Institut für Informatik
Technische Universität München

Ball Tracking and Action Recognition of Soccer Players in TV Broadcast Videos

Dissertation

Murat Duruş

Lehrstuhl für Bildverstehen und wissensbasierte Systeme
Institut für Informatik
Technische Universität München

Ball Tracking and Action Recognition of Soccer Players in TV Broadcast Videos

Murat Duruş

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. Hans Michael Gerndt
Prüfer der Dissertation:	1.	Univ.-Prof. Michael Beetz, Ph.D., Universität Bremen
	2.	Univ.-Prof. Dr. Daniel Cremers
	3.	Univ.-Prof. Dr. Martin Lames

Die Dissertation wurde am 27.06.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 05.03.2014 angenommen.

To my Mum, Sunay...

Abstract

Video processing has found many applications in team sports analysis, statistics collection and video archiving. Rapid developments of imaging technology and increasing video processing power have been intensively improving the development of sports video analysis tools. Those tools include the extraction of player technique, tactics and furthermore the decision making about the game. Therefore sports videos provide an excellent means of observing and analyzing the games.

In ball games more complex rules are applied to determine the winner. Therefore annotating video indexing becomes a more difficult task which requires the position of the ball as well as the position of the players. During a typical game the attention of the audience as well as the players is always focused on the ball. The position of the ball determines the movement of the players which has direct effect on the game events and their outcome. On the other hand, the possession of the ball provides the information on which team has better control of the game that also conclude in the decision of a team applying their tactics in a better way.

In this thesis we investigate the techniques for (1) detection and (2) tracking of the ball and (3) the detection of the players in digital video frames of soccer games. Novel resampling approaches are proposed in tracking with the smart filtering techniques for small and single target tracking in team sports. Besides, the (4) extraction of the ball actions are studied with the new measures defined. Therefore, the tracking problem studied is challenging since we investigate the way the ball is played beyond the physical properties of the ball. (5) The developed algorithms are also integrated to an existing ASPOGAMO Web Based Soccer Analysis System.

Kurzfassung

Bildverarbeitung hat viele Anwendungen in der Analyse von Teamsportarten, der Erstellung von Statistiken und der Videoaufbewahrung. Rapide Entwicklungen in der Aufnahmetechnik und steigende Leistungen bei der Bildverarbeitung haben die Entwicklung von Werkzeugen zur Analyse von Sportvideos deutlich beschleunigt. Diese Werkzeuge umfassen die Auswertung von Spielertechniken, Taktiken und die Entscheidungsfindung in Bezug auf das Spiel. Daher stellen Sportvideos eine hervorragende Basis zur Beobachtung und Beurteilung des Spiels dar.

Bei Ballspielen werden komplexe Regeln angewendet, um den Gewinner zu ermitteln. Daher wird das Annotieren von Videos eine komplexe Aufgabe, die sowohl die Position des Balls als auch der Spieler erfordert. Während eines typischen Spiels ist die Aufmerksamkeit der Zuschauer und der Spieler stets auf den Ball gerichtet. Die Position des Balls beeinflusst die Bewegung der Spieler, was direkten Einfluss auf den Ausgang des Spiels hat. Auf der anderen Seite bietet der Ballbesitz die Information darüber, welches Team die bessere Kontrolle über das Spiel hat, was auch bedeutet, das dieses Team seine Taktik besser anwendet.

In dieser Arbeit untersuchen wir Techniken zur (1) Detektion und (2) Verfolgung des Balles, sowie (3) zur Detektion der Spieler in digitalen Videobildern von Fußballspielen. Neuartige resampling Ansätze zur Objektverfolgung werden vorgeschlagen mit intelligenten Filtertechnik zur Verfolgung von kleinen und einzelnen Zielen in Teamsportarten. Weiterhin wird (4) die Erkennung von Ballaktionen untersucht mit Hilfe des neudefinierten Messverfahrens. Daher ist die Aufgabenstellung der Objektverfolgung herausfordernd, da wir über die physischen Eigenschaften des Balles hinaus untersuchen wie der Ball gespielt wird. (5) Das vorgestellte System ist auch in die bestehende Webapplikation ASPOGAMO integriert und die Herausforderungen werden demonstriert.

Acknowledgements

I am indebted to many people for making the time working on my Ph.D. an unforgettable experience. First and foremost I want to thank my advisor Prof. Michael Beetz, PhD. It has been an honor to be his Ph.D. student. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit.

I would also like to thank Prof. Dr. Bernd Radig for his ideas and support during my Ph.D. work and steady influence throughout my Ph.D. career. This dissertation would not have been possible without the guidance and the help of him. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been invaluable.

Furthermore, I would also like to thank Prof. Dr. Daniel Cremers for his feedbacks and for many motivating discussions.

I am very grateful to my reviewer Prof. Dr. Martin Lames for insightful comments during my work in this thesis.

I would like to thank the chairman of my committe Prof. Dr. Hans Michael Gerndt for the support and improving efforts during the examination of my thesis.

In addition, I feel very privileged to get to know and collaborate with the members of the Intelligent Autonomous System Group Dr. Zahid Riaz, Dr. Jan Bandouch, Dr. Andreas Fedrizzi and Dr. Lars Kunze. I would like to express my special appreciation to ASPOGAMO project group Dr. Suat Gedikli, Dr. Nicolai v. Hoyningen-Huene, Dr. Francisco Siles, Dr. Christoph Mayer, Hang Zhou, Bernhard Kirchlechner and Andreas Bigontina for the valuable discussions and reviews during this work. I have learned a lot from our computer vision discussions with Dr. Suat Gedikli and probabilistic tracking discussions with Dr. Nicolai v. Hoyningen-Huene. Our collaboration about the web applications with Dr. Franciso Siles and Dr. Christoph Mayer provided me new insights for the web technology.

I would like to thank to our department secretary Sabine Wagner and Manuela Fischer-Russenberger for the professional support they has provided during the administrative processes.

I would like to thank to Prof. Dr. Aytul Ercil for the support she has provided through my academic career from the beginning.

I would like to present my deep thanks to my mother, Sunay and my sister, Dilek. I would like to thank all my family for the invaluable support they provided through the thesis.

Munich, June 2013

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

"A picture shows me at a glance
what it takes dozens of pages of a book to expound."

(Ivan Turgenev: in Fathers and Sons in 1862*)*

Video processing has found many applications in sports, such as slow motion replay, pattern analysis, statistics collection and video archiving. The development of high-speed digital cameras and video processing has increased people's attention in sports video analysis [Babaguchi et al., 2002, Ballan et al., 2008, Duan et al., 2003]. In addition, the requirements in broadcasting and analysis has created a need for developing automatic systems and tools for content-based video information retrieval. Sports videos have been used for the analysis of player technique, tactics and decision making. Sport videos belong to a special domain where player movements and actions play a crucial role, since they pose direct effects in the result of the game. Sport videos provide an excellent means of recording, observing, analysing, evaluating and checking performance. Videos can be recorded and watched as many times as desired to evaluate the game and the performances. The video information provides all types of details which might be missed by the technical team or the audience. It can also be used for dedicated reasons such as offside cameras etc. The playbacks also provide the coaches and the audiences see the events again in order to understand the actions happened. It provides the team with all the details of he visual information needed for the feedback of the game.

Advances in multimedia technology have allowed the recording of videos in different perspectives and with various details. The decrease in the costs of the hardware devices like cameras and storage devices leads to a higher demand in automatic video indexing and summarization. The summarization and indexing of videos usually refer to the extraction of the content to store and analyze the video in a more efficient and intelligent way. As the activities within the video form the major content of the videos, they form the most interesting part for the audiences and analysts. Therefore interpretation of the content and recognition of activities within a video becomes the most inevitable feature for efficient video indexing and analysis. A detailed survey of the sports video analaysis is presented in chapter 2.

The automatic observation of sports games gain increased attention in the video analysis community. The content of sports videos can be summarized using positions and trajectories of the players as well as the tools or components used within the game i.e. ball, puck. Sports video analysis consists of extraction of the spatio-temporal data with the corresponding athlete as well as a higher level abstraction of the events. In track-and-field athletics the competitor who passes the winning post first, jumps the highest or longest, or throws the ball becomes victorious. All the competitors can be judged by their finishing position, or by the time taken on by the distance covered during performance. Such metrics also apply to other branches like swimming, rowing, cycling, skiing and other sports. The automatic game analysis systems provide quicker and more detailed and accurate results than any other resources. The statistics and game events can be stored on a database for interpretation or the creation of game and player statistics.

In ball games more complex rules are applied to determine the winner, for instance by points scored, sets won or goals scored. Therefore annotating video indexing becomes a more difficult task which requires the position of the ball as well as the position of the players. The attention of the audience as well as the players is always focused on the ball. The position of the ball usually influences the movement of the players which also effect game events and their outcome. The possession of the ball shows which team has better control of the game from which one can conclude which team applies their tactics in a better way.

In this thesis we investigate the techniques for tracking the ball and the detection of the players in digital videos of soccer games. The ball is the key element of the game and many actions take place around it. Tracking of the ball can be conducted by localization of the ball over time. We face challenges here due to two aspects: the physical properties of the ball and the way it is played during the game. To tackle this challenges, we develop efficient and robust algorithms to detect and track the ball during the game. In order to extract the actions, the localization of the players is also implemented through the thesis. As the tracking of the players is beyond the scope of the thesis, only the detection of the players is investigated.

1.1 Sports Video Analysis

The advances in digital technology render the recording and analysis of most of the events in our life possible as well as the sport games. Audiences watching live sports events on TV can experience a game at several levels, making their own choice of viewing direction, hearing the roar of the crowd by the help of modern broadcasting technologies. Hence, the constant pressure on producers to augment their coverage with new and unusual viewing angles, effectively transporting television viewers much closer to the action than would ever be possible. The sports video analysis makes the game more interesting for the audience with more interesting and better represented data. It has direct impact on the result of the game by providing the decision support for the trainers and helps them to be able to explain the situation to the players so that they absorb and understand the required information easily. Automatic observation and analysis of sports gain increasing attention due to its application in entertainment, highlight extraction, content insertion, tactical analysis, tracking and decision support.

An overview of the system is depicted in figure 1.1. The first part of the system is the acquisition of data through broadcast cameras. In the second phase, the main results of the computational problem is depicted. The algorithms run on the video in order to detect and track the ball. The uppermost figure shows the tracked ball. In the same phase, additionally the positions of the players are extracted for the further action representations. Player and ball positions are the fused in order to define events and extract ball actions. In the middle and the lower part of the second phase, pass and possession actions are depicted, respectively. The rightmost last phase of the image shows the general output visualization of the system. This image shows the fusion of both information on the soccer field, i.e. the position of the players and the ball trajectory.

1.1.1 Audience Entertainment

As the market gets bigger the broadcasting companies try to attract new audiences using advanced techniques of representing the video content. Those include instant replays of important situations, better views or visualizations of the events. Animation of the sports games is also gaining more attention since they are used to attract the audience to the game and sometimes entertain the audience better than the real game. In some cases the audience can also make changes in the video during watching those animations and they see the artificial game in the form of a computer game.

For instance, recently the BBC has unveiled an animated advert for the London 2012 Olympic Games which shows the UK as a giant stadium with Olympic athletes preparing and competing in a range of landscapes [BBC News, 2012]. In this advert, they showed the scenes including swimmers battling it out in lanes created by buoys on a net cast by a fisherman, a BMX rider preparing at the edge of a cliff, track cyclists racing around quarries and sprinters and gymnasts going through their paces on streets. Those type of adverts definitely attracts people's attention for sport games, which then increases the broadcasters ratings and

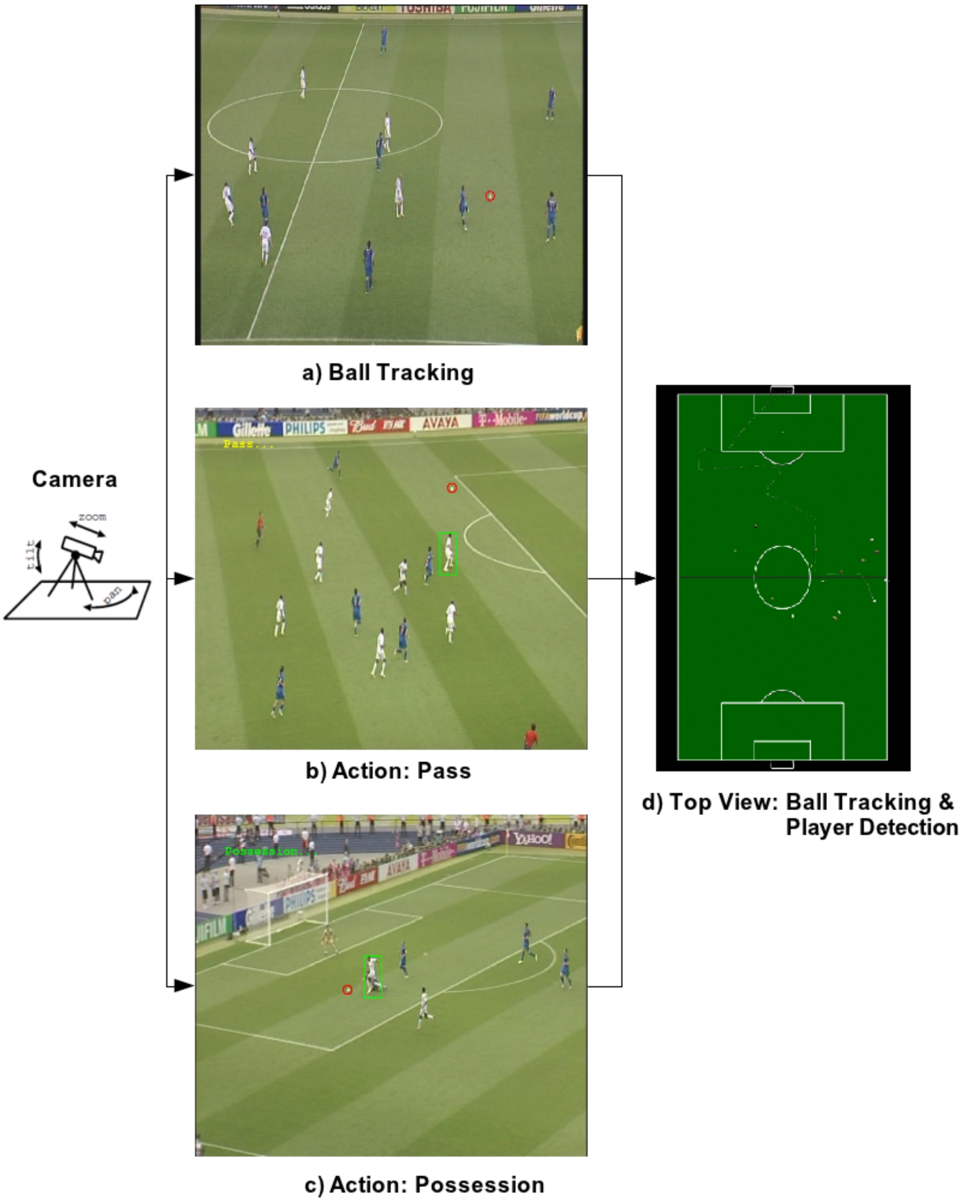


FIGURE 1.1. Overview of the System

share in the market.

Nowadays mobile applications in sports video analysis become an emerging market with the development in mobile devices [Farin, 2005]. The applications vary from online analysis tools to any type of applications which generate highlights for the audience so they they can quickly and instantly see what is happening during the game. Other than the classical replay selection and generation, mobile devices nowadays allow the users to see a virtual view or in program augmentation of the game or see the analysis or the reports of the game from a reporter supporting their team. Other type of applications provide the audience with instant game scores and information [Beetz et al., 2005, Wan et al., 2003]. Those type of applications find place mostly for the audiences who bet on the game results.

1.1.2 Highlights Extraction

Sports highlight extraction is one of the most important applications in sports video analysis. Most of the research has been conducted on the detection of highlights, using audio or visual features of the video material [Wan et al., 2005, Xiong et al., 2003].

Classification of events and extraction of highlights during the game or after the game makes it easier to interpret and analyze the game. The largest part of the analysis is being done manually nowadays, which makes the automatic extraction of the crucial or the most time consuming data for the game analysts a very important application. Thus it is a profitable idea to visualize information and automatically generate data using the machine vision algorithms. This saves a lot of time for game analysts and makes the process cheaper while they are able to focus on processing other data during the game. The results generated can be used in applications varying from online video advertising and video summarization to interactive TV and other multimedia applications.

1.1.3 Content Insertion

Advances in multimedia technologies made it feasible for real-time content insertion in sports videos [Thomas, 2006]. Content insertion is mostly used to increase the audience's attention and for advertising [Wan et al., 2005]. In video analysis, the applications take place in insertion of content while visualizing critical situations or highlighting the events. In other sports than soccer, for instance, blending of a virtual line showing the position of the athlete who won the last world record during the sports competitions makes the viewing experience more interesting for the audience. Another application is the insertion of the offside line in soccer games, which is mostly done manually during the replays in sports videos. In Sydney 2000 Olympic

Games, Orad Hi-Tec Systems Ltd., introduced the first use of virtual imaging in Olympic competition. Its Virtual World Record Line debuted during Australian Nine Network's television and Web coverage of the Olympic qualifying swimming trials. The line, connected directly to the electronic timing in the event pool, featured a superimposed line on the water's surface and graphics depicting existing world records.

Sports videos are watched by billions of people everyday from different parts of the world. Due to such amount of viewers, advertising directly becomes very important since it is the key income of broadcasting companies. For instance, it is reported by FIFA that the 2010 FIFA World Cup in South Africa was shown in every single country and territory on Earth, generating record-breaking viewing figures in many TV markets around the world. The in-home television coverage of the competition reached over 3.2 billion people around the world, or 46.4 percent of the global population, based on viewers watching a minimum of over one minute of coverage. In this big market, advertising becomes one of the most important aspects of the game. The brands shown on the screen might not be always global brands known all over the world but local names known in a region or country. Active advertisement or content insertion allows the brands shown during the game to be customized according to the region or country [Tanwer and Reel, 2010]. For instance a brand known very well in the US might not be that known or marketed in Europe, which makes advertising that brand during the broadcast in Europe useless. By analyzing the sport games, the application developers and the broadcast companies detect where and when to place their advertisements. For instance, an appearance of a brand just after a scoring event would take more attention than the normal advertisement but on the other hand placing advertisements on the screen during important situations might annoy the audience.

1.1.4 Tracking

Automated tracking of moving objects is an interesting research topic for scientists and professionals with its applications including video surveillance, scene analysis, abstraction. Sport games consist of actions of athletes using the tools of the game in order to compete with the opponent athletes or teams. Therefore the localization of the players and the game components, like ball, puck etc., plays a crucial role in sport game analysis. Extensive research has been conducted on tracking the athletes using active and passive sensors. As active sensors are not allowed by most of the sport federations, video becomes the most popular and effective tool for analysis. Player and ball trajectories are extracted from video analysis mostly by manual processes by sports analysis companies. A survey on sport analysis which focuses on the tracking of the players is studied in [Gengembre and Pérez, 2008, Li et al., 2005]. The tracking of the athletes within the game is one of the most important applications, which provides the localization and the actions as well as the performance analysis data.

1.1.5 Decision Support and Computer Aided Referee

In most sports, the athletes perform the game in the absence of referees in the start of their career during the amateur plays. At the beginning of their career they usually come to the decisions by agreement with the other athletes or team members. At higher levels they play under the observation and decisions of the referee. The referee and his assistants, usually chosen by the sport federation, organization or by the teams involved in the tournament. Therefore each athlete sees himself sharing the role of the referee up to some extent, which might lead to arguments during the game. The justice by the referee cannot be applied with 100% of objectivity since sometimes the referee and the assistants might also miss the position. The lack of the view, communication problems or the instant motivation problems can directly affect the result of the game. For instance, in the case of ghost goals where it occurs when the ball passes the goal line but jumps straight back out. According to the rules of the game it should be a goal, but it might be very likely misinterpreted by the referee. This wrong decisions may not only decide the outcome of a game but the fate of a team in a given tournament. A wrong decision may represent hundreds of millions of dollars won or lost for a team. It is now known by the whole world that the situation was a goal but not counted due to those problems, when Frank Lampard scored for England against Germany in South Africa ’10, or Spain’s Michel kicked the ball inside Brazil’s goal in Mexico ’86. Therefore nowadays in many sports replay judging and computer aided referee is commonly used or systems are being developed in order to increase the performance of the referees [Leo et al., 2008c, Wong and Dooley, 2010]. The real-time video analysis provide the referee with objective information during the game. Most of this methods are based on manual or automatic player and ball tracking algorithms.

1.1.6 Tactical Analysis

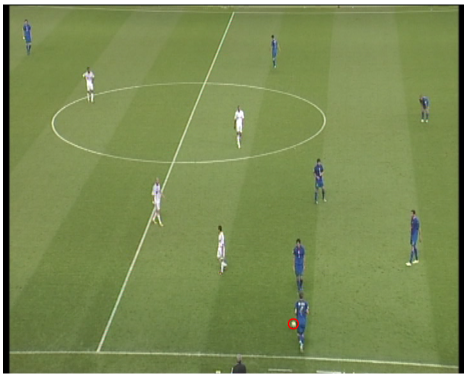
Video information provides the analysts and trainers with the option to redesign the game in order to see the positions and results. Trainers are the most interested audiences on the automated analysis of sports games. The real-time analysis of the game lets them to quickly see the mistakes made by their own team as well as understanding the tactics in an objective way. By using the generated analysis result they can also show the athletes their mistakes on the generated video. As a famous Japanese philosopher said: "A picture is worth a thousand words." Since the athletes are under to much stress during the game, it is often difficult to explain their mistakes to them by show on the paper. Most of the trainers still use pen and paper to explain their athletes the mistakes and tactics which is visually a really poor way regarding the online video data. For instance, the trainer of the Japan Women's National Team Masayoshi Manabe called the London Olympics 2012 Games as "the toughest environment I've had to coach in". Because the Earls Court venue didn't provide the data-crunching coach with the WiFi network he needed to analyze real time data of his opponents. Manabe called the omission of an internet connection on court a "huge surprise" that shows the inevitable need of the real time analysis for coaches [Zhu et al., 2007, 2009]. He stated that the delay on the information was one of the key factor causing them lose the game against South Korea.

