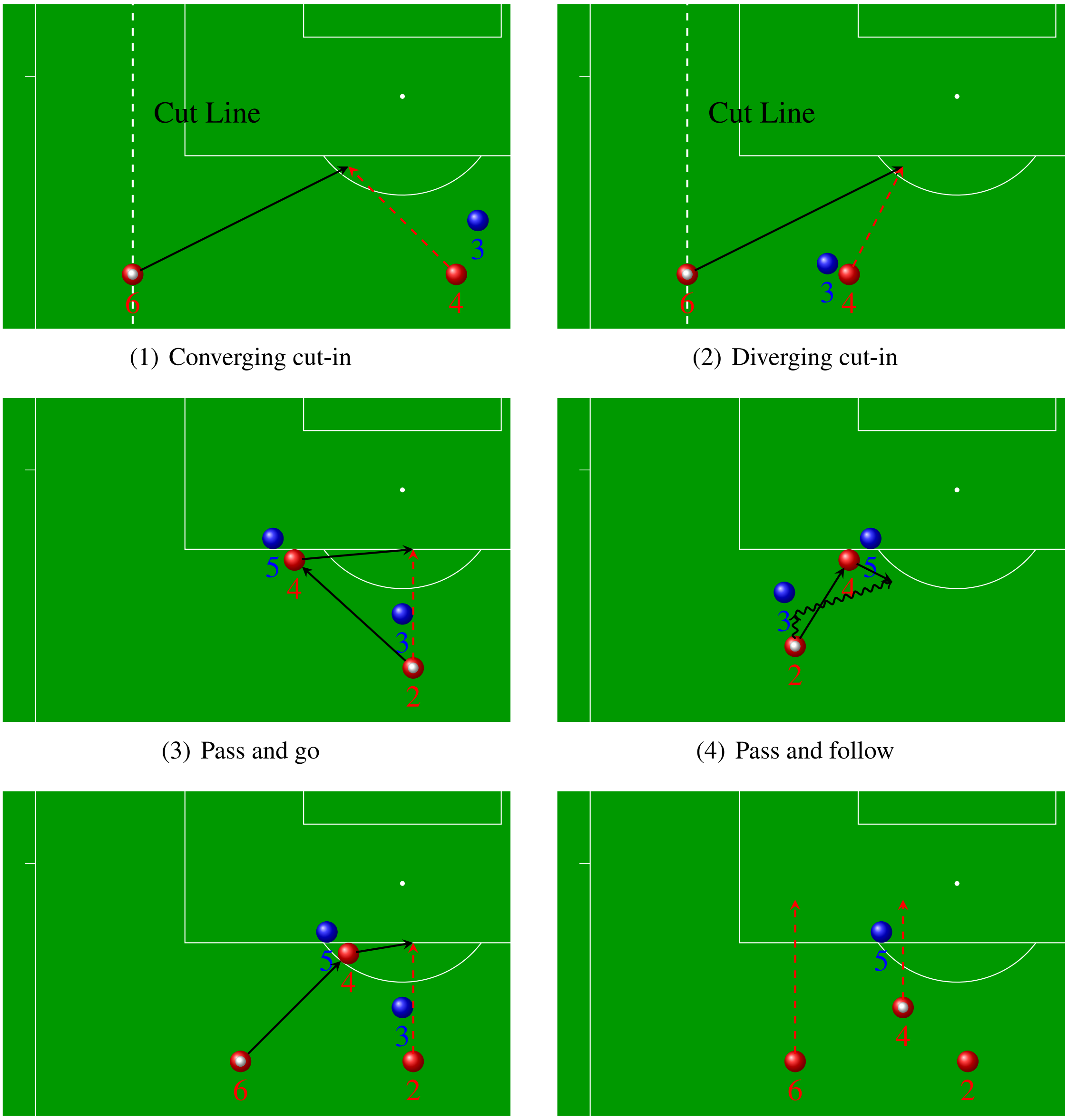
FIGURE 5.4 *Preventing Covering. The players behind the* ball line *are in preventing covering. The preventing covering players bear the pc label.*

the wall pass, this player is facing his own goal, and in the overlapping, the opposite situation occurs. For examples of the types of assisting, see figure 5.5.

#### 5.1.2.2 Defensive Player

The intention behind *tackling the ball* is to directly or indirectly gain *possession*, or to stop an attacking player from doing what was intended. *Forcing* occurs when the defence pushes the attacking team to play in a specific area of the field, or drives them to play at a time, when gaining control of the ball is facilitated. In *space forcing*, for example, the defence makes the offensive team play on the *wings*, where the attacker with the ball can be *double team*ed in order to gain the ball. In *time forcing*, the defender marking the player with the ball, might try to gain some time until another defender arrives, so again, the player with the ball can be *double team*ed. With an *interception*, a defender tries to get to the ball before it reaches the receiver, in an attacking pass attempt. If the interception is successful, this would immediately change the team in *possession*. If the interception fails, a team-mate of the defender may still be in position of winning the ball. Finally, the *interruption* can be any action that leads to a stop in the attack, for example a deflection of the ball out of the pitch, a ball clearance, a tactical foul, a hand touch of the ball, or an *offside trap*.

The actions open for a defender without explicitly pursuing to gain ball control, that is *with-*



(1) Converging cut-in

(2) Diverging cut-in

(3) Pass and go

(4) Pass and follow

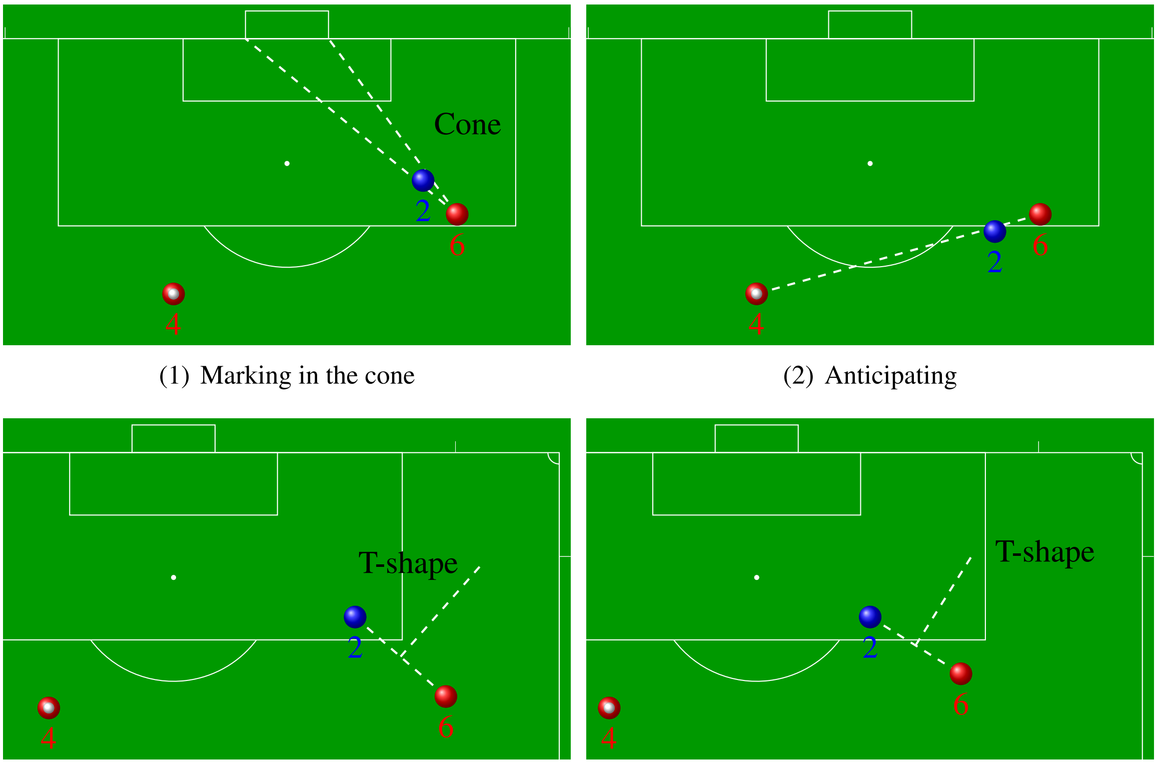
(5) Wall pass

(6) Overlapping

FIGURE 5.5 *Assisting Categories. The first row shows cut-ins, the left one is a converging one, since both, the ball and the player are moving towards each other, while the right one is a diverging cut-in, where both the receiver and the ball are going away of the* ball line*. The subfigure* (3) *and subfigure* (4) *are combinations to get rid of a strong marking. The pass and go is similar to the wall pass, and the pass and follow is a feint used to confuse the defenders, while the attacker (labelled 2) keeps the forwarding advance. The subfigure* (5) *is a wall pass, where the aim of the attacking players is to get rid of the defender labelled 3, with a manoeuvre where the defender cannot control the movement of the ball and all the players at the same time. Also in the subfigure* (6)*, the defender faces the dilemma of which player to tackle. The player with the ball has now many opportunities, a dribble is possible to get rid of player 5, a back pass is possible to player 2, or a pass to player 6 is another option.*

*out tackling the ball* are: marking, doing tactical work, recovering position and being reference for counter-attack. While *marking* a player, the defender have in general three possible positions to assume, depending on the objectives of the defence. To *mark in the* cone means that the defender must be placed inside the triangle formed between the base of the goal and the opponent being marked. Usually a defender should be located inside the cone while marking the attacker which is nearest to the one having the ball. If the ball is being pass to the attacker being marked, then the defender can move to be in *anticipation*, which is when it is located in the line segment joining the player with the ball and the attacker begin marked. If the defender arrives to the receiver zone on time, this might lead to a successful interception, and to the consequent possession. Finally, a player can *mark in T-shape*, which is the case when the defender has the option of close down an attacker which receives a pass, but also allows the defender to close down in the inner side of the defensive area, making the defence more compact and reducing the open spaces to the opponents. The danger of the T-shape marking is that if the defender closes down too much, then the vertical line of the T-shape will be inside the goal area, and the attacker being marked can receive a pass, which is a risky situation. The diagrams corresponding to the different types of markings are shown in figure 5.6.

*Tactical work* refers to covering or closing spaces. The defender executing *covering* (or *diagonal defence*) must be ready to go for the ball, in case the defender assigned to the attacker with the ball has been surpassed. *Closing spaces* is to keep the position as defender between the ball and the own goal, in order to difficult the use of possible free spaces that the attacker might want to use. A *recovering position* action is required when a defender has been leaved behind the *ball line*, and then this player must try to go back in 'action', and find a position in which the defending team can be supported in the recovering of the ball control. Those defenders that deliberately remain ahead of the *ball line*, are known as *references for counter-attack*, they must be ready to receive a pass and they must begin with a counter-attack. This references are very useful, since if their team recovers *possession*, then a long pass can be emitted, and a quick counter-attack can lead to a good opportunity to score. The figure 5.7 shows an example of the references for counter-attack.



(1) Marking in the cone

(2) Anticipating

(3) Marking in T-Shape

(4) Marking in T-Shape (dangerous case!)

FIGURE 5.6 *Marking Types. In subfigure (1) a defender marking in the cone reduces the shot angle of the marked attacker. If a pass is tried with the marked attacker as receiver, the defender can anticipate before the ball reaches the receiving zone and with this try to gain* possession. *The last row shows two examples of T-shape marking, in which the defender is ready to close down to the attacker if it is expected to get a pass, or to close down to the middle of the field to keep the defence compact. The T-shape marking on subfigure (4) is dangerous, since the marked attacker might get a pass inside the goal area, along the line corresponding to the vertical side of a T.*

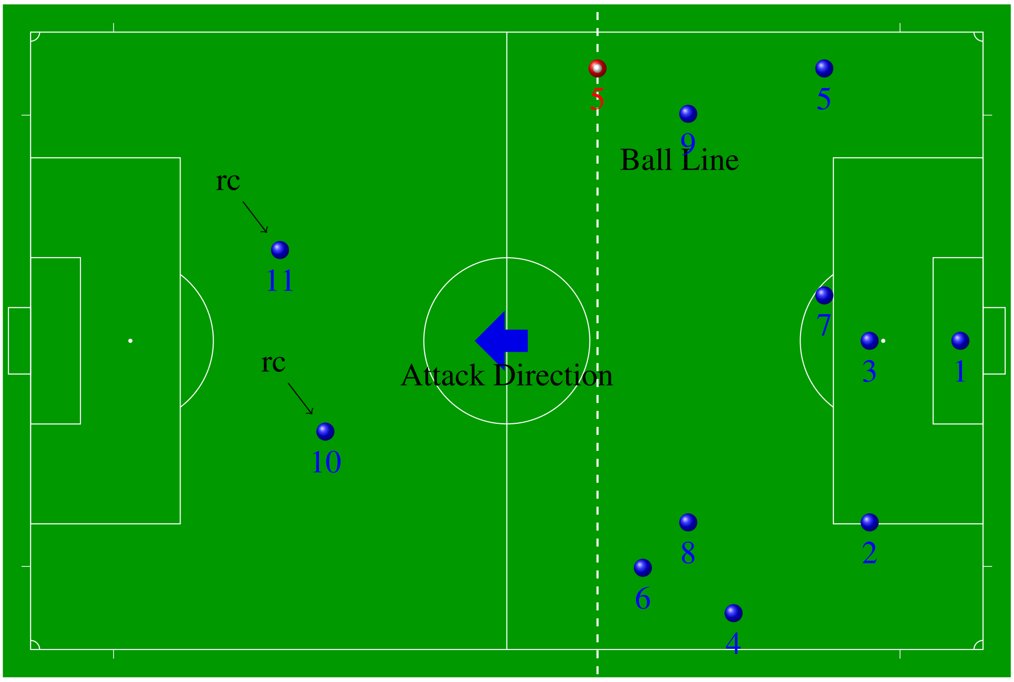


FIGURE 5.7 *References for counter-attack. The players marked with the rc label are the references for counter-attack. Those players must be ready to receive a pass and start a counter-attack.*

## 5.2 ASPOGAMO **Abstract Model**

Since football is such a complex dynamical system, the approach followed in this work to try to comprehend it, is to segment it into smaller meaningful pieces, in order to reduce the task to several manageable parts. The way that was chosen to represent these different parts is a hierarchical model, in which each representational level models a specific part of the game. In every representational layer only few aspects of the game are addressed, permitting with this structure, the more tractable development of the interpretation tasks. Also, this hierarchical structure permits to keep a simple representation of the information that corresponds to each level, and in case that more complex information is required in a particular layer, then lower layers are queried to gather the necessary information to complete the task. The mechanisms for querying lower layers from upper layers are also provided, and with such information flow, it is possible to construct more elaborated interpretations of the game. The image representing the proposed ASPOGAMO abstract model can be seen in the figure **??**, and a brief description of the actual information stored in each layer can be seen in the table 5.1.

Another important aspect to note is that the proposed model has to be constructed using observable measurements, since the idea is to populate the model with the required foundational data in an automated manner. To fulfil the above objective, and since some of the states of the system are unobservable, then the model must choose the state with the maximum probability of being true. For example, if a player kicks the ball and this is being received by a team-mate, the system will select the pass as a representation of the current action, since otherwise, it would be very difficult to precisely know without uncertainty that the intention of a player was not a pass, but, lets say, a failed attempt to shoot a goal. In the particular example given above, it will be assumed that if the ball travels from a player to a team-mate, then it is because a pass was intended. Velocity constraints must also be checked to classify the action as an actual pass.

The developed system is not only capable of managing the automatic information provided by the perception modules (tracker subsystem), but is able to deal with loaded information gathered using other commercially available systems too, and is capable of reading in, manually annotated information about a game. This is another advantage of such a hierarchical model, namely that the system can very easily and efficiently acquire a new game.

For the development, analysis, and evaluation of the algorithms designed as part of the present work, a ground truth dataset was used. This ground truth dataset comprises 16 games from the Germany *1. Bundesliga* (first league) during the season 2008/2009. The games feature the *FC Bayern München* as home team, and were played in the *FC Bayern München* home

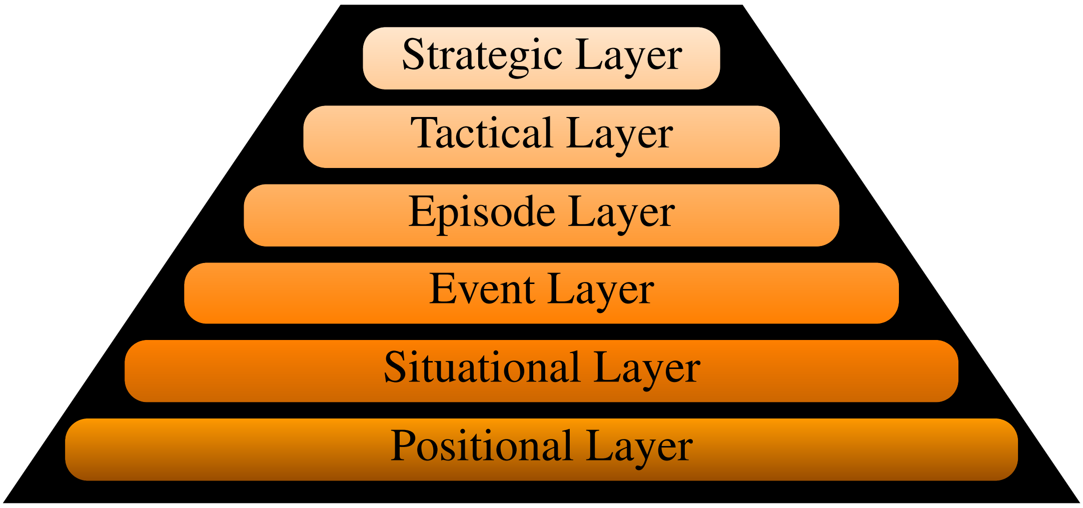


FIGURE 5.8 ASPOGAMO *Abstract Model*

stadium, the *Allianz Arena*[2]. For ease of reference, during this work, this dataset will be called the *Bayern Games*. The position of the players, referee and the ball are known during each of the games, as well as the actions performed by the player with the ball and other events such as referee directives: fouls, game interruptions and the like. The dataset was originally generated using a commercial system called Amisco, which our project-partner *FC Bayern München* have installed in its home stadium. A description of the Amisco system can be found in chapter 2. The table 5.2 details each of the mentioned games.

In the rest of the section the abstract model will be explained, and the way to populate it will be described in the section 5.3.

The lowest layer corresponds to the **Positional Layer**, which constitutes a representation of the kinematic information of the objects of interest during the game. In the above chapters **??** and **??**, greater effort was dedicated into the extraction of the positional information of the objects of interest in the game, from the digital video. The remaining layers are built up on top of this fundamental layer, and as mentioned above, mechanisms for querying the positional layer from upper layers are available to construct more complete game representations.from the digital video, In this layer, the positions, velocities and accelerations of each of the players, ball and referees are stored, such that it is always possible to query the system to obtain such information in a quickly manner, and perform more complex calculations in a simple way. Information such as the ball status can be read from this layer, which will prove necessary for the later layers.

This layer also allows the interpolation of the position of a particular object of interest even

[2] This stadium was also used during the inaugural game for the *WC* 2006 between Germany and Costa Rica.

| Layer | Information stored |
|---|---|
| **Positional** | positions, velocities, accelerations, interpolations |
| **Situational** | ball line, preventive covering, counter-attack references, build-up rhombuses, distances to goal, distance to ball |
| **Event** | actions: passes, shots, centres, clearances, tackles, dribbles, ...; events: kick-off, throw-ins, free kicks, ... |
| **Episode** | stages distribution, phases distribution, attack episodes: build-up play, final touches, shooting; defend episodes: forcing, intercepting; transition episodes: attack → defend, defend → attack, offensive restarts, defensive restarts |
| **Tactical** | passing: depth, meeting, encompassing, penetration, back-pass; defending: anticipation, cone or T-shaped marking |
| **Strategic** | style of play: defensive or attacking, possession football or direct football; team recognition: Brazilian versus English versus German football, ... |

TABLE 5.1 *The table shows examples of the information contained in each of the layers of the* ASPOGAMO *Abstract Model.*

if the observations taken from it are only partially complete, in space and time. This is very helpful in the generation of the trajectories of the objects of interest, since it is assumed that the objects do not appear or disappear from the field in a non-continuous way.

The information that belongs to the positional layer is generated by the tracker subsystem described in chapter **??**, which generates the positions of the objects of interest. Functions to estimate the velocities and accelerations based on the positions of those objects are incorporated to the database interface, such that this information is also available for its use in the subsequent layers of the model.

The **Situational Layer** represents situative information in the sense of expressing the location of things in comparison with another ones. For example, the *ball line*, dividing the pitch into two parts: front and back, the players in *preventive covering* and *counter-attack reference* during offensive and defensive episodes, *build-up rhombuses* for preparing passes, distances and angles to opposing goal, distances of players to ball, as well as others. With this data encoded, it is easy to provide information to the upper levels in order to generate more complex representations of the analysed situations.

Another example of information codified in this layer is the number of opponents marking a player with the ball, since this could serve as a measure of pressure imposed over that player. Detecting the offsides can be done by comparing the relative positions of the attacking players

| Game id | Date | Home Team | Away Team |
|---|---|---|---|
| 1 | 2008.11.01 | FC Bayern München | Arminia Bielefeld |
| 2 | 2008.11.22 | FC Bayern München | FC Energie Cottbus |
| 3 | 2009.02.08 | FC Bayern München | Borussia Dortmund |
| 4 | 2009.02.21 | FC Bayern München | FC Köln |
| 5 | 2009.04.11 | FC Bayern München | Eintracht Frankfurt |
| 6 | 2008.08.15 | FC Bayern München | Hamburger SV |
| 7 | 2009.03.07 | FC Bayern München | Hannover 96 |
| 8 | 2008.08.31 | FC Bayern München | Hertha BSC Berlin |
| 9 | 2008.12.05 | FC Bayern München | 1899 Hoffenheim |
| 10 | 2009.03.21 | FC Bayern München | Karlsruher Sport Club |
| 11 | 2009.05.12 | FC Bayern München | Bayer Leverkusen |
| 12 | 2009.05.02 | FC Bayern München | Borussia Mönchengladbach |
| 13 | 2009.04.25 | FC Bayern München | FC Schalke 04 |
| 14 | 2009.05.23 | FC Bayern München | VfB Stuttgart |
| 15 | 2008.09.20 | FC Bayern München | Werder Bremen |
| 16 | 2008.10.25 | FC Bayern München | VfL Wolfsburg |

TABLE 5.2 Bayern Games *dataset. Dataset of 16 Bayern Games during the season 2008/2009, used for the development, analysis and evaluation of the current work.*

and the defenders. Several interesting information is contained in this layer, and this will served to feed the higher level layers to build more semantically-valued information about the matches.

This layer is mainly a conceptual intermediate layer between the *physical raw storage* of the data in the positional layer, and the high-level event and episode layers. The information contained in this layer is required in order to classify the actions in the upper layer, as well as to sustain the tactical layer. Some of the information in this layer is explicitly stored, and other is derived from the information stored, using database functionality.

The next level is the **Event Layer**, for which the concept of event is required. An *event* is defined as a segment of time at a given location which can be conceived to have a beginning and an end [ZT01]. The use of *events* to describe more complex activities helps in reducing this complexity so the analysis becomes a more tractable task. For the particular case of football, an *event* $e_k$ could be an *action* $a_k$ performed by the player with the ball, such as pass, clear, dribble, or shot. For example, a pass in this layer is defined as a player kicking the ball with the intention to give it to a team-mate. Also, an *event* $e_k$ could be a game *interruption* $i_k$ resulting from a ball going out of the pitch or when the referee halts the game. Other logical events of the game are included in this layer, such as kick-off, throw-ins, free kicks, and others. As an example of the modelling power of this layer, consider a match represented by a sequence of events $M = \{i_0, a_1, a_2, a_3, i_4, i_5, a_6, \dots, i_N\}$, where the $a$'s corresponds to action events, and the $i$'s to interruption events. This sequence could be interpreted([Hof79]) by *kick-off*, pass, dribble, pass, foul, indirect kick, pass, ..., game end, for example. Of course all performed *actions* occur in the dynamic stage and the *interruptions* in the static stage.

The **Episode Layer** combines events from the event layer into consecutive sequences in logical order as to express the becoming of a game in a time lap. To provide a more meaningful description of the match, in terms of higher level activities and not only in terms of individual *events*: *actions* and *interruptions*, the *episode* concept is defined. In very general terms, an *episode* is a happening that is distinctive in a series of related *events* and depends on a larger context for their sense and importance. In the football case, this concept can be used to model groups of consecutive *events* that represent team intentional activities, for example, an action *episode*: $E_k = \{a_p, \dots, a_{p+q}\}$ can be used to represent an offensive attack of a team, in which build-up play, final touches and shooting are addressed. Mainly there are two types of episodes to describe the game, which are the offensive episodes, including build-up play, final touches, and shooting; and the defensive episodes, including forcing and intercepting. Another kind of episodes are the transition episodes, which are instantiated by defensive restarts, offensive restarts or the transition between attacks and defends stages[3].

Even if is technically possible to model the sequences of consecutive *interruptions* as interruption *episodes* too, this makes no much sense, since the interest of the present work is related to football *tactics* and *strategy*, and not to general activity representation. Therefore, the *interruptions* will be used only as boundary delimiters of the more interesting action *episodes*, which from now on will be simply referred to as *episodes*. In summary, an *episode* $E_k$ will be understood as a sequence of *actions*: $E_k = \{a_p, \ldots, a_{p+q}\}$ meaning a higher level activity of a team. Our definition of *episode* differs from the ones given in other works: [Sta03, vHH05, vHHKB07]. There, an episode means a ball contact, marked as successful if two consecutive ball actions are performed by the same team. This definition can be confusing, for example in a dribble, since a dribble consist of a single action being performed by a single player, and it does not matter if for moving the ball several ball touches have to be done, the action is a still a single event. Therefore, we humbly believe, that our definition is more appropriate and intuitive, and it facilitates the better understanding of the underlying process.

As an example of an *episode*, a diagram showing the 19-th *episode* of the 14-th *Bayern Game*[4] (BG14E19) is presented in figure 5.9. There, an attack of the home team (red) is performed, where it starts with a pass after an *indirect free kick* was granted due to a committed foul, and ends with the ball being stolen by the away team (blue). At the beginning of the episode, the player 6 passes the ball to the player 21, which runs with the ball to perform a pass to player 11, which attempts a pass to player 7, but the ball gets stealth by an interception of the player 35 of the away team.

The first of the highest levels of description of the model is the **Tactical Layer**, which aims to describe the tactical work performed by the players and the team. This layer does not model the isolated player actions per se, but rather the implicit intention in the performed actions, in order to reach the planned team-objectives for the game. For example, during a pass, the desired action can be a depth pass, a meeting pass, an encompassing movement or the like, which directly reflects the tactical work required to fulfil specific targets. Another example is during defending, where anticipation, cone or T-shape covering are chosen as the tactic used to exert pressure on the offensive team.

Finally, the **Strategic Layer** spans over styles of play of the teams: more defensive or

[3] The difference between restarts and transitions is that in the former, an interruption of the game has occurred, while in the later, the ball was always *in play*. An example of the first is a ball-out-of-bounds followed by a throw-in, and an example of the second is a ball being stealth after a successful interception.

[4] In the following, for abbreviation purposes, the $n$-th *episode* of the $m$-th *Bayern Game*, will be referred to as BG$m$E$n$.

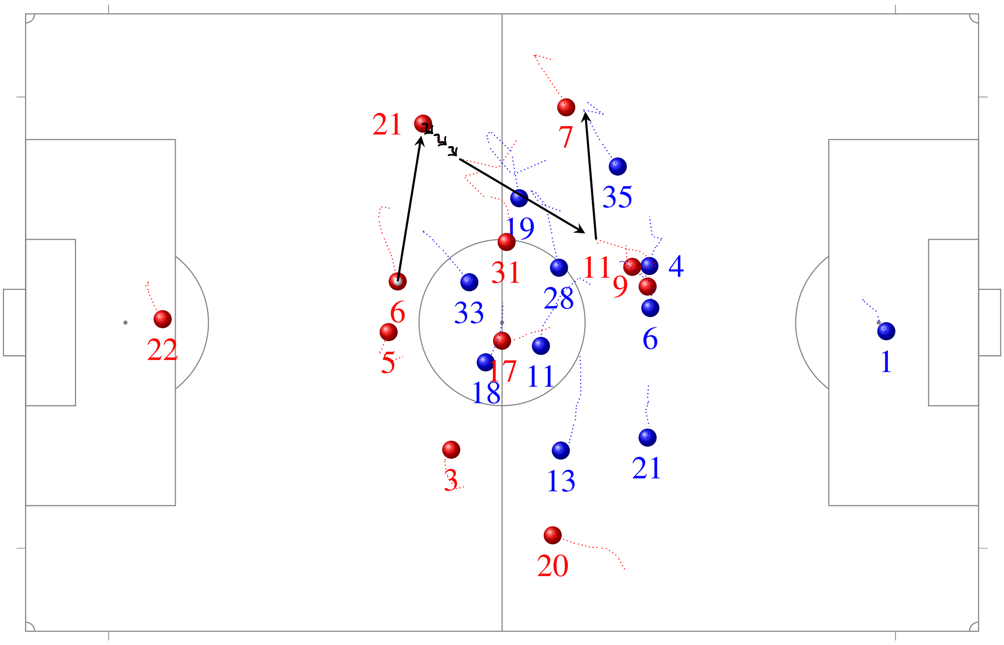


FIGURE 5.9 Bayern Game *14, Episode 19. The episode is conformed of 4 actions: a pass, a dribble, and two passes, where the last one was unsuccessful.*

offensive, possession football or direct football. Based on this layer, different squads can be classified and recognized. This layer uses the tactics found in the Tactical Layer in order to describe the behaviour of a team in game (or series of games), in order to determine the global strategy utilized to play the game.

## 5.3 ASPOGAMO **Abstract Model Population**

As the first step in the population of the ASPOGAMO Abstract Model, there is the storage of the positional data inside the database, which is accompanied with database functionality to allow the calculation of the velocities and accelerations of the objects of interest. Most of this procedure is done after the perception modules, when the tracking subsystem estimates the positions of the objects based on the segmentation over the frames in the TV video.

In the case of the situational layer, some information is generated and stored explicitly, for example, the distances and angles to each player with respect to their opponents and teammates. Other information is provided by means of database interface functionality, such as the ball line (in case an offside is to be search for), the counter-attack references, as well as others. Since the situational layer is more a conceptual layer serving as a bridge between the raw positional data and the event and episode layers, this mixed methodology seems to be the right approach for storing the information in this level. This level will mostly transform the queries from the upper layers into direct queries about relative positions of particular objects in the field.

As explained in the section 5.1, the football is a game of phases with two stages: *dynamic stage*, and *static stage* (see figure 5.1), depending on the ball status. Therefore, a natural step in the game interpretation, is the detection of the times when the ball is *out of play*, and *in play*, which basically represent each of the mentioned stages, respectively. This information belongs to the **Episode Layer**, which, from the semantic point of view, is closely related to the **Event Layer**, since the episodes' delimitation will be very useful for the subsequent action classification. From the logical modelling point of view, the event layer is below the episode layer, since an episode is composed of several related individual actions, nevertheless, the episode layer will be explained first, since from an initialization point of view, the episode layer is populated first than the event layer. This though, is a matter of an implementation detail and does not affect the semantics of the representation.

For example, a counter-attack episode of a team after an offside error of the opposing team, can be described as a sequence of actions like: offside, indirect free kick, pass, dribble, pass, pass, dribble, centre, shot, and might culminate into a goal. For reaching such a description, not only the actions are to be recognized individually, but the boundaries of the logical sequences are to be detected. For the detection of such logical boundaries, a football game is mathematically modelled as a stochastic process, described by a *HMM*, where the states are the game stages already described in section 5.1. A first level *HMM* describing the possible stage transitions is shown in figure 5.10. A *HMM* is chosen since the states of the game are not directly observable, but only some positional characteristics about the objects of interest. Besides, a Markov property is assumed for the relation between the consecutive state transitions of the model. That is, the system is memoryless, and the complete history till time $t-1$, is not necessary to obtain the probability of being in a particular state at time $t$. For example, to know that before a ball out of bounds through the touch line sent by an opponent, there was a tackle, and before that two dribbles, and before that three passes, is not relevant to state that the following action, being a throw-in, belongs to a static stage state. Using this *HMM*, the counter-attack example mentioned above: offside, indirect-free kick, pass, dribble, pass, pass, dribble, centre, shot, goal, will be described as the following state transitions:

$$S \to S \to D \to D \to D \to D \to D \to D \to D \to S \tag{5.1}$$

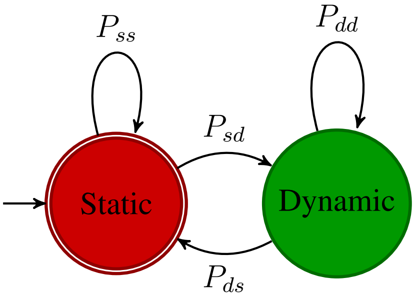


FIGURE 5.10 HMM *describing the stage transitions during the game. The transition probabilities are: from static to static $P_{ss}$, from static to dynamic $P_{sd}$, from dynamic to static $P_{ds}$, and from dynamic to dynamic $P_{dd}$.*

where $S$ corresponds to a static stage state and $D$ to a dynamic stage state. The corresponding transition probabilities between the stages states can be estimated based on the relative frequencies of the different transitions observed using the ground truth. The estimated probabilities for the *Bayern Games* are as follows: $P_{dd} = 0,901$, $P_{sd} = 0,047$, $P_{ds} = 0,047$, and $P_{ss} = 0,005$. Since the transitions are probabilities, then their sum must equal one. In the table 5.3 and in the figure 5.11 the accumulated observed transitions are shown for the *Bayern Games*.

Since after every opening static stage must come another closing static stage, for example, the game starts in a static stage when the referee has given the initial whistle, and it also ends in another static stage at the end of the game by the referee. Another example is when an offside is committed by a team, the game is stopped, then going to a static stage, is is again restarted with a indirect free kick, another static stage. Therefore, the dynamic-static and static-dynamic transition must be symmetric in number, and the transition probabilities between the static and dynamic stages must be equal, as obtained. As can be easily noticed from figure 5.11, the majority of the transitions are from the dynamic stage state to itself. That is, most of the actions performed during the game are actions with the aim of keeping the ball *in play*.

The fact that almost all of the stage transitions in a game are dynamic-dynamic, does not mean that also most of the game time is being spent on the dynamic stage state too. This is a different aspect of the model that can be observed, and from its analysis, a very surprising fact can be obtained. It was observed that the effective *ball in play* time of a game is around 54 minutes in average, that is only 60% of the total time. The remaining time is spent in game interruptions, for example, in the time that the game gets interrupted after a foul, or similar. A graphical representation of this is shown in the figure 5.12. The empirical probability density functions (not normalized) for the duration of each single event in each stage is shown in

| **Game id** | **dynamic-dynamic** | **dynamic-static** | **static-dynamic** | **static-static** |
|---|---|---|---|---|
| 1 | 2055 | 131 | 131 | 17 |
| 2 | 2667 | 105 | 105 | 9 |
| 3 | 2426 | 143 | 143 | 14 |
| 4 | 2338 | 121 | 121 | 13 |
| 5 | 2438 | 112 | 112 | 12 |
| 6 | 2782 | 118 | 118 | 17 |
| 7 | 2435 | 132 | 132 | 12 |
| 8 | 2691 | 120 | 120 | 13 |
| 9 | 2140 | 127 | 127 | 13 |
| 10 | 2637 | 121 | 121 | 17 |
| 11 | 2123 | 126 | 126 | 17 |
| 12 | 2615 | 117 | 117 | 14 |
| 13 | 2346 | 125 | 125 | 23 |
| 14 | 2052 | 124 | 124 | 16 |
| 15 | 2346 | 119 | 119 | 10 |
| 16 | 2268 | 131 | 131 | 14 |
| **Total** | **38 359** | **1 972** | **1 972** | **231** |
| **Average** | **2 397,44** | **123,25** | **123,25** | **14,44** |
| **Relative frequency** | **0,9018** | **0.0464** | **0.0464** | **0,0054** |

TABLE 5.3 *Frequencies of the stage transitions per game for the* Bayern Games. *For example, a dynamic-dynamic transition would be a pass followed by a dribble; a dynamic-static transition can be a failed pass followed by a throw-in; a static-dynamic transition is represented by a indirect kick followed by a pass, and finally an offside followed by an indirect kick can be view as a static-static transition. As can be observed, more than 90% of the transitions correspond to dynamic-dynamic transitions.*

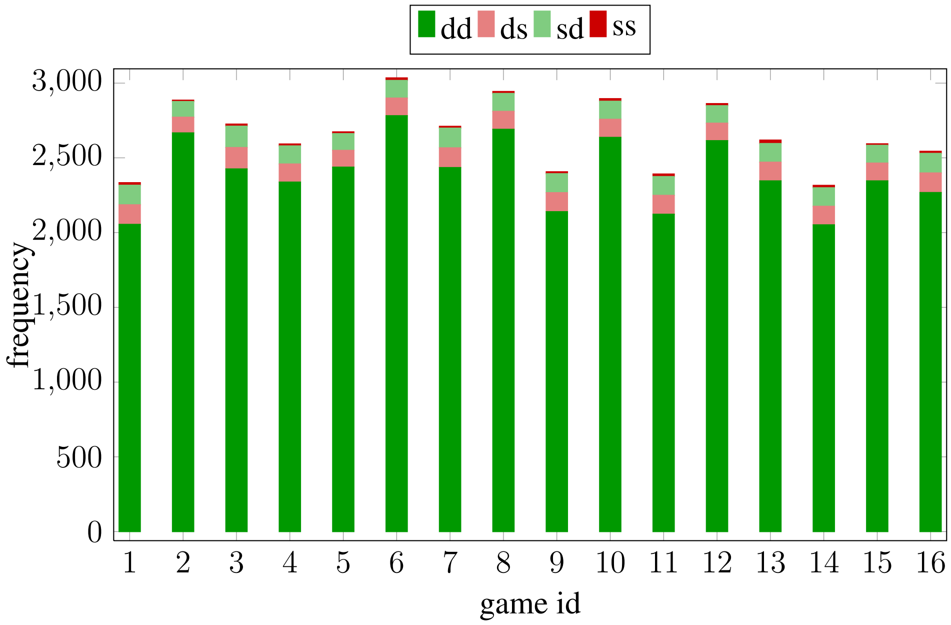


FIGURE 5.11 *Relative frequencies of the stage transitions per game for the* Bayern Games. *Each vertical bar corresponds to a game taken from the* Bayern Games, *where the first component corresponds to the actions classified as dynamic-dynamic transitions, then the next two components describe the static-dynamic and dynamic-static transitions, and finally the static-static transitions. For example, a dynamic-dynamic transition would be a pass followed by a dribble; a dynamic-static transition can be a failed pass followed by a throw-in; a static-dynamic transition is represented by a indirect kick followed by a pass, and finally an offside followed by an indirect kick can be view as a static-static transition. As can be observed, most of the transitions correspond to dynamic-dynamic transitions (blue areas).*

figure 5.13(1) and figure 5.13(2). The average interruption time is about 17 seconds, and the mean duration of a dynamic action is approximately $1,3$ seconds. This values were obtained considering all the events from home and away teams in the *Bayern Games*. As can be seen, the individual actions performed by the players are very quick, for example a pass, or a shot; while the static interruptions take in average, several seconds, for example to attend a hurt player after a foul.

After each state transition of the *HMM* (see figure 5.10) that conduces to a particular state $x_k$, there is a non-observable output $o_k$, for which only indirect positional information $y_k$ of the objects of interest is observable, and from it, the actual output $o_k$, and the hidden state $x_k$ are to be estimated. Therefore, the estimation problem can be stated as follows: let a match $M$ be represented as a sequence of state transitions:

$$X = \{x_k : x_k \in [S, D], \forall k \in [0, N]\} \tag{5.2}$$

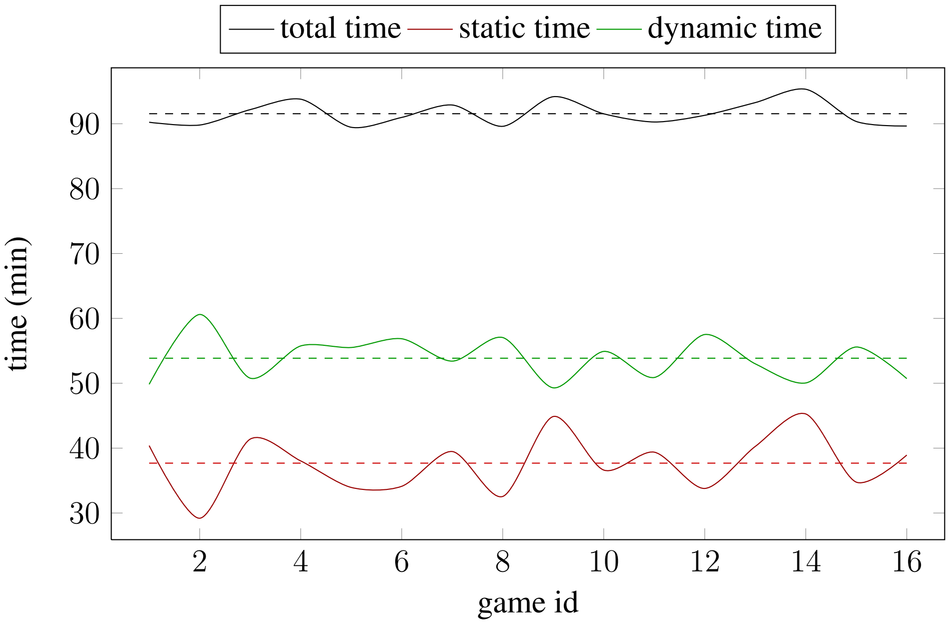
where $N$ is the time step at which the last transition to the static stage signalising the end of the game occurs, $S$ represents the static stage and $D$ the dynamic stage. An example of how such a state transition sequence would look like is $M_z = \{S, D, D, D, S, S, D, \ldots, S\}$. As can be seen, the initial and final states of a match belong to the static stage, since the games start with the whistle of the referee, and end in a similar way.

To a state transition sequence, there is an associated output sequence:

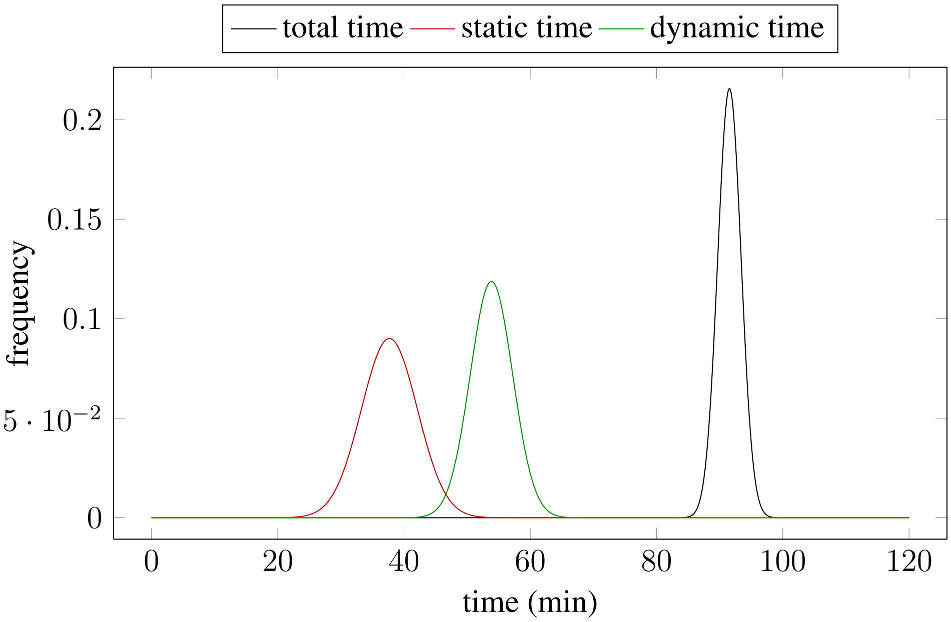
$$O = \{o_k : o_k \in E, \forall k \in [0, N]\} \tag{5.3}$$

where $E$ is the set of events that can occur during the game.