1.2 Problem Description

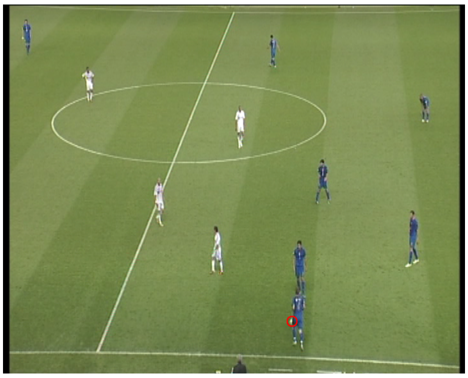
The characteristics of most team sports include the interaction of players with each other and the ball. Sports games are transferred to the digital domain using different type of sensors. The most common data acquisition tools used for recording sports games are the digital cameras. Even though multimedia technology is growing very rapidly, it might still not be possible to acquire high quality images due to costs or communication issues. As it is mostly not possible to choose the right hardware for the analysis, problems like quantization, motion blur, overlays and reflections have to be faced. Since data is transferred to the main analysis unit or the broadcasting modules through various protocols and environments, this also causes the loss of data. Upon that ball is the smallest element of the game and such data acquisition issues directly affect the performance of ball tracking. Some of the ball shapes are shown in figure 1.2.

Additionally, the detection of players becomes difficult due to the loss of color and shifts and losses within the frame. The damages and interlacing methods of the broadcast frames might lead to the difficulties in detecting the players throughout the game.

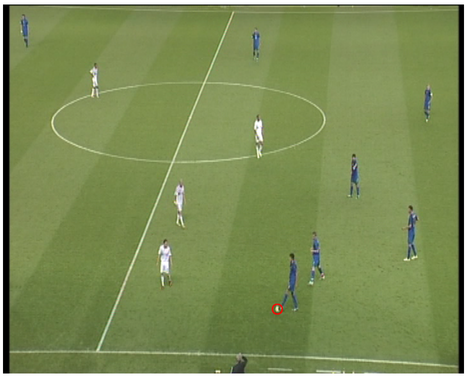
Ball is the main element of most sports games which is mostly a round, usually spherical but sometimes ovoid object. In ball games, the progress of the game follows the state of the ball as it is hit, kicked or thrown by players. Since ball is the main tool used to beat the other team, all interesting actions take place around it. Therefore images showing the ball paths have become indispensable for analyzing scenes, since they provide information about the tactical situation on the field. In order to get more comprehensive assessment of the games, the interaction between ball actions, game situations and the effect of ball actions are required. Those more comprehensive methods give the opportunity to the trainers and sport scientists to create new opportunities in sport to get insights into the process instead of comparing the final



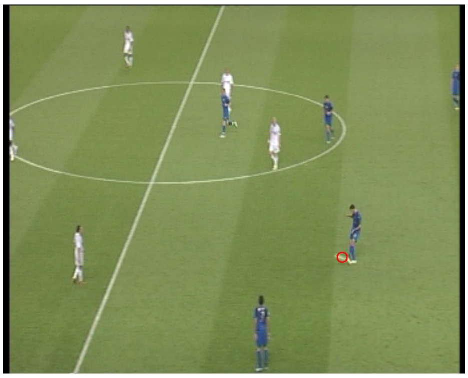
(A)Slight occlusion



(B)Heavy occlusion



(C)Confusion with player



(D)Player shot

(E)Fast camera motion blur

(F)Shape deformation due to fast speed

(G)Confusion with field lines

FIGURE 1.2. Typical ball appearances during the game.

results only after the end of the game.

The real time tracking of the ball and detection of ball actions also provide the opportunity for online analysis of the game. Detection and tracking of the ball is a necessary step for a comprehensive analysis of the game and it is one of the challenging problems due to the size and the motion dynamics of the ball. Even though the temporal velocities during the pass actions seem to be mostly constant, rapid movements of the ball during the actions as well as camera movements generate difficulties in detecting and tracking the ball. The mentioned difficulties occur due to camera motion, size, occlusion and the lack of camera field of view when the ball is kicked high in the air. Ball detection is usually very difficult to achieve since the cameras are moving and the settings are being changed. Additionally, the ball is always very small in size with respect to other objects within the frame and it is mostly occluded by players during the game, since at most of the game time the ball is at the foot of the players. Another issue is also that when the ball is kicked high in the air, it is out of the field of view of the camera. The ball is usually moving with high speed and mostly confused with the parts of players and line markings. This rapid and unpredictable acceleration also makes the detection more difficult.

The existing web based sports analysis systems are very limited in number and in terms of their functionalities. A web based system which is accessible from all over the world and which is always available is an inevitable need. Especially for the ball tracking on the web, the system is challenging since many systems do not provide automated ball tracking algorithms in their systems.

1.3 Contributions

The main philosophy of all ball games is to possess the ball by applying the own skills in order to achieve the tactics to score a goal. Throughout a typical ball game, the ball is always at the focus of the game and most of the actions take place around it. Therefore, providing tracks of the ball and the players is a necessary step for comprehensive and deeper analysis of soccer games.

Understanding the movement of the ball is essential for analysis of the match. Most of the methods so far reveal semantics in broadcast material, that have been already encoded by humans, or provide only low-level statistical information based on trajectories and containing far less semantics. Few enrich very small parts of the video with semantics that are interesting for sport scientists and trainers beside limited classifications.

In this work we propose novel enhancements in the detection [Huang et al., 2007, Pei et al., 2009] and tracking of the ball [Kim and Kim, 2009, Liu et al., 2010, Ren et al., 2009] and we further analyze the ball semantics.

- The detection algorithm proposed in this thesis uses a more efficient method during the search for the ball within the game which has a direct effect on the performance increase of the tracking system. The first contribution of this thesis includes novel algorithm development for detection of the ball during the soccer game as an application. The algorithms developed so far in the literature implemented rule based [Pei et al., 2009] methods or employed Hough transform to detect the ball candidates [Huang et al., 2007]. However, rule based methods are not reliable in a complex environments like soccer match and Hough transform is not applicable to very small size ball candidates. [Miura and Shimawaki] integrated a route estimation method in order to detect the ball candidates on a very limited and small amount of data. The method is not practical since only tries to detect the ball independent of its environment which leds to many false positives as seen in their work. In our method, we implemented a ball detection method based on not only general shape and color characteristics but also the awareness from the environment. The environment awareness refers to the determination of more probable regions for the ball detection. The methods developed so far is also evaluated on a limited number of data on selected high quality frame. Our algorithm is evaluated in long broadcast game sequences selected randomly which strengtens the generability of the proposed method.

- We have proposed a ball tracking method thats is beyond the already proposed methods for long ball sequences. The methods studied so far have mostly focused on the tracking of the ball using color [Yu et al., 2003b], shape features [Huang et al., 2007, Yu et al., 2003b] or multiple camera views provided [Ren et al., 2009]. As the shape and color information might not always be available, this causes the performance of those systems decreases radically. In this work, in addition to the state of the art techniques, we implemented an extended particle filter using ball dynamics and ball motion characteristics. The extension to the particle filter for ball tracking includes smart region filtering in image space and resampling methods in state space. As a novel approach resampling of the particle filtering approach is modeled according to the ball dynamics and the ball motion characteristics. In contribution to the current literature we track the ball not only as a single element within the game but also a main element of the game, which is actively and highly probable involved in actions. We tested the system on the final game of FIFA World Cup 2006 and games from FIFA World Cup 2010 and observed better performance than the already proposed methods (see Chapter 4.9.4). The algorithm is also evaluated on hockey videos from the World Series Hockey 2012. Up to the best of our knowledge our method is the first vision based method to be evaluated on field hockey video data.

- The third contribution of this thesis is the automatic recognition of ball actions merged with player positions. Sor far in the literature, the ball player interaction has not been studied extensively. The methods developed so far by [Leo et al., 2008b, Niu et al., 2010, Zhu et al., 2009] on the ball action recognition do not include player localizations. They try to apply methods using the position of the ball on the field, but the ball and player interaction is not investigated. In our method, we defined novel measures based on ball dynamics and developed and Hidden Markov Model based approach for ball action recognition. In this method, we used a novel technique in which we treate the ball like another player with a different importance level and we investigate the results of the interaction of the ball with the players. We do not only extract the actions using only the position of the ball on the field but also its interaction with respect to the players.

- The fourth contribution of the thesis is the novel trajectory analysis methods. As the ball does not always stay on the field during the game but can be also played in flying state, we proposed a new technique in order to detect rolling and flying states using the ball trajectory. The flying and rolling state classification of a soccer ball using 2D broadcast video image is proposed. The free style flying of an object is always affected by the gravity. This causes the occurence of curve shaped trajectories in the image domain during a soccer game. In this work, we developed methods for the detection of the 2D curves in order to classify the ball actions. Most of the trajectory characteristic analysis are in the field of physics [Ghista and Liu, 2008, Goff and Carré, 2012]. In this work, we focus on the motion of the ball in image plane and develop a novel algorithm to detect the rolling and flying ball states from the trajectory information.

- Another contribution of the thesis is the development a web based Ball Observation analysis tool for ball tracking and its integration to the ASPOGAMO Web System [Bigontina, 2011]. We discussed the system design and architecture and explained the system operation in detail.

The development explained in this thesis is a complete closed system for the automated ball tracking and analysis of games. The methods mentioned so far are composed to a Ball Observation System developed under the ASPOGAMO project [Beetz et al., 2009]. In this thesis, we investigate the problem of small size and fast moving object tracking and we test our system on broadcast sports video domain. This thesis discusses all steps required for the implementation of a Ball Observation System for computer aided soccer game analysis. We propose novel and robust algorithms for tracking of the ball and analysis of ball motions. The system includes modules ranging from the video input to higher level ball action extraction. The input to the system can be either from multi or single camera system. In addition, up to our best knowledge there is no comparable system in the literature for ball detection, tracking and action recognition.

1.4 Thesis Outline

In Chapter 1, an introduction to sports video analysis presented and the problem description given. The main application areas of sports video analysis are presented with the up to date developments and applications. Major application areas range from the entertainment for the audience to comprehensive and tactical analysis methods for coaching purposes. Later in the chapter, the novelties discovered and the methods proposed are explained. The chapter concludes with a general outline of the thesis.

Chapter 2 includes a detailed survey of the current state of the art methods in soccer game and video analysis. Developments ranging from commercial tools in soccer industry to up to the date sports video analysis algorithms in the scientific literature are analyzed and explained. The detailed literature review is described for each module of a typical video based soccer video analysis. This includes the recent research and work that have been done on ball detection, ball tracking and ball action recognition.

In Chapter 3, single target tracking methods are investigated in detail. The most common tracking methods explained along with the derivations and applications. The most common method of Nearest Neighbor (NN) algorithm is explained and the non-Bayesian and Bayesian approaches are investigated. Kalman Filter (KF) and Particle Filter (PF) approaches are presented and comparisons of the existing methods are discussed.

In Chapter 4, major novelties of the thesis are presented. In the first part, the structure of the ball games and the challenges in detection and tracking of the ball are explained in order to give an insight to the topic. Then the methods developed in player detection are explained which aims the ball player interaction analysis. After that the features of the ball are given, the ball novel detection and filtering algorithms are investigated. Filtering of the ball hypotheses and advanced the ball tracking methods proposed and explained. A novel that we invented for extraction of flying and rolling ball situations are discussed. The accuracy of the proposed methods are depicted in the tables and figures and the outcomes are discussed. The summary of evaluations and detailed results for broadcast soccer and hockey videos concludes the chapter.

Chapter 5 focuses on the high level information extraction and summarization based on the ball tracks and player positions gathered in chapter 4. For the recognition of the hidden states and Hidden Markov Model based approach is proposed and implemented. Player ball interaction scenarios that are studied and explained in detail, followed by the ball action analysis. We conclude the chapter presenting the evaluations in broadcast soccer videos and hockey videos.

In Chapter 6, a web based Ball Observation System is described. The mentioned system is integrated to the existing ASPOGAMO Web based soccer video analysis system. The technical details of the system as well as the user interface design is presented.

In Chapter 7, we give an outlook to the conclusion of our work and give an interpretation of the study. After the short summary of the thesis, a discussion on the possible future insights are discussed.

CHAPTER 2

Soccer Game and Video Analysis

Sports video analysis has rapidly developed due the growth of the attention of the audiences and the investments in the market. The analysis of soccer videos can be classified as pre game, in game or post game data processing. Soccer video analysis includes the recording and the examination of the video data with the goal of event extraction and summarization. Extraction and summarization results are presented by various graphical representations. Those events refers to the player actions, ball actions and the interaction of the players with the ball as well as with each other. The activity of the players with the ball constitutes the most important part of the analysis.

Because the players engaged with the ball becomes in power of making the movements according to the position and the tactic in order to have the control of the game. For the analysis, the most important data includes the ball and player position, players involved, the action taken with the ball, and the outcome of the activity. The soccer video analysis system consists of data acquisiton, preprocessing, feature extraction, detection, tracking and action recognition layers. In this thesis, a general framework for the analysis of soccer games is proposed and developed. The modules include the preprocessing of the input data, extraction of the features for ball and player detection and tracking of the ball in order to extract the ball actions.

The work presented through this thesis is a part of the ASPOGAMO [Beetz et al., 2009] project developed by [Beetz et al., 2009]. The general flow chart of the approach is shown in figure 2.1. The video observation acquired through the cameras are fed to the observation and knowledge base. The input is mostly the broadcasted material from various TV channels. The observation system is responsible for the application of the computer vision techniques in order to estimate camera parameters, segment and track the ball and the players. The output of the computer vision algorithms are then fed to the knowledge base in order to build a higher level description of the game. The users to the mentioned ASPOGAMO system is mostly

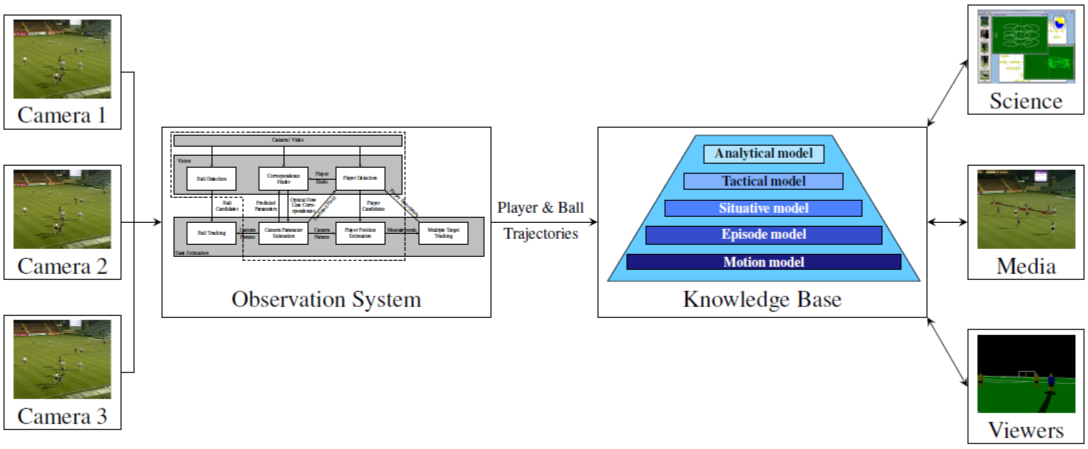


FIGURE 2.1. Flow Chart of the ASPOGAMO Approach [Beetz et al., 2009].

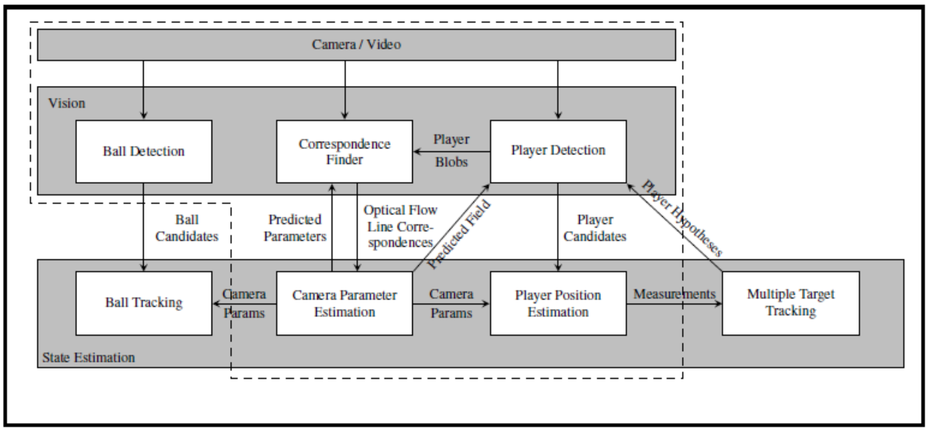


FIGURE 2.2. ASPOGAMO Observation System [Beetz et al., 2009].

trainers, sports scientists and the audience [Beetz et al., 2005, 2004, 2007].

2.1 Related Work

Over the last decades, the attraction in soccer and the drive for forming successful teams has increased the motivation of researchers and engineers. The analysis has been conducted on a limited time of the video data where the automatic analysis remains quite brief. In order to get an overview of commercial soccer analysis tools, we present a survey of the commercial systems and the research work that has been conducted in the following sections.

2.1.1 Commercial Soccer Match Analysis Tools

The computerized soccer video analysis technology has advanced rapidly in the last decade with the decrease in hardware costs and increasing attention in soccer video statistics. Most of those systems are designed to asssist coaches after the game. They are based on providing some statistics based on the position of the players. In most systems manual annotation is done or bought from other companies focused on the data generation and the further statistical analysis are applied using the already annotated data. Those commercial soccer analysis tools can be analyzed in three different categories; statistical analysis systems, vision-based tracking systems and electronic player tracking systems. The first group of systems are the most common products in soccer analysis. They present the statistical data depending on the position provided manually or automatically from other sources. DartTrainer[Dartfish, 2012] of Dartfish uses the tagging of the events during the game manually in order to extract the highlights and the scenes through the game. Focus X2[Elite Sports Analysis, 2012] system of Elite Sports Analysis has also a similar application providing the statistics to manage the replays and emphasize event highlights. Similarly, Scanfoot[Scanball, 2012] system of Scanball has operators entering manually all the players' relations with the ball, as well as the referee's breaks. Scanball system motivates on providing the statistics and analysis to clubs, newspapers, television channels. They also focus on web apllications and mobile application development for soccer game statistics. SoccerScout[Touch-Line Data Systems Ltd., 2012] of Touch-Line Data Systems Ltd. is one of the leading supplier of global soccer content to online and mobile media.

The second category of systems include the video-based player tracking systems. Those tools implement the machine vision methods and they use manual interaction of the operators as well. Prozone3[Prozone Sports Ltd., 2012] is one of the leading player tracking systems and provides a physical, technical and tactical coaching tool using multiple cameras installed within a stadium. It is mostly intended for coaches and management staff to analyse physical, technical and tactical performance information through an interactive and engaging coaching tool. Prozone3 uses a data capture system incorporating 8-12 digital IP cameras installed throughout the stadium in order to track the players through the game. Epsis[Symah Vision, 2012] of Symah Vision also uses a video-based player tracking system using the center-circle lines or goal-areas as image patterns. They also implemented the camera tracking by mounting a motion sensor on the camera in order to measure the pose of the camera. Amisco[SportsUniversal, 2012] of Sports Universal SA also uses image undertanding techniques to track the players in the video data. The Amisco system relies on a set of 3 to 6 cameras installed on the stadium. They implement model based calibration and data structure to built event models. Another well known video-based tracking systems is Vis.Track[Cairos Technologies AG, 2012] from Cairos Technologies AG. Vis.Track uses two cameras and track the players using the image processing algorithms and with the help of an operator.

The third group of soccer video analysis systems include the tracking of the players or the ball using electronic chips attached to the player shirts or within the ball. Of course this method provides the most accurate measuremet results but it is not preferred by most of the football federations since it is regarded as interfering within the game. Additionally the use of active sensors makes the players feel uncomfortable during the training of the match. The inmotiotec [inmotiotec GmbH, 2012] system from inmotiotec GmbH tracks the players using a transponder and responds with a pre-defined signal within a real-time network is connected to several base stations, namely measuring stations. Digital Sports Information system of Trakus Inc. uses radio tags attached to the players and antennas installed in the stadium in order measure the position of the players.

The systems explained so far have found many applications in stadiums and they are used by most of the leading clubs and broadcast companies. They are mostly composed of many cameras, other sensors or antennas installed in the stadium where the processing is done using a bunch of computers. Data is usually entered by salaried operators during the game and the accuracy sometimes stays at a very rough level. For example, the ball positions are not provided in real-time, but they are provided as an extra package which is provided to the trainer or the broadcaster a few hours after the game.

2.1.2 Soccer Video Analysis

Video is the richest source of information so far in the soccer game analysis. Video data provides records, observation and analysis including the visual and audio features.

In the scientific literature many studies have been conducted so far in the vision based analysis of soccer games [Armatas et al., 2005, D'Orazio and Leo, 2010, Park and Yilmaz, 2010, Sousa et al., 2011]. [Sousa et al., 2011] described an hierarchy in his work to classify different automatic soccer video event recognition systems. The analysis are classified as low, middle and high depending on the structure of the methods and the information provided. Similarly, [D'Orazio and Leo, 2010] presented a survey of vision based soccer analysis systems. The survey paper structured in a way to present the review of vision based systems in three different classes, i.e, video summarization, provision of augmented information and high-level analysis. Additionally, the weakness and strengths of the methods are investigated and the feasibility of the techniques for the extensive and real time soccer video analysis are examined.

Another approach is to analyse the video data output as a social network [Park and Yilmaz, 2010]. In [Park and Yilmaz, 2010], authors studied the interactions between the objects within the video and creted models of small communities using modularity. Within this social interaction network, they generated the nodes for each individual where the most important individuals put in the center and the priority decreased from center to other sides.

[Barris and Button, 2008] presented a detailed report on the vision-based motion analysis in sports in three groups, i.e., manual, automated and commercial systems. The main aim of the systems mentioned in this work is to filter the appropriate video sequences in which the players can robustly labeled in a cluttered and interacting environment. [Barris and Button, 2008] then proposed potential applications like planning tactics and strategies. In [Armatas et al., 2005], the authors used a commercial sports analysis program and analysed the numerical data to extract. In their work, they examined the way how the counter attacks started and comcluded that most of the successful counter attacks were started from stealing the ball where many of them started around the central axis.

[Coldefy and Bouthemy] used very high level visual features like dominant color features and camera motion analysis associated with the speech detection system for the abstraction of the soccer video.

The first systematical framework in machine vision for soccer analysis and summarization was by [Ekin et al., 2003]. In this work authors used low level and kinematic features for the detection of some events through the video sequence, i.e., slow motion detection, goal detection, referee detection and penalty box detection.

In [Babaguchi and Jain, 1998], the authors followed a different way and proposed a method to recognize the changes in the score legend on the TV screen to detect the events. Instead of focusing the motion information and detailed analysis, authors focused here on the score changes and used this information to conclude on the results.

Most of the soccer video analysis systems depend on the images gathered from multiple cameras. [Iwase and Saito, 2004] developed a machine vision system to track players and referees using 15 cameras at 15 frames per second. They provided the results on 500 frames of providing objective measurements about the players. [Barros et al., 2007] used for cameras to process the players'for around 10 hours of video. In [Junior and Anido, 2004], the authors used five cameras and 66 compters for a distributed real-time player tracking system. The duration of the analysis where tracking has been utilized is usually quite short with respect to the length of the game. For example, in [Iwase and Saito, 2004] 33.3 seconds, in [Utsumi et al., 2002] 30 seconds and in [Junior and Anido, 2004] of real soccer game data is analyzed.

2.1.3 Ball Detection

Ball detection is the first step in ball tracking and action recognition. It is necessary to detect the position of the ball in order to track it. The measurements and the localisation of the ball plays the main factor in the tracking performance and detecting the main events of the game. Extensive research has been conducted on the detection of the soccer ball in video frames. Some of those methods include single view mono cameras where the others include multiple and mostly dedicated cameras for the detection of the ball i.e., the offside cameras. In [Hossein-Khani et al., 2011], ball detection is implemented in a 5-step approach with the aim of detecting the corner events using the detection of the playfield and field lines and the ball. Cleaning morphological methods are used to detect the ball and the method is said to perform better than the circular Hough transform. Another method include the detection of the ball for the refinement of the placed kicks i.e., free kick, corner kick, penalty [Li et al., 2009]. In [Li et al., 2009] work ball detection results are used in combination with the detected field lines using Hough transform in decision making.

In [Pei et al., 2009], a ball detection framework for ball detection from broadcast soccer video is proposed. In this work, authors detected the ball candidates using a rule based algorithm and Kalman filter is employed for filtering the ball candidate locations. They employed two different search modes named rough and detailed search according to the output of the matching results.

In [Huang et al., 2007], authors proposed a method based on Hough transform and shape analysis. For ball detection very straightforward methods are used like the color, roundness and eccentricity. Trained color histograms are used for playfield and ball detection and ball detection is employed very simple based on the assumptions on its position according to the players.

Soccer ball moves very fast and it is mostly occluded by players. In [Shimawaki et al., 2006], ball route estimation is explained using simple linear prediction in order to increase the ball detection performance. Ball detection is implemented using a seperability filter [Fukui and Yamaguchi, 1998].

In [Yu et al., 2003b], ball size information is estimated using the player size within the video frames. [Yu et al., 2003b] proposed a ball detection method by using the inference of the ball size from such information and non-ball region removal. Kalman filter is used to filter out the ball candidates in candidate feature events. The authors then improved their method and used a two phase trajectory based algorithm in order to detect the ball positions in broadcast soccer videos [Yu et al., 2003d].

2.1.4 Ball Tracking

Ball tracking constitutes the main part of the ball motion analysis. It has been researched extensively so far [D. Jones, 2011]. There are two different methods used in the literature for automatically tracking the soccer ball. Those methods include the usage of electronic components and the development of image processing and video processing algorithms where a tracking method is also employed.

In [Holthouse and van de Griendt, 2009] and [Orenstein and J. Maune, 1999] an apparatus for the detection and tracking of the ball in sports are invented. The first system used electronic beacons to track the ball and this information is later used to detect the player kick, pass, bounce, strike or carrying a ball events. In the system of [Orenstein and J. Maune, 1999], directional antennas are used to track the moving ball.

The authors in [Pingali et al., 2000] claims a real-time ball tracking system to generate virtual replays of tennis games. They used a specific machine vision system which consists of 6 monochrome cameras which are very close to the ball. Auto iris lenses lenses are used to deal with the ligting changes. Even though the paper mentiones about the tracking of the tennis ball, the method explained in this work is filtering based on motion intensity and shape. In other words, ball localisation is implemented using frame differencing in a specific camera system. The work proposed in [Wong and Dooley, 2010] on ball tracking is focused on the tracking of the soccer ball for umpiring applications. Here the camera system for data acquisiton is not clearly stated and the ball tracking is implemented using size, shape and motion information.

Similarly, [Liu et al., 2010] uses frame differencing and size information to detect the ball using frame differencing. A linear motion estimation is employed along with the ball detection.

A dynamic and adaptive Kalman filter is proposed in [Kim and Kim, 2009] to track the ball robustly. The dynamic Kalman filter is explained to have better results by controlling the velocity of the state vector whereas the adaptive Kalman filter uses player infromation to overcome occluded situations of the ball.

Some other methods mentioned in [Ren et al., 2009] uses model based approaches for tracking the soccer ball on a video input of multiple fixed cameras. The fixed cameras have the advantage since they provide the whole pitch area during the game and do not need to be calibrated every time before the game. In this work, the removal method is used through the detection phase similar to [Yu et al., 2003b] using the field and motion models. Velocity, size and color information is used along with the Kalman filter algorithm. Similar to [Yu et al., 2003b], [Liang et al., 2005] uses the Kalman filter based on template matching. In this work the ball detection is implemented using weighted graphs and Viterbi algorithm to filter the ball candidates.

Authors in [Yu et al., 2003c] and [Yu et al., 2003d] proposed a tracking framework for broadcast soccer video using Kalman filter based trajectory mining. Here the candidate collection is explained to be done using a heuristic false candidate reduction. They have also implemented manually the team possession information in [Yu et al., 2003d].

[Kim et al., 1998] proposed a physics based 3D tracking of the ball for a limited number of frames using a similar approach to [Wang and Li, 1997]. For puck tracking, [Cavallaro, 1997] developed a system using 10 IR cameras, some of which also connected to broadcast cameras. They integragted an electronic circuit inside the puck and used the signal to track the puck.

Many methods explained above are tested only on a limited number of frames. In this thesis, we proposed and developed a tracker and tested it on a much larger datasets of various soccer games as well as on the field hockey videos. Our approach has also better perspective in the generality and the robustness of the algorithm.

2.1.5 Ball Action Recognition

Semantic analysis of soccer games has been extensively studied using the video content or data from external sources. Ball is the main component of the game and ball position and dynamics are the most crucial information sources while defining the actions. The main motivation of a player during the game is to possess the ball and show his skills using the ball in a team. Ball action recognition has been studied in the area of pyhsics, robotics and and computer vision.

The studies in physics investigates the physical aerodynamics of the soccer ball [Ghista and Liu, 2008, Goff and Carré, 2012]. [Goff and Carré, 2012] used a ball launcher and high-speed cameras to determine the drag and lift coefficients of the ball using trajectory analysis. They have analyzed the characteristics of the physical actions happening while the soccer ball is flying. Similarly, in [Ghista and Liu, 2008] the ball spin action is investigated using the physical approach and manually annotated data of the high-speed cameras. Physics based ball action recognition has also found applications in 3D reconstruction of the ball trajectory from single camera volleyball sequences [Chen et al., 2011]. The authors in [Chen et al., 2011] studied the motion equations and parameter estimation to set up the definition of the 3D trajectories based on physical characteristics. The estimated ball actions are used to automatically generate free viewpoint virtual replays.

In robotics [Castillo et al., 2011] developed a real-time optical flow algorithm to track the ball and get the speed of the ball in Robocup's league. Here the speed estimation of the ball is fed to the robot to guide the action of the robot.

In [Kataoka and Aoki, 2011], a combination of particle filter and real adaboost classifier are used to track the players as it provides robustness to occlusion. Ball localisation is implemented using the likelihood edge and color features. The projection of the players and the ball onto the pitch is done using the single camera image.

[Niu et al., 2010] and [Zhu et al., 2009] proposed an attack pattern recognition algorithm based on ball's state and real-world trajectory. Ball states are clustered and the recognition of the attack states are investigated. Ball detection is implemented using size and color features where they are filtered using the Viterbi algorithm to utilize particle filter afterwards [Zhu et al., 2007, 2009].

[Leo et al., 2008b], [Leo et al., 2008a] and [Yu et al., 2003a] took the results of the analysis one step ahead and proposed a real-time system to undertand the interactions between the ball and the players. In [Leo et al., 2008a,b], they used an input data provided by 6 different cameras set up in the stadium where the player motions are extracted using human body configuration analysis with an innovative neural approach based on a counterlet representation of human silhouette data. The team possession analysis are recognized using the local minima of the ball velocity and Support Vector Machines are utilized to find out which team has the ball possession.

The authors in [Hashimoto and Ozawa, 2006] proposed a multi-camera machine vision system for automatic judgement of offside events in soccer games. The first phased for the recognition of the event is implemented by tracking the players and calculating the offside line. The second decision phase uses the calculation of the 3D trajectory of the ball.

Finding the interesting pass patterns is one of the research topics in the ball action analysis and recognition [Hirano and Tsumoto, 2004]. The irregularity of the data and the requirements of the multi-scale matching is used to extract the interesting pass patterns during the game. Manually annotated data is used to cluster the ball positions based on multi-scale matching.

2.1.6 Player Detection

In order to provide summarization and higher level abstraction of soccer videos, the detection of the players and the referee becomes inevitable. Usually the player detection and player tracking methods are combined. In this section we present the literature that only focused on the detection of the players. The detection of the players are usually implemented using the color and shape features [Jia Liu, Xiaofeng Tong, Wenlong Li, Tao Wang, Yimin Zhang, 2009, Nunez et al., 2008]. The players are treated as standard predefined sized objects. The colors are learned using various color learning algorithms and the search has been done based on those constraints. Those appearance models [Jia Liu, Xiaofeng Tong, Wenlong Li, Tao Wang, Yimin Zhang, 2009] or player templates [Choi et al., 1997, Xu and Shi, 2005b] are then used to classify the players belonging to different teams and the referee. [Ming et al., 2009] used background subtraction algorithm based on Gaussian Mixture Models (GMMs) to detect the players.

2.2 Soccer Video Structure

Soccer broadcast has one of the biggest share in the TV broadcast industry. Recognition and the capture of the interesting scenes in real-time is required for informative and entertainment purposes.

The hardware used during the broadcast is transported to various stadiums and due to that reason on game days, a large staff is required to operate the various systems within the stadium. An example framework is explained in [Extron Electronics, 2012]. Here multiple broadcast cameras are located within the stadium to provide optimum coverage from various angles and locations. Professional camera operators positioned in strategic locations catch the interesting scenes during continuous game and crowd coverage. The camera output is sent to the server or the equipment room where all the views of different cameras are synchronized and pre-processed through frame grabbers. Server output is then distributed to various systems for broadcasting or display and further post-processing inside the stadium. In another room the production staff then edits and manages the content to feed the video data to broadcast over satellite or different displays within the stadium. It is possible in some stadiums the fusion of the signals from different sensors. Usually broadcast truck lots are located around the stadium for the satellite connection. The efficient transfer of huge amount of data via wired and wireless connections comprises another important problem in soccer broadcast industry.

Soccer video provides us with the permanent record of the game for further analysis. For that reason coaches and the football analysts have the chance to go through the game again and again in order to note the missing point during the match. From the player's perspective, he gets the chance to see his mistakes from a different view in an objective way. It is also easier and more comfortable for the coaches and players to analyse the event in the light of the video information. The presentation of the video information has also direct effect on the increase of the learning skills of coaches and players. Because the learning time becomes quicker and the editing and the observation of the events become easier. Portable and practical visual illustrations improve the efficieny in the analysis of soccer games while improving the experience of the players.

The soccer video information for analysis can be acquired via three different sources. The first source is the multiple static camereas installed in the stadium. In the static camera systems the cameras are installed in the stadium as shown in Figure 2.3. Those type of cameras usually do not provide a high resolution since they are not intended mostly for the image processing operations. The cameras installed in the stadium can also be used to observe the behaviour of the audience for security reasons. The frames acquired from those cameras include the different fields-of-view of the stadium and two different type of approaches are used for the analysis. In the first method, each camera image is processed separately and the output data is associated with the knowledge of the camera positions and the overlapping fields-of-view. The synchronisation of the cameras is the first step in multiple single field-of-view processing. This is achieved by request response method or using predefined time-stamps to grab the image. Additionally, multiple single camera views increase the reliabilty of the measurements and make the system robust to occlusions. Multiple measurements from different single-views are associated in the last part to filter the measurements.

In some cases the full 3D reconstruction might have higher cost in the complexity where the interpolation of some views are used for image analysis. Image mosaicing is the method of basically blending together the different single view inputs to form a larger and radiometrically balanced image so that the output image looks like one large image from a single camera [Hansen et al., 1994, Morimoto and Chellappa, 1997]. Blending can be done using user specified cutting edges or using the matching of the key point in two images. [Choi et al., 1997] and [Kim and Hong, 2000] applied in their work state of the art view interpolation to the field of professional soccer videos. The problem of image mosaicing is a combination of three problems namely, correcting geometric deformations using image data and/or camera models, image registration using image data and/or camera models, eliminating seams from image mosaics.

The last type of video source is the broadcast soccer video from the TV cameras installed in the stadium. The output of the TV cameras are processed in the production room and the video after the production is directly broadcasted using the satellite connection as explained above. The broadcast video information mentioned in this thesis is the output of the production room in the stadium same as the frames are delivered on the TV screen.

The broadcast is transmitted from production room to TV screen as a combination of different shots which is basically sequences of frames taken uninterruptedly by one camera. Shot transition detection is used to split up a film into basic temporal units called shots where a shot is a series of interrelated consecutive pictures taken contiguously by a single camera and representing a continuous action in time and space [Jokovic and Orevic, 2009, Lin et al., 2006, Urhan et al., 2006]. The transition between two consecutive shots, in other words the cuts, can happen in two types, namely abrupt and gradual transitions. Abrupt transition is a sudden transition from one shot to another where one frame belongs to the first shot and the next frame

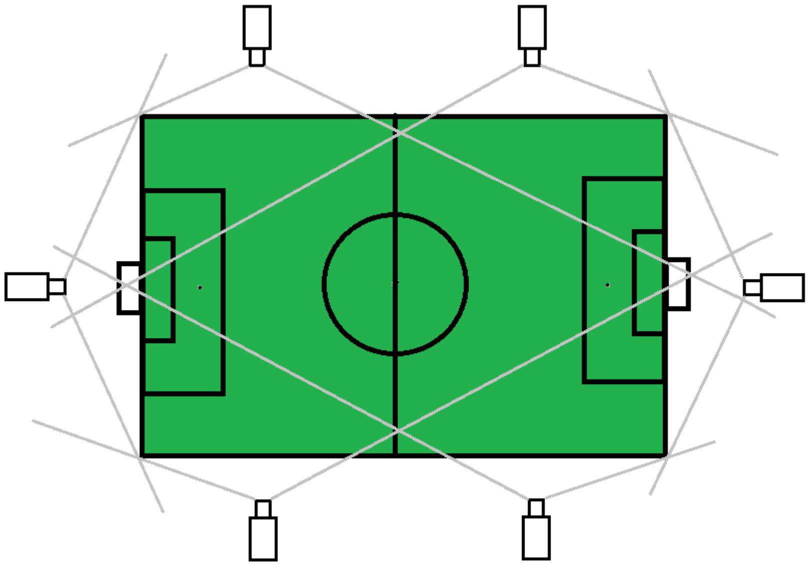


FIGURE 2.3. Multi static camera positioning in the stadium.

belongs to the second shot. In the literature, Abrupt transitions are also known as hard cuts or simply cuts.

Other type of transition is the Gradual Transitions where the two shots are gradually replaced by each other. These transitions are done by using chromatic or spatial effects. Gradual Transitions are known as soft cuts and can be found in various types, e.g., dissolves, fades, wipes and effects .

A dissolve happens when the frames from the last sequence fade out and the frames from the next sequence fade in. Here fade-out and fade-in and are used to describe a transition to and from a blank image, respectively. The dissolve is used to ovelap the two shots using the effects in between. Dissolves typically have a duration of 1 to 2 seconds ($25 - 50$ frames) for videos but a dissolve in a broadcast soccer video takes less than a second ($15 - 20$ frames) [Gedikli, 2008].

A wipe is a type of soft cut where the transition from one shot to another occurs by sweeping from one side of the frame to another by using special sweeping shapes. Effects are also used during the wipes in soccer broadcasts at the beginning and the end of replay sequences. Figure 2.4 shows the illustrations of some shot transitions, i.e. cuts.

The broadcast soccer video analysis becomes more difficult than static camera views since many effect are added by the production during the broadcast. Those effects include cuts, dissolves and effects between two consecutive shots.

In the literature, the detected shot transitions are called a hit, the non detected ones are called

(A)Hard cut

(B)Dissolve

(C)Effect

FIGURE 2.4. Shot boundary examples, hard cut, dissolve and effect [Gedikli, 2008].

a missed hit and the falsely detected cuts where actuall no hit happens are called false hit. In the literature, most algorithms achieved good results in detecting the hard cuts [Cernekova et al., 2006, Priya and Domnic, 2011] but many failed in detecting the soft cuts[Porter et al., 2000, Tanwer and Reel, 2010]. It becomes difficult to detect the Gradual Transitions since they happen slowly and gradually. Through the shots the hard cut occurs between frame i and $i+1$ while the soft cut occurs between frames i and j. Detection of cuts becomes a non-trivial task for computers unless it is provided with strong artificial intelligence algorithms.

In this work, we rely on the scene iterator developed in [Gedikli, 2008] to search and move through the sequences captured from the TV cameras. In his work, [Gedikli, 2008] developed a system to iterate through the video and detect the new video segments in shot boundaries in the broadcast soccer video. He designed two independent classifiers on the broadcast soccer video segments. The first classifier is to detect field-view or non-field-view where the second one is for slow-motion-replay or non-slow motion-replay. In his work, he proposed a feature-based method to detect hard cuts, and color histogram comparison to detect soft cuts like dissolves, fades, wipes and effects. He assumes the length of the shot as 20 frames and the shot boundary detection is disabled after the first detection for the next 20 frames to prevent multiple detections.

Using this secene iterator, we are able to skip all non-field-view and slow-motion-replay since they would not have any meaning while evaluating the player detection, ball detection and tracking. We also skipped the constant regions like channel logo and score legend on the screen and excluded those regions in the processing. The remaining sequences are used for the player detection, ball detection, ball tracking and ball action recognition modules.

2.3 Soccer Video Processing

Soccer video processing refers to the objective computer-based examination and processing of frames in soccer broadcast videos. It focuses on the detection and the tracking of the objects as well as the higher level event detection and extraction either manually or automatically. The application might vary from advertising to action recognition or graphical interpretation as explained in the introduction chapter. The aim is to extract information from the video frames using manual or automated tools and represent them at different levels of interpretation.

In this thesis, we focus on the analysis of the game from a novel perspective by using the ball motions. Most of the currently existing soccer video analysis systems focus on the raw data extraction based on players' movements. In other words, most of the computer-aided methods provide the user or the audience with the image coordinates of the players at discrete time stamps. A very less number of analysis methods are capable of providing the ball activity and movements during the soccer game. Most of the methods provide the interpolation of active players' positions as the ball positions. Ball motions are always assumed to be always at the same speed and the ball is always assumed to be rolling which is mostly misleading for the deeper analysis of the game.

However the deeper analysis of the game requires the actual ball track and actions. For example, pass actions between the players does not have to be always on the ground. If a coach looks at the data provided by only interpolation her cannot discriminate if the opponent team is playing on the ground or doing more passing where the ball is flying in the air. This might lead to wrong tactical interpretations of the game.

Ball is very difficult to track within the frame using only the image understanding techniques if it is treated as a single object. In our approach, we propose a novel tracking method for single and very small target that are in interaction with the environment.

To sum up, the soccer video analysis provide the coaches and audience what is going on during the game and provide the deeper analysis using the advanced video and image processing methods.

CHAPTER 3

Tracking Small Single Targets in Clutter

Tracking small single targets deals with the problem of fusing various measurements from different sources with the predicted state of a small single target to estimate its position. Major advances in hardware and software increased the signal processing capabilities in sensing mechanisms. This has increased the level of information in the measurement data for tracking systems as well as the accuracy and precision of the results. Tracking is associating the measurements obtained from a sensor in order to estimate the current state of the target. Basically, most of the major tracking methods consist of two consequtive steps, i.e., prediction and update. In the prediction step, the aim is to state about the next future step based on the motion model and current measurements. In the update step, the aim is to combine the predicted next future step with the associated current measurements. The state of the target of interest usually evolves according to the equation

$$x(k+1) = F(k)x(k) + v(k) \tag{3.1}$$

where the correct measurements are represented by

$$z(k) = H(k)x(k) + w(k) \tag{3.2}$$

where v and w are zero mean mutually independent, white, Gaussian noise sequences with known covariance matrices $Q(k)$ and $R(k)$, respectively.

Tracking of the ball proposed in this thesis generally includes the processing of the measurements from the soccer TV broadcast camera with the aim of estimating the current state of the ball. Here the state of the ball represents the ball kinematics, i.e., ball position, velocity and acceleration. Here the measurements from the TV broadcast camera related to the state of the ball in an intelligent way using,

- measurement success rate measure

- efficient pyramidal ROI
- player kinematics

The ball and player position candidates gathered in each frame is also filtered using temporal consistency to get better estimation results. Kinematic components of the ball and the players are used to maintain the estimate of the ball.

In this chapter, we present the general single target tracking methods and we discuss why we have proposed and extended version of particle filtering methods using an intelligent and efficient way of sampling for team-game ball tracking. Our approach increases the efficiency of the small-single target tracking which is mostly studied and implemented by using Kalman filtering methods.

3.1 Related Work

Tracking a single small target deals with the problem of data association in cluttered environment. The estimation of the target state and the modeling of the state vector and filtering of the measurements. In a tracking framework, the set of validated measurements consist of the incorrect measurements as well as the correct ones. The measurement error consists of errors happen around the target or through the mesurement sensor itself. These type of detections or errors are named as clutter. In image processing the clutters might occur due to the input camera paramaters,camera motion, occlusion by other objects and the kinematic characteristic of the target. The Nearest Neighbor filtering [Alouani et al., 1990, Cover and Hart, 1967, Li, 1993] is a commonly used method which is explained in the next sections. [Bar-Shalom and Fortmann, 1987] proposed non-Bayesian (Track-Splitting Filter) and Bayesian approaches (Probabilistic Data Association Filter (PDAF)) for single target tracking. The most widely know linear state estimation technique used also within the algorithms mentioned is the Kalman filter [Kalman, 1960, Kalman and Bucy, 1961]. Among all mentioned above the Particle Filter [Arulampalam et al., 2002] is the most suitable solution for soccer ball tracking due to its features and kinematics. Therefore we used an extended version of Particle Filter to track the soccer ball in broadcast soccer videos. The details of the method is explained in detail in the next chapter.

3.2 The Nearest Neighbor Standard Filter

The Nearest Neighbor Standard Filter (NNSF) is the most commonly used method in single target tracking [Cover and Hart, 1967, Farina and Pardini, 1978]. The measurement which is the nearest to the prediction is used in the update phase of the state [Alouani et al., 1990]. The distance measure used for the nearest neigbor detection is given

$$\begin{aligned} d^2(z) &= [z - \hat{z}(k|k-1)]' S^{-1}(k+1)[z - \hat{z}(k+1|k)] \\ &= v'(k+1) S^{-1}(k+1) v(k+1) \end{aligned} \tag{3.3}$$

where S represents the covariance matrix of the innovation for true measurements.