The process by which the relevant events are found in a game is called *event segmentation*. Event segmentation is the mechanism by which a continuous activity is separated into individual meaningful events [ZT01, KZ08]. For the football particular case, it is required to find out the times when the game has gone to a static state, since this will mark the boundaries of most of the relevant events. In simpler terms, it is necessary to find the *interruptions*, which correspond to the ball going out of the pitch, or to the game being stopped by the referee. As can be seen from table 5.4, almost 70% of the interruptions are due to the ball going out of the pitch. This includes events such as *ball out of bounds* through the *touch lines*, or the *goal lines*. This interruptions are always followed by events such as *throw-ins*, *corner kicks*, *goal kicks*, *goals* and *auto-goals*, depending on the affiliation of the last player who has touched the

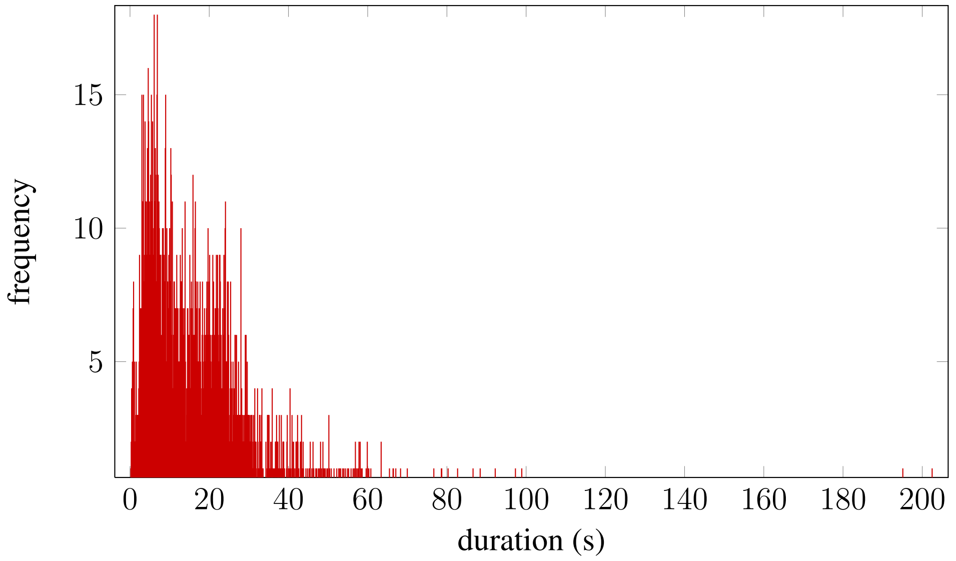


(1) Time duration per stage

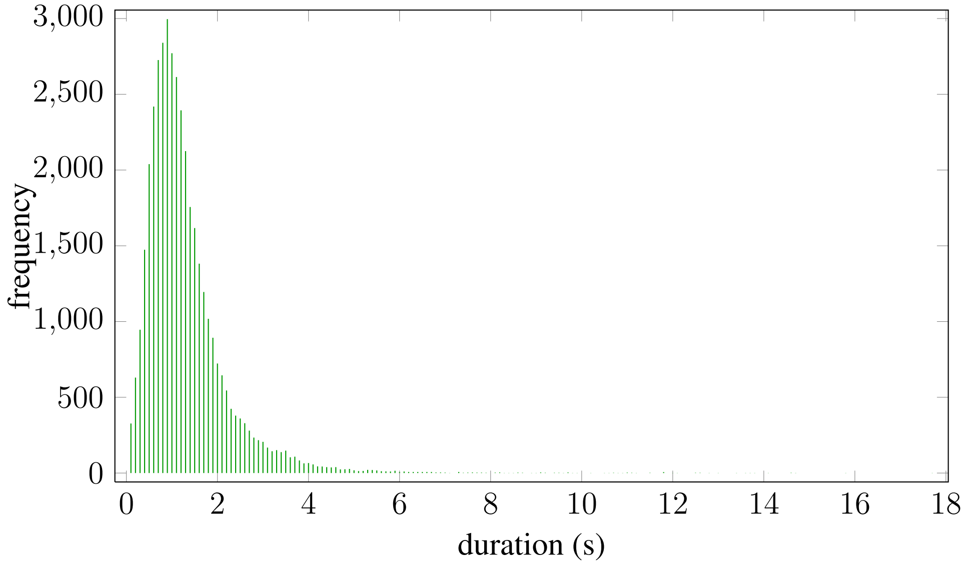


(2) Gaussian models for the stages' time distribution

FIGURE 5.12 *Time durations per stage per game The chart in subfigure (1) shows how the time is distributed along the game stages for the* Bayern Games. *The dashed lines show the averages. The diagram in subfigure (2) shows the Gaussian models use to represent the time distribution along the game stages for the* Bayern Games. *The mean and standard deviations are* $(91,55;1,85)$, $(37,69;4,43)$, *and* $(53,86;3,36)$ *for the total time, and the static and dynamic stages respectively.*



(1) Time distribution of the events in the static stage



(2) Time distribution of the events in the dynamic stage

FIGURE 5.13 *The figure shows the time distributions of the events in the static stage and the dynamic stage for the* Bayern Games.

ball.

Situations when the ball is out of bounds, can be detected by using the information contained in the positional layer, and the functions on the databases design for this purpose. This permits the classification of interruption events, which act as delimiters for the game actions. If the positional data is accurate enough, the detection of the interruptions due to a ball out of bounds, reduces to the localization of the frames where the ball has at least one coordinate out of the demarcated field, that is:

$$\frac{W}{2} + r < |x| \vee \frac{H}{2} + r < |y| \tag{5.4}$$

where $W$, $H$ are the width and height of the field of play respectively[5], $r$ the ball radius, and $(x, y)$ the ball position. In the case of the *Bayern Games*, the table 5.5 shows the classification results. As can be seen, if the information about the objects of interest is precise, then the classification can reach perfect performance. In the ball out of bounds classification task, the $1284$ ball out of bounds events in the *Bayern Games*, have a precision, recall and F1-score of $100\%$ in average. When the ball re-enters the pitch, the interruption end frame is then found.

Since the referee must decide if the whole of the ball crossed any of the lines [FIF11], in order to consider it out of the pitch and to declare a ball out of the bounds, then the task is affected by subjectivity, and therefore, a slight difference with respect to the results just presented may occur. For example, due to human error, a ball might be declared out of bounds even if it was not the case, or vice-versa. Similarly, if the positional data has an error in the detected ball positions, a false positive or a false negative might be produced. This issues will be problematic only in those cases, where the decision is very close, that is, when the ball did not went out clearly by a far distance margin, which in general, are the lesser-occurring cases. In such situations, manual interaction from the system operator is required to correct the misclassification.

The remaining 30% of the game interruptions (see table 5.4) are due to referee directives: *fouls*, defective ball being replaced, attention to an injured player, or any reason not mentioned elsewhere in the *Laws of the Game* [FIF11]. This events are more difficult to detect automatically, since they do not explicitly and solely depend on positional data, and therefore are not directly observable. As part of the current project, classification experiments based on the deformation of the contour shapes of the players during the game were performed, in order to try to automatically detect some of these events (see [Lad10]). No definitive results were

[5] The field lines belong to the areas to which they are boundaries [FIF11]. This is why the lines width is not being taken into account for the calculations.

| **Game id** | **Interruptions** | **Relative Frequencies** | |
|---|---|---|---|
| | | **Ball Out of Bounds** | **Others** |
| 1 | 130 | 0,71 | 0,29 |
| 2 | 104 | 0,71 | 0,29 |
| 3 | 141 | 0,77 | 0,23 |
| 4 | 120 | 0,68 | 0,32 |
| 5 | 110 | 0,66 | 0,34 |
| 6 | 117 | 0,56 | 0,44 |
| 7 | 130 | 0,72 | 0,28 |
| 8 | 118 | 0,67 | 0,33 |
| 9 | 126 | 0,54 | 0,46 |
| 10 | 119 | 0,60 | 0,40 |
| 11 | 125 | 0,68 | 0,32 |
| 12 | 116 | 0,66 | 0,34 |
| 13 | 125 | 0,69 | 0,31 |
| 14 | 124 | 0,59 | 0,41 |
| 15 | 118 | 0,66 | 0,34 |
| 16 | 130 | 0,60 | 0,40 |
| **Total** | 1953 | | |
| **Average** | **122,06** | **0,66** | **0,34** |

TABLE 5.4 *Relative frequencies for the incidence of game interruptions for the* Bayern Games. *The column named 'Interruptions' is the number of interruptions per game, which in average is approximately 122. The last two columns represent the relative frequencies per game of the interruptions due to ball out of bounds (*throw-ins, corner kicks, goal kicks, goals, auto-goals*), and due to other reasons (e.g.* fouls, offsides*) respectively.*

| Game id | Ball out of bounds | False Positives | False Negatives | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|---|---|
| 1 | 92 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 2 | 74 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 3 | 109 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 4 | 82 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 5 | 73 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 6 | 66 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 7 | 94 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 8 | 79 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 9 | 68 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 10 | 71 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 11 | 85 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 12 | 76 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 13 | 86 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 14 | 73 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 15 | 78 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| 16 | 78 | 0 | 0 | 100,00 | 100,00 | 100,00 |
| Total | 1284 | 0 | 0 | | | |
| **Average** | **80,25** | **0** | **0** | **100,00** | **100,00** | **100,00** |

TABLE 5.5 *Evaluation of the detection of ball out of bounds, based on the ball positional data for the* Bayern Games. 1284 *ball out of bounds events were correctly classified.*

obtained, which clearly suggests that more research in this area is still required. Other events occurring during the static stage, such as player substitution, yellow or red card showing, are not explicitly counted as *interruptions* by themselves, since they must be performed during the lapse of another already triggered *interruption*, as stipulated in the *Laws of the Game* [FIF11]. Therefore, this information is to be manually annotated by the system operator.

The detection of ball out of bounds interruptions as event boundaries corresponds only partially to all the possible delimiters that exist. The remaining event boundaries are represented by other situations, such as when an opposing player steals the ball, and with his action changes the ball possession. Therefore, the *possession* has to be determined in order to completely delimit the *episodes*, since the ball control will serve as an indicator of the team membership of each *episode*. The detection of the ball out of bounds was already addressed and the results were presented in the table 5.5. For the detection of these other boundaries, the ball contact is to be used. The ball contact means that the episode changes membership any time the ball gets *touch* by a player of the opposing team. In the figure 5.14, a representation of the ball possession for each team is presented. The continuous periods of time, when a particular team is in possession, represents the search episodes. As can be seen from the figure, the home team episodes are delimited by game interruptions as well as by away team episodes, and vice-versa.

For the *Bayern Games*, an average $F1$-score of $86,57$ was obtained in the segmentation of the episodes based on the ball contacts. With respective recall of $93,48$ and precision of $80,64$. It can be also concluded from the table that more than 6 thousand episodes were correctly detected. These figures demonstrate a good performance in the classification of the episodes in general terms. The big number of false positives shown as results on the table 5.6 are a little bit misleading, since the way to calculate the false positives is by finding frames that do not match between the ground truth episodes and the detected episodes, that is: suppose a ground truth episode has a time span of $[f_{gt_b}, f_{gt_e}]$, and the nearest detected episode has a time span of $[f_{dtb}, f_{dte}]$, such that $f_{dtb} < f_{gt_b} \wedge f_{dte} > f_{gt_e}$, therefore there are two false positives present in this case, with spanning times $[f_{dtb}, f_{gt_b} - f_{dtb}]$ and $[f_{dte} - f_{gt_e}, f_{dte}]$, even if those fake episodes have only $1$ frame of duration, due to imprecisions in the ground truth as to when exactly does an episode starts or ends. Therefore, the results on the table must be considered as the worst case scenario for the episode segmentation, and better results are indeed being obtained. Other error detection mechanisms can be used in order to be more clear about the detected episodes, such as number of detected episodes in a detection window spanning a fixed number of frames to the left and to the right. As mentioned above, this will reduce the number of false positive episodes detected and will show the better performance of the classification.

| **Game id** | **Episodes** | **False Positives** | **False Negatives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 1 | 607 | 148 | 38 | 80,40 | 94,11 | 86,71 |
| 2 | 555 | 155 | 40 | 78,17 | 93,28 | 85,06 |
| 3 | 614 | 128 | 39 | 82,75 | 94,03 | 88,03 |
| 4 | 596 | 128 | 35 | 82,32 | 94,45 | 87,97 |
| 5 | 514 | 154 | 37 | 76,95 | 93,28 | 84,33 |
| 6 | 542 | 167 | 49 | 76,45 | 91,71 | 83,38 |
| 7 | 528 | 140 | 43 | 79,04 | 92,47 | 85,23 |
| 8 | 539 | 124 | 35 | 81,30 | 93,90 | 87,15 |
| 9 | 598 | 124 | 50 | 82,83 | 92,28 | 87,30 |
| 10 | 579 | 114 | 50 | 83,55 | 92,05 | 87,59 |
| 11 | 550 | 131 | 40 | 80,76 | 93,22 | 86,55 |
| 12 | 522 | 145 | 33 | 78,26 | 94,05 | 85,43 |
| 13 | 599 | 151 | 39 | 79,87 | 93,89 | 86,31 |
| 14 | 573 | 141 | 43 | 80,25 | 93,02 | 86,17 |
| 15 | 578 | 120 | 33 | 82,81 | 94,60 | 88,31 |
| 16 | 582 | 107 | 29 | 84,47 | 95,25 | 89,54 |
| Total | 9076 | 2177 | 633 | | | |
| **Average** | **567,25** | **136,06** | **39,56** | **80,64** | **93,48** | **86,57** |

TABLE 5.6 *Evaluation of the detection of episodes, based on the ball contacts data for the* Bayern Games. 6266 *episodes were correctly classified.*

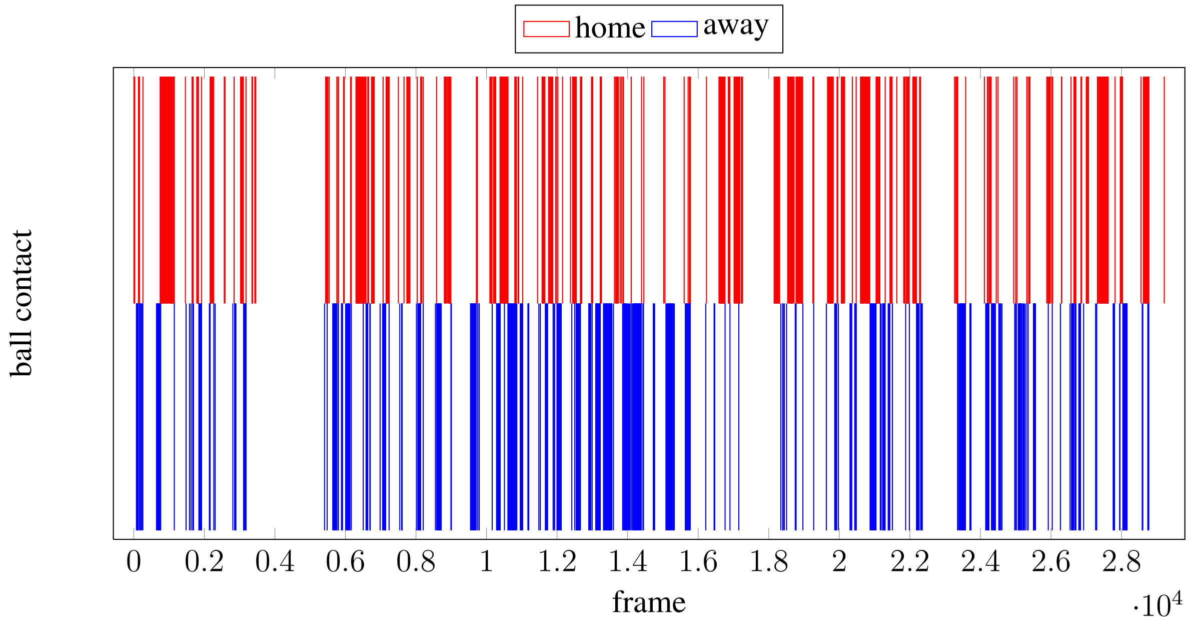


FIGURE 5.14 *Ball possession representation for the fourth game (first half) of the* Bayern Games. *The abscissa represents the frame number, and the ordinate represents team ball possession. The home team possession corresponds to the upper part (red) and the away team possession to the lower part (blue). The white regions represent the times when no team is in possession due to game interruptions.*

Once the episodes for each team have been found, the next step is to classify each action inside the sequence of actions that forms the episode. To fulfil this task, the ball contacts information will be used. In the table 5.7, it can be seen that almost all the actions correspond to passes and dribbles (around 80%), while the remaining actions are intereceptions and shots (around 20%). For the case of passes, they are detected based on the cases where a ball contact of a player is followed by another ball contact of a different player but both belonging to the same team. The results of the detection are presented in the table 5.8. For the generation of the errors, the observed passes were compared against the ground truth passes based on their frame span match, and as with the case of the episodes detection, this might lead to higher errors due to only a small miss-match in the frame boundaries of the events. In average, the values for precision, recall and F1-score are respectively $92,67$, $99,67$, and $95,82$, and as explained before, this shall be considered as the worst case scenario for the classification. The same reasoning applies to the other actions: dribbles, interceptions and shots. Once the passes are detected, then a series of features are obtained such as: team association, time duration, initial and final spatial positions, length, and angle. All those calculated features are very useful in posterior classifications of the passes for tactical analysis purposes.

In the case of dribblings, a dribble is defined as the acion when a player drives the ball for a determined time period, and where no change of player possesion occurs. The classification of dribbles can be performed based solely on the ball contacts. The results over the more than $650$ dribbles in the *Bayern Games* are in average $91,59$ for the precision, $93,37$ for the recall, and $92,47$ in the case of the $F1$-score, see table 5.9.

The interceptions, even not being a type of action that a player in possession can perform, is also detectable using the ball contacts approach. The results of the classification are depicted in the table table 5.10. Approximately, $400$ interceptions where classified, from which a precision, recall and F1-score of $92,58$, $97,63$, and $95,03$ respecctively, were obtained.