While deciding on the measurement results which are nearest neigbor to the prediction, with a some probability the correct measurement might not be taken. This occurs due to the the prediction and the sum of estimation errors cumulated till the current state. This errors happens since the filter calculated error covariance matrix does not account for the possibility of processing an incorrect measurement. This is the weak part of the NNSF and mostly causes the divergence of the tracking performance from the correct results [Li, 1993, Singer and Sea, 1973]. Therefore the NNSF might rely on the incorrect measurements as they are correct and continue the estimation of next steps depending on that. This causes possible performance degradation in the filtering process. An evaluation of the posterior probability of nearest neigbor was first proposed in [Jaffer and Bar-Shalom].

3.3 The Track-Splitting Filter

As the name implies the Track-Splitting Filter generates the hypothesis tracks by splitting the track into seperate hypothesis after the initialization phase. The Track-Splitting filter is firstly described in [Smith and Buechler, 1975].

The algorithm is a recursive approach which can be successfully utilized in environments with unknown number of targets. New tracks can be initialized using the nearest neigbor or other similarity measures. As a non-Bayesian approach the algorithm deos not yield a probability that the sequence is correct. In the Track-Splitting filter one hypothesis track is generated for each expected measurement around the location starting from the first prediction. This means that for each measurement seperate updated state and covariance are computed and followed using the Kalman Filter equation in section 3.6. The process is repeated for the second measurement and then repeated in the same manner.

On of the problems of this method is the exponential increase in the number of hypothesis tracks since the expected measurement regions always grows. This causes a problem like the degeneracy phenomenon of Particle Filter. To prevent the exponential grow of the tracks a likelihood measure is defined and the unlikely split tracks are eliminated. If we represent the l^{th} sequence of the measurements up to time k as

$$Z^{k,l} \triangleq \{z_{i_{1,l}}(1), ..., z_{i_{k,l}}(k)\} \tag{3.4}$$

where $z_i(j)$ is the i^th measurement at time j.
Given that this specific sequence is a track

$$\theta^{k,l} \triangleq \{Z^{k,l}\ is\ a\ correct\ track\} \tag{3.5}$$

Then the likelihood function of this specific track sequence is given as

$$\begin{aligned} \Lambda(\theta^{k,l}) &= p[Z^{k,l}|\theta^{k,l}] \\ &= p[z_{i_{1,l}}, ..., z_{i_{k,l}}(k)|\theta^{k,l}] \end{aligned} \tag{3.6}$$

Replacing Z^K the sequence of measurements up to time k results in the joint probability density function as

$$\Lambda(\theta^{k,l}) = \prod_{j=1}^{k} p[z_{i_{j,l}}|Z^{j-1}, \theta^{k,l}] \tag{3.7}$$

With the linear-Gaussian assumption,

$$\begin{aligned} p[z_(j)|Z^{j-1}, \theta^{k,l}] &= N[z(j); \hat{z}(j|j-1), S(j)] \\ &= N[v(j); 0, S(j)] \end{aligned} \tag{3.8}$$

where $v(j)$ is the innovation term or the measurement residual.

$$v(k+1) = z(k+1) - \hat{z}(k+1|k) \tag{3.9}$$

Using equations 3.7 and 3.8

$$\Lambda(\theta^{k,l}) = c_k exp[\frac{-1}{2}\sum_{j=1}^{k} v'(j)S^{-1}(j)v(j)] \tag{3.10}$$

where c_k is a constant

$$c_k = \prod_{i=1}^{k} |2\pi S(i)|^{\frac{-1}{2}} \tag{3.11}$$

Here it is assumed that the target detection probability is unity. This means that the missing measurement sequences are ignored. If the target detection probability is smaller than unity the best technique is to choose a track which has at least enoough amount of that provides an acceptable goodness of fit.

Then the log-likelihood function becomes

$$\lambda(k) \triangleq -2log[\frac{\Lambda(\theta^{k,l})}{c_k}] = \sum_{j=1}^{k} v'(j)S^{-1}(j)v(j) \tag{3.12}$$

It can be computed recursively as

$$\lambda(k) = \lambda(k-1) + v'(k)S^{-1}(k)v(k) \tag{3.13}$$

The second term in equation 3.13 has a chi-square density with n_z degrees of freedom and as the innovations are independent the log likelihood function at time k is chi-square distributed with kn_z degrees of freedom. Equation 3.13 is also a measure of the goodness of fit. The track is accepted iff

$$\lambda(k) \leq a \tag{3.14}$$

where a follows from the chi-square tables for kn_z degrees of freedom.

$$P\{\chi^2_{kn_z} > a\} = \alpha \tag{3.15}$$

Here α is the probability of a true track to be rejected and it is typically taken like 10^{-2}

The usage of the algorithm is usually not very practical for long sequences since the likelihood becomes more dominated by the old measurements than the new measurements in time. The usual solution to this problem is utilizing a sliding window or a fading memory for the implementation. The Track-Splitting filter also undergoes a problem of lack of competition between the same measurements since only the likelihood of a track is evaluated. The most important problem of the Track-Splitting approach is the high computational and memory costs as it is an exponentially growing operation.

3.4 The Probabilistic Data Association Filter

The Probabilistic Data Association Filter (PDAF) is a suboptimal Bayesian method which assumes a single target of interest that is modeled by equation 3.1. The PDAF algorithm which uses all the validated measurements with their posterior probabilities was proposed and described in [Bar-Shalom and Tse, 1975] In [Li and Bar-Shalom, 1991], authors present an effective approach of a hybrid nature to the nonsimulation performance evaluation of the probabilistic data association filtering (PDAF) method for tracking in clutter. [Colegrove and Davey, 2000] described in their work the state estimation equations for a probabilistic data association filter (PDAF) which is updated by a fixed number of nearest neighbours.

The probabilistic data association filter (PDAF) builds a validation region each sampling time which is probably the one related to the target among several measurements. The false measurements are modeled as independent identically distributed (IID) random variables with uniform spatial distributions.

The set of validated mesurements at sampling time k is denoted

$$Z(k) \triangleq \{z_i(k)\}_{i=1}^{m_k} \tag{3.16}$$

where m_k is a random variable that represents the number of measurements in the validation region. Then the cumulative set of measurements are given by

$$Z^k \triangleq \{Z(j)\}_{j=1}^{k} \tag{3.17}$$

The PDAF differs from the optimal approach as it decomposes the estimation with respect to the origin of each element of the latest set of measurements as see in equation 3.16. The optimal Bayesian algorithms decomposes the estimation with respect to each sequence of the measurements as in equation 3.17. In the PDAF the state is assumed to be normally distributed according to the latest estimate and the covariance matrix.

$$p[x(k)|Z^{k-1}] = N[x(k); \hat{x}(k|k-1), P(k|k-1)] \tag{3.18}$$

Here the $Z(k)$ contains measurements from the elliptical validation region denoted

$$\begin{aligned}\tilde{V}_{k+1}(\gamma) &\triangleq \{z : [z - \hat{z}(k+1|k)]' S^{-1}(k+1)[z - \hat{z}(k+1|k)] \leq\} \\ &= \{z : v'(k+1) S^{-1}(k+1) v(k+1) \leq \gamma\}\end{aligned} \tag{3.19}$$

where v is the innovation term of equation 3.9 and γ is the paramater gathered from the chi-square distribution. The weighted norm of innovation above 3.19 is chi-square distributed with number of degrees of freedom equal to the dimension n_z of the measurement. The square root of $g = \sqrt{\gamma}$ is called the number of sigmas.

Following the equation of assumption in 3.18 if we have the events

$$\theta(k) \triangleq \{z_i(k)\ is\ the\ target\ originated\ measurement\},\ i = 1, ..., m_k \tag{3.20}$$

$$\theta(k) \triangleq \{none\ of\ the\ measurements\ is\ target\ originated\} \tag{3.21}$$

associated probabilities

$$\beta(k) \triangleq P\{\theta_i(k)|Z^k\},\ i = 0, 1, ..., m_k \tag{3.22}$$

Based on these assumptions the events are mutually exclusive which implies

$$\sum_{i=0}^{m_k} \beta_i(k) = 1 \tag{3.23}$$

The total probability theorem with respect to above events lets us represent the conditional mean of the state at time k as

$$\begin{aligned} \hat{x}(k|k) = E[x(k)|Z^k] &= \sum_{i=0}^{m_k} E[x(k)|\theta_i(k), Z^k] P\{\theta_i(k)|Z^k\} \\ &= \sum_{i=0}^{m_k} \hat{x}_i(k|k)\beta_i(k) \end{aligned} \tag{3.24}$$

where $\hat{x}_i(k|k)$ is the updated estimate which is conditioned on the event $\theta_i(k)$ that the i^{th} measurement is correct. Then the estimate posterior becomes

$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)v_i(k)\ \ i = 1, ..., m_k \tag{3.25}$$

where the corresponding innovation is

$$v_i(k) \triangleq z_i(k) - \hat{z}(k|k-1) \tag{3.26}$$

Here the gain is

$$W(k+1) = P(k+1|k)H'(k+1)S^{-1}(k+1) \tag{3.27}$$

since there is no measurement uncertainty. If there is no correct measurements then the estimate becomes

$$\hat{x}_0(k|k) = \hat{x}(k|k-1) \tag{3.28}$$

Then the state update equation of PDAF can be rewritten as

$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)v(k) \tag{3.29}$$

The equation 3.29 has a non linear characteristic due to the nonlinear innovation terms. The so called combined innovation term is derived as follows.

$$v(k) \triangleq \sum_{i=1}^{m_k} \beta_i(k)v_i(k) \tag{3.30}$$

Then the error covariance associated with the updated state estimate becomes

$$P(k|k) = \beta_0(k)P(k|k-1) + [1-\beta_0(k)]p^c(k|k) + \tilde{P}(k) \tag{3.31}$$

where

$$\tilde{P} = W(k)[\sum_{i=1}^{m_k} \beta_i(k)v_i(k)v_i'(k) - v(k)v'(k)]W'(k) \tag{3.32}$$

and the covariance of the state estimate at time k updated with the correct measurement.

$$P^c(k|k) \triangleq [I - W(k)H(k)]P(k|k-1) \tag{3.33}$$

The first term in equation 3.31 shows that none of the measurements are correct with the respective weighting. The second term shows that with the mentioned weight the correct measurement is available and the third term is used to increase the covariance of the updated state. This happens due to the effect of measurement origin uncertainty.

The advantage of the PDAF is that the estimation accuracy also depends on the data encountered which makes the estimation more realistic and accurate. As seen above the covariance is dependent of the measurements. This is one of the main characteristics of the nonlinear filters. The prediction of the state and the measurement of time $k+1$ is utilized as in equations 3.1 and 3.2 given as

$$P(k+1|k) = F(k)P(k|k)F'(k) + Q(k) \tag{3.34}$$

One of the novelities of the PDAF is the evaluation of the association probabilities in equation 3.24. For the data association the first step is the classifying the conditioning into the past data Z^{k-1} and the latest data. Here the data mentioned consists of m_k measurements at locations $z_1(k), ..., z_{m_k}(k)$. The probabilities of events can be written as

$$\begin{aligned}\beta_i(k) &\triangleq P\{\theta_i(k)|Z^k\} \\ &= P\{\theta_i(k)|Z(k), m_k, Z^{k-1}\},\ i = 0, ..., m_k\end{aligned} \tag{3.35}$$

Applying Bayes'rule, we get

$$\beta_i(k) = \frac{1}{c}p[Z(k)|\theta_i(k), m_k, Z^{k-1}]P\{\theta_i(k)|m_k, Z^{k-1}\}, \quad i = 0, 1, ..., m_k \tag{3.36}$$

where c represents the normalization constant. The joint density of the validated measurements is the product of the normal probability density function of the correct measurements and the uniform probability density function of the incorrect measurements. Here the assumption is that the incorrect measurements are uniformly and independently distributed within the validation region. Then the probability density function of the correct measurement is represented as

$$\begin{aligned}p[z_i(k)|\theta_i(k), m_k, Z^k] &= P_G^{-1}N[z_i(k); \hat{z}(k|k-1), S(k)] \\ &= P_G^{-1}N[v_i(k); 0, S(k)] \\ &= P_G^{-1}|2\pi S(k)|^{\frac{1}{2}}exp[\frac{-1}{2}v_i'(k)S^{-1}(k)v_i(k)]\end{aligned} \tag{3.37}$$

Here P_G is called the gate probability which represents the correct measurements that fall in the validation gate. Then the probability density function of equation 3.36

$$p[Z(k)|\theta_i(k), m_k, Z^{k-1}] = \begin{cases} V_k^{-m_k+1}P_G^{-1}N[v_i(k); 0, S(k)], & i = 1, ..., m_k \\ V_k^{-m_k}, & i = 0 \end{cases} \tag{3.38}$$

Here the V_k is the volume of the validation region explained in equation 3.19. In equation 3.19 the volume of the elliptical validation region can be computed as

$$V_k = c_{n_z}|\gamma S(k)|^{\frac{1}{2}} = c_{n_z}\gamma^{\frac{n_z}{2}}|S(k)|^{\frac{1}{2}} \tag{3.39}$$

where n_z is the dimension of the measurement z and c_{n_z} is the volume of the n_z dimensional unit hypersphere.

The prior probabilities of the events θ_i conditioned only on the number of validated measurements are

$$\begin{aligned}\gamma_i(m_k) &\triangleq P\{\theta_i(k)|m_k, Z^{k-1}\}\\ &= P\{\theta_i(k)|m_k\}\\ &= \{ \begin{matrix} \frac{1}{m_k}P_DP_G[P_DP_G + (1-P_DP_G)\frac{\mu_F(m_k)}{\mu_F(m_k-1)}]^{-1}, & i=1,...,m_k \\ (1-P_DP_G)\frac{\mu_F(m_k)}{\mu_F(m_k-1)}[P_DP_G + (1-P_DP_G)\frac{\mu_F(m_k)}{\mu_F(m_k-1)}]^{-1}, & i=0 \end{matrix}\end{aligned} \tag{3.40}$$

where $\mu_F(m_k)$ is the probability mass function (PMF) of the number of false measurements, i.e, clutters and P_D is the correct target detection probability.

Usually two models are used for the PMF in the literature, namely parametric and nonparametric models.

In the parametric model the PMF is represented by a Poisson density with parameter γV_k

$$\begin{aligned}\mu_F(m_k) &= P\{m_k^F = M_k\}\\ &= e^{-\lambda V_k}\frac{(\lambda V_k)^{m_k}}{m_k!},\ m_k = 0,1,2,...\end{aligned} \tag{3.41}$$

where λ is the spatial density of false mesurements and λV_k is the expected number of false measurements in the validation region or the gate. The parametric model yields

$$\gamma_i(m_k) = \begin{cases} \frac{P_DP_G}{P_DP_Gm_k+(1-P_DP_G)\lambda V_k}, & i=1,...,m_k \\ \frac{(1-P_DP_G)\lambda V_k}{P_DP_Gm_k+(1-P_DP_G)\lambda V_k}, & i=0 \end{cases} \tag{3.42}$$

In the nonparametric model the PMF is represented by a diffuse prior as

$$\mu_F(m_k) = \frac{1}{N},\ \ m_k = 0,1,...,N-1 \tag{3.43}$$

which yields

$$\gamma_i(m_k) = \begin{cases} P_D P_G/m_k, & i = 1, ..., m_k \\ 1 - P_D P_G, & i = 0 \end{cases} \tag{3.44}$$

In other way the nonparametric model can be derived by replacing the Poisson parameter with the sample spatial density of validated measurements, i.e,

$$\lambda = \frac{m_k}{V_k} \tag{3.45}$$

Using the equations above, the PDAF with the Poisson clutter model can be gathered as

$$\beta_i(k) = \frac{e_i}{b + \sum_{j=1}^{m_k} e_j}, \quad i = 1, ..., m_k \tag{3.46}$$

$$\beta_0(k) = \frac{b}{b + \sum_{j=1}^{m_k} e_j}, \quad i = 1, ..., m_k \tag{3.47}$$

where

$$e_i \triangleq exp\{\frac{-1}{2} v_i'(k) S^{-1}(k) v_i(k)\} \tag{3.48}$$

$$\begin{aligned} b &\triangleq \lambda |2\pi S(k)|^{\frac{1}{2}} (1 - P_D P_G)/P_D \\ &= (2\pi/\gamma)^{n_z/2} \lambda V_k c_{n_z} (1 - P_D P_G)/P_D \end{aligned} \tag{3.49}$$

Similarly the nonparametric PDAF is given by

$$b = (2\pi/\gamma)^{n_z/2} m_k c_{n_z} (1 - P_D P_G)/P_D \tag{3.50}$$

To sum up, even if the estimate in equation 3.1 seems to be linear, the PDAF is a nonlinear filter since the associated probabilities is dependent on the innovations according to equations 3.46 and 3.47. The details of the parametric and nonparametric versions of PDAF is described in more detail in [Y. Bar-Shalom]

3.5 The Optimal Bayesian Approach

The Bayesian filter is theoretically the best possible estimation technique possible for a single target. Here for the single target modeled by equations 3.1 and 3.2 and arbitrary number of measurements. The Bayesian approach represents the decomposition of the state in terms of all measurements from the initial to the present time. The Optimal Bayesian Approach presented here is firstly proposed by [Singer et al., 1974]. Thus the estimation covers all the time states from the beginning till present which is normally the ideal case regarding also the computational cost.

The sequence of measurements is given as

$$Z^{k,l} \triangleq \{z_{i_{1,l}}(1), ..., z_{i_{k,l}}(k)\} = \{Z^{k-1,s}, z_{i_{k,l}}(k)\} \tag{3.51}$$

which results in total number of measurement history at time k as

$$L_k = \prod_{j=1}^{k}(1 + m_j) \tag{3.52}$$

where m_j is the number of measurements at time j. Here the first term in the paranthesis accounts for the case where no measurements are correct. Then con ditioning on all the data up to time k we get

$$\beta^{k,l} = P\{\theta^{k,l}|Z^k\} \tag{3.53}$$

Here the first term in the conditional probability means an event that the l^{th} history at time k is the correct sequence of measurements. Since those events are mutually exclusive the conditional mean of the state at time k is denoted

$$\begin{aligned}
\hat{x}(k|k) &= E[x(k)|Z^k] \\
&= \sum_{l=1}^{L_k} E[x(k)|\theta^{k,l}, Z^k]P\{\theta^{k,l}|Z^k\} \\
&= \sum_{l=1}^{L_k} \hat{x}(k|k)\beta^{k,l}
\end{aligned} \tag{3.54}$$

where $\hat{x}(k|k)$ is called the history conditioned estimate. For each history then we have a standars filter as

$$\hat{x}(k|k) = \hat{x}^S(k|k-1) + W^l(k)[z_{i_{k,l}}(k) - \hat{z}^S(k|k-1)] \tag{3.55}$$

where $z_{i_{k,l}}(k)$ is the measurement at time k in sequence l. Here $\hat{z}^S(k|k-1)$ is the predicted measurement which corresponds to history $Z^{k-1,S}$ with covariance $S^S(k)$. Similarly, the gain of the filter is

$$W^l(k) = P^S(k|k-1)H'[S^S(k)]^{-1} \tag{3.56}$$

Then the covariance of the history conditioned update state can be expressed as

$$\begin{aligned} P^l(k|k) &= E\{[x(k) - \hat{x}^l(k|k)][x(k) - \hat{x}^l(k|k)]'|\theta^{k,l}, Z^k\} \\ &= [I - W^l(k)H]P^S(k|k-1) \end{aligned} \tag{3.57}$$

which yields the covariance associated with the combined estimate as

$$P(k|k) = \sum_{l=1}^{L_k} \beta^{k,l} P^l(k|k) + \sum_{l=1}^{L_k} \beta^{k,l} \mu^l(k) \mu'^l(k) - \mu(k)\mu'(k) \tag{3.58}$$

where

$$\mu^l(k) \triangleq \hat{x}^l(k|k) - \hat{x}(k|k) \tag{3.59}$$

$$\mu(k) \triangleq \sum_{l=1}^{L_k} \beta^{k,l} \mu^l(k) \tag{3.60}$$

As seen from the algorithm above the memory and computation requirements increase with time which has the advantage of optimality but disadvantage of the computational costs.

Suboptimal versions of the optimal Bayesian approach exists by combining all tracks that have identical histories for a specific number of scans before the present one. This approach is called the N-scan-back filter.

3.6 Kalman Filter

The Kalman filter or also known as linear quadratic estimation (LQE), is an algorithm which uses a series of measurements taken over time with other inaccuracies to maintain an estimate of unknown variables [Kalman, 1960, Kalman and Bucy, 1961]. Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state. The filter is named for Rudolf (Rudy) E. Kalman, one of the primary developers of its theory.

The Kalman filter has various applications in guidance, navigation and control of aircraft

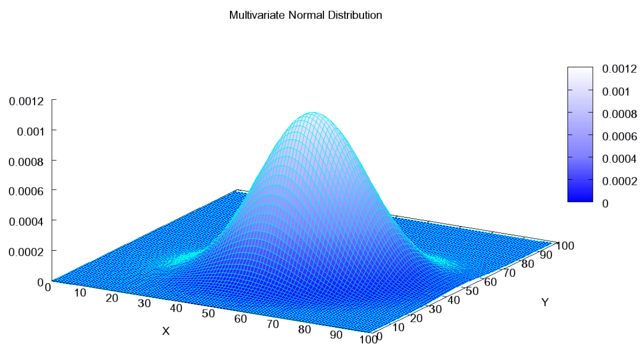


FIGURE 3.1. The multivariate normal (or Gaussian) distribution probability density function.

and spacecraft vehicles and many other scientific problems of time series analysis. Furthermore, the Kalman filter is a widely applied concept signal and image processing for tracking single targets.

The algorithm works in a two-step process, i.e., prediction step and update step. In the prediction step the Kalman filter maintains and estimate of the current state variables along with their uncertainties. The update phase starts once the outcome of the next measurement is gathered. Here the measurements are observed (necessarily with an error and noisy) the estimate is updated using a weighted average where more weight is associated with the estimates of higher certainty. The algorithm is a recursive process and that's why the system can run in real time using only the previous state and current measurements. The assumption of the Kalman filter is that the underlying system is a linear dynamical system where all error terms and measurement have a multivariate Gaussian distribution.

The multivariate Gaussian or Normal distribution is given by

$$p(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \tag{3.61}$$

where mean $\mu = \bar{x}$ and covariance $P = \sigma^2$.

The probablity density function of univariate and bivariate normal distributions are shown in 3.1

The normal distribution is the most prominent probability distribution since it very tractable analytically meaning that a large number of measurements involved in this distribution can be derived in explicit form. In addition to that, the most important characteristic of the normal distribution arises from the central limit theorem. The central limit theorem states that the mean of a large number of random variables independently drawn from the same distribution is also distribution. In other words, the sum of sufficiently large number of independent random variables approaches a Gaussian or normal distribution in the limit. This characteristic gives the Gaussian or normal distribution an exceptionally wide application in sampling. For these reasons, the normal distribution is commonly used in signal processing specifically traking as a simple model for complex scenarios.