Finally, the shots are obtained from the remaining actions not yet classified into passes, dribbles, or interceptions, and by manually removing referee actions such as game halts and the like. The precision obtained was $73,71$, the recall $100,00$, and the total $F1$-score $84,79$. Undetected actions like fouls affect the performance of the classifier, and more research is required in this area, as mentioned before.

| **Game id** | **Actions** | **Relative Frequencies** | | | |
|---|---|---|---|---|---|
| | | **Passes** | **Dribbles** | **Interceptions** | **Shots** |
| 1 | 1703 | 0,39 | 0,34 | 0,24 | 0,02 |
| 2 | 2109 | 0,43 | 0,36 | 0,19 | 0,02 |
| 3 | 1776 | 0,39 | 0,35 | 0,23 | 0,03 |
| 4 | 1779 | 0,40 | 0,35 | 0,23 | 0,02 |
| 5 | 1909 | 0,44 | 0,36 | 0,19 | 0,01 |
| 6 | 1944 | 0,41 | 0,38 | 0,19 | 0,01 |
| 7 | 1899 | 0,43 | 0,37 | 0,18 | 0,01 |
| 8 | 1983 | 0,43 | 0,37 | 0,19 | 0,01 |
| 9 | 1623 | 0,37 | 0,35 | 0,26 | 0,02 |
| 10 | 1875 | 0,41 | 0,36 | 0,22 | 0,01 |
| 11 | 1684 | 0,41 | 0,35 | 0,22 | 0,02 |
| 12 | 2000 | 0,43 | 0,38 | 0,18 | 0,01 |
| 13 | 1740 | 0,39 | 0,35 | 0,24 | 0,02 |
| 14 | 1527 | 0,39 | 0,34 | 0,26 | 0,01 |
| 15 | 1832 | 0,41 | 0,35 | 0,22 | 0,02 |
| 16 | 1813 | 0,41 | 0,36 | 0,21 | 0,02 |
| **Average** | **1824,75** | **0,41** | **0,36** | **0,22** | **0,02** |

TABLE 5.7 *Relative frequencies of the offensive and defensive player actions with the ball for the* Bayern Games*, The column named 'Actions' presents the total count of the actions performed by the players of both teams (home and away). The remaining columns show the relative frequencies of the individual actions under the classification presented in section 5.1, The individual actions are sorted from left to right, depending on its descending occurrence, being the pass the most common action, and shot the least.*

| **Game id** | **Passes** | **False Negatives** | **False Positives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 1 | 670 | 57 | 1 | 92,16 | 99,85 | 95,85 |
| 2 | 910 | 61 | 6 | 93,72 | 99,34 | 96,45 |
| 3 | 696 | 72 | 3 | 90,63 | 99,57 | 94,89 |
| 4 | 705 | 70 | 2 | 90,97 | 99,72 | 95,14 |
| 5 | 833 | 46 | 1 | 94,77 | 99,88 | 97,26 |
| 6 | 797 | 74 | 2 | 91,50 | 99,75 | 95,45 |
| 7 | 824 | 57 | 3 | 93,53 | 99,64 | 96,49 |
| 8 | 849 | 59 | 2 | 93,32 | 99,76 | 96,43 |
| 9 | 599 | 80 | 4 | 88,22 | 99,34 | 93,45 |
| 10 | 774 | 52 | 5 | 94,23 | 99,41 | 96,75 |
| 11 | 688 | 59 | 2 | 92,10 | 99,71 | 95,76 |
| 12 | 868 | 42 | 2 | 93,45 | 99,67 | 96,46 |
| 13 | 676 | 62 | 0 | 93,33 | 100,00 | 96,55 |
| 14 | 590 | 63 | 2 | 92,47 | 99,74 | 95,97 |
| 15 | 747 | 55 | 3 | 92,48 | 99,56 | 95,89 |
| 16 | 737 | 80 | 2 | 89,58 | 99,71 | 94,38 |
| Total | 11963 | 989 | 40 | | | |
| **Average** | **747,69** | **61,81** | **2,50** | **92,67** | **99,67** | **95,82** |

TABLE 5.8 *Evaluation of the detection of passes, based on the ball contacts data for the* Bayern Games*, More than* 10934 *passes were correctly detected,*

| **Game id** | **Drib-bles** | **False Negatives** | **False Positives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 1 | 580 | 64 | 42 | 90,06 | 93,25 | 91,63 |
| 2 | 766 | 71 | 46 | 91,52 | 94,33 | 92,90 |
| 3 | 625 | 55 | 55 | 91,91 | 91,91 | 91,91 |
| 4 | 618 | 66 | 48 | 90,35 | 92,79 | 91,56 |
| 5 | 686 | 66 | 33 | 91,22 | 95,41 | 93,27 |
| 6 | 747 | 60 | 51 | 92,57 | 93,61 | 93,08 |
| 7 | 700 | 51 | 41 | 93,21 | 94,47 | 93,83 |
| 8 | 729 | 48 | 45 | 93,58 | 93,96 | 93,77 |
| 9 | 576 | 53 | 71 | 91,57 | 89,03 | 90,28 |
| 10 | 669 | 71 | 35 | 91,13 | 95,42 | 93,22 |
| 11 | 596 | 72 | 39 | 89,22 | 93,86 | 91,48 |
| 12 | 751 | 40 | 35 | 93,51 | 94,27 | 93,89 |
| 13 | 609 | 57 | 41 | 92,95 | 94,82 | 93,88 |
| 14 | 518 | 59 | 44 | 91,90 | 93,83 | 92,85 |
| 15 | 643 | 61 | 44 | 90,90 | 93,26 | 92,06 |
| 16 | 655 | 67 | 68 | 89,89 | 89,76 | 89,83 |
| Total | 10468 | 961 | 738 | | | |
| **Average** | **654,25** | **60,063** | **46,135** | **91,59** | **93,37** | **92,47** |

TABLE 5.9 *Evaluation of the detection of dribbles, based on the ball contacts data for the* Bayern Games*, More than* 8769 *dribbles were correctly detected,*

| **Game id** | **Intercep-tions** | **False Negatives** | **False Positives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 1 | 417 | 39 | 6 | 91,45 | 98,58 | 94,88 |
| 2 | 391 | 40 | 11 | 90,72 | 97,26 | 93,88 |
| 3 | 407 | 52 | 9 | 88,67 | 97,84 | 93,03 |
| 4 | 417 | 39 | 10 | 91,45 | 97,66 | 94,45 |
| 5 | 369 | 24 | 9 | 93,89 | 97,62 | 95,72 |
| 6 | 373 | 23 | 11 | 94,19 | 97,14 | 95,64 |
| 7 | 348 | 25 | 8 | 93,30 | 97,75 | 95,47 |
| 8 | 378 | 29 | 9 | 92,31 | 97,48 | 94,82 |
| 9 | 421 | 32 | 7 | 92,94 | 98,36 | 95,57 |
| 10 | 412 | 26 | 9 | 93,56 | 97,67 | 95,58 |
| 11 | 373 | 28 | 9 | 93,02 | 97,64 | 95,27 |
| 12 | 358 | 29 | 8 | 93,56 | 98,14 | 95,79 |
| 13 | 423 | 35 | 8 | 91,09 | 97,81 | 94,33 |
| 14 | 399 | 22 | 15 | 94,93 | 96,49 | 95,70 |
| 15 | 411 | 31 | 11 | 93,17 | 97,47 | 95,27 |
| 16 | 389 | 28 | 11 | 93,02 | 97,14 | 95,03 |
| 16 | 389 | 99 | 11 | 79,03 | 97,14 | 87,15 |
| Total | 6286 | 502 | 151 | | | |
| **Average** | **392,88** | **31,38** | **9,44** | **92,58** | **97,63** | **95,03** |

TABLE 5.10 *Evaluation of the detection of interruptions, based on the ball contacts data for the* Bayern Games*, More than* 5633 *interceptions were correctly detected,*

| **Game id** | **Shots** | **False Negatives** | **False Positives** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
|---|---|---|---|---|---|---|
| 1 | 36 | 6 | 0 | 85,71 | 100,00 | 92,31 |
| 2 | 42 | 11 | 0 | 79,25 | 100,00 | 88,42 |
| 3 | 48 | 19 | 0 | 71,64 | 100,00 | 83,48 |
| 4 | 39 | 13 | 0 | 75,00 | 100,00 | 85,71 |
| 5 | 21 | 8 | 0 | 72,41 | 100,00 | 84,00 |
| 6 | 27 | 11 | 0 | 71,05 | 100,00 | 83,08 |
| 7 | 27 | 11 | 0 | 71,05 | 100,00 | 83,08 |
| 8 | 27 | 11 | 0 | 71,05 | 100,00 | 83,08 |
| 9 | 27 | 15 | 0 | 64,29 | 100,00 | 78,26 |
| 10 | 20 | 8 | 0 | 77,14 | 100,00 | 87,10 |
| 11 | 27 | 8 | 0 | 77,14 | 100,00 | 87,10 |
| 12 | 23 | 10 | 0 | 72,97 | 100,00 | 84,38 |
| 13 | 32 | 9 | 0 | 71,88 | 100,00 | 83,64 |
| 14 | 20 | 8 | 0 | 71,43 | 100,00 | 83,33 |
| 15 | 31 | 11 | 0 | 74,42 | 100,00 | 85,33 |
| 16 | 32 | 10 | 0 | 72,97 | 100,00 | 84,38 |
| Total | 479 | 169 | 0 | | | |
| **Average** | **29,94** | **10,56** | **0** | **73,71** | **100,00** | **84,79** |

TABLE 5.11 *Evaluation of the detection of shots, based on the ball contacts data for the* Bayern Games*,* 310 *shots were correctly detected,*

## 5.4 Semantic Analysis Examples

In the following section, particular examples of the use of the ASPOGAMO Abstract Model are discussed. In the first place, game sheets will be discussed. The games sheets are used for representation of game episodes and as a game summarisation tool. In the second place, passes analysis for the detection of fail passes and its reasons, and particularly those situations considered as disastrous passes. The system is not limited to the above mentioned applications, of course, since many more queries can be answered by means of the representing ASPOGAMO Abstract Model, but they represent general applications of the model to answer specific question that might be posed.

### 5.4.1 Game Sheets

The *game sheets* are one example of the higher level usage of the ASPOGAMO Abstract Model. A game sheet is a tool for game summarisation, that can be automatically generated for each episode in a game, by means of the automated semantic annotation system discussed in the present work. A game sheet consists of an image, where the actions performed by the players are depicted using special symbols (see page xxv). The game sheets are a quick first view of the happenings in a game, and summarises the whole game in a few images. Besides the performed actions, also the trajectories of the players are depicted in the games sheets. As a reference, the *Bayern Games*, contain in average approximately $567$ episodes per game, as can be seen from the table 5.6.

An example of a game sheet for a particular episode is shown in figure 5.15. Having a closer look at the $BG1E10$ depicted, it can be observed that it is a very interesting episode. The ball is being possessed by the away player #30 (blue), who passes the ball, avoiding loosing it under the intense marking of the two home players #6 and #7 (red). The away player #8 receives the pass and starts a dribble by the right side, being accompanied by team-mates #30 and #10, as well as by opposing players #7, #15 and #16. The player with the ball then reaches the shoot position, and being blocked by player #7 and #16, shoots aiming the goal.

In the figure 5.16, the first 20 episodes of the first game of the *Bayern Games* can be observed. Each sheet corresponds to an episode in the game, from the time when the game starts with the kick-off in the episode 1, see subfigure (1), followed by a series of passes and dribbles, until the ball is caught by the opposing team. In the subfigure (2) the away team (blue) tries to do a dribble to move the ball to the opposing goal, but then the ball is stealth again by the home team (red). From the episodes #11 and #12 (subfigure (11) and subfigure (12)), which correspond to the actions following the events described in figure 5.15, it can be seen

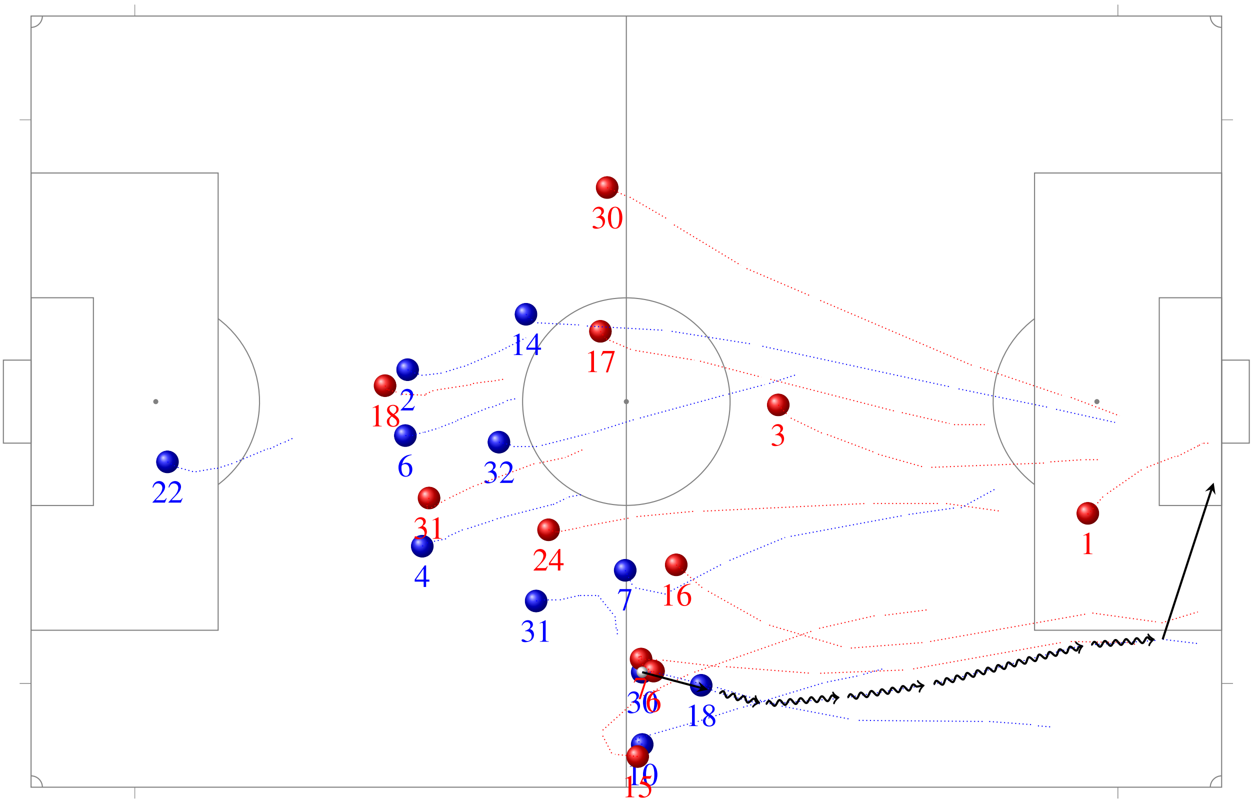
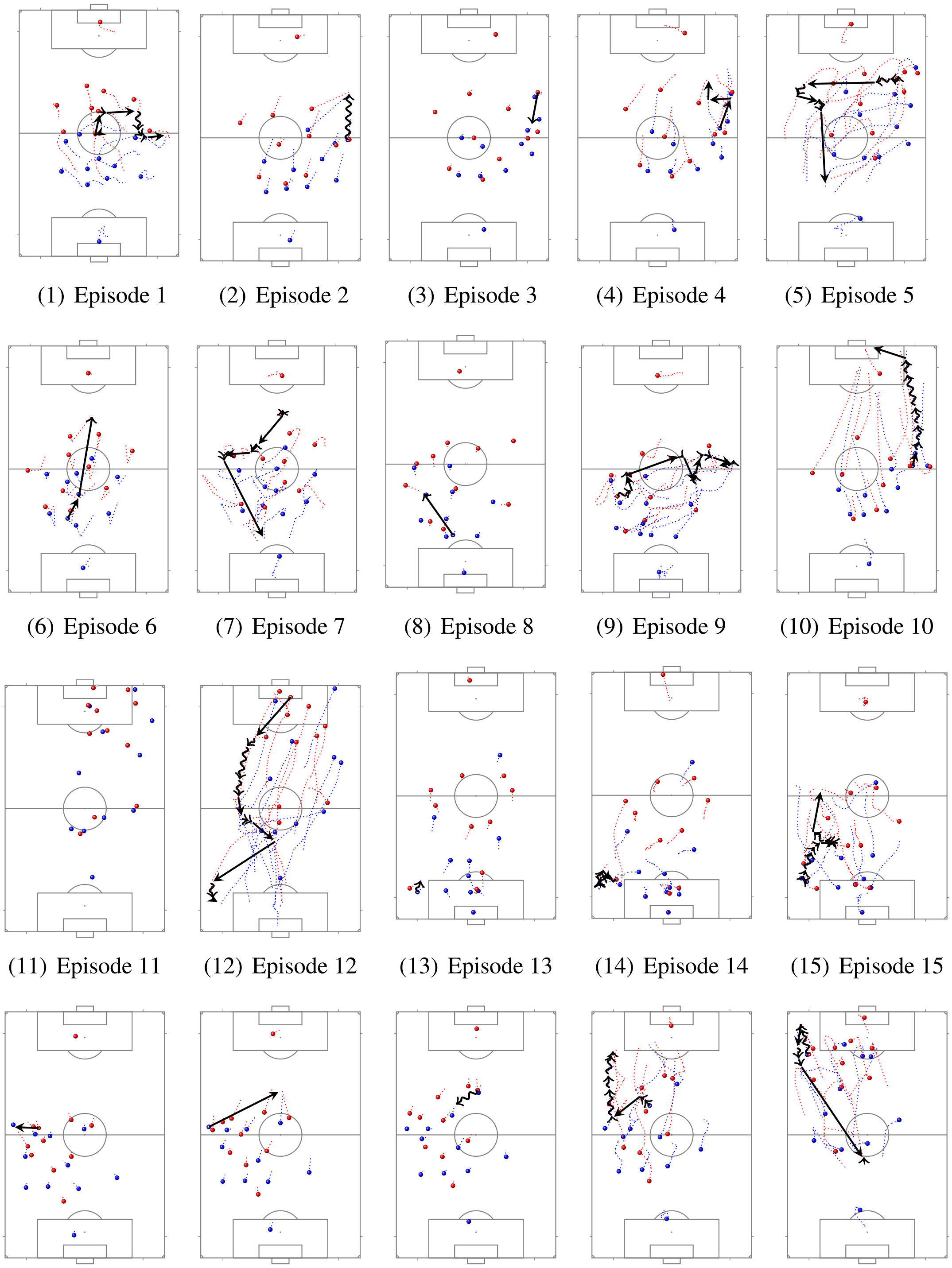


FIGURE 5.15 *Game Sheet corresponding to the episode #10 of the first game of the* Bayern Games

that a failing shot happened, since the next action was a goal-kick after the ball going out of bounds through the touch line. A counter-attack attempt is performed by the home team after the corresponding goal-kick.

For some particular audiences (players, referees, trainers, journalists, and others) not all of the episodes are interesting, so the game sheets provide a way to filter out those uninteresting episodes. In that case, a selection is made, for example, by choosing only those episodes with a minimum number of actions (avoiding singled-action episodes), or with a particular type of action (shots), or involving a particular player, or containing actions when the attacking team reaches the opposing third of the field, as well as others. As an example, in the figure 5.17, all the episodes containing shots from the first half-time of the game #1 of the *Bayern Games* are shown. Such filtering allows for the easy and quick interpretation of a game, or part of it.

More specific information can be obtained out of the games sheets, for example satistical informaton about how many actions of a specific kind there are, how many passes where performed by a particular player, what are the most common episodes resulting in a goal, and many more. Besides, from the video database, the video section containing the interesting episode, codified in the game sheet, can be easily extracted and presented to the user of hte system, for posterior analysis, like in the debriefing sessions.



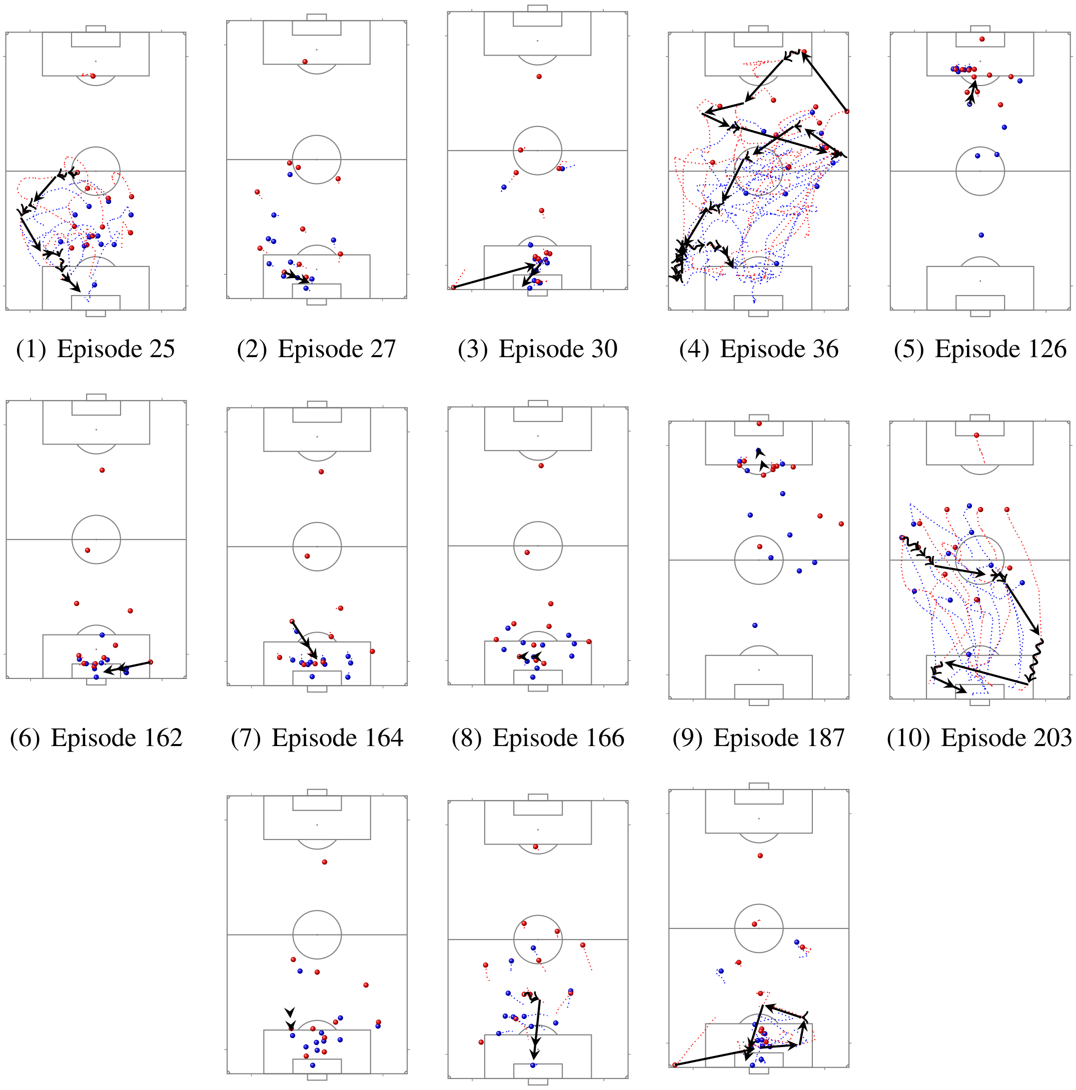
(1) Episode 1 (2) Episode 2 (3) Episode 3 (4) Episode 4 (5) Episode 5

(6) Episode 6 (7) Episode 7 (8) Episode 8 (9) Episode 9 (10) Episode 10

(11) Episode 11 (12) Episode 12 (13) Episode 13 (14) Episode 14 (15) Episode 15

(16) Episode 16 (17) Episode 17 (18) Episode 18 (19) Episode 19 (20) Episode 20

FIGURE 5.16 *Game Sheets corresponding to the first 20 episodes of the first game of the* Bayern Games.



(1) Episode 25

(2) Episode 27

(3) Episode 30

(4) Episode 36

(5) Episode 126

(6) Episode 162

(7) Episode 164

(8) Episode 166

(9) Episode 187

(10) Episode 203

(11) Episode 210

(12) Episode 232

(13) Episode 235

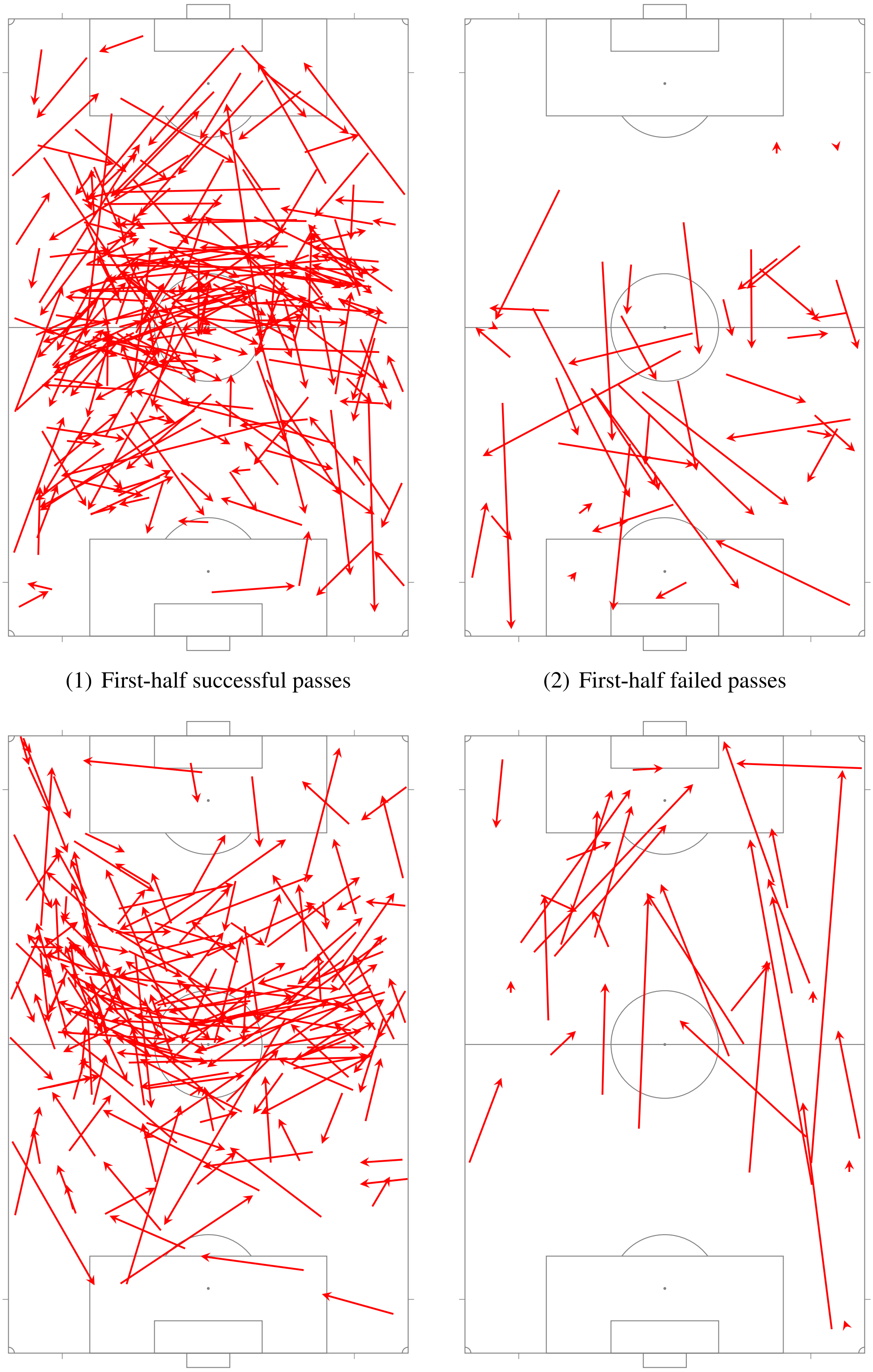
FIGURE 5.17 *Game Sheets corresponding to the 13 episodes containing shots for the first half of the first game of the* Bayern Games.

### 5.4.2 Passes Analysis

A pass is defined as the action performed between two players of the same team, with the aim of moving the ball from one field region to another, and of course with the desire to keep the ball possession. The ASPOGAMO Abstract Model allows also for the extraction of *passes charts*, which in a similar way as the games sheets, tries to summarise a particular aspect of the game, in this case, the passes of the teams. Examples of the passes charts are shown in the figure 5.18 and figure 5.19.

The pass chart allows for a quick review of the performance with respect to passes during the game, not only with respect to each half period, but also with respect to the other team. For example, in the figure 5.18 (1) a larger number of passes is observed, when compared with the figure 5.18 (3), which indeed represented an increase of around 20% between the first-half and second-half. Also, it can be easily noted that the home team have dominated the game regarding the passes, since the density of passes lines is higher in the figure 5.18 (1), than in the figure 5.19 (1). With respect to the failed passes, both teams in both half-times look qualitatively similar. Another aspect that can be observed is that the home team have played passes more in the middle-field, than the away team. Other similar charts for other actions (dribbles, shots, and others) can be generated for such interpretations of the game. Finally, in a complementary way, histograms of the pases lengths can be observed in the figure 5.20, where again, the domination of the home team is explicit (due to the larger counts for the passes).

The games sheets and passes charts are only examples of the kind of information that can be generated out of the ASPOGAMO Abstract Model. Depending on the special needs and specific requirements of the system's users, the required functionality can be easily assembled based on the established modules of the system, and if required, straightforward extensions can be quickly implemented.



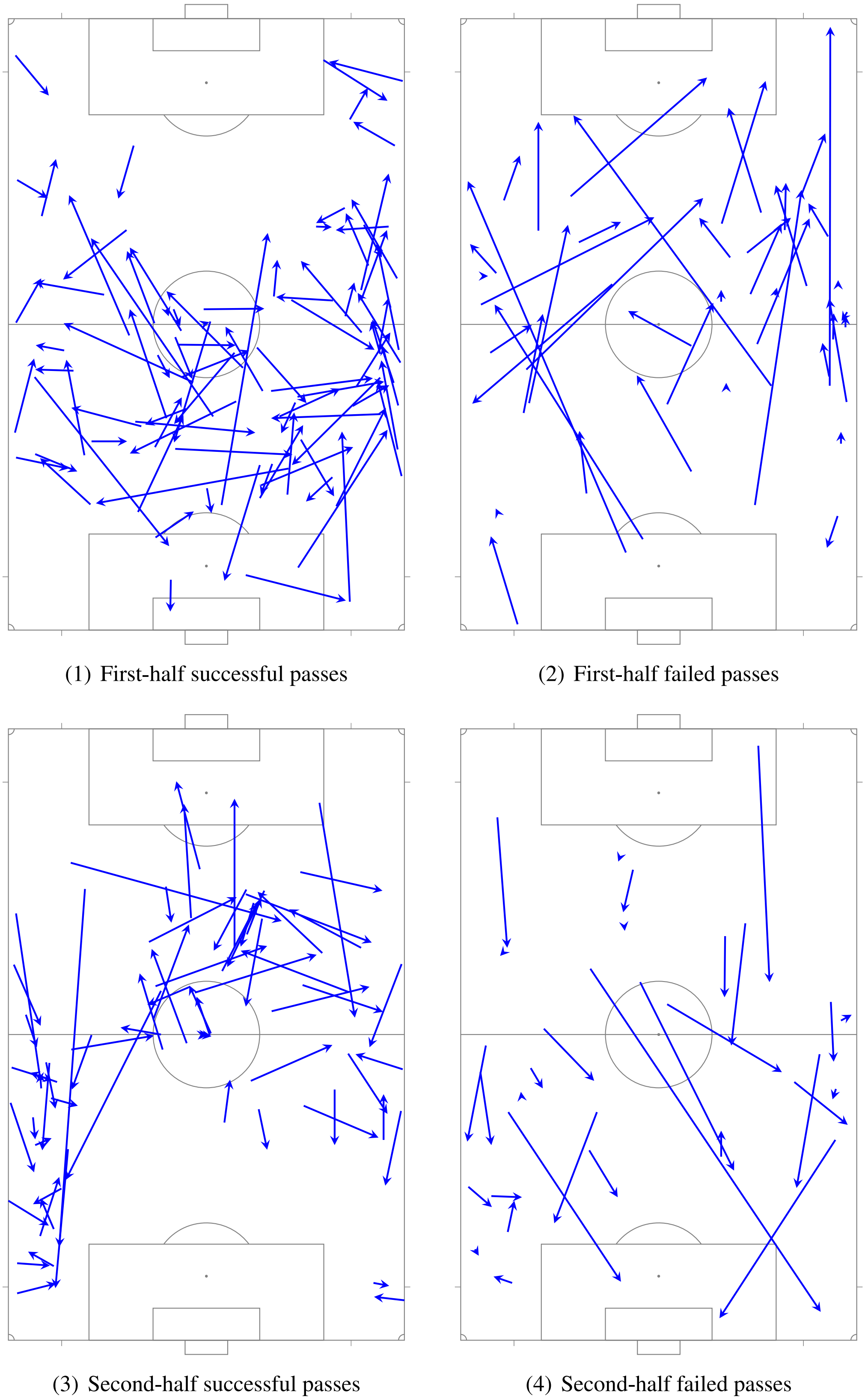
(1) First-half successful passes

(2) First-half failed passes

(3) Second-half successful passes

(4) Second-half failed passes

FIGURE 5.18 *Passes chart for the home team in the game #1 of the* Bayern Games. *The upper part of the figure corresponds to the first-half period, and the lower part to the second half-period. The left part of the figure corresponds to the successful passes, and the right part to the failed passes.*



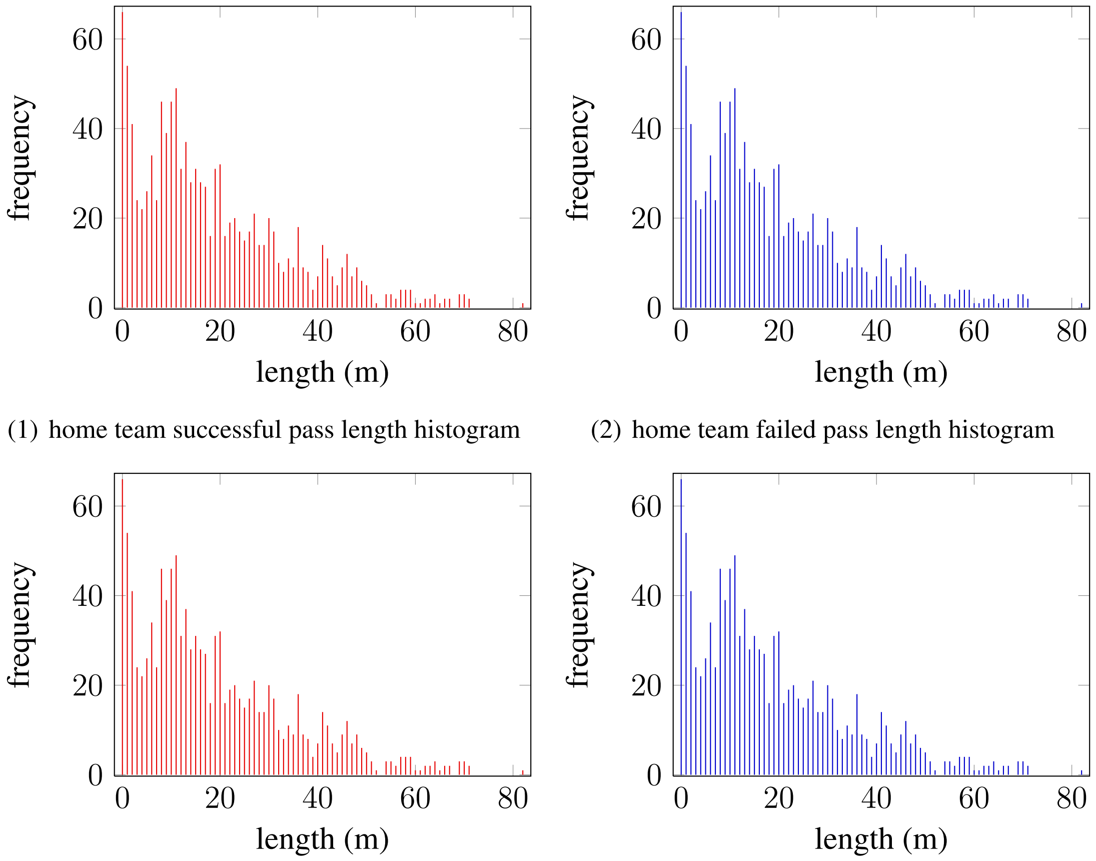
(1) First-half successful passes

(2) First-half failed passes

(3) Second-half successful passes

(4) Second-half failed passes

FIGURE 5.19 *Passes chart for the away team in the game #1 of the* Bayern Games. *The upper part of the figure corresponds to the first-half period, and the lower part to the second half-period. The left part of the figure corresponds to the successful passes, and the right part to the failed passes.*



(1) home team successful pass length histogram

(2) home team failed pass length histogram

(3) away team successful pass length histogram

(4) away team failed pass length histogram

FIGURE 5.20 *Pass length histograms. The upper part of the figure corresponds to the pass length histogram of the home team (first-half and second-half period). The lower part of the figure corresponds to the pass length histogram of the away team (first-half and second-half period). The left part of the figure corresponds to the successful passes, and the right part to the failed passes.*

# Chapter 6

# Conclusions

He who has a ’why’ to live, can bear with almost any ’how’.

*(Friedrich Nietzsche)*

## 6.1 Epitome

Association football is one of the diffused, richest and complex team sports currently practised, and since the recent development of more powerful computer systems, the interest in devising and implementing computational systems for analysing it have increased. Several applications could be supported using such computational systems, for example, game summarisation using play highlights, strategical, tactical, and statistical analyses, helping sports scientists to devise better training strategies, multimedia annotation, such as 2D and 3D video enhancements, helping in verifying referee decisions, and many more. A method for performing an objective analysis of football games is required to fulfil the needs of several football-related groups, such as coaches, journalists, sport scientists, fans, and others. Due to this necessities, the main objective of this thesis was to investigate mechanisms for the creation of a software application for the objective automated analysis of football games from TV broadcast video signals.

For the realisation of such a system, an abstract model was created, such that it stored the game-related data for the appropriate extraction and presentation of the relevant information to the football interested groups. This model allowed the answering of queries about the game, in such a way as to minimise the subjectivity present in the human interpretation of the game. Various questions could be answered, for example about the score chances of a player in a particular situation, the passes of a team and its characteristics, the important episodes occurred during the game, the performance of teams and players, and many more. The main idea was to provide a computational system with which generate an objective analysis of football games when required.

The hypothesis of the thesis was that the required abstract model could be created based on the positional information of the football targets: players, referees, and the ball. For fulfilling such objective, the input video was analysed, in order to extract only those frames corresponding to far-view scenes, which correspond to scenes were the targets positions could be extracted. It was presented how the trajectories of the targets could be generated using a video tracking strategy, where an algorithm of target detection and target localisation were presented, implemented, and evaluated. After that, every layer of the hierarchical abstract model was built, based on the information of the layers bellow each one, until the lowest level – the positional layer – was reached, and our hypothesis proved correctly. It was shown that with the ASPOGAMO abstract model, a game was represented as a sequence of actions forming episodes, with its own semantic meaning. Thus, the sought rich representation of a football game was found, and with it, the concrete implementation of an automated semantic annotation system for association football from TV broadcast video was achieved.

As a comparison with our system, the products from several existing companies (and some already fused or vanished) were presented, as well as some of the proposed methods found in the professional literature. The main objective of the present work was focused on the association football, nevertheless, some advances in other sports were revised. The review of this systems, helped in structuring an understanding of others' efforts, and motivated us to solve some of the existing problems, and to propose our presented integral solution. Our proposed solution was presented, and the computational framework used to accomplish it was also described.

In particular, for the *temporal segmentation* of the input video, the diverse TV editing effects were analysed. A new dissimilarity measure for detecting the editing effects for filtering them was presented and evaluated. Also, once the shot boundaries were found, the generated frame sequences were classified. Once the interesting and informative video shots corresponding to the far-view scenes were selected, the positional data had to be generated in a process called *spatial segmentation*. For accomplishing the spatial segmentation, a method of target detection and localisation was devised. With this data filling the lower level of the abstract model, then it was possible to estimate the trajectories of the interesting targets, which are the fundamental input to the remaining modelling and analysis. In order to model the football game, its logical structure was presented. This description was used to support the identification of a suitable game model based on measurable observations. Once identified, the ASPOGAMO abstract model was proposed and explained, in which several layers codifies different aspects of the game and gave rise to a *semantic segmentation* of the game. Last but not least, some examples of the use of the model were presented as illustrations of the capabilities of the system.

## 6.2 Contributions

The implementation of a software platform for the automated generation of a semantic annotation of football games from TV broadcast video, is the principal contribution of the current work. The semantic annotations are aimed at the interpretation of the game,to serve different uses in the football-related community. For example, the recognition of offensive episodes in an automatic manner, or the game summarisation by means of the proposed game sheets, are two specific abilities of the developed system. Also, more advanced game analysis are possible, and for that a series of analysis tools were implemented as part of the system, see table 7.1 and table 7.2.

For the implementation of the semantic annotation system, the required perception mechanisms producing the positional data out of the input video were devised, implemented, and evaluated. In the particular case of the shot boundary detection in chapter 3, a new dissimilarity measure was used for the scene representation. The proposed measure combines colour information, as well as texture of the scenes, producing a very robust feature, capable of reaching an average F1-score of $94,48\%$, for the classification. The temporal segmentation of the original video allows for different presentations of the information, for example, indexing of the highlights of the game can easily accessed though our video database system.

Another contribution lies in the new *spatial segmentation* proposed method for the detection and localisation of the targets. The algorithm uses colour and texture for the classification of the foreground and background, proving a robust method against strong illumination changes and other adverse conditions. The algorithm uses for the estimation of the state vector of the system a minimum matching of a bipartite graph, which despite its simplicity, it has proven to be very convenient, even when partial or total occlusions of the targets were present.

The implementation of the functionality required to build the abstract model in the *semantic segmentation*, was also a contribution of the present work. The model allows for the retrieval of useful information to answer posed queries in an attempt to interpret football games. Many functions and procedures were supplied, and mechanisms for the incorporation of new functionality were added, to update or expand the model. As a test bench, a simple web interface for the system was created, so users could remotely process their own videos and generate interesting analyses.

Many games were used for the evaluation of the system, for example from the FIFA's World Cups: 2002, 2006, and 2010; also local games (*Bayern Games*) were tested. Comparing to the results found in the professional literature (see for example [DL10]), we believe that a good game covering was achieved.

During the multidisciplinary ASPOGAMO project, a great cooperation between computer scientists and sport scientists was generated, in order to develop the game model, and to test their functionality.

## 6.3 Future Directions

In the next paragraphs, some ideas resulting after the finalisation of the current work will be delineated, as to provide some clues about how to achieve new or better characteristics for the system. Some of the presented ideas have been in some way, already explored as part of the current thesis, but not completely implemented or evaluated, so we believed they are a good starting point for the continuing development of the system.

As mentioned in the chapter 3, the problem of detecting the long dissolves is still open, so it is important to develop new algorithms for the detection of such editing events. Also, it is necessary to integrate parallelisation techniques into the routines used to populate the abstract model, especially those processing images from the input video, since that code sections consume most of the resources, due to the massive amount of data.

Also, as mentioned in chapter 5, more research is required to develop classification mechanisms for the recognition of actions, or events based on the deformation of the contour shapes of the players during the game. The individual player action recognition for the detection of: runs, walks, jumps, and others is a desired feature. Some classification experiments were conducted, as described in [Lad10]. In the same vein, it is desirable to be able to recognise actions not depending solely on the positional data of the targets, for example, fouls, free kicks, penalties, penalisation cards, and others. This will probably required the fusion with other information sources other than the trajectories followed by the targets.

Since the incompleteness of the bipartite graph for the target localisation is a useful property, then the previous classification of the players to each corresponding team would eliminate some of the bi-graph edges. This increase in the decoupling of the bipartite graph, will speed up the minimum matching calculation, and of course, if the team association is correctly performed, then better results in the target localisation will be reflected.

Even, if the system was thoroughly tested in many football games (several games from *WCs*, as well as the *Bayern Games*), it is important to investigate how the system can be adapted to other sports, such as hockey, rugby, and others. Besides, capabilities to deal with other weather conditions like snow are to be incorporated to the system, for example during the game Costa Rica - United States of America, in the CONCACAF qualification round for the *WC*2014, played on 2013.03.23. The system can be extended to the modelling and analysis of other human activities, as was already showed for cognitive robotics in [TB08].

In the case of the target localisation algorithm, the Markov property was assumed, in the sense that the current state vector $\mathbb{X}_k$ only depended on the last state vector $\mathbb{X}_{k-1}$. But it might be the case, that extending the past time being considered for the estimation of the current state, the results can be improved, specially in the case of occlusions or clutter of the targets. That is, a longer sequence of past state vectors: $\{\mathbb{X}_{k-1}, \cdots, \mathbb{X}_{k-n}\}$, $n \in \mathbb{N}$ should be considered for obtaining the current state vector.

Due to time constraints, our proposed target localisation algorithm was not tested together with the system developed in [vHH10], also as part of the ASPOGAMO project. In that work, good results were presented for the tracking of the targets for a whole game, but the old template matching was still used for the target localisation, and we would like to believe that if better positional data is given to the tracker, then much better results can be achieved. Also, testing of both systems working together on the evaluation data used in the present work (several complete games) will be useful.

# Bibliography

[AC97] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. In *Nonrigid and Articulated Motion Workshop*, pages 90–102. IEEE, 1997.