The Kalman filter consists of a set of recursive equations for the prediction and update steps in order to track the target. The recursiveness makes the estimation feasible only from the state of the previous time step and the measurement from the current step.

In the prediction step, predicted state estimate and the estimate covariance is given in equations 3.62 and 3.63, respectively.

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k \tag{3.62}$$

and

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k \tag{3.63}$$

In the update step, the updated state estimate and the state covariance is as given in equations 3.64 and 3.65, respectively.

$$\hat{x}_{k|k-1} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \tag{3.64}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \tag{3.65}$$

Here the measurement residual and the innovation covariance is represented by the following equations.

$$\tilde{y}_k = z_k - H_k \hat{x}_{k|k-1} \tag{3.66}$$

$$S_k = H_k P_{k|k-1} {H_k}^T + R_k \tag{3.67}$$

The Kalman filter is a minimum mean-square error estimator. The Kalman gain is the parameter which yields the minimum mean-square error estimates. The Kalman gain is computed as 3.68

$$K_k = P_{k|k-1} {H_k}^T S_k^{-1} \tag{3.68}$$

In Kalman filter, typically the two phases alternate, with the prediction advancing the state until the next scheduled observation, and the update incorporating the observation. However, in some cases the observations and the measurements might not be available. In such cases the update is not necessary and then the update might be skipped where only multiple prediction steps are performed. Similarly if multiple independent observations or measurements are available at the same time then multiple update steps may be performed.

Extensions and generalizations to the method also have been developed such as the Extended Kalman Filter and the Unscented Kalman filter for nonlinear systems. The Extended Kalman Filter [Jazwinski, 1972] essentially linearizes the non-linear model function around the current estimate utilizing the Taylor expansion. [Julier and Uhlmann, 1997] developed the Unscented Kalman Filter which uses a deterministic sampling technique known as the unscented transform to pick a minimal set of sample points, also called sigma points, around the mean. The sigma points are then propagated through the nonlinear functions, from which the mean and the covariance of the estimate are covered. The Unscented Kalman Filter captures the true mean and covariance more accurately and removes the requirement to explicitly calculate Jacobians.

The Kalman-Bucy filter is the continuous time version of the Kalman filter [Bucy and Joseph, 1968, Jazwinski, 1972]. There are also Hybrid Kalman filter derivations where the model is continuous but the measurements are discrete time.

3.7 The Particle Filter

The particle filter is a tracking method, also known as Sequential Monte Carlo method (SMC), bootstrap filtering, survival of the fittest or the Condensation, that can handle more sophisticated models than other methods. [Arulampalam et al., 2002] presented a detailed summary of the method and its application in various methods. Generally speaking, the filtering usually refers to determining the distribution of a random variable at a specific time given all the observations up to that time. Particle filter is so named since they implement approximate filtering using a set of particles, also known as weighted samples of the distribution. The particle filters are usually used to estimate Bayesian models in which the latent variables are connected in a Markov chain.

In Sequential Importance Sampling particle filtering, the posterior probability density function is approximated by a weighted sum of N_p random samples x_k^i

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{N_p} w_k^i \delta(x_k - x_k^i) \tag{3.69}$$

where δ represents the Dirac delta function [Dirac, 1958] which is zero everywhere except zero and with an integral of one over the real line. The importance weights are normalized $\sum_i w_k^i = 1$ so that the integral of the probability density function satisfies the maximum value of one.

The particle filter consist of sampling and filtering steps. Particles are distributed or sampled according to a proposal $q(.)$ This sampling is called an importance density. The new weights depend on the choice of importance density and computed through the filtering step. The recursive equation for the weights with the normalizing constant factor then reads as

$$w_k^i \sim w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)} \tag{3.70}$$

As seen in equation 3.69 if N_p goes to infinity the approximation approaches to the true density.

[Doucet, 1998] showed in his work that the optimal importance density function that minimizes the variance of the true weights conditioned on x_{k-1}^i and z_k is

$$q_{opt}(x_k|x_{k-1}^i, z_k) = p(x_k|z_k, x_{k-1}^i) = \frac{p(z_k|x_k, x_{k-1}^i)p(x_k|x_{k-1}^i)}{p(z_k, x_{k-1}^i)} \tag{3.71}$$

The optimal importance density in equation 3.71 requires sampling directly from the real distribution, making the weights redundant. But this can hardly be achieved in real system applications. In most computer vision applications, the prior $p(x_k|x_{k-1}^i)$ is used as importance density leading to a weight update according to the likelihood $w_k^i \sim w_{k-1}^i p(z_k|x_k^i)$.

The density at a specific point can be calculated as a sum of the weights of all particles at the same point. Alternatively, the same density can be represented by a single particle with a high weight or multiple particles at the same location with proportional lower weights.

One of the biggest problems in SIS particle filter is the degeneracy phenomenon. The degeneracy phenomenon happens after a few iterations where one particle always began to have high weights and the other particles has negligible weights. The solution to degeneracy phenomenon is the resampling by eliminating the particles with negligible weights and duplicating the particles with larger weights. This method is called the Sequential Importance Sampling (SIR) particle filter. The resampling step of the SIR particle filter generates a new particle set $\{x_k^i\}$ by resampling N_p times from the current particle set and initializing new weights with $\frac{1}{N_p}$. The SIR algorithm is explained in algorithm 3.1

Particle filters are often an alternative to Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF) with sufficient samples. Particle filters approach the Bayesian optimal estimate with sufficiently large samples which makes them more accurate than Kalman Filter

Algorithm 3.1 SIR Particle Filter

Require: $\{(x^i_{k-1}, \theta^i_{k-1}), w^i_{k-1}\}^N_{i=1}, u_{k-1:k}, z_k$
  return $\{(x^i_k, \theta^i_k), w^i_k\}^N_{i=1}$
  for $i = 1$ to N **do**
    $\theta^i_k \sim p(\theta_k|\theta^i_{k-1})$
    $x^i_k \sim p(x_k|x^i_{k-1}, \theta^i_{k-1}, u_{k-1})$
    $w^i_k \leftarrow p(z_k|x^i_k, \theta^i_k, u_k)$
  $W \leftarrow \sum^N_{i=1} w^i_k$
  for $i = 1$ to N **do**
    $w^i_k \leftarrow w^i_k/W$
  $\{(x^i_k, \theta^i_k), w^i_k\}^N_{i=1} \leftarrow Resample(\{(x^i_k, \theta^i_k), w^i_k\}^N_{i=1})$

variations.

In [Chen and Liu, 2000], the approaches are combined by using a version of Kalman filter as a proposal distribution for the particle filter. [Chen and Liu, 2000] developed a special Kalman filter called the Mixture Kalman Filter where they track the mean and covariance of Gaussians for the target position by using Kalman filters inside each particle instead of the position itself. The technique of replacing the state by parameters of a model describing the entire state pdf is known as marginalization or Rao-Blackwellization [Casella and Robert, 1996]. Rao-Blackwellization improves the computational efficiency of particles since sampling can partially be replaced by and analytical solution. [Hoyningen-Huene, 2011] developed a novel Rao-Blackwellized Resampling Particle Filter (RBRPF) for multi-target tracking applications in soccer games. In his work, he proposed a novel resampling method by modeling the players positions as Gaussian distributions. He validated the performance of his algorithm in player tracking in broadcast soccer videos.

CHAPTER 4

Ball Detection and Tracking

Ball games, especially soccer, has different characteristics than other games with respect to the measures of performance and metrics applied for success. The winner is determined by goals scored or sets won, which means that the performance of the team might not be always inferenced from the result of the game. Those mentioned characteristics makes soccer more difficult to analyse and define success measures. Throughout a soccer game, ball always stay as the major element and the tool used to achieve success for a team. Therefore, performances taken around and with the ball usually constitutes the important actions in grading the game performances.

In this chapter we introduce novel ball motion analysis methods which represents ball detection methods based on a robust algorithm and real time ball tracking method which employs novel resampling methods regarding the ball motion characteristics. The overview of the framework is depicted in figure 4.1.

The first step in the image analysis is the segmentation of the field in order to eliminate noise and filter out the false positives. Then in the second step the detection of the players is investigated. One of the major parts of the thesis investigates the detection and the tracking of the ball which is presented in detail. The proposed approach not only detects or tracks the ball as a

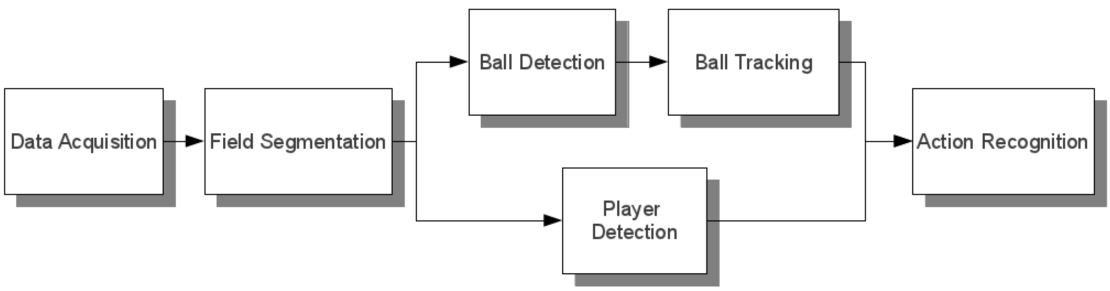


FIGURE 4.1. General System Overview

single object but also take into account the ball motion and game characteristics. Major methods proposed for ball detection in the literature focused on the elimination of non-ball regions and then applying a rule based exhaustive search in order to extract the ball regions [Huang et al., 2007, Pei et al., 2009]. We propose a novel approach based on the shape characteristic and additionally the motion characteristics of the ball and its role within the game. After presenting the details of the ball detection, the proposed novel ball tracking algorithm is proposed along with the derivations and perspectives. Many of the tracking methods developed so far only focused on the ball tracking regarding it independent of the environment. They have been tracking the ball with the assuming that it is only a moving object [Liu et al., 2010] or developing Kalman Filter based methods with linear assumptions that use multiple cameras in order to solve the occlusion problems [Kim and Kim, 2009, Ren et al., 2009]. We propose a smart resampling based ball tracking method that is evaluated on longer sequences than the mentioned methods and performs more accurate results on the broadcast soccer videos. The generality of the algorithm is also proved by the experiments on hockey sports videos. After the detection and the tracking algorithms have been explained, a novel ball state classification algorithm is presented on the following sections. This method provide promising results in classifying the ball states as flying or rolling on the broadcast soccer video data. The chapter concludes with the exprimental results and conclusion of the methods developed.

4.1 Structure of Ball Games

Ball games have a history which goes back to ancient Chinese times and the game tsh-chu [Christopher Carling, Mark Williams, 2006]. At that time the goal of the game was to propel a ball into a net suspended between two poles. Another game was invented by Japanese culture where they has two players who try to keep the ball of the ground using their feet as long as possible. Later in the history the game developed by Italians as calcio and by the British as mob football. Those details about the evolution of the football is presented in more detail in [Christopher Carling, Mark Williams, 2006]. In such games, the goal was to move the ball behind predefined boundary points.

In the second half of the nineteenth century, institutionalisation of soccer was founded by the Football Association in England, where clear rules and standards are firstly defined. These rules are later adopted by European countries and then worldwide up to 1904 where the internationalization of the game formalized under a governing body of Fédération Internationale de Football Association (FIFA) [Christopher Carling, Mark Williams, 2006]. After mentioned regulations and standards of the game have been established, the attention of the audience in-

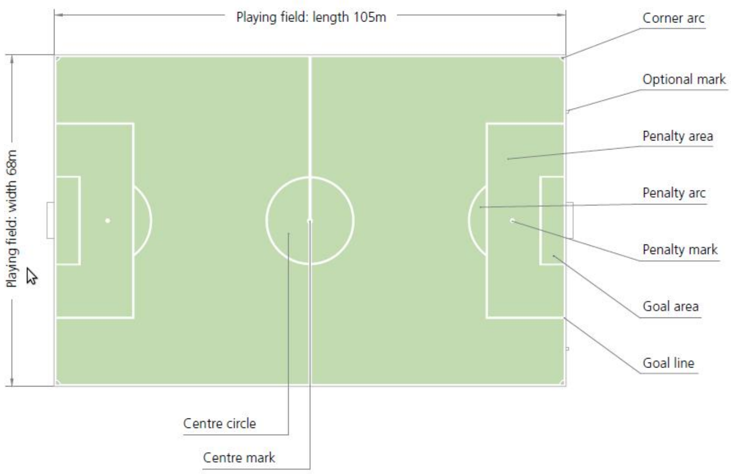


FIGURE 4.2. The outline of a standard pitch [FIFA, 2007].

creased more and more each decade. Since the last decade the formal computerized analysis of soccer games took attention with the development of the computing industry and increasing trend in the formal analysis of the game. Computers and video analysis methods assured the objective statistical and event based analysis to be more feasible.

Ball games have different structures than other games with respect to indefinite performance measures. In ball or puck games, the aim is to use the mentioned equipment in order to achieve the goal while belonging to the rules of the game. The movements of the ball or puck might be accomplished using the body parts (football, handball) or external tools (hockey sticks). Ball is exchanged between the players to achieve and pass beyond the goal area. Passing of the ball is done at high speeds in order to keep the possession of the ball and to propel the ball to the goal area as soon as possible. For this reason, the ball is usually in contact with the players and that makes the ball detection and tracking a more diffcult and challenging task to achieve. Players train a lot to make the passes quicker and more accurate while moving ball to the opponents goal area since they play in a limited time and in a predefined area. A typical soccer pitch with predefined areas indicated is as shown in figure 4.2. Additionally, the team who scored first gets the physicological advantage, positive feeling and motivation about their own tactics.

4.2 Challenges

Typical observation during a ball game is the players interacting with the ball. During the whole time of the game ball is in play and in focus of attention. Despite the higher quality hardware power and imaging devices, it might still be diffcult to get good quality of images due

FIGURE 4.3. Low quality frame examples.

FIGURE 4.4. Low quality and skipped frame examples.

to communication concerns, video formats and the broadcasting channel capacity. Therefore the first problem arises as preprocessing the image in order to enhance the features to be used in detection and to increase the measurement accuracy. Especially for the video information, some of the frames have higher quality but some of them are skipped or broadcasted at very lower quality due to communication issues. The examples of the some frames for broadcast soccer video is depicted in figures 4.3 and 4.4.

Ball is the main element of the soccer which is normally in spherical shape but however it might diverge from the standard shape and observed as blurred due to its rapid motions. During the game the ball state changes as it is hit, kicked or thrown by players. In other words, any contact with the players provides the ball another state.

Since the ball is the smallest element of the game, the explained data acquisition problems directly affect the performance of ball detection and tracking. Some of the ball shape examples are shown in figure 4.5. In those frames, it is observed that ball rarely appears in a perfect circular shape. Therefore the static features should be reinforced by additional static and dynamic features.

Many methods developed so far only focus on the color and shape features of the ball [Huang et al., 2007, Liang et al., 2005, Yu et al., 2003b]. In this thesis, we depend not only

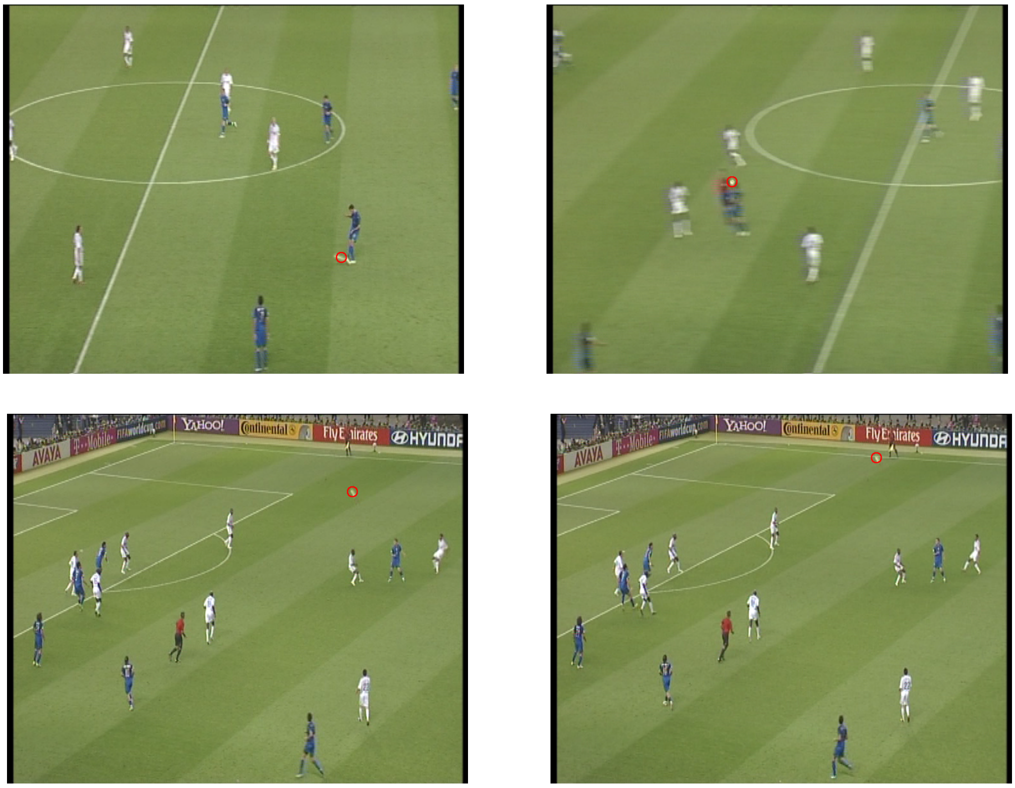


FIGURE 4.5. Ball shape deformations during the play.

on the color and shape features but also take into account the ball dynamics with its relative position and interaction with the players. The challenging problems investigated in this work mostly occur due to,

- fast camera movements
- lack of camera view
- fast ball kinematics, i.e. rapid and unpredictable acceleration
- small size
- occlusion by the players
- confusion with player parts and field lines

Therefore it makes the prediction of the ball position difficult to achieve since the measurements might also be misinterpreted due to deviations from the standard color and shape characteristics.

The main goal of this work is to analyze and summarize the soccer game based on the ball motions. However, for the further analysis, in order to extract the ball interactions, we need the player positions. Players have bigger sizes regarding the ball but they might still be difficult to detect due to

- fast camera movements
- lack of camera view
- occlusion by other players
- confusion with field lines

The details of the proposed player detection algorithm is explained in section 4.4.

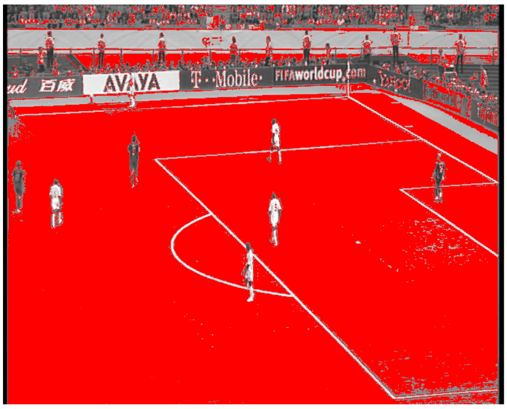
4.3 Segmentation of the Field

First step in the system development is the detection of the cuts and the filtering of the effect free frames in order to apply the image understanding algorithms developed in this thesis. For the filtering of those frames we rely on the system mentioned in [Gedikli, 2008]. The details of the system is explained in section 2.2

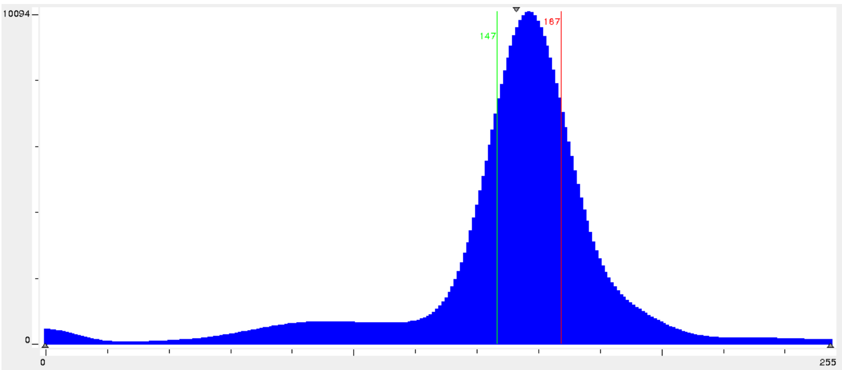
The field region is the biggest smooth area within a frame. The segmentation of the field is accomplished by using the dominant color distribution [Ekin et al., 2003, Xu and Shi, 2005a]. However, color intensities may vary from stadium to stadium which needs a training phase beforehand. The main assumption in the extraction of the dominant color region requires the camera mostly focusing on the field and thus the dominant color feature makes the pitch segmentation succesful. The associated color value is normally observed at different intensities of green channel. Therefore the first step in field segmentation is the decomposing the frame to its 3-channels in order to get the gray values of the green channel. In this step we used the peak value feature of the histogram as also proposed in [Ekin et al., 2003]. An example frame and its histogram distribution is shown in figure 4.6

However [Ekin et al., 2003] used a manual linear distance measure in their work for the dominant color distance measure threshold. In our method, we defined an automated method to define the field region. For the purpose of the automatic classification, the mean of the peak values of first 1000 frames are used to train for the mean peak value of the field color. This training phase is always available during the broadcast sports games since TV channels start the programs around half an hour before the game. In such programs they show the field and

(A)Original Image.



(B)Field Segmentation.



(C)Green channel histogram of the image.

FIGURE 4.6. An example frame and its corresponding histogram distribution filtered within $|\sigma|$ distance.

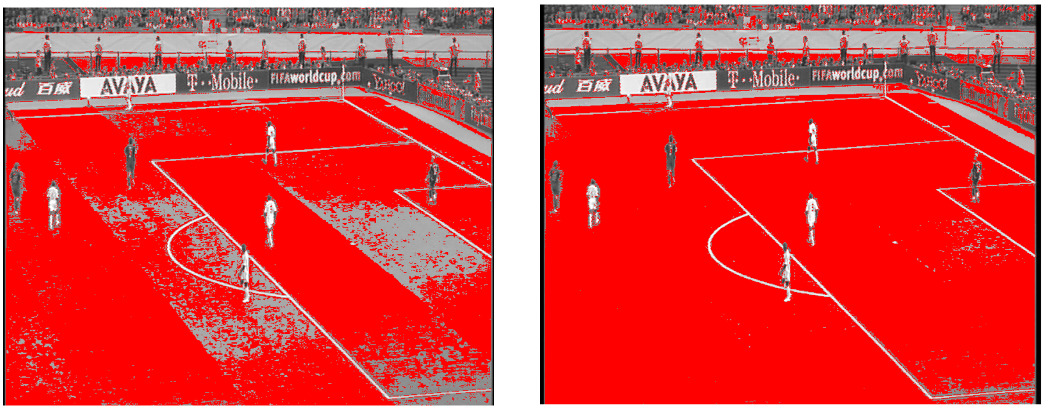


FIGURE 4.7. Normal distance measure and the Gaussian fitted σ distance measure results.

the training players while commentators discuss about the game. The learned peak value is computed by

$$m_p = \frac{\sum_{i=1}^{1000} G_{peak}(i)}{1000} \tag{4.1}$$

where m_p is the mean of the peak intensity values of the histogram and i is the frame number. $G_{peak}(i)$ is the gray value of the corresponding histogram's highest bin.