[AG06] Y.Demiris A.Dearden and O. Grau. Tracking football player movment from a single moving camera using particle filters. In *The 3rd European Conf. on Visual Media Production (CVMP 2006)*, 2006.

[AKM03] Z. Aghbari, K. Kaneko, and A. Makinouchi. Content-trajectory approach for searching video databases. *Multimedia, IEEE Transactions on*, 5(4):516 – 531, dec. 2003.

[ana] The Analysis Zone. http://www.theanalysiszone.com/products.asp.

[AR11] J. K. Aggarwal and M. S. Ryoo1. Human activity analysis: A review. In *Journal ACM Computing Surveys (CSUR)*, volume 43, ACM New York, NY, USA, 2011.

[Asc11] Ascensio System Ltd. Ascensio system. http://www.footballsoftpro.com/, 2011.

[Aut] AutomaticAthlete. Profile: the vis.track system. http://sportsactivated.com/vistrack.

[AWKG05] E.L. Andrade, J.C. Woods, E. Khan, and M. Ghanbari. Region-based analysis and retrieval for tracking of semantic objects and provision of augmented information in interactive sport scenes. *Multimedia, IEEE Transactions on*, 7(6):1084 – 1096, dec. 2005.

[Bad90] A. D. Baddeley. Psychology Press, 1990.

[BBG$^+$06a] Michael Beetz, Jan Bandouch, Suat Gedikli, Nico von Hoyningen-Huene, Bernhard Kirchlechner, and Alexis Maldonado. Camera-based observation of football games for analyzing multi-agent activities. In *Proc. of Int. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

[BBG$^+$06b] Michael Beetz, Jan Bandouch, Suat Gedikli, Nico von Hoyningen-Huene, Bernhard Kirchlechner, and Alexis Maldonado. Camera-based observation of football games for analyzing multi-agent activities. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

[BBK05] Lluis Barceló, Xavier Binefa, and John R. Kender. Robust methods and representations for soccer player tracking and collision resolution. In *Proc. of Int. Conf. on Image and Video Retrieval (CIVR)*, pages 237–246, 2005.

[BBR96] John Boreczky, , John S. Boreczky, and Lawrence A. Rowe. Comparison of video shot boundary detection techniques. pages 170–179, 1996.

[BDH04] Adrien Bartoli, Navneet Dalal, and Radu Horaud. Motion panoramas. *In Journal of Computer Animation and Virtual Worlds*, 15(5):501–517, nov 2004.

[BGB$^+$07] Michael Beetz, Suat Gedikli, Jan Bandouch, Bernhard Kirchlechner, Nico v. Hoyningen-Huene, and Alexander Perzylo. Visually tracking football games based on TV broadcasts. In *Proc. of Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2007.

[bHPV04] Jean bernard Hayet, Justus Piater, and Jacques Verly. Robust incremental rectification of sports video sequences. In *In: British Machine Vision Conference*, pages 687–696, 2004.

[BI06] Bruce R. Barringer and R. Duane Ireland. *Entrepreneurship: Successfully Launching New Ventures*. Pearson Prentice Hall, 2006.

[BKL05] Michael Beetz, Bernhard Kirchlechner, and Martin Lames. Computerized real-time analysis of football games. *IEEE Pervasive Computing*, 4(3):33–39, 2005.

[BLM00] A. Bonzanini, R. Leonardi, and P. Migliorati. Semantic video indexing using mpeg motion vectors. In *Proc. EUSPICO' 2000*, pages 147–150, Tampere, Finddland, September 2000.

[BOWT07] Paul Bradley, Peter O'Donoghue, Blake Wooster, and Phil Tordoff. The reliability of prozone matchviewer: a video-based technical performance analysis system. In *International Journal of Performance Analysis of Sport-e*, number 7 in 3, pages 117–129, 2007.

[BSF88] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*, volume 179 of *Mathematics in Science and Engineering*. Academic Press, 1988.

[BvHHK$^+$09] Michael Beetz, Nicolai v. Hoyningen-Huene, Bernhard Kirchlechner, Suat Gedikli, Francisco Siles, Murat Durus, and Martin Lames. ASPOGAMO: Automated Sports Game Analysis Models. *Int. Journal of Computer Science and Sports (IJCSS)*, 2009.

[Cai11] Cairos AG. Cairos AG. http://www.cairos.com, 2011.

[Chr05] Christopher Carling, A. Mark Williams and Thomas Reilly. *Handbook of Soccer Match Analysis*. Routledge, 2005.

[CS05] Kyuhyoung Choi and Yongduek Seo. Tracking soccer ball in tv broadcast video. In Fabio Roli and Sergio Vitulano, editors, *Image Analysis and Processing - ICIAP 2005*, volume 3617 of *Lecture Notes in Computer Science*, pages 661–668. Springer Berlin / Heidelberg, 2005.

[CSCZ04] Shu-Ching Chen, Mei-Ling Shyu, Min Chen, and Chengcui Zhang. A decision tree-based multimodal data mining framework for soccer goal detection. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 265 – 268, june 2004.

[DFG] DFG Deutsche Forschungsgemeinschaft. http://gepris.dfg.de/gepris/OCTOPUS/;jsessionid=E70F11C0F66C71EBDEAEACF39FA0F018?module=gepris&task=showDetail&context=projekt&id=51462600.

[DFL] DFL Deutsche Fußball Liga GmbH. Ligavorstand vergibt Auftrag zur Erhebung offizieller Spieldaten. http://www.bundesliga.de/de/liga/news/2010/index.php?f=0000176864.php.

[DL10] T. D'Orazio and M. Leo. A review of vision-based systems for soccer video analysis. In *Pattern Recognition*, volume 43, pages 2911–2926, New York, NY, USA, 2010. Elsevier Science Inc.

[Dug] Addy Dugdale. Why fifa refuses to sanction goal-line technology. http://www.fastcompany.com/1664627/why-fifa-refuses-to-sanction-goal-line-technology.

[EBMM03] A.A. Efros, A.C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *Proc. of Int. Conf. on Computer Vision (ICCV)*, volume 2, pages 726–733, 2003.

[Eli11] Elite Sports Analysis. Elite Sports Analysis - Sports statistics and Performance analysis. http://www.elitesportsanalysis.com/index.htm, 2011.

[Emi11] Emilio Maggio and Andrea Cavallaro. *Video Tracking*. Routledge, 2011.

[ETM03] Ahmet Ekin, A. Murat Tekalp, and Rajiv Mehrotra. Automatic soccer video analysis and summarization. *IEEE Trans. on Image Processing*, 12(7):796–807, 7 2003.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[FIF07] FIFA. The fifa world cup tv viewing figures. http://www.fifa.com/mm/document/fifafacts/ffprojects/ip-401_06e_tv_2658.pdf, October 2007.

[FIF11] FIFA. Laws of the game, 2011.

[FLB$^+$04] Pascual Figueroa, Neucimar Leite, Ricardo M. L. Barros, Isaac Cohen, and Gerard Medioni. Tracking soccer players using the graph representation. In *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, volume 4, pages 787–790, 2004.

[FLB06a] Pascual J. Figueroa, Neucimar J. Leite, and Ricardo M. L. Barros. Tracking soccer players aiming their kinematical motion analysis. *Compututer Vision and Image Understanding*, 101(2):122–135, 2006.

[FLB06b] P.J. Figueroa, N.J. Leite, and R.M.L. Barros. Background recovering in outdoor image sequences: An example of soccer players segmentation. *Image and Vision Computing*, 24(4):363–374, April 2006.

[Fre00] Jan-Gerd Frerichs. *Entwicklung eines In-situ-Mikroskops zur bildgestützten Online-Überwachung von Bioprozessen*. PhD thesis, Universität Hannover, 2000.

[GB00] J.M. Garbarino and E. Billi. A video computerised tool for analysing the trajectories of players in team game. Videotape at Symposium Technology and Sport, L. Katz (Chair), Calgary, Canada, June 2000.

[GBvHH$^+$07] Suat Gedikli, Jan Bandouch, Nico von Hoyningen-Huene, Bernhard Kirchlechner, and Michael Beetz. An adaptive vision system for tracking soccer players from variable camera settings. In *Proc. of Int. Conf. on Computer Vision Systems (ICVS)*, 2007.

[Ged09] Suat Gedikli. *Continual and Robust Estimation of Camera Parameters in Broadcasted Sports Games*. PhD thesis, Technische Universität München, 2009.

[GN08] P. Geetha and Vasumathi Narayanan. A survey of content-based video retrieval. In *Journal of Computer Science 4*, 2008.

[GP08] Nicolas Gengembre and Patrick Pérez. Probabilistic color-based multi-object tracking with application to team sports. Technical Report 6555, INRIA, May 2008.

[Heg] Hego Group. Graphics solutions and services for enhancing tv and sport. http://www.hegogroup.com/.

[HLB07] Yu Huang, Joan Llach, and Sitaram Bhagavathy. Players and ball detection in soccer videos based on color segmentation and shape analysis. In N. Sebe, Y. Liu, and Y. Zhuang, editors, *Int. Workshop on Multimedia Content Analysis and Mining (MCAM)*, number 4577 in Lecture Notes in Computer Science (LNCS), pages 416–425. Springer, 2007.

[Hof79] Douglas R. Hofstadter. *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, twentieth-anniversary edition edition, 1979.

[HSC06] Chung-Lin Huang, Huang-Chia Shih, and Chung-Yuan Chao. Semantic analysis of soccer video using dynamic bayesian network. *Multimedia, IEEE Transactions on*, 8(4):749 –760, aug. 2006.

[HWL$^{+}$09] X.L. Han, L.F. Wu, X.S. Liu, Z.H. Cheng, and Y. Gong. Offense-defense semantic analysis of basketball game based on motion vector. In *Proc. of Int. Symp. on Image Analysis and Signal Processing (IASP)*, pages 146–149, 2009.

[HWTH11] Zhipeng Hu, Yaowei Wang2, Yonghong Tian, and Tiejun Huang. Selective eigenbackgrounds method for background subtraction in crowed scenes. In *IEEE International Conference on Image Processing*. IEEE, 2011.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2003.

[iCo] iCoast. Welcome to the InternetCoast. http://www.internetcoast.com/index.php?src=news&srctype=detail&category=News&refno=166.

[Impa] Impire AG. News. http://www.bundesliga-datenbank.de/de/80/.

[Impb] Impire AG. Vis.track. http://www.impire.de.

[Ins] Institut für Spielanalyse. Homepage. http://www.spielanalyse.org/.

[IS07a] N. Inamoto and H. Saito. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *Multimedia, IEEE Transactions on*, 9(6):1155 –1166, oct. 2007.

[IS07b] N. Inamoto and H. Saito. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *IEEE Transactions on Multimedia*, 9(6):1155–1166, 2007.

[Jan10] P.K. Janert. *Data Analysis with Open Source Tools*. O'Reilly, 2010.

[JC03] Gaël Jaffré and Alain Crouzil. Non-rigid object localization from color model using mean shift. In *Proc. of the International Conf. on Image Processing (ICIP)*, pages 317–320, 2003.

[JdOA04] Bruno Müller Junior and Ricardo de Oliveira Anido. Distributed real-time soccer tracking. In *Proc. of ACM Int. Workshop on Video Surveillance and Sensor Networks*, 2004.

[JH00] Bernd Jähne and Horst Haußenecker. *Computer Vision and Applications*. Academic Press, 2000.

[Jon09] Jonathan Wilson. *Inverting the Pyramid*. Orion Books Ltd., 2009.

[KC01] Irena Koprinska and Sergio Carrato. Temporal video segmentation: A survey. In *Signal Processing: Image Communication*, pages 477–500, 2001.

[KCM03] Jinman Kang, Isaac Cohen, and Gerard Medioni. Soccer player tracking across uncalibrated camera streams. In *IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (PETS)*, 2003.

[KDD99] V. Kobla, D. DeMenthon, and D. Doermann. Detection of slow-motion replay sequences for identifying sports videos, 1999.

[KH00] Hyunwoo Kim and Ki Sang Hong. Soccer video mosaicing using self-calibration and line tracking. In *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, volume 1, pages 592–595, 2000.

[KZ08] Christopher A. Kurby and Jeffrey M. Zacks. Segmentation in the perception and the memory of events. In *Trends in Cognitive Sciences Volume 12, Issue 2, February 2008, Pages 72-79*, February 2008.

[KZW$^+$08] Y. Kong, X.Q. Zhang, Q.D. Wei, W.M. Hu, and Y.D. Jia. Group action recognition in soccer videos. In *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[Lad10] Franz Ladurner. Auswertung von algorithmen zur automatischen erkennung von spielaktionen in sportspielen (evaluation of automatic player action recognition algorithms used in sports games). Master's thesis, TUM Germany, 2010.

[Las] Lastdownload Software. Ascensio match expert 2.1. http://ascensio-match-expert.lastdownload.com/.

[LI94] John Chung-Mong Lee and Dixon Man-Ching Ip. A robust approach for camera break detection in color video sequence. In *in Proc. IAPR Workshop on Machine Vision Application (MVA'94*, pages 502–505, 1994.

[Lie99] Rainer Lienhart. Comparison of automatic shot boundary detection algorithms. pages 290–301, 1999.

[LK81] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 674–679, 1981.

[LKH$^+$02] Dario G. Liebermann, Larry Katz, Mike D. Hughes, Roger M. Bartlett, Jim McClements, and Ian M. Franks. Advances in the application of information technology to sport performance. *Journal of Sports Sciences*, 20(10):755–769, 2002.

[LLHG06] Yang Liu, Dawei Liang, Qingming Huang, and Wen Gao. Extracting 3d information from broadcast soccer video. *Image and Vision Computing*, 24(10):1146 – 1162, 2006.

[LMP03] R. Leonardi, P. Migliorati, and M. Prandini. Semantic indexing of sports program sequences by audio-visual analysis. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 1, pages I – 9–12 vol.1, sept. 2003.

[LMS$^+$08] Marco Leo, Nicola Mosca, Paolo Spagnolo, Pier Luigi Mazzeo, Tiziana D'Orazio, and Arcangelo Distante. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In *Proc. of Int. Conf. on Content-based Image and Video Retrieval (CIVR)*, volume 1, pages 525–534, Niagara Falls, Canada, 2008. ACM.

[LSA01] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5), May 2001.

[LTL$^+$07] Jia Liu, Xiaofeng Tong, Wenlong Li, Tao Wang, Yimin Zhang, Hongqi Wang, Bo Yang, Lifeng Sun, and Shiqiang Yang. Automatic player detection, labeling and tracking in broadcast soccer video. In *Proc. of British Machine Vision Conf. (BMVC)*, 2007.

[LTSH07] GuoJun Liu, XiangLong Tang, Da Sun, and JianHua Huang. Robust registration of long sport video sequence. In *Proc. of Int. Conf. on Computer Vision Systems (ICVS)*, 2007.

[Mar] Jeffrey Marcus. Fifa president apologizes for errors. http://www.nytimes.com/2010/06/30/sports/soccer/30ref.html?hp.

[Mas00] Massimo Lucchesi. *Soccer Tactics, An Analysis of Attack and Defense*. Reedswain Inc., 2000.

[Mas01] Massimo Lucchesi. *Attacking Soccer, A Tactical Analysis*. Reedswain Inc., 2001.

[McG] Liam McGrath. Goal-line technology: Crossing too many lines for fifa? http://www.geekweek.com/2010/04/goal-line-technology-cross-too-many-lines-for-fifa-1.html.

[McK] Noel McKeegan. The adidas intelligent football. http://www.gizmag.com/adidas-intelligent-football/8512/.

[MHK06] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90 – 126, 2006. Special Issue on Modeling People: Vision-based understanding of a person's shape, appearance, movement and behaviour.

[MRG02] Ram Mylvaganam, Neil Ramsay, and Frederic De Graca. Sports analysis system and method. International application published under the Patent Cooperation Treaty. International Publication Number WO 02/071334 A2, 2002.

[Nat] National Media Group. Lucent technologies: Lucent serves up global platform. http://www.nmgsports.com/lucent.htm.

[NB01] Chris J. Needham and Roger D. Boyle. Tracking multiple sports players through occlusion, congestion and scale. In *Proc. of British Machine Vision Conf. (BMVC)*, 2001.

[Nee03] Christopher James Needham. *Tracking and Modelling of Team Game Interactions*. PhD thesis, University of Leeds, October 2003.

[NSC06] Peter Nillius, Josephine Sullivan, and Stefan Carlsson. Multi-target tracking - linking identities using bayesian network inference. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 2187–2194, 2006.

[Opt11] Opta Sportsdata Ltd. Opta sportsdata. http://www.optasports.com/, 2011.

[Ora11] Orad. Trackvision. http://www.orad.tv/products/trackvision, 2011.

[ORP00] N. M. Oliver, B. Rosario, and A. P. Pentland. Eigenbackgrounds: A bayesian computer vision system for modeling human interactions. In *IEEE Trans. on Patt. Anal. and Machine Intell.*, volume 22(8), pages 831–843. IEEE, 2000.

[Osh] Jeremiah Oshan. Mls joins forces with opta sports: More statistics will soon be available. http://www.sbnation.com/soccer/2011/3/4/2029956/mls-opta-sports-statisics-chalkboards.

[Pan] Panini Digital. Digital Soccer Projects: Match Analysis Sample, World Cup2006. http://www.paninidigital.com/.

[Pan11] Panini Digital. Panini Digital. http://www.paninidigital.com/, 2011.

[PBS01] H. Pan, P. Van Beek, and M. Sezan. Detection of slow-motion replay segments in sports video for highlights generation, 2001.

[PF08] Donovan H. Parks and Sidney S. Fels. Evaluation of background subtraction algorithms with post-processing. In *IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, pages 192–199, 2008.

[Pic04] Massimo Piccardi. Background subtraction techniques: a review. In *Proc. of Int. Conf. on Systems, Man and Cybernetics*, pages 3099–3104. IEEE, 2004.

[PJC98] Gopal Pingali, Yves Jean, and Ingrid Carlbom. Real-Time Tracking for Enhanced Tennis Broadcasts. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 260–265, 1998.

[PJC99] Gopal Sarma Pingali, Yves Jean, and Ingrid Carlbom. Lucent vision: A system for enhanced sports viewing. In *Proceedings of the Third International Conference on Visual Information and Information Systems*, VISUAL '99, pages 689–696, London, UK, 1999. Springer-Verlag.

[PJOC00] Gopal Pingali, Yves Jean, Agata Opalach, and Ingrid Carlbom. Converting real world events into multimedia experiences. In *International Conference on Multimedia and Expo 2000*, 2000.

[PMMS08] V. Pallavi, Jayanta Mukherjee, Arun K. Majumdar, and Shamik Sural. Ball detection from broadcast soccer videos using static and dynamic features. *J. Visual Communication and Image Representation*, 19(7):426–436, 2008.

[POJ00] Gopal Pingali, Agata Opalach, and Yves Jean. Ball tracking and virtual replays for innovative tennis broadcasts. In *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, page 4152, Washington, DC, USA, 2000. IEEE Computer Society.

[POJC01] Gopal Pingali, Agata Opalach, Yves Jean, and Ingrid Carlbom. Visualization of sports using motion trajectories: Providing insights into performance, style, and strategy. In *Proc. of 12th Annual IEEE Visualization Conf. (Vis 2001)*, pages 75–82, 2001.

[Pro] ProZone Sports Ltd. Delivering Performance insights. `http://www.prozonesports.com/folder/prozone/pdf/page_2.pdf`.

[Pro11] ProZone Sports Ltd. MatchInsight. `http://www.prozonesports.com/index.html`, 2011.

[RCKL06] Myung-Cheol Roh, Bill Christmas, Joseph Kittler, and Seong-Whan Lee. Robust player gesture spotting and recognition in low-resolution sports video. In *Proc. European Conf. on Computer Vision (ECCV 2006)*, 2006.

[Rob99] Robyn Jones and Tom Tranter. *Soccer Strategies*. Reedswain Inc., 1999.

[ROTJ04] J.R. Renno, J. Orwell, D. Thirde, and G.A. Jones. Shadow classification and evaluation for soccer player detection. In *Proc. of British Machine Vision Conf. (BMVC)*, 2004.

[RRG$^+$04] J. Rymel, J. Renno, D. Greenhill, J. Orwell, and G.A. Jones. Adaptive eigenbackgrounds for object detection. In *ICIP 2004*, Digital Imaging Research Centre School of Computing and Information Systems, Kingston University, Kingston upon Thames, Surrey, KT1 2EE , UK, 2004.

[RRH$^+$05] T. Rodriguez, I.D. Reid, R. Horaud, N. Dalal, and M. Goetz. Image interpolation for virtual sports scenarios. *Machine Vision and Applications*, 16(4):236–245, September 2005.

[Rus06] Rusell Mclean, editor. *Football Encyclopedia*. Kingfisher Publications Pic, 2006.

[Saa] Saab Group. Defence and security. http://www.saabgroup.com/en/.

[SC04] Francisco Siles-Canales. Estimación de la forma y textura celular para microscopía in-situ (Estimation of the shape and texture of cells for in-situ microscopy). Tesis de licenciatura, Universidad de Costa Rica, 2004.

[SC13] Francisco Siles-Canales. Temporal Segmentation of Association Football from TV Broadcasting. In *INES 2013 – IEEE 17th International Conference on Intelligent Engineering Systems*, 2013.

[SCBC06] Valter Di Salvo, Adam Collins, McNeill Barry, and Marco Cardinale. Validation of prozone ®: A new video-based performance analysis system. In *International Journal of Performance Analysis of Sport-e*, number 6 in 1, pages 108–119, 2006.

[Set03] Daniel Setterwall. Computerised video analysis of football - technical and commercial possibilities for football coaching. Master's thesis, KTH Stockholm, 2003.

[SG99] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR)*, page 252, 1999.

[SH05] Marc Strickert and Barbara Hammer. Merge SOM for temporal data. *Neurocomputing*, 64:39–71, 2005.

[Sin06] James Sinclair. Categorisation in knowledge contexts. http://www.jrsinclair.com/academic/MidTermReview.php, 2006.

[Sit64] R. W. Sittler. An optimal data association problem in surveillance theory. *IEEE Trans. on Military Electronics*, 8:125–139, April 1964.

[SM10] Scott Murray. *Football For Dummies*. John Wiley and Sons, Ltd., 2010.

[SMND07] P. Spagnolo, N. Mosca, M. Nitti, and A. Distante. An unsupervised approach for segmentation and clustering of soccer players. In *Proc. of Int. Machine Vision and Image Processing Conf. (IMVIP)*, 2007.

[Spoa] Sport Universal. Amisco. http://213.30.139.108/sport-universal/uk/amiscopro.htm.

[Spob] Sport Universal. Amisco. http://www.unice.fr/ufrstaps/lamhes/articles.php.

[Spoc] Sport Universal. Amisco. http://www.sport-universal.com/news/PROZONE-and-AMISCO-join-forces-to-advance-sports-performance-analysis.html.

[Spod] Sportstec. SportsCode User Manual v7. http://www.sportstec.com/software/manuals/SportsCode_V7_WEB.pdf.

[Spoe] Sportstec. SportsCode User Manual v8. http://www.sportstec.com/UserFiles/PDFs/Manual-SportsCodeV8.pdf.

[Spo11a] Sport Universal. Amisco. http://www.sport-universal.com/, 2011.

[Spo11b] Sportstec. Sports Video Analysis Software Video Performance Analysis. http://www.sportstec.com/, 2011.

[SSMS06] T. Shimawaki, T. Sakiyama, J. Miura, and Y. Shirai. Estimation of ball route under overlapping with players and lines in soccer video image sequence. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 359 –362, 0-0 2006.

[Sta03] Thomas Stammeier. Automatische erwerb von bewegungs- und episodenmodellen für die spielanalyse im fußball. Master's thesis, TUM Germany, 2003.

[Str06] R. L. Streit. The PMHT and related applications of mixture densities. In *Proc. of Int. Conf. on Information Fusion (FUSION)*, Florence, Italy, July 2006.

[Sü05] Süddeutsche Zeitung AG. Verzögerung beim Chipball. Süddeutsche Zeitung AG, June 2005.

[TB08] Moritz Tenorth and Michael Beetz. Towards practical and grounded knowledge representation systems for autonomous household robots. In *Proc. of Int. Workshop on Cognition for Technical Systems*, Munich, Germany, October 2008.

[TCP04] D.W. Tjondronegoro, Yi-Ping Phoebe Chen, and Binh Pham. Classification of self-consumable highlights for soccer video summaries. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 579 –582 Vol.1, june 2004.

[TCSU08] Pavan Turaga, Rama Chellappa, V. S. Subrahmanian, and Octavian Udrea. Machine recognition of human activities: A survey. In *IEEE Transactions on Circuits and Systems for Video Technology*, 2008.

[Thea] The Guardian. Guardian interactive chalkboards. http://www.guardian.co.uk/football/chalkboards.

[Theb] The Guardian. World cup 2010 statistics: every match and every player in data. http://www.guardian.co.uk/news/datablog/2010/jul/09/world-cup-2010-statistics.

[Thec] The IET Faraday 2009. What is it? - opta sports data. http://faraday09.theiet.org/football/whatisit.cfm.

[Tho07] Graham Thomas. Real-time camera tracking using sports pitch markings. *Journal of Real-Time Image Processing*, 2:117–132, 2007. 10.1007/s11554-007-0041-1.

[Tra] Trakus. Trakus Homepage. http://www.trakus.com.

[TRA11] TRACAB. Tracab image tracking system. http://www.tracab.com/, 2011.

[VDP03] J. Vermaak, A. Doucet, and P. Pérez. Maintaining multi-modality through mixture tracking. In *Proc. of Int. Conf. on Computer Vision (ICCV)*, pages 1110–1116, 2003.

[vHH05] Nicolai v. Hoyningen-Huene. Grounding Action Models: Combining Data Mining and Knowledge Representation. Master's thesis, TUM Germany, 2005.

[vHH10] Nicolai v. Hoyningen-Huene. *Real-time Tracking of Player Identities in Team Sports*. PhD thesis, Technische Universität München, 2010.

[vHHKB07] Nicolai v. Hoyningen-Huene, Bernhard Kirchlechner, and Michael Beetz. GrAM: Reasoning with Grounded Action Models by Combining Knowledge Representation and Data Mining. In *Towards Affordance-based Robot Control*, 2007.

[VMP03] Nicolas Vandenbroucke, Ludovic Macaire, and Jack-Gerard Postaire. Color image segmentation by pixel classification in an adapted hybrid color space.

application to soccer image analysis. *Computer Vision and Image Understanding (CVIU)*, 90:190–216, may 2003.

[WADP97] C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. In *IEEE Trans. Pattern Anal. Machine Intell.*, volume 19(7), pages 780–785. IEEE, 1997.

[Wika] Wikipedia. Goal-line technology. http://en.wikipedia.org/wiki/Goal-line_technology.

[Wikb] Wikipedia. Wikipedia 2010 fifa world cup. http://en.wikipedia.org/wiki/2010_FIFA_World_Cup.

[WMZL05] Fei Wang, Yu-Fei Ma, Hong-Jiang Zhang, and Jin-Tao Li. A generic framework for semantic sports video analysis using dynamic bayesian networks. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 115 – 122, jan. 2005.

[WP04] J. R. Wang and N. Parameswaran. Survey of sports video analysis: Research issues and applications, 2004.

[WXC$^+$08] Jinjun Wang, Changsheng Xu, Engsiong Chng, Hanqing Lu, and Qi Tian. Automatic composition of broadcast sports video. *Multimedia Syst.*, 14(4):179–193, 2008.

[XLO04] Ming Xu, Liam Lowey, and James Orwell. Architecture and algorithms for tracking football players with multiple cameras. In *IEEE Intelligent Distributed Surveillance Systems (IDSS)*, pages 2909–2912, 2004.

[XXC$^+$04] Lexing Xie, Peng Xu, Shih-Fu Chang, Ajay Divakaran, and Huifang Sun. Structure analysis of soccer video with domain knowledge and hidden markov models. *Pattern Recogn. Lett.*, 25:767–775, May 2004.

[XZZ$^+$08] Changsheng Xu, Yi-Fan Zhang, Guangyu Zhu, Yong Rui, Hanqing Lu, and Qingming Huang. Using webcast text for semantic event detection in broadcast sports video. *Multimedia, IEEE Transactions on*, 10(7):1342 –1355, nov. 2008.

[YCK98] Yusseri Yusoff, William J. Christmas, and Josef Kittler. A study on automatic shot change detection. In *ECMAST '98: Proceedings of the Third European Conference on Multimedia Applications, Services and Techniques*, pages 177–189, London, UK, 1998. Springer-Verlag.

[YLC04] Yao-Quan Yang, Yu-Dong Lu, and Wei Chen. A framework for automatic detection of soccer goal event based on cinematic template. In *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, volume 6, pages 3759 – 3764 vol.6, aug. 2004.

[YlJBkY02] Ho-Sub Yoon, Young lae J. Bae, and Young kyu Yang. A soccer image sequence mosaicking and analysis method using line and advertisement board detection. *ETRI Journal*, 24(6):443–454, December 2002.

[YLXT06] X. Yu, H.W. Leong, C. Xu, and Q. Tian. Trajectory-based ball detection and tracking in broadcast soccer video. *Multimedia, IEEE Transactions on*, 8(6):1164 –1178, dec. 2006.

[YWX$^+$07] Jinhui Yuan, Huiyi Wang, Lan Xiao, Wujie Zheng, Jianmin Li, Fuzong Lin, and Bo Zhang. A formal study of shot boundary detection. In *Circuit and Systems For Video Technology, 2007, IEEE Transaction on*, pages 168–186, 2007.

[YYLL08] Xinguo Yu, Xin Yan, Liyuan Li, and Hon Wai Leong. An instant semantics acquisition system of live soccer video with application to live event alert and on-the-fly language selection. In *CIVR*, pages 495–504, 2008.

[ZHX$^+$07] Guangyu Zhu, Qingming Huang, Changsheng Xu, Yong Rui, Shuqiang Jiang, Wen Gao, and Hongxun Yao. Trajectory based event tactics analysis in broadcast sports video. In *Proc. of MULTIMEDIA*, pages 58–67, New York, USA, 2007.

[ZT01] Jeffrey M. Zacks and Barbara Tversky. Event structure in perception and conception. In *Psychological Bulletin*, volume 127, pages 3–21, 2001.

[ZXH09a] Guangyu Zhu, Changsheng Xu, and Qingming Huang. Sports video analysis: From semantics to tactics. In A. Divakaran, editor, *Multimedia Content Analysis, Signals and Communication Technology*, page 295. Springer, 2009.

[ZXH$^+$09b] Guangyu Zhu, Changsheng Xu, Qingming Huang, Yong Rui, Shuqiang Jiang, Wen Gao, and Hongxun Yao. Event tactic analysis based on broadcast sports video. *IEEE Trans. on Multimedia*, 11(1):49–67, January 2009.

[ZXZ$^+$08] Guangyu Zhu, Changsheng Xu, Yi Zhang, Qingming Huang, and Hanqing Lu. Event tactic analysis based on player and ball trajectory in broadcast video. In

*Proceedings of the 2008 international conference on Content-based image and video retrieval*, CIVR '08, pages 515–524, New York, NY, USA, 2008. ACM.

# Chapter 7

# Appendix

> Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.
>
> *(Antoine de Saint Exupéry)*

## 7.1 Football Ontology

The football ontology used through the current work while modelling the semantics of the game, is shown in the figure 7.1.

## 7.2 EER Diagrams

The EER diagram used for the implementation of the functionality of the ASPOGAMO Abstract Model is depicted in figure figure 7.2, and figure 7.3.

The functionality of the Database in order to implement the ASPOGAMO Abstract Model required a series of procedures in order to generate the data in the different layers of modelling. The functions implemented with a companion descripton are depicted in the

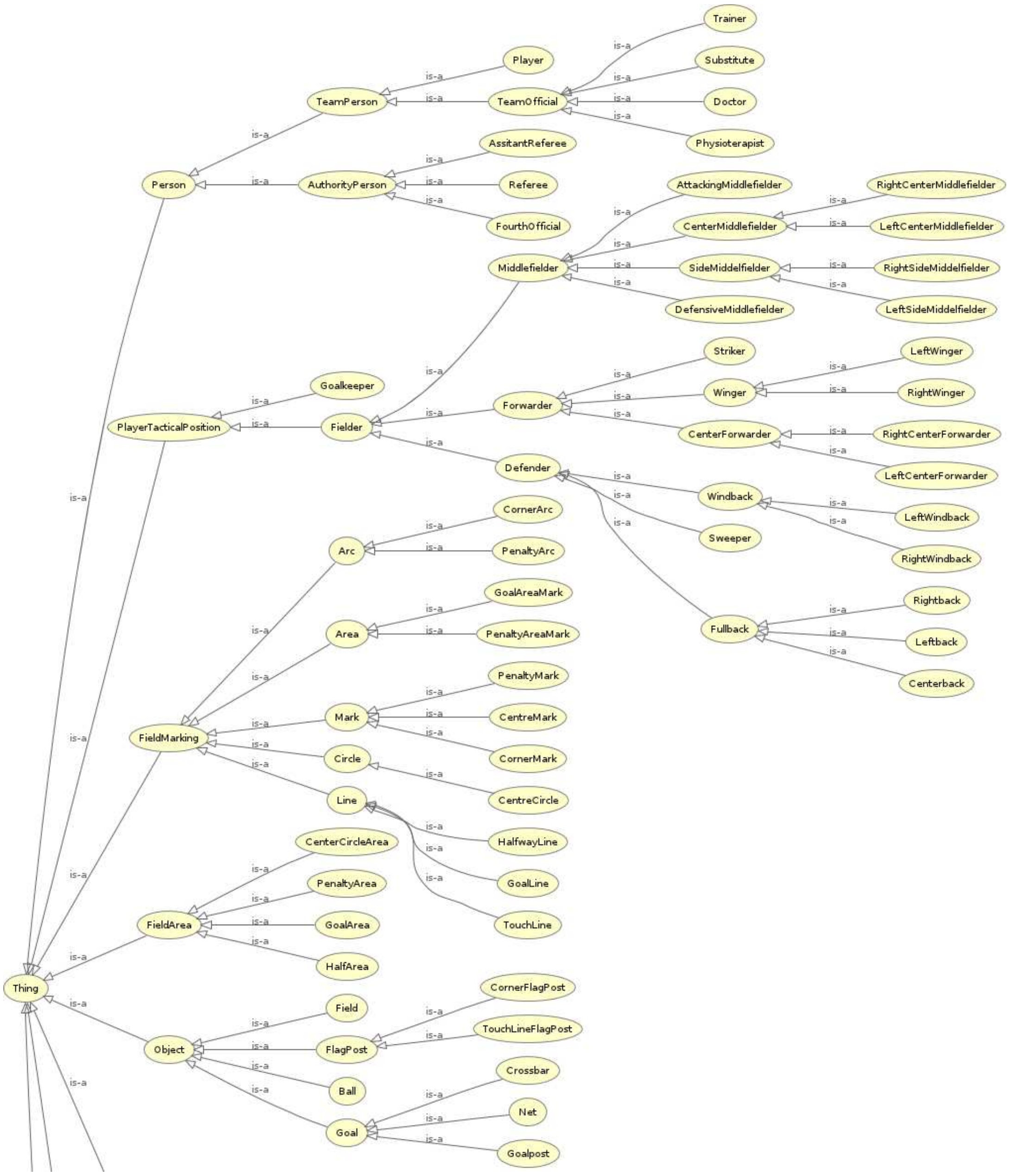


**FIGURE 7.1** *The asserted Football Ontology describing some of the objects and relations is depicted.*

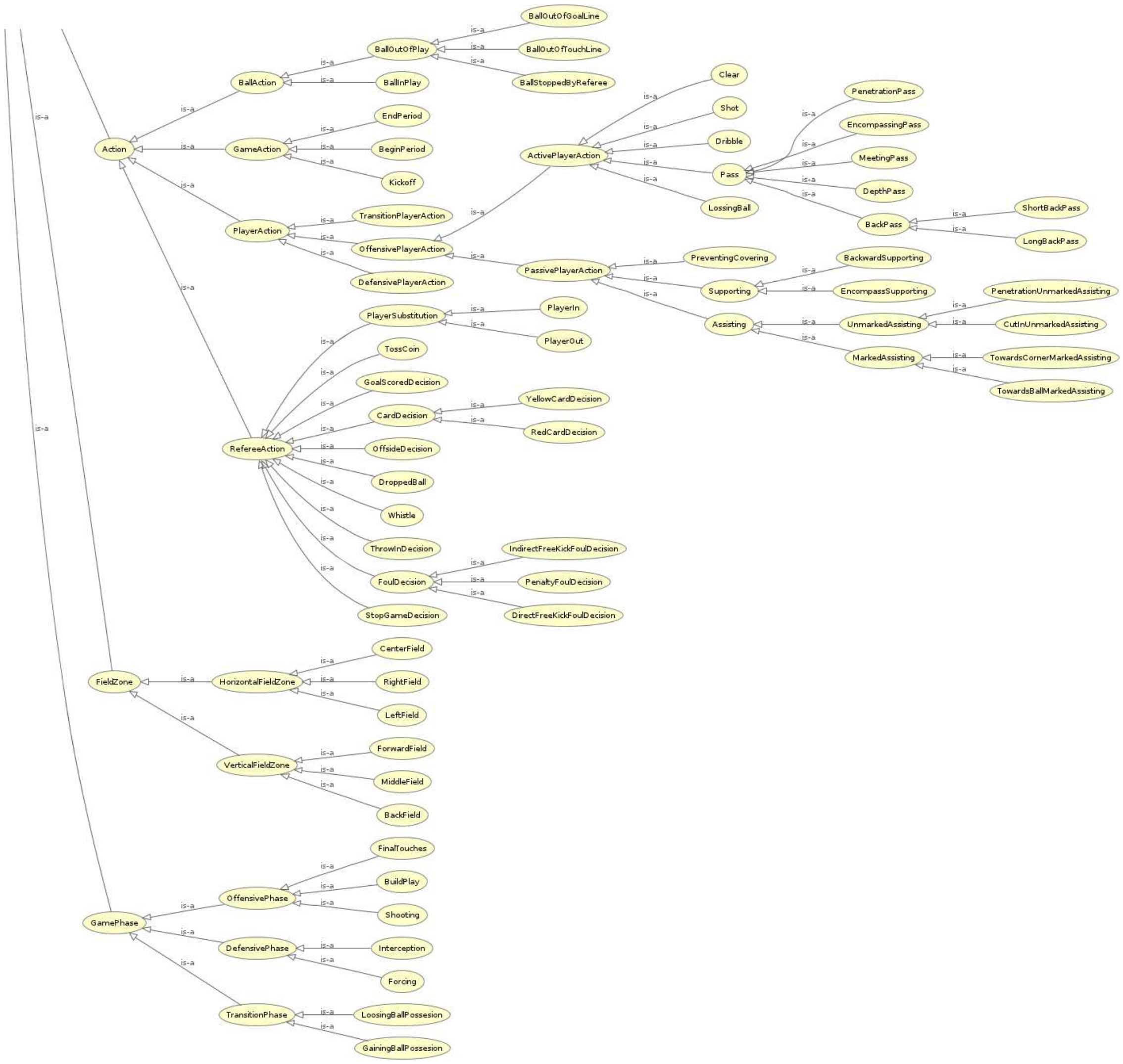


FIGURE 7.1 *(cont. 1.) The asserted Football Ontology describing some of the objects and relations is depicted.*