Then we fitted a Gaussian to the histogram of the green channel and filtered the pixels which is σ distance from the peak value. For the selection of the field regions, we defined an automatic distance measure with respect to m_p as

$$d_{intensitygreencomponent}(j) = |m_p - I(j)| \tag{4.2}$$

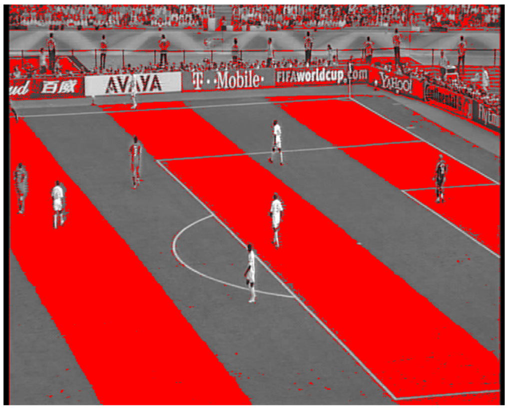
where $d_{intensitygreencomponent}$ is the distance measure of the corresponding intensity value $I(j)$ of the j_{th} pixel. The pixels that are filtered as field regions are selected as

$$d_{intensitygreencomponent} < |\sigma| \tag{4.3}$$

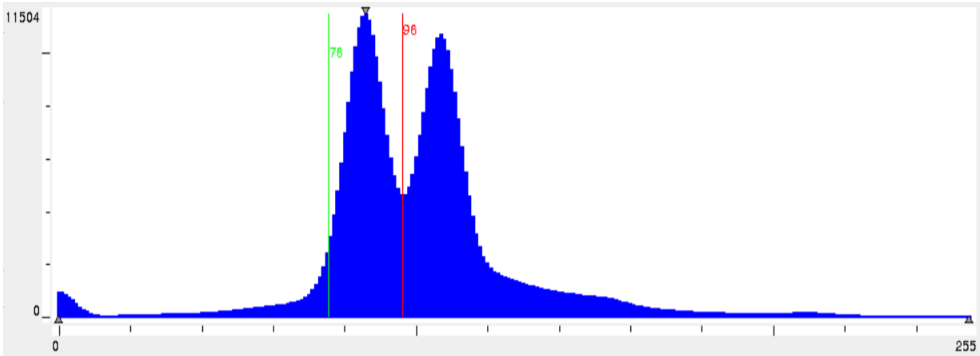
The mentioned Mahalanobis distance showed better results than normal linear distance calculation as shown in figure 4.7

The green color is seen as dark and light green in most of the football stadiums. This color difference happens due to different mowing directions. The stripes seen on the football field are made by blending the blades of grass in different directions during mowing. In addition to that, the direction that the grass is bent during mowing determines the light or dark colored stripe. When the blades of the gras are bent away from the observer, the grass appears lighter since more light is reflected. If the grass is bent away from the camera it appears lighter because

(A)Original Image.



(B)Darker Lawn Region Segmentation.



(C)Darker Lawn Region Filtered in the histogram of the blue channel.

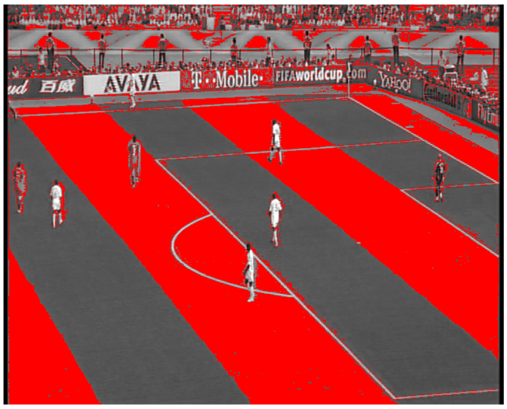
FIGURE 4.8. The blue channel histogram filtering and darker field region segmentation of two lawn regions.

the light reflects off the wide lengthy part of the blade. If the grass is bent away towards the camera it appears darker because the tips of the blades are seen and shadows occur under the grass. Finally, as a result the color of the grass region depends on the direction it is mowed and observed.

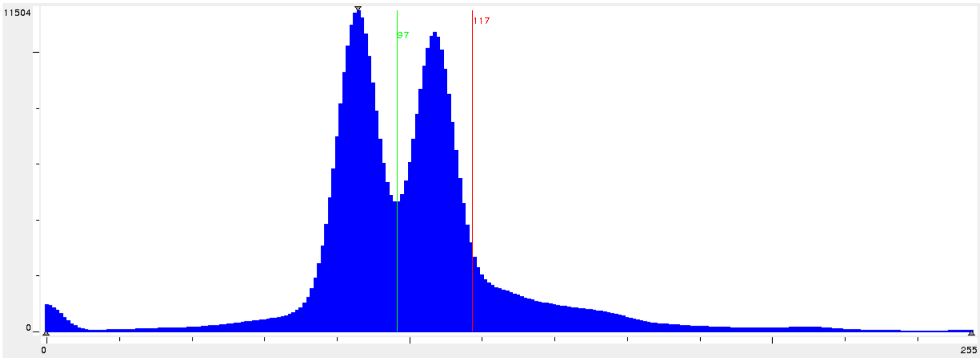
As a result of our experiments, we also found out a characteristic feature in the blue channel of the corresponding soccer broadcast videos. The observed feature was that as the field is composed of light and dark green regions, they form the two peak bins in the histogram of the blue channel. An example of that case is as shown in figures 4.8 and 4.9.

Two peaks in the blue channel histogram indicate the light and dark parts of the grass regions. We seperated those two peak regions from their middle point and did the same Gaussian fitting procedure and automatic distance measure as explained above for the green channel. The training in order to get the peak values of histogram bins are done in a similar manner for the same first 1000 frames. But the difference in the blue channel is that the largest two values are searched and the mean values are found between the smaller and the larger values. The

(A)Original Image.



(B)Darker Lawn Region Segmentation.



(C)Lighter Lawn Region Filtered in the histogram of the blue channel.

FIGURE 4.9. The blue channel histogram filtering and lighter field region segmentation of two lawn regions.

computation is evaluated as

$$\begin{aligned} L_p &= \frac{\sum_{i=1}^{1000} G_{largerpeak}(i)}{1000} \\ S_p &= \frac{\sum_{i=1}^{1000} G_{smallerpeak}(i)}{1000} \end{aligned} \tag{4.4}$$

In above equations, L_p and S_p are the mean value of the larger and smaller of two highest bins of the histogram, respectively. Similarly, $G_{largerpeak}(i)$ and $G_{smallerpeak}(i)$ represent the larger and smaller of two peak values of the histogram of i_{th} frame.

We then investigated the combination of the two segmented regions by extracting the intersection of the biggest region in the green channel with the union of the detected grass regions in the blue channel. Here the biggest region detected in the first method is unified with the grass regions detected using the second method. The resulting field region is used as a mask to run the player detection as well as ball detection and tracking.

Segmentation of the field regions decreases the noise and clutters within the image especially the regions where spectators have unexpected rapid motions. For the case of player detection, it also increases the accuracy and decreases the computational costs since we only focus our interest on the field regions. Player detection problems sometimes occur when the players are around the field borders. We handle this problem by decreasing the player size constraints around the borders. If a player region candidate is found around the border by using color feature but the size constraint is not satisfied, we still process the mentioned region as player region.

If players stand still on the border, it is not a problem except the field lines. Therefore the field region is extended using the dilation method in order to get a secure and stable region for detection and tracking. This is employed to be sure that the players and the ball lies within the region of interest (ROI).

Ball is the fastest moving element of the game, and it is not only in rolling situations when it is played but it might also be flying when it is played in the air. Ball is not only played as passes on the ground but it might also be passed through the air. For that reason, ball might not be found on the field region during the detection or tracking. This is one of the main problems discussed in this thesis and the solution proposed is explained in detail in sections 4.6 to 4.7.

4.4 Player Segmentation

Player regions need to be segmented and detected in order to localize the identities and extract the actions regarding ball player interaction. Players can be classified as the most important identities of the game after the ball. Therefore the localization of the players is inevitable for the higher level analysis and recognition of the game. Players are easier to classify regarding the ball in means of size and motion. They cannot move as rapid as the ball and have bigger sizes. Their smooth motions and bigger regions makes it more feasible to detect than the ball. However tracking of the players is still an interesting topic studied in many other studies [Gengembre and Pérez, 2008, Hoyningen-Huene, 2011, Kataoka et al., 2011, Sato and Aggarwal, 2005]. In this thesis, we focus on the abstraction and summarization of the game focusing on ball detection and tracking. Therefore the tracking of the players is beyond the scope of this thesis. For ball player interaction scenarios, we only detected the player positions without any further tracking methods. However a detailed study about player tracking in soccer games can be found in [Hoyningen-Huene, 2011] which was also developed under the ASPOGAMO project.

In this thesis, player detection has been developed as a separate module and the positions are fed to the ball action recognition system for higher level analysis. In the first step of player detection, the image is transferred from an arbitrary color space to the RGB color space. In field games, the field background forms a smooth enough area for the size of players and the player jerseys are always different than the smooth field green region.

In our approach we used the motion cues with the color deviations in order to detect the players. In the previous section 4.3 the field detection is explained in detail. Player detection is based on the region of interest (ROI) classified as field region. It always makes sense to search the players within that region since the game is always played on the field. This filtering also reduces the noise related to the spectators and other objects around the field. The following operations are all done within the mentioned ROI which is the field region for soccer games.

Player jerseys have different color values than the field color and players are mostly in action in order to grab the ball or position according to their tactical line up. In order to extract the player regions, we have used the standard deviation of the green channel gray value distribution. A rectangular mask has been chosen for this task since player regions are found in rectangular shape forms.

In our experiments, we have tested many mask sizes for different videos and different situations through the game. As a result of those experiments, we have found out that a $5x5$ rectangular mask provided better results. Therefore a $5x5$ mask has been chosen as a deviation mask experimentally. The resulting deviation image is depicted in figure 4.10.

As seen in figure 4.10 the values the range of gray values does not seem to be distributed at a larger range due to the size of the mask chosen. Using a larger size mask might generate gray values distributed at a better range but that would lead to misclassifications. In order to solve this problem, we have multiplied the resulting gray values by 2 to have a better distribution of gray values as shown in figure 4.10. This allowed us to be able to classify the player regions easily.

Since we have frames of size $720x576$ they are changed to the next odd values, namely, $721x577$. Additonally, we have also mirrored the gray values at the image borders.

After getting the deviation image we have used the hysteresis threshold [Canny, 1983] in order to extract the player regions on the field. The idea this work is to simply classify image points in three classes. The values higher than the upper threshold value are directly accepted and the values less than the lower threshold value are automatically rejected. The values that fall in between those values are called as potential points.

Potential points are classified according to their distance to secure points. For player detection we have chosen a distance measure of 20 pixels in our experiments. We have chosen this value because the distance between the players is usually more than 20 pixels and with a 20

FIGURE 4.10. Deviation image of a sample frame using 5x5.

pixel distance we guarantee that we have all the player parts segmented. We use that influence of secure points to strengthen our classification accuracy and thus the detection rate. The result of hysteresis threshold is depicted in figure 4.11

After the hysteresis threshold, the problem is to determine the regions connected in order to extract the player borders. We have used an 8-neighborhood connected components method in order to determine the unified regions. 8-neighborhood provided accurate results for the foreground segmentation.

In this step, the field lines might be problematic since they have high deviation values from the field green. In this case we rely on the field model developed by [Gedikli, 2008] and we subtract the detected field lines from the current frame. The mentioned operation results are depicted in figure 4.12.

The remaining regions are classified using the size constraints. The players are sometimes confused with the field lines and they are detected as two smaller sperate regions depending on their position. We solve this problem automatically by detecting a mean size of the detected regions and unifying the regions smaller than half of the mean value of the detected regions. An example scene is shown in figure 4.13

Another problem we faced here was the association of the regions. For that purpose we

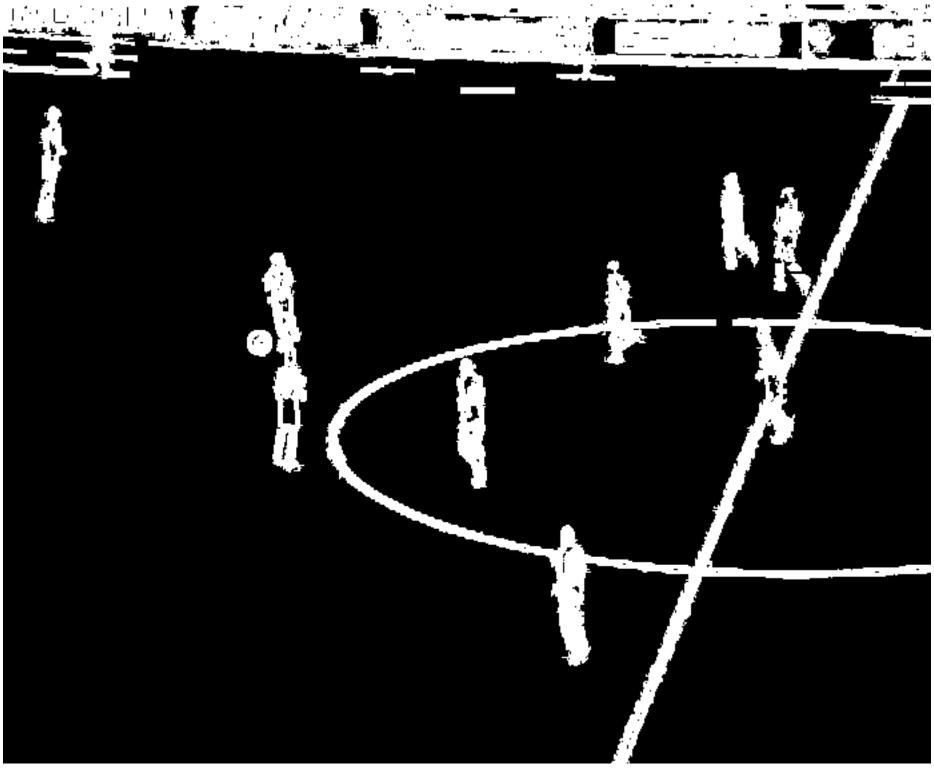


FIGURE 4.11. Segmented Regions.

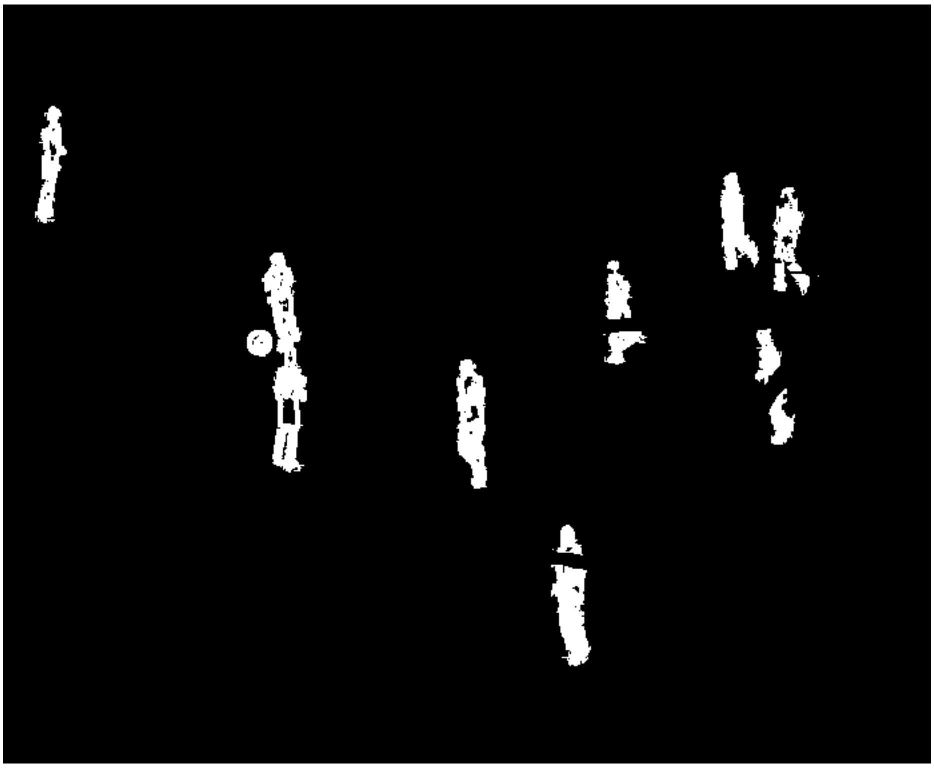


FIGURE 4.12. Image and field lines subtracted.

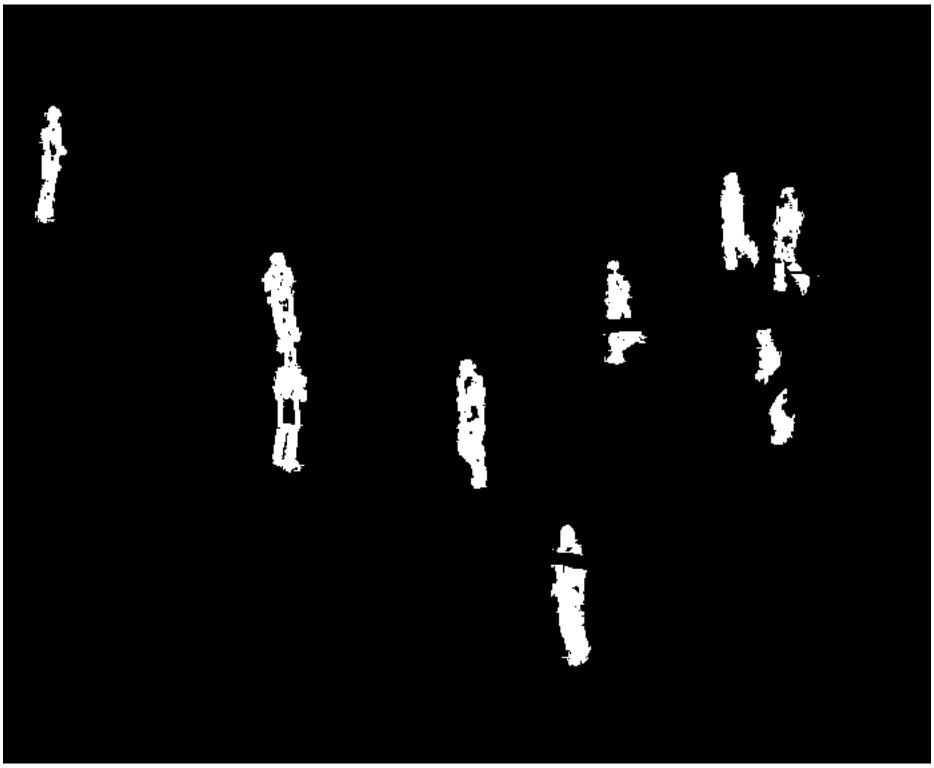


FIGURE 4.13. Segmented regions including small falsely detected regions.

unified the regions according to a dynamically defined distance measures. The dynamically defined distance measures refers to the connection of the regions that are in the neighborhood of half of their width or height. The larger of width and height features of a region is taken into account while determining the distance measure for small regions.

If all the player regions are segmented at their full size appearances than this step is skipped as explained in the algorithm. The resulting detected players are shown in figure 4.14.

In our player detection algorithm, we assume a specific size and rectangularity range at the beginning. Since the players are mostly in standing position during the game the rectangularity stays in a predetermined range.

The player regions are saved with their corresponding positions as (x_{ij}, y_{ij}) where i stands for the corresponding player region index and j is the frame number. In this part no recognition or other classification is carried out other than detection since the detected player positions are enough for ball player interaction analysis. Players are detected without classifying their teams or tracking them since our work focuses on extracting the ball actions rather than tracking the players or identifying teams. We focus on the summarization of the game regarding the ball as the central object of actions and other object interacting with it. Our method includes a novel approach by investigating and modeling the game based on the ball actions.

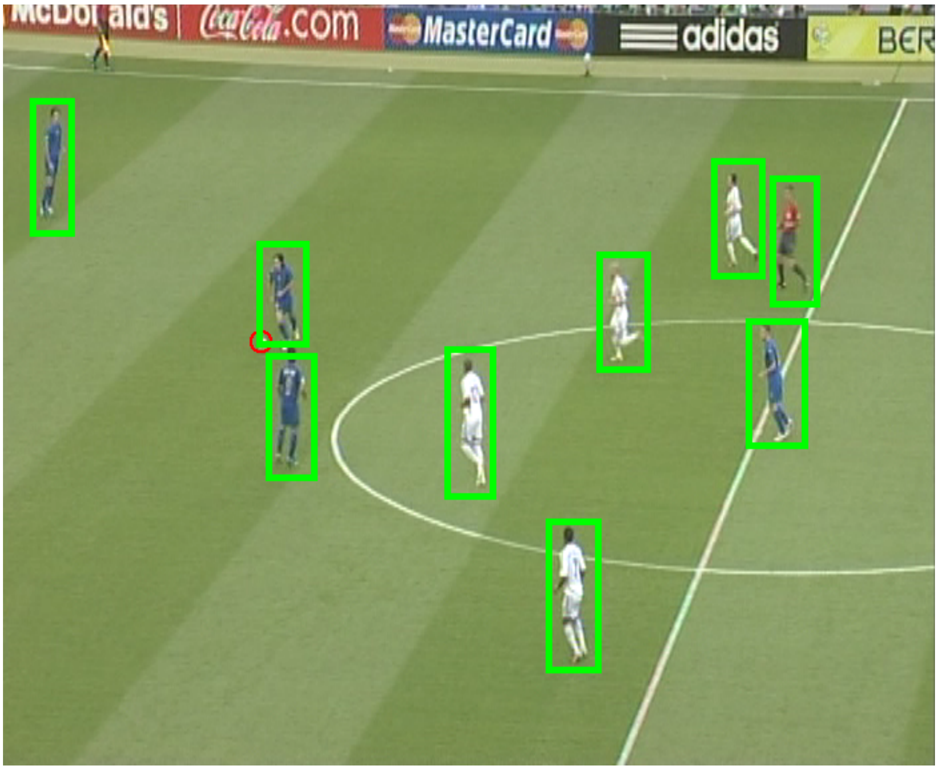


FIGURE 4.14. Association of falsely segmented regions and the results of the detection.

4.5 Ball Feature Extraction

Detection and tracking of the ball requires the determination and the extraction of the features of the ball. For ball feature extraction we used the static and dynamic features. The static features of the ball depend on the shape and color characteristics of the ball where the dynamic features represent the ball motions and kinematics. A soccer ball is an inflated sphere used to play soccer. All the standard FIFA authorized soccer balls are tested for the following specifications [FIFA, 2013];

- consistent circumference
- permanent roundness
- consistent rebound
- minimal water absorbption
- weight
- minimal pressure loss

FIGURE 4.15. An example of Adidas©Jabulani.

- shape and size retention
- balance

The standards are very strict about the ball specifications and its duration through the whole game since even a very small defect on the ball might affect the situations occured during the game. An example of a single standard soccer ball is as shown in figure 4.15. The ball in figure 4.15 was used as the official match ball of the 2010 FIFA World Cup. Ball circumference $69cm \pm 0.5cm$, weights $420 - 445g$, inflated to a pressure of $0.6 - 1.1$ atmospheres, and covered in leather.

Knowing the above standards, firstly we treate the ball as a single still object, we observe the color and shape features of the object. The first feature of the soccer ball is that it is consistently a round object. As it should provide the above standards it should normally expected to be a round object. Because of the motion of the ball and data acquisition systems we can very rarely observe the ball in a perfect round shape. This problem is discussed in detail in 4.6. Here we focus on the general features that might be used to recognize the ball.

The second feature is the color of the ball. Before the development of leather production and lighting techniques, the ball was colored in dark colors. After the advances in the floodlights and the leather production, the white ball is permitted to be used to help the spectators see the ball easier. Nowadays the color of the soccer ball is almost always white. Most of the ball types authorized by FIFA are colored white. They also have some figures and text on the data but for the broadcast soccer cameras it is very difficult to recognize those text and the ball is seen as white. Exceptionally in very rare cases, where the field becomes white due to heavy snow, the ball might be in orange color.

Additionally ball is the smallest element of the game on the field. As this might be an advantage while filtering the ball candidate regions but however is usually becomes a disadvantage since ball might be confused with man objects around because of its size. Some examples of this cases are the field lines, penalty shooting point, players parts etc. Some example cases of various ball appearances are shown in figure 4.5

The static features of the ball might not be always observed due to the rapid ball motions, quick movements of the broadcast camera. Thus we also employed dynamic features of the ball in association with the static features. Fortunately, ball is almost always in motion and it has a constant velocity while being played between the players. Also if it is not played between the players it is always in possession of the players or out of the game depending on the situations. We have additionally employed that kinematics in order to recognize the ball during the game. Those features can be also observed in example frames of figure 4.5

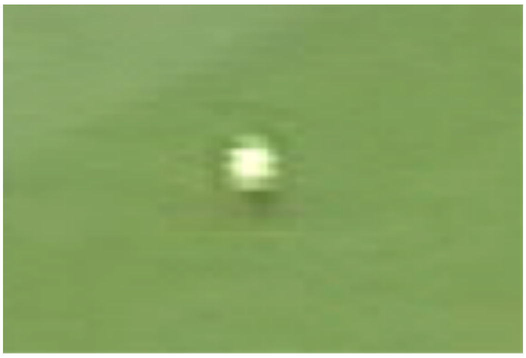
4.6 Ball Detection

Ball detection is usually very difficult to achieve since cameras are always moving and their settings are always adjusted according to those movements. As one can obviously realize that the ball is very small in size regarding the other objects within the frame and it is mostly occluded by players during the game since within the most time of the game the ball is at the foot of the players. Another issue is also that when the ball is kicked high in the air it is out of the field of view of the camera and as it is usually moving at high speed and confused with the parts of players and line markings. This rapid and unpredictable acceleration also makes the detection of the ball a challenging task. In the next two sections the segmentation of the ball using static features with the filtering of the ball candidate regions are discussed and presented.

4.6.1 Segmentation of Ball Regions

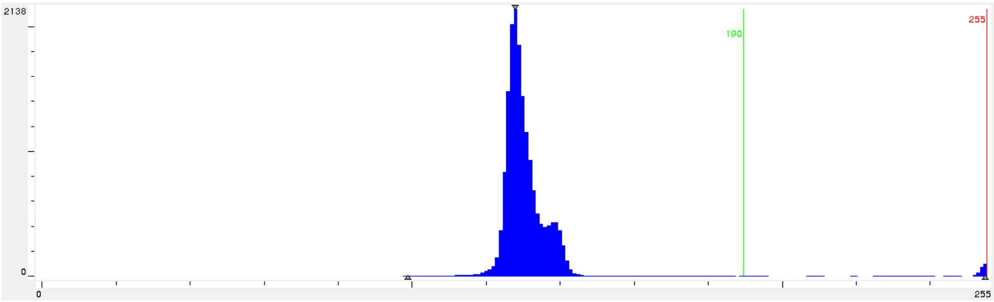
In the literature, the segmentation of the ball regions are implemented using only the static features [Hossein-Khani et al., 2011, Pei et al., 2009]. Generally shape and color features are employed for segmenting the ball regions through the game. However the fusion of that features is not investigated extensively so far in the current litearture.

In our method, the color feature of the ball is the firstly searched feature during the detection. Ball is usually colored white and even if it is very small and hardly seen the color feature still remains a discriminative feature. Color distributions of some ball sample regions are shown in figure 4.16



(A)Original Image (zoomed).

(B)Ball Region Segmentation (zoomed).



(C)Green channel histogram of the original image.

FIGURE 4.16. Ball color distribution example and filtering regions based on color.

In a soccer video game frame white colored regions are the ball, field lines and markings and sometimes some parts of the players. Based on that general information, for the first step of the segmentation we have utilized the color feature and employed a simple filtering on the frame. Those filter provided the candidate regions based on the color feature. Through our experiments we have observed the most of the false positives happening at the field lines and player parts.

Therefore we then filtered the ball candidates based on the size constraints of the detected regions. The size of a sample ball is entered to the system manually in the offline phase while training the ball templates. The ball template generation is explained in detail in the ball tracking section. The diameter of a standard size soccer ball is around 22 cm which stays constant through the game. It might be also feasible to use the size of the players as a reference to detect the ball based on the assumption on the size of the players since player sizes are easy to extract.

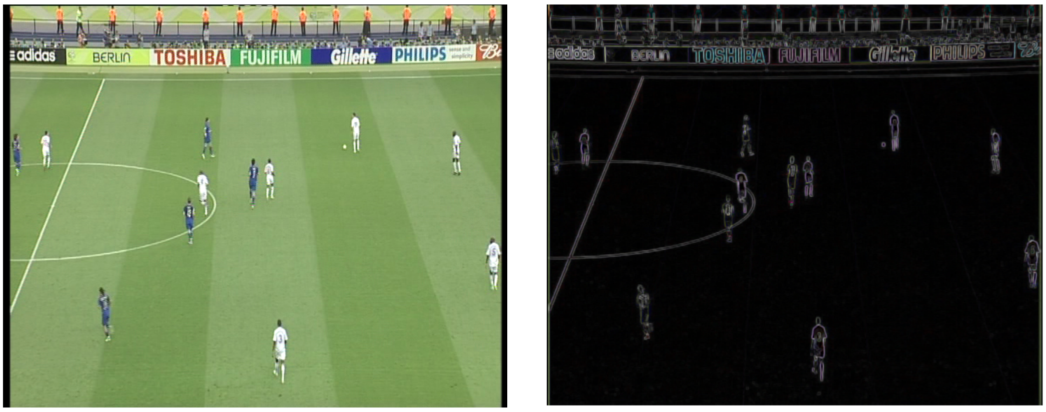
Afterwards we filtered out the non-ball regions based on their size or in other words filtered the ball region candidates. For that purpose we applied the shape constraints and filtered the output of the size based filtering. The measure we use for the ball region detection is the most characteristic feature of its shape, namely the circularity given by

$$
\begin{aligned}
f_{circ} &= \frac{\mu_c}{\sigma_c} \\
\mu_c &= \frac{1}{N}\sum_{i=1}^{N-1} \parallel (x_i, y_i) - (\bar{x}, \bar{y}) \parallel \\
\sigma_c &= \frac{1}{N}\sum_{i=1}^{N-1} (\parallel (x_i, y_i) - (\bar{x}, \bar{y}) \parallel - \mu_c)^2
\end{aligned}
\tag{4.5}
$$

where f_{circ} is the circularity, N is the number of contour pixels, (x_i, y_i) are the positions of the pixels on the contour and $(\bar{x}, \bar{y})$ are the center points of the region.

As mentioned above, most of the problems we encountered were in confusion of ball candidates with player parts and field lines. We solved the first problem by filtering out the candidates in the neighborhood of the players. In other words, if a ball candidate is found in the same region with a player it is filtered out and that region is classified as player part. This might be interpreted as ignoring the situations when the player has the ball possession but in most cases even when the player has possession of the ball is mostly apart from the player himself. Of course here the cases when the ball is occluded by player is missed but according to our observations and game statistics we take into account those costs. Additionally, the philosophy of the detection is the extraction of those particular ball information from a large frame without specific cooperation which means the simple presence based on the color and shape features as mentioned above. The confusion with the lines are eliminated using the shape features since the line parts are classified as long rectangular regions.

In this thesis, to deal with confusion problems and to increase the accuracy of the detection, different than the state of the art methods we included the ball dynamics in addition to the shape and color characteristics of the ball. To detect the ball motions, standard deviation of gray values in the green channel of the image is employed in association with the above explained color and shape constraints. Since ball is mostly in motion during the game, the deviation of the pixels gives us a reliable information in detecting the ball. The standard deviation is calculated using a window based approach. The window is employed around the last detection and if no candidate is found then the detection is run for the whole field region. So for the region of interest for the detection two states are defined, one based on the previous detection and if no reliable candidate is found within that window another search within a whole frame in employed. If no detection found using the two search states, then the frame is skipped and the search in continued in the next frame with the whole region.



(A)Original Image. (B)Deviation Image.

FIGURE 4.17. standard deviation images based on a $3x3$ mask.

The standard deviation of the pixel values are computed within a rectangular mask of size $3x3$ pixels similar to the method employed for player detection. As the ball is moving at higher speed and has smaller size it has higher deviation than the players and other objects which was the reason for employing a smaller mask. To highlight the ball regions and have better use of the range of gray values available in the output image, the result is multiplied by 2 and at the image borders the gray values are mirrored. Here the employment of the smaller mask also assures the ball size constraints mentioned as the second feature for the detection of the ball. A bigger mask could result in missing the ball regions since it is very small and the deviation within a bigger mask cannot provide us the influence by such a small sized ball. The results of the standard deviation based search is depicted in figure 4.17.

4.6.2 Filtering of the Ball Candidates

In the next phase, we employed a rule based classification method in order to filter the ball candidate regions. The classification is done by the fusion of the three separate arrays associated to ball candidate regions. Those three arrays are generated by using the aforementioned ball detection features, namely color, shape and standard deviation. The fusion of the ball candidate region information is employed using a weight function defined for each corresponding candidate region. Based on the constraints, each corresponding candidate is then associated a weight. The weight function is defined as linear function which depends on the results of the three ball features examined. Each feature fulfilled by the ball candidate adds a value of 1 to the corresponding region.

If a candidate region satisfies all color, shape and standard deviation constraints then it has a weight measure of 3 and classified as a reliable ball region. The process is then moved to the next frame.

If a candidate region satisfies only color and shape constraints then it has a weight measure of 2. It means that the ball is not moving but staying at constant position which usually happens before free kick positions.

If a candidate region satisfies only color and standard deviation constraints then it has a weight measure of 2. This also happens when the ball is kicked fast by a player or if the camera is moving. Because when the ball is kicked hard or if the camera is moving fast the ball appears more in an elliptical shape rather than circular.

An image region satisfying shape and standard deviation constraints is not observed during the game. This case does not happen because when the deviation is usually high, meaning the ball is in motion, the circularity is then no more satisfied due to disturbances.

Additionally, searching the whole frame has computationally very high costs. For this reason, we proposed a dynamic, search window employed detection based on the previous position detection of the ball and the result on the standard deviation image. The search windows are converged around both positions. If a ball position is associated a weight of 3 then it is classified as a reliable candidate and taken into account while defining the search window.

If no ball candidate is found within the frame, we then move to the next frame by increasing the search window to whole frame till a reasonable candidate is found. When a reasonable candidate is found then we reduce the search window again and continue the ball detection loop. The block diagram the process is shown in figure 4.18

In our experiments, we used a search window of size $100x100$ pixels. This led to the decreased processing time and reduced the computational costs as expected.

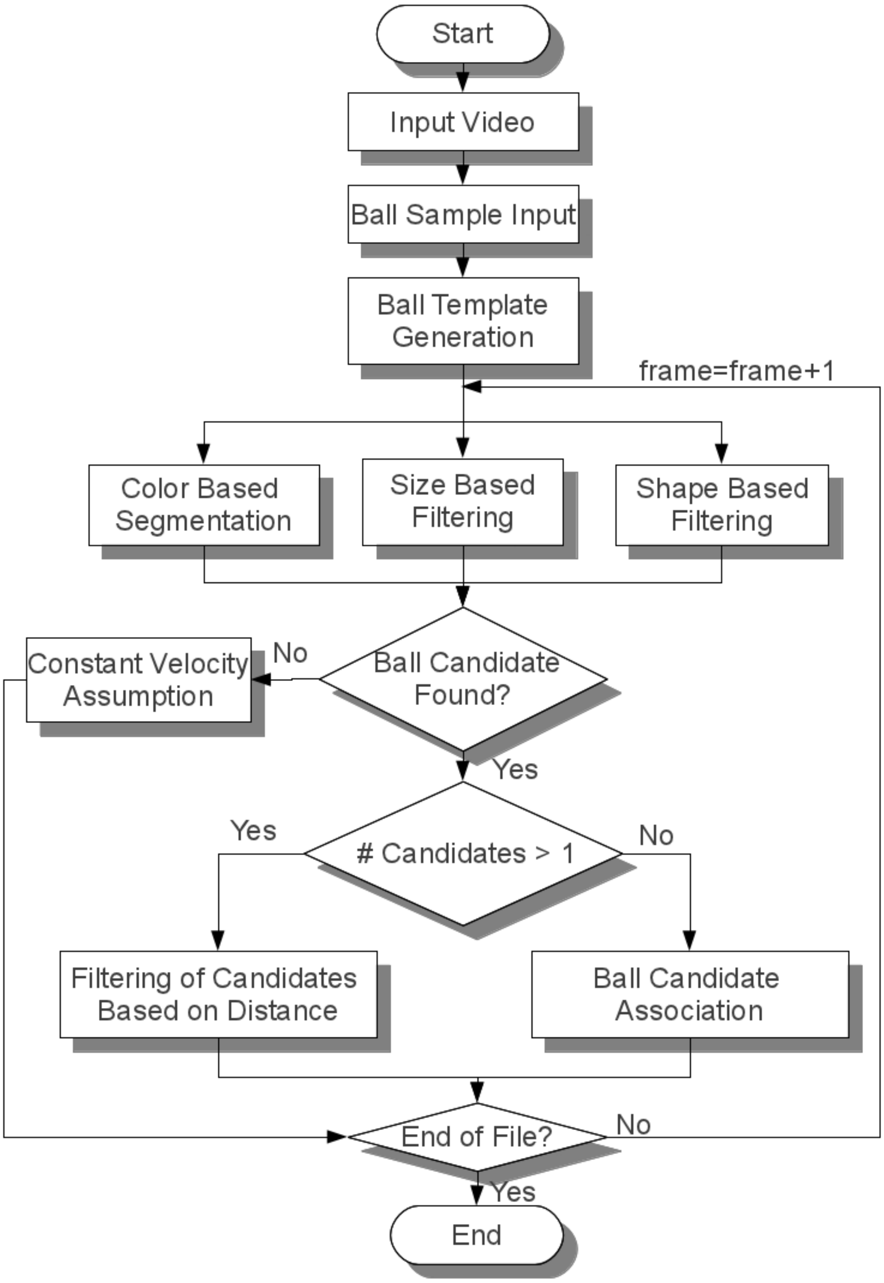


FIGURE 4.18. Ball Detection Algorithm.

4.7 Ball Tracking

The particle filter is a Sequential Monte Carlo (SMC) method that can handle more sophisticated models than other methods [Arulampalam et al., 2002]. For the purpose of ball tracking, we implemented a novel SMC approach which represents a recursive estimator of the ball position. In other words, the particle filter method proposed approximates the posterior over all ball candidates. In this method, previous estimate of the ball position is further moved to the next measurement scan by prediction of the location according to a motion model. Particles are distributed between the prediction and the current measurements proportional to the assigned former weights. The weights are determined by the frequency of the samples that resulted in the same association. The tracking method proposed follows the position of the ball in the image domain. A measurement is assigned to the target is analytically limited in order to prevent over sampling. The algorithm is linear in the number of particles and measurements.

Small and single target tracking approach is proposed and developed through this thesis along with the investigation of dynamic number of particles. As the frequency of the samples is increased, we began to decrease the number of particles to another different predetermined level. This approach leads to improvement of the accuracy and reduced computational costs. In other words, if the frequency of the samples increases more than a specific value, we then decrease the number of particles to another predefined level to in order to reduce the computational costs. In other words, we can sufficiently state the correctness of the measurements then we reduce the number of particles distributed. When it becomes again the case that the confidence of the measurements decreases then we began to increase the number of particles used in order to improve the accuracy of the tracker.

The tracking of the ball is implemented in every frame following an extended version of particle filter framework [Arulampalam et al., 2002]. We have extended the usual particle filter according to the ball dynamics. The novelties of the algorithm developed in this thesis lies in the resampling of the particles following the ball dynamics. This approach can be generalized for most of the other small object tracking applications where the kinematics might be predicted.

The novelties of the proposed algorithm include beyond the state of the art resampling methods developed through the sports games, especially soccer, for identity tracking. Our approach is built upon the design of a tracker regarding the target characteristics. The target characteristics do not only refer to the kinematics but also to the interaction of the target with the objects around it. Since soccer games are played in a controlled environment according to predefined rules, the soccer broadcast video analysis domain provides a suitable application domain for

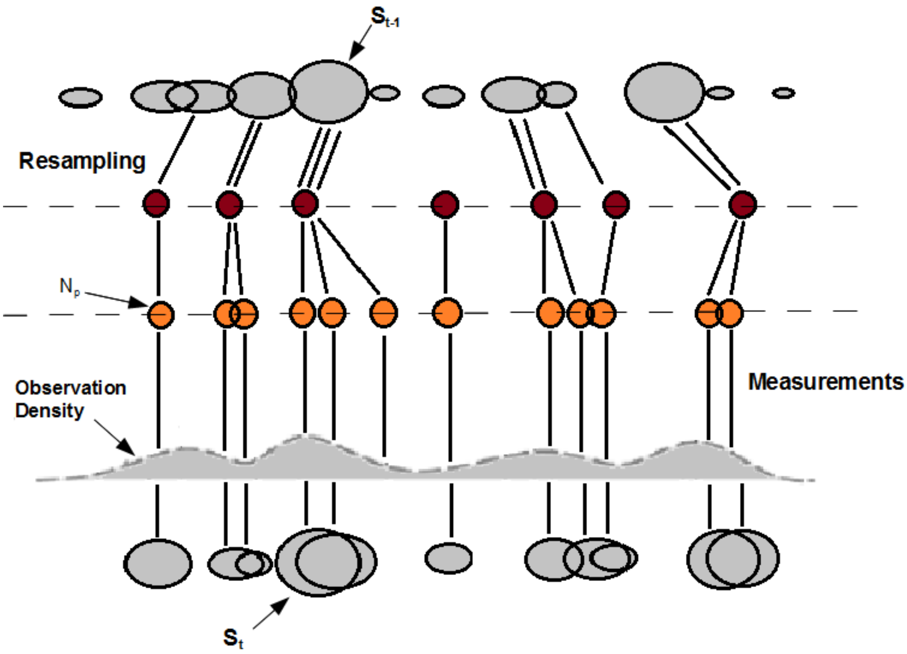


FIGURE 4.19. Particle Filter at a Glance.

the implementation and realization of our approach. Our approach can be generalized to most of the other small object tracking applications in sports domain where the kinematics might be predicted. One example for the generalization of the idea is depicted with the experiments on field hockey videos. Since field hockey is also a team sport with a predefined structure we have tested the algorithm on a game from Bridgestone World Series Hockey 2012. The experimental details are presented in the next sections with the acquired results.

4.7.1 Assumptions

A typical soccer the game is always played with one ball, meaning a single target to track. The visibility of the ball changes through the game due to its speed, interaction with players and rapid camera movements. Ball is replaced a few times during the game when it is kicked outside of the field. If the ball is kicked far away from the field it is immediately replaced by a new ball in order to provide the continuity of the game. Even if the ball is replaced a few times during the game, it always retains the standard characteristics as explained in section 4.5. By standard characteristics, other than the static ball features we also refer to the way the ball is played and the general ball kinematics. Therefore the way how the ball is played and the general movements of the ball stay constant during the game.

Even though the ball represents a rapidly moving identity in the game, the temporal velocity of the ball stays constant through the game. The ball usually moves rapidly but the path it might travel in 40 milliseconds (for a typical broadcast video of 25 frames per second) is very limited in distance. The fastest recorded speed of a soccer ball so far is around 200 km/h which makes the ball move around 2 meters maximum distance per frame. Additionally, the fastest movements of the ball occur during shots and shots occur very rarely around the goal areas which rarely happen in most of the time during a game. Regarding the characteristic of the game, the ball has rapid direction and acceleration changes but when it is played between the players normally the ball speed stays constant. Even if the ball speed is quite high between the players, it stays almost constant at this passing time interval. And also if the ball is in possession of a player the speed becomes relatively very low. For all the mentioned reasons, a linear motion model is chosen in our approach for the modeling of the ball kinematics. Even though actual sport games include lots of unexpected motions, for a broadcast video data, the linear and constant velocity model is a sufficient approach.

In the proposed method, ball states are expressed in position, velocity and acceleration. This is the common practice in tracking the single targets. Additionally, the estimation is done on the two-dimensional position on the playing pitch since the 3D position of the ball can hardly be extracted from sports videos. Also those 3D estimation results would not be precise enough regarding the size of the ball.

Our approach is explained for a motion model with a constant temporal velocity. Even if it might not be true for complex cases, for most of the situations the assumption holds true. As explained above, the ball can travel at most 2 meters per frame. In cases where the proposed algorithm might not be enough, the motion model can easily be extended to a nonlinear motion model.