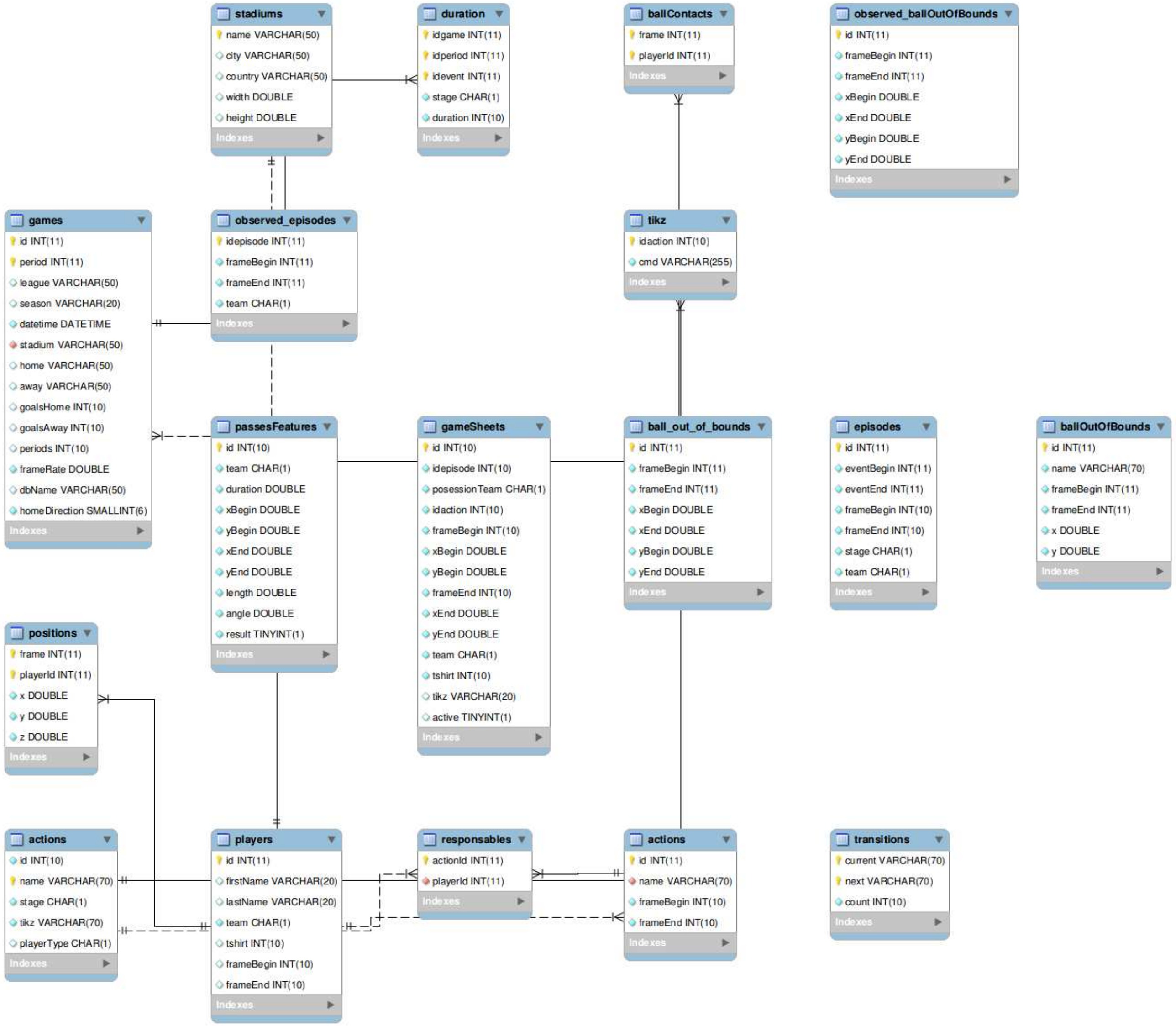


FIGURE 7.2 *EER Diagram Sketch*

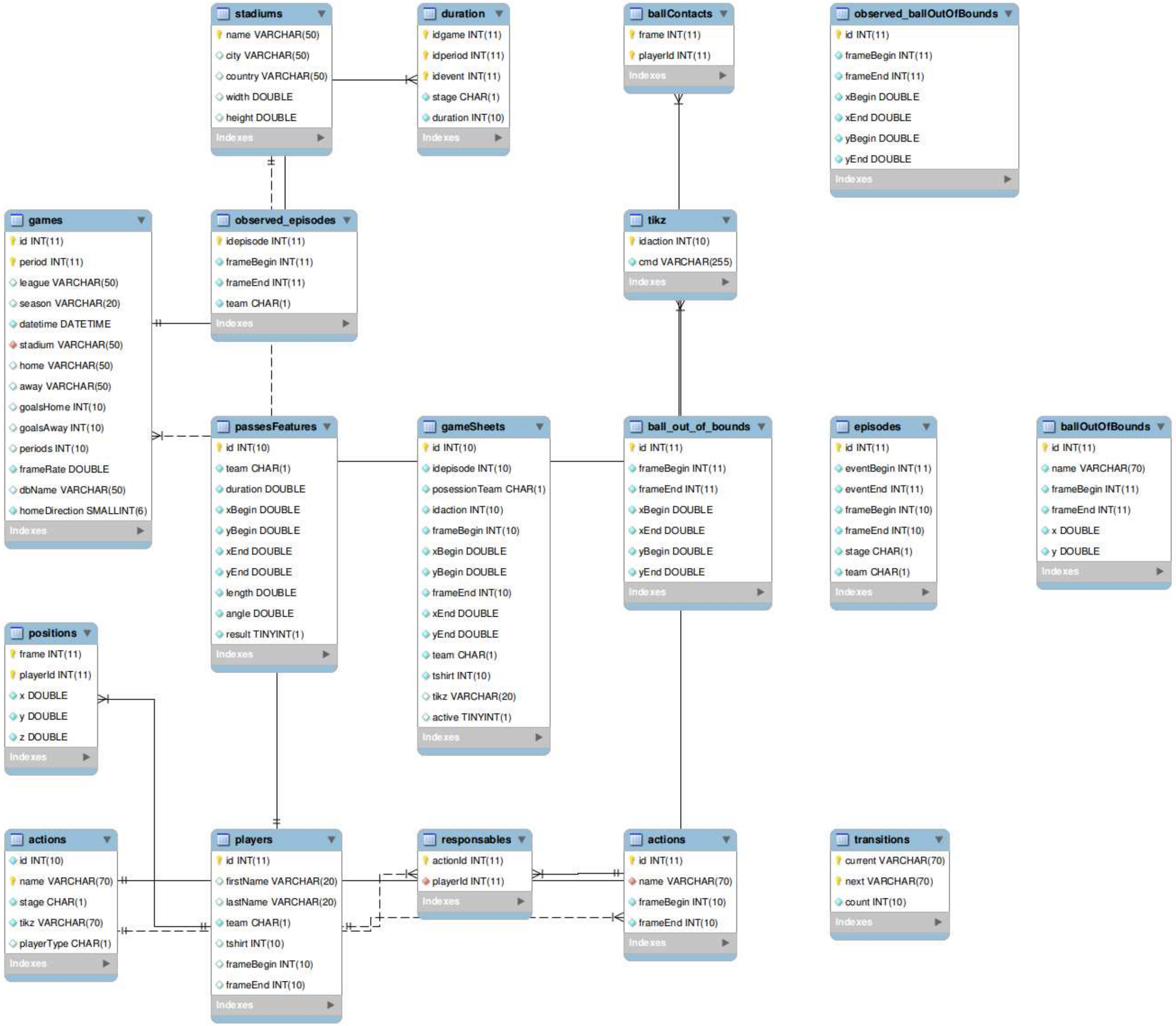


FIGURE 7.3 *EER Diagram expanded*

| **name** | **param_list** | **returns** |
|---|---|---|
| euclidDistance2D | x DOUBLE, y DOUBLE | double |
| euclideanDistance2D | x1 DOUBLE, y1 DOUBLE, x2 DOUBLE, y2 DOUBLE | double |
| frame2time | frame INT, frameRate DOUBLE | varchar(20) CHARSET utf8 |
| getFrameRate | gameDB CHAR(64) | double |
| isFrontBallLine | homeDirection INT, playerTeam CHAR(1), xPlayer DOUBLE, xBall DOUBLE | tinyint(1) |
| score | gameDB CHAR(64) | char(5) CHARSET utf8 |
| squaredEuclideanDistance2D | x1 DOUBLE, y1 DOUBLE, x2 DOUBLE, y2 DOUBLE | double |
| team2string | team CHAR(1) | varchar(20) CHARSET utf8 |
| time2frame | minutes INT, seconds INT, tenths INT,frameRate DOUBLE | int(11) |

TABLE 7.1 *DB Functions for implementation of the* ASPOGAMO *Abstract Model.*

| name | param_list |
|---|---|
| actionCount | IN gameDB CHAR(64), IN actionName CHAR(64) |
| actionsCount | IN gameNumber INT |
| actionsStages | IN gameDB CHAR(64) |
| actionsStagesFSM | IN gameDB CHAR(64) |
| actionsTransitions | IN gameDB CHAR(64) |
| actionTransitions | IN gameDB CHAR(64), IN actionName CHAR(64) |
| actionTransitionsRegexp | IN gameDB CHAR(64), IN actionName CHAR(64) |
| attackDirection | OUT ad DOUBLE, IN gameDB CHAR(64), IN team CHAR(1) |
| attackDirectionHome | IN gameDB CHAR(64) |
| ballLine | OUT x DOUBLE, IN gameDB CHAR(64), IN frame INT |
| ballOutOfBounds | IN gameDB CHAR(64) |
| ballOutOfBoundsNumber | IN gameDB CHAR(64), OUT N INT |
| compareDribblesErrors | IN gameDB CHAR(64) |
| compareEpisodes | IN gameDB CHAR(64) |
| compareEpisodesErrors | IN gameDB CHAR(64) |
| compareInterceptionsErrors | IN gameDB CHAR(64) |
| comparePassesErrors | IN gameDB CHAR(64) |
| compareShotsErrors | IN gameDB CHAR(64) |
| createTableBallContacts | IN gameDB CHAR(64) |
| createTableBallContactsPlayers | IN gameDB CHAR(64) |
| createTableGameSheets | IN gameDB CHAR(64) |
| createTableGTDribbles | IN gameDB CHAR(64) |
| createTableGTInterceptions | IN gameDB CHAR(64) |
| createTableGTNoShots | IN gameDB CHAR(64) |
| createTableGTPasses | IN gameDB CHAR(64) |

TABLE 7.2 *DB Procedures for implementation of the* ASPOGAMO *Abstract Model.*

| **name** | **param_list** |
|---|---|
| createTableGTShots | IN gameDB CHAR(64) |
| createTableObservedDribbles | IN gameDB CHAR(64) |
| createTableObservedEpisodes | IN gameDB CHAR(64) |
| createTableObservedInterceptions | IN gameDB CHAR(64) |
| createTableObservedPasses | IN gameDB CHAR(64) |
| createTableObservedShots | IN gameDB CHAR(64) |
| createTablePassesFeatures | IN gameDB CHAR(64) |
| effectivePlay | OUT rate DOUBLE, IN gameDB CHAR(64) |
| effectivePlayFrames | OUT frames INT, IN gameDB CHAR(64), IN team CHAR(1) |
| effectivePlayTime | IN gameDB CHAR(64), IN team CHAR(1) |
| effectiveTeamPlay | OUT rate DOUBLE, IN gameDB CHAR(64), IN team CHAR(1) |
| effectiveTeamPlayFrames | OUT frames INT, IN gameDB CHAR(64), IN team CHAR(1) |
| episodeActions | IN gameDB CHAR(64), IN ep INT |
| episodeActionsType | IN gameDB CHAR(64), IN ep INT, IN action VARCHAR(64) |
| episodePlayers | IN gameDB CHAR(64), IN ep INT |
| episodePlayersNumber | IN gameDB CHAR(64), IN ep INT |
| frame2time | IN frame INT |
| gameSheet | IN gameDB CHAR(64), IN ep INT, IN text INT |
| gameSubSheet | IN gameDB CHAR(64), IN action INT |
| getScore | OUT score CHAR(5), IN gameN INT |
| goalEpisodes | IN gameDB CHAR(64) |
| insertEventDurations | IN gameDB CHAR(64) |
| interruptedAttacks | IN gameDB CHAR(64), IN team CHAR(1), IN stage CHAR(1) |

TABLE 7.2 *(cont 1.) DB Procedures for implementation of the* ASPOGAMO *Abstract Model.*

| name | param_list |
|---|---|
| noBallGameInterruptionsNumber | IN gameDB CHAR(64), OUT N INT |
| noShotEndingEpisodes | IN gameDB CHAR(64), IN team CHAR(1) |
| passesChart | IN gameDB CHAR(64), IN team CHAR(1), IN result INT(1) |
| passesChartCompare | IN gameN INT, IN team CHAR(1), IN result TINYINT, IN forceside TINYINT(1) |
| pitchSide | OUT side INT, IN gameDB CHAR(64), IN team CHAR(1) |
| playersBallLine | IN gameDB CHAR(64), IN frame INT, IN side CHAR(1), IN team CHAR(1) |
| playersTeamBallLine | IN gameDB CHAR(64), IN frame INT, IN side CHAR(1), IN phase CHAR(1), IN team CHAR(1) |
| possesionTeam | OUT team CHAR(1), IN gameDB CHAR(64), IN frame INT |
| shotEndingEpisodes | IN gameDB CHAR(64), IN team CHAR(1) |
| stageDistribution | IN gameDB CHAR(64) |
| timeDistribution | IN gameDB CHAR(64) |

TABLE 7.2 *(cont 2.) DB Procedures for implementation of the* ASPOGAMO *Abstract Model.*

## 7.3 Passes Decision Trees

Some examples of exploratory decision trees for the classification of passes using extracted features are shown in the figure 7.4, figure 7.5, and figure 7.6.

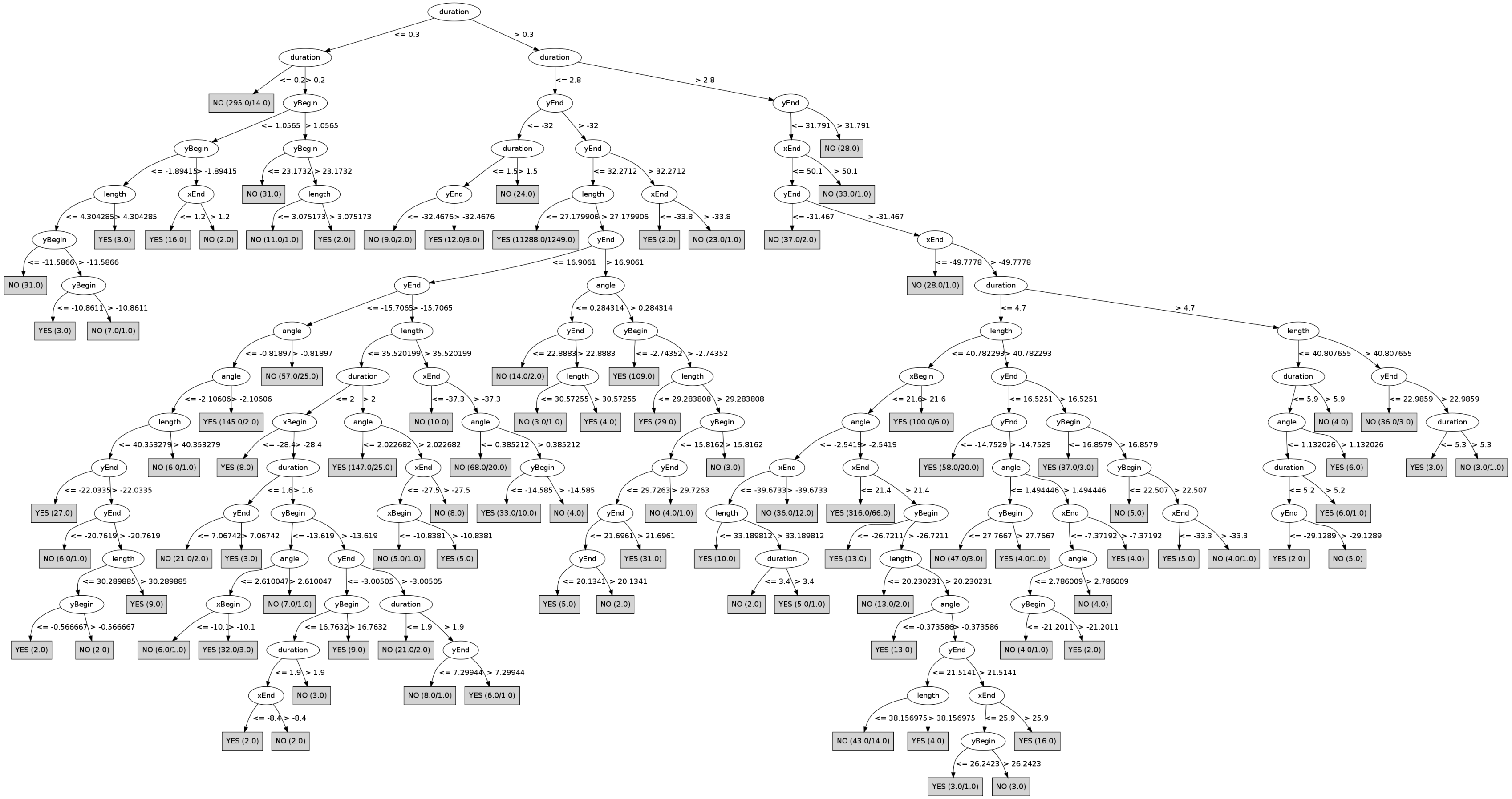


FIGURE 7.4 *Passes features decision tree classification for all the passes of the Game # 1 of hte* Bayern Games *(both periods).*

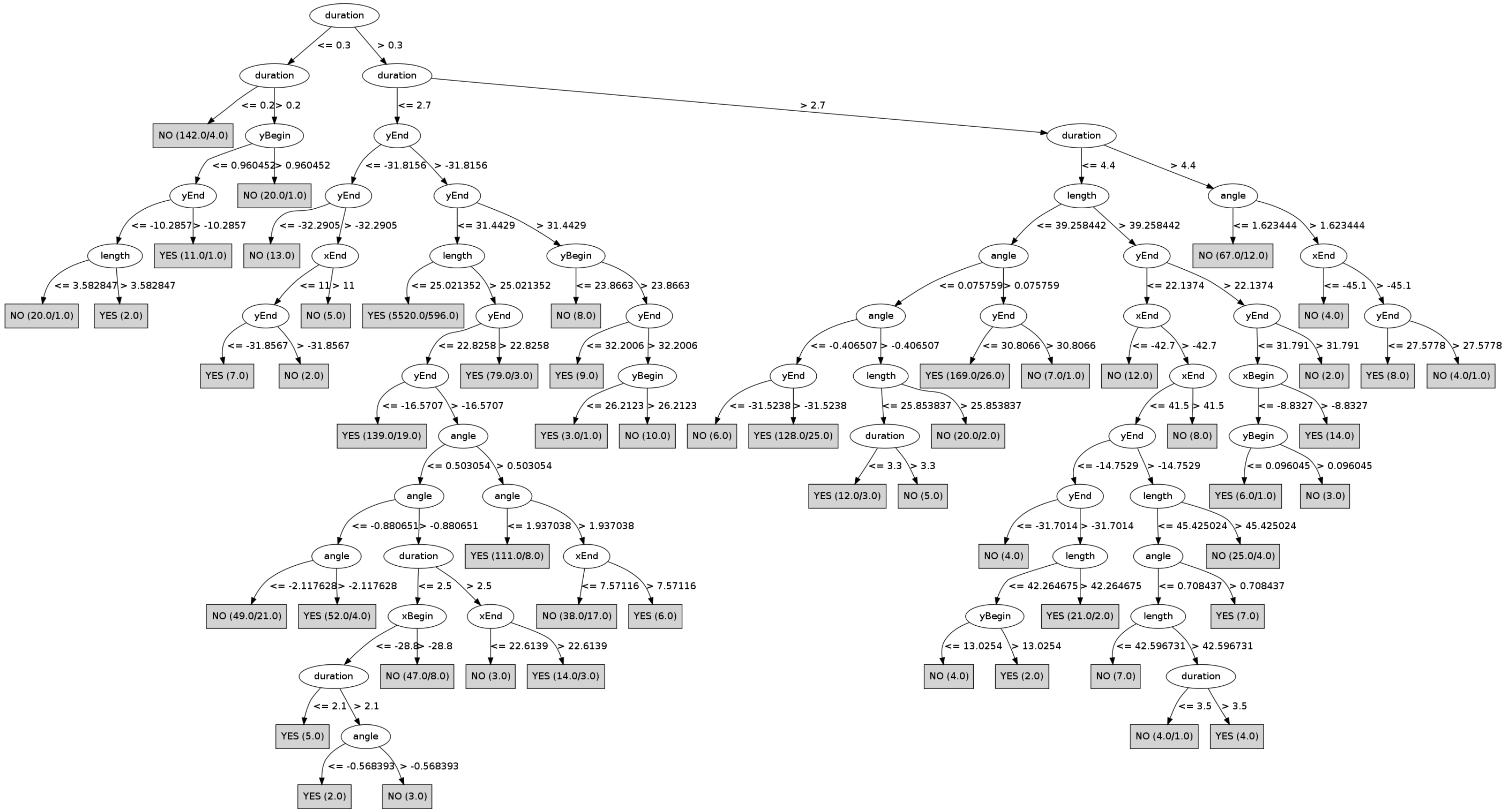


**FIGURE 7.5** *Passes features decision tree classification for all the passes of the Game # 1 of hte* Bayern Games *(first half-time period).*

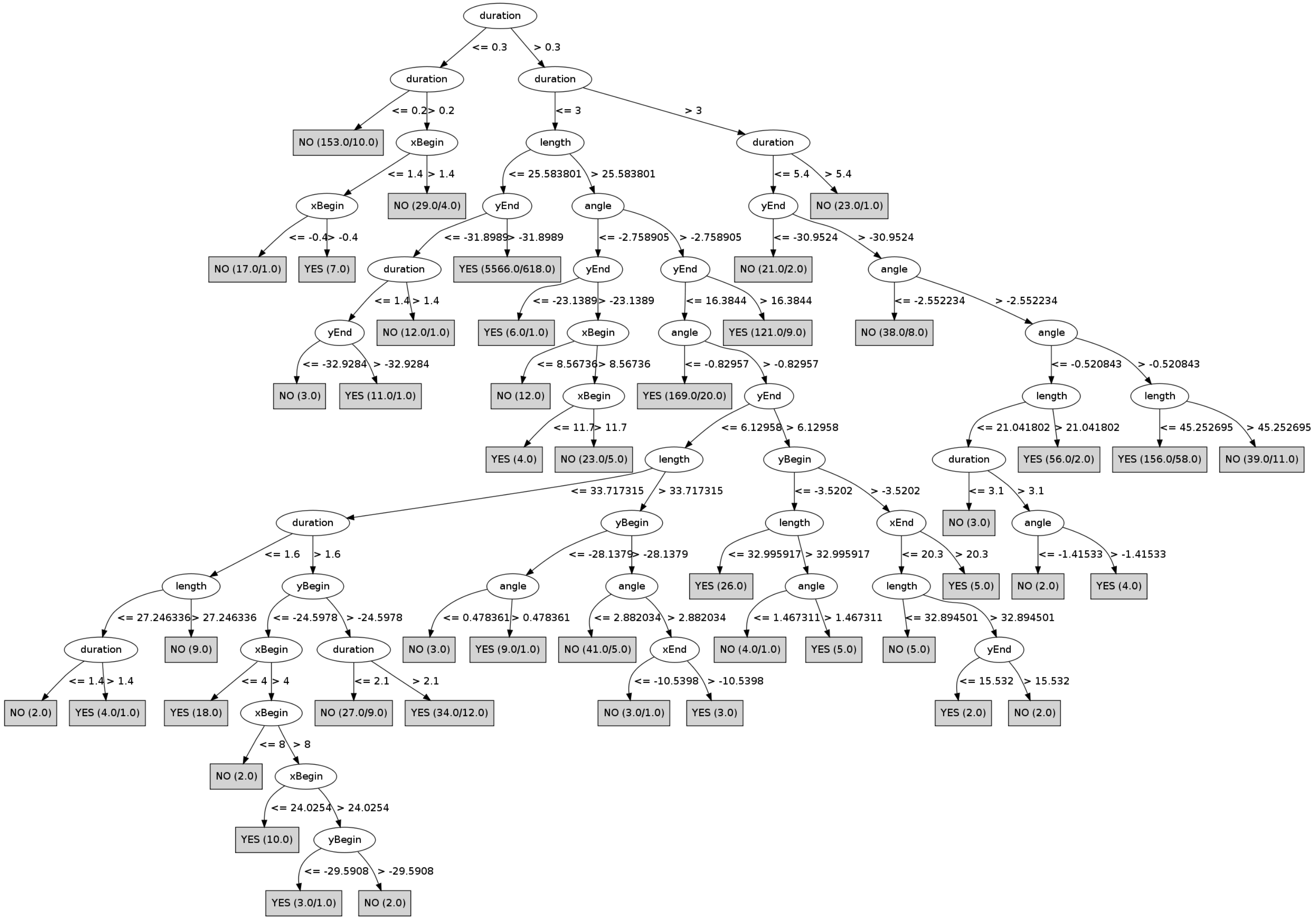


FIGURE 7.6 *Passes features decision tree classification for all the passes of the Game # 1 of hte* Bayern Games *(second half-time period).*