The measurements are observed once each time step and the confidence measure for each measurement is used in order to detect the correctness of the measurements. The confidence measure is computed by a template matching algorithm. Measurements are modeled to be independent from other measurements of the same time step. For the association, we state that the multiple measurements might result from one target candidate as well as a single measurement resulting from a single target.

4.7.2 State Space

As explained in section 3.7, the ball tracks are described by hypothetical ball states, x_k^i, including position, velocity and acceleration. The current state is updated according to the ball dynamics assuming a constant velocity motion model. Here x_k^i is a state and associated with non-negative numerical factors w_k^i called importantance weights, which sum up to 1 . Briefly, at each time step, the particle filter receives a sample set S_{t-1} representing the previous posterior of the ball states and the measurement z_k. Then N_p samples are generated to represent the posterior.

For the ball the state model might be written as

$$m_{k,j} = (x_{pos}, y_{pos}, x_{pos}\prime, y_{pos}\prime, x_{pos}\prime\prime, y_{pos}\prime\prime) \tag{4.6}$$

In this model, the constant velocity consequently returns acceleration assumed to be 0 which then leads to

$$m_{k,j} = (x_{pos}, y_{pos}, x_{pos}\prime, y_{pos}\prime) \tag{4.7}$$

The posterior of the ball state is approximated by following the particle filter equation 3.69 for the set of N_p samples.

4.7.3 Resampling Step

First step in generating samples is the resampling which refers to drawing the state from previous posterior. In the basic particle filter approach, the sample is drawn with probability according to the sample weight. Most of the methods apply the minimal variance for resampling known also as the deterministic selection [Arulampalam et al., 2002, Kitagawa, 1996] or stochastic universal sampling [Baker, 1987]. In this step, the main problem is to decide which sample to draw from the previous set. In the typical particle filter approach, the common method is to draw the samples using residual resampling [Liu and Chen, 1998]. In this method, the samples are drawn only according to their weights and replicated deterministically given

$$N_i = [w_{k-1}^i N_p] \tag{4.8}$$

Here the particles with higher weights are sampled many times and the particles that have negligible weights are discarded. We have also used a variable number of particles after the resampling as also stated in [Hoyningen-Huene, 2011]. This method is not commonly used in the literature but on the other hand there is no limitation for the number of particles used at each time step. After number of particles is adjusted according to the scene, the weights reinitialized at identical values. In addition to the weights we also use the confidence measure to determine the number of samples drawn as depicted in equation 4.9.

The novelties of our method improve the resampling by dynamic number of particles and the dynamic search space. In this approach, the number of particles is a dynamic variable and it is optimized according to the confidence level of the measurements and resampling rate. The dynamic search space represents an interest region for measurements of a target based on the target and environment characteristics. In the typical approach [Arulampalam et al., 2002], the number of samples always stays constant. However it can be deduced that if the confidence level of the measurements are high enough then we do not need to sample all particles or in case the observations are not very reliable then we can increase the number of particles in the resampling phase. In our method, we propose a confidence measure C, as follows in equation 4.9, in order determine the number of particles resampled dynamically. Here a Gaussian distribution is fit to the set of values to be resampled and the variance is used as a confidence variable. For the ball tracking application, we exploited three particle levels of 50, 100 and 200 particles during the sampling. To define the number of particles to be resampled, we depend on the variance of the previous samples and their Euclidean distance to the previous ball state position. The confidence variable depending on the variance of the samples is defined as

$$\frac{1}{C} \sim \sigma_k = \frac{\sum_{i=1}^{N_p}(x_{k-1}^i - \overline{x_{k-1}})}{N_p} \tag{4.9}$$

If the variance is higher than the predefined threshold it means that there are more than one strong hypothesis and therefore we increase the number of particles sampled in order to get better observations. While increasing the samples we take into account the distance to the previous time step and more new samples are drawn from the candidate which has smaller Euclidean distance to the previous state. On the other hand if we have a smaller variance and distance to the previous state, we then began to decrease the number of particles used in sampling. This might sound to the reader that the method might suffer from the degeneracy phenomenon but we overcome this problem by defining a limit of samples drawn for each value and using the Euclidean distance. If one of the measures, confidence or distance, is not satisfied then the samples are reinitialized.

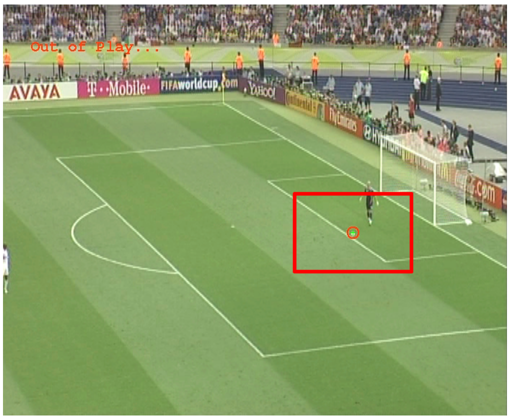
In addition to dynamic number of particles, the other contribution of our work is the smart window based approach for tracking the ball dynamics. In soccer ball tracking, the target is very small and does not provide us many textural or shape information throughout the game period. Therefore we propose a novel window based resampling method which is an efficient search approach to accurately track the kinematics of the target. By using the confidence measure defined in equation 4.9 we adjust the size of the search window around the target dynamically. Ball is played between the players and stays on the field throughout the game. As explained above, the maximum distance it can travel in one frame is around 2 meters which only happens during shooting. Based on our assumptions about the nature of the target, we also know that the ball is mostly played by the players and it is found on the field.

If no reliable measurement can be gathered through a frame we then increase the size of the search window. This process is continued in a pyramidal approach since a reliable measurement can be observed from the target with a bigger confidence level. In some cases when the goalkeepers restart the game with long shots, the ball flies for long time and usually out of the view. Especially in such situations, the ball search window is extended to the size of the image until the samples began to converge to a reliable value. An examples scenario is illustrated in figure 4.20.

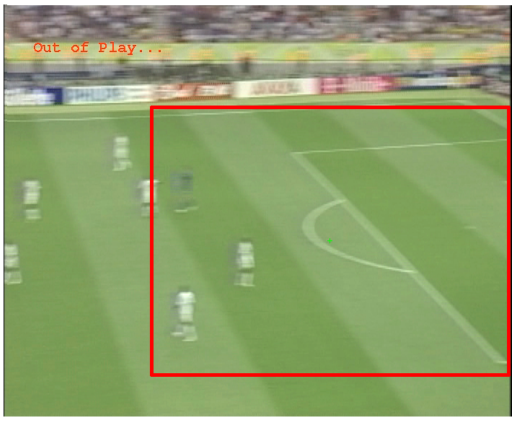
Enlarging and the shrinking of the search window automatically based on the observations increases the performance and the accuracy of the tracking directly as expected. The biggest problem in small target tracking, especially ball tracking application in this case, is the suffering from lots of clutters and false positives as the target is relatively very small and the motion is rapid. The minimization of the search window do not only decrease the number of clutters and but also minimizes the false positives encountered.

In our work, we proposed a smarter sampling method for the ball games played in teams. In this approach, we employed the ball motion information in addition to the ball kinematics. This is an efficient method while dealing with targets that have less texture but high deviations in motion dynamics. Here the idea was to distribute the particles in an efficient way regarding the dynamical characteristics of the target. Using the definition of the smart region of interest concept proposed, we are then able to focus on the most valuable observations. Compared to the other methods, our method has advantage of the efficient search and therefore provides more valuable observations throughout the filtering. Our assumptions about the nature of the sensor data increase the accuracy of the method. [Kwok et al., 2004] developed another method in order to increase the computational efficiency of the target tracking by proposing a static window depending on the previous states. But better than the proposed method we adjust the search window according to the target kinematics that has better performance than a static window.

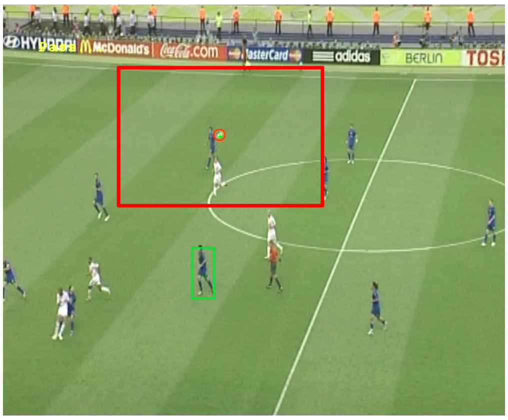
Two novelties mentioned above increase the accuracy of the estimation and decrease the computational performance especially in less cluttered environments by defining the search window in a smart way depending on the nature of the target and changing the number of samples depending on the confidence level of the estimation. Those dynamic resampling technique can be also generalized to many other small single target environments where target kinematics can be inferred. As the sports games are played in a controlled environment with predefined rules, we can then deduce the characteristic of the identities within the game.



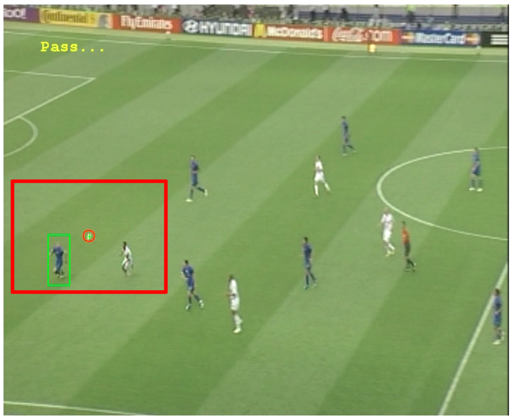
(A) High confidence.



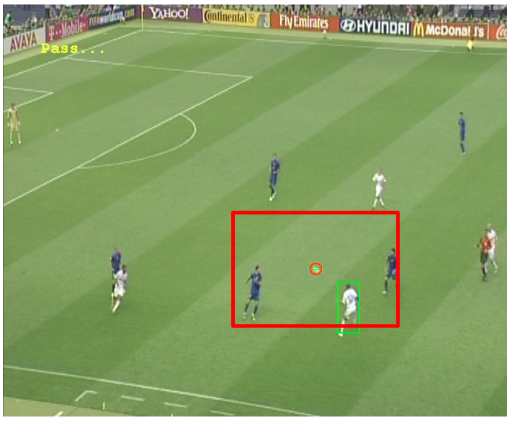
(B) Low confidence.



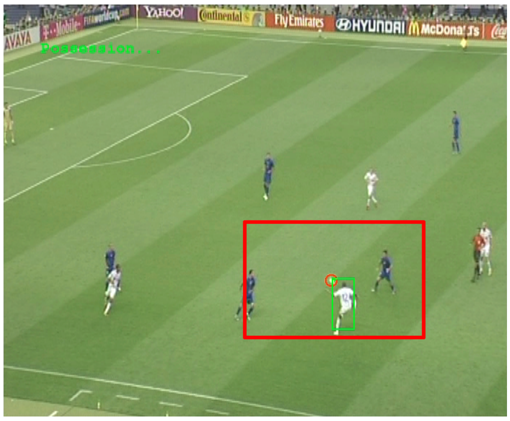
(C) Middle confidence.



(D) High confidence.



(E) High confidence.



(F) High confidence.

FIGURE 4.20. Various search window sizes during a goalkeeper shot.

4.7.4 Prediction Step

Prediction step is also known as the sampling step in particle filtering. Once a sample index is drawn in the resampling step, the corresponding sample is then used to predict the next state x_k^i. Sampling x_k^i is done from

$$p(x_k|x_{k-1}) \tag{4.10}$$

conditioned on the previous state x_{k-1}^i where the system dynamics are included. This equation is also called the prediction model. The samples drawn correspond to a prediction of the ball states. For the ball we have a state model as explained in equation 4.7. In this step, the goal is to generate samples to the posterior and thus the importance sampling is applied for the target from the likelihood and the prediction,

$$p(z_k|x_k^i)p(x_k|x_{k-1}^i) \tag{4.11}$$

using the proposal distribution of equation 3.71 then we get

$$q(x_k^i|x_{k-1}^i, z_k) \tag{4.12}$$

The division of equation 4.11 to 4.12 then gives us the importance weight. This derivation follows from equation 3.70.

As in most computer vision applications, we have taken the prior $p(x|x_{k-1}^i)$ as importance density and then we get the weight update as

$$\begin{aligned} w_k^i &\sim w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)} \\ w_k^i &\sim w_{k-1}^i p(z_k|x_k^i) \end{aligned} \tag{4.13}$$

The sample is then fed to the sample set and when the all N_p samples are drawn, the importance weights are normalized to have a sum of 1 to satisfy the total probability theorem.

As seen from the equations and the derivations, the particle filter is an approximate, sample based approach of the Bayes Filter. In particle filtering, we try to approach the Bayes estimation by using the generated N_p samples in discrete domain. As the limit of the number of samples approaches infinity, then we can converge to the true posterior.

$$\lim_{N_p \to \infty} \Rightarrow \text{True Posterior (Bayes Filter)} \tag{4.14}$$

4.7.5 Measurement Step

The measurements are carried based on the ball templates generated. During the initialization phase, we created two ball templates and those templates are used for the experiments mentioned through the thesis. The template images have been chosen such that they contain very small number of pixels of the changing background but mostly belong to the ball region. Fortunately the soccer field is a smooth region which lets us avoid the background clutter problem due to the nature of the target environment. The template generated is in a circular shape since it better represents the shape of the ball and deviations around the border. Since ball has only one side, the one sided, circular, template represents the target perfectly. This operation is done once manually during the initialization and for the experiments we presented the results here based on only those two ball templates generated in two different lawn regions. The generated pattern is passed as an image for the template matching. In our method, we define an identity for each template where its size and shape can be chosen arbitrarily in random parts of the soccer field.

Matching of the template is implemented in a way following a pyramidal approach investigated in [Hofhauser et al., 2008, Steger, 2002]. Thus during the initialization various templates are generated at different levels and the matching implemented at those levels smaller than or equal to the template level. We use the sum of the differences as a feature which is very stable and fast.

The error measure used is defined as

$$Error[x_{pos}, y_{pos}] = \frac{\sum_{u,v} |Image[x_{pos} - u, y_{pos} - v] - Template[u, v]|}{Area(Template)} \tag{4.15}$$

where u and v are the pixel points on the current part of the image and the template.

For the tracking applications in changing environments the gray values need to be normalized. But since the illumination does not change drastically during a typical soccer game, due to the constant lighting in the stadium, we directly employ the sum of differences. We also optimize the used pattern for runtime. This method leads to more stable matching results and the possibility of detection is increased. In this technique several distorted versions of the same pattern are generated and used in addition to selection of the significant gray values described as in [Steger, 2002].

The reference position of the template is its center of gravity, which is also the value returned from the error measurement and matching process. Therefore it is practical in the initialization to provide the templates where the ball is positioned in the center. For this reason, the system

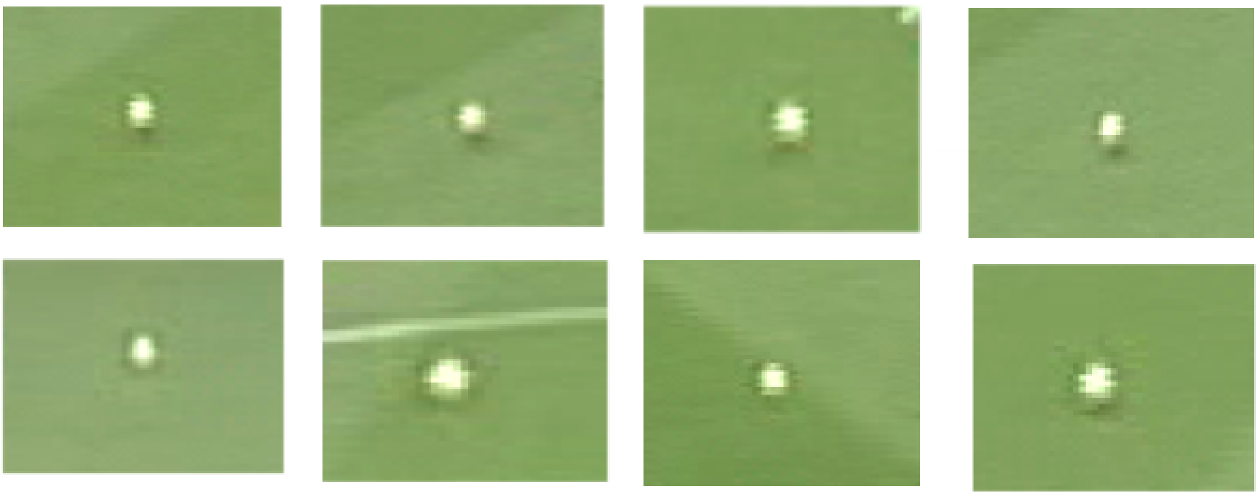


FIGURE 4.21. Ball template examples.

asks the user circle shaped templates in order to increase the matching score by providing a template suitable with the shape and nature of the target.

Applying [Steger, 2002] approach we also compute the difference between the found position and the center of gravity as a correction vector. However this is mostly not necessary in our application since we have a less textured target with a symmetric circular pattern. Therefore in our experiments the correction vector was assumed as null vector.

Hereby the template is moved over the pixel points of the measurement region to get the error value so that the template always lies completely inside the candidate interest region of tracking. The matching for one position is stopped if the error is too high. This technique leads to a reduced runtime but on the other hand, it might cause to miss correct matches.

In our method, the measurements are gathered from the search domain which is proposed in the smart sampling method. All the measurements which show a matching error smaller than the predetermined threshold are returned in the output as ball candidates. For the sake of simplicity we set this value to around $\%50$ of the target size.

Some of the ball template examples used through the experiments are depicted in figure 4.21. As seen in template examples, the ball holds very less textural information and the shape characteristics are heavily distorted. However, employing the smart resampling and the filtering of the measurements assures accurately tracking the ball even in highly occluded cases as shown in section 4.9.4. Here we do not only track the kinematics but also integrate the knowledge of the target specifications.

4.7.6 Filtering Step

Tracking of the ball is employed in 2D image domain and action representation measures are also proposed in the image domain. If the ball is not detected by its shape and textural characteristics in the first cycle, we then run a coarser search based on the prediction of the ball position. Therefore, in this step the ball kinematics are also involved for determining the search space.

After the sampling phase, the weights for new particles must be set appropriately by equation 4.13. The sampled particles are assumed to give robust associations given the samples and the measurements. The filtering of the target is determined by the constraints on the template matching and the nearest neighbor algorithm to the already existing track.

Final estimate is chosen by filtering the particle with the maximum probability. In common it is also applicable to filter according to the weighted mean of all particles. The disadvantage of that method is the so called ghost phenomenon. The ghost phenomenon, mentioned by [Blackman and Popoli], is the appearance of a non-existing target at the mean of different modes of the distribution in cases where the target is not known with a high confidence.

4.8 Ball Trajectory Estimation

In the previous part of the chapter, the tracking methods and the approaches we have developed are explained and investigated. Depending on accurate and efficient tracking of the ball, we are then able to continue further analysis of the game and provide higher level actions based on the ball localizations. In this section, we present a novel approach for the classification of the ball states based on the trajectory analysis. This method includes a novel technique in the classification between the rolling and flying ball states based on the 2D ball localizations in the image domain. The invariant trajectory features are invented and investigated to extract the different ball states.

4.8.1 Classification of Ball States: Flying / Rolling

In many sports games ball might be in flying or rolling state different than any other game elements. As the players are always stay on the ground through the game, the ball might be present either in one of the mentioned states. The flying state of the ball makes it an interesting and difficult target to infer about in some situations through the game. Additionally, the duration of the ball played through the air and the actions through the air or on the field provides inevitably valuable information for the trainers to infer about improving their tactics

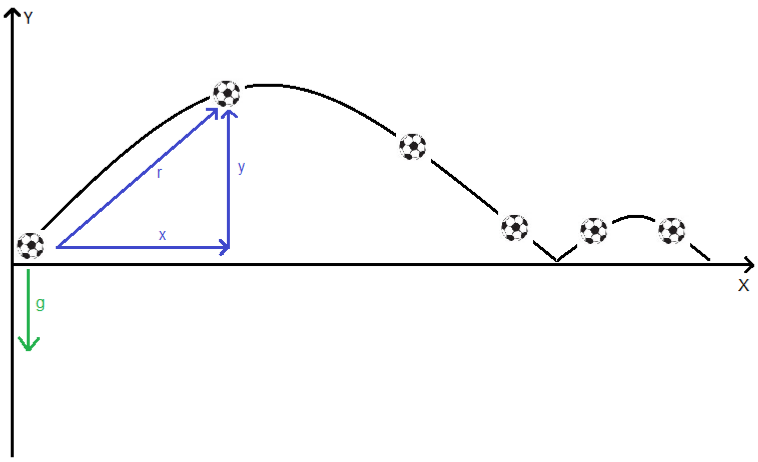


FIGURE 4.22. Projectile Motion of a ball in 2D image plane.

or analyzing the opponent team. In our approach, we developed a method to classify the flying and rolling states of the ball which leads to the deeper analysis of the game. Ball trajectory analysis has been studied extensively so far, but up to the best of our knowledge no work has been done on the flying or rolling ball state classification based on the trajectory analysis.

Detecting the ball states if the ball is rolling or flying using the broadcast video data is a difficult task. Through the game ball is mostly played on the ground, but the information when it is flying or rolling provides the trainers deeper knowledge to infer about the tactics of the opponent team. For example, if the opponent team is attacking by passing on the ground or if they are trying to approach the goal with long or short passes through the air. Based on the mentioned reasons, we invented some invariant features and implemented a trajectory mining method to discriminate between the flying and the rolling ball states.

When the ball is kicked by a player through the air, it follows a curvilinear path because the action of the gravity. This motion where the movement along the curved path under the gravitational force results from the kick obliquely near earth 's surface is called as the projectile motion. An example of a projectile motion is shown in figure 4.22.

For the purpose of ball state classification, a physical motion model of the ball is built and analyzed. The features that we used during the classification of the ball states depend on the curvilinear trajectory mining and the comparison of the actual trajectory to the minimum distance between the ball possession states.

When the ball is played through the air, it does not follow the shortest path between the two player possessions or two ball positions. As seen also in figure 4.22, the kicked or thrown object follows a curved path through its motion in the air. Even though the motion happens in three dimensional world coordinates, when it is mapped to the image plane ball trajectory always forms a curvilinear path.

When a player passes the ball on the ground it almost follows a linear path, which is almost the shortest path between the two players, depicted in figures 4.23 and 4.24. However while it is flying it does not follow the shortest path but it follows a curved path through the air as seen in figures 4.26 and 4.27.

The difference between the shortest path and the actual path in terms of trajectory and distance is the one of the features that we use in the detection of the flying ball states. The difference between the actual path and the shortest path is depicted in figure 4.28. As we have investigated in our experiments, the proposed measure performs accurate results for the classification of the flying ball states and classifies the rolling and flying ball states successfully.

If the ball is kicked by the player with initial velocity v_0 and with an angle θ as in figure 4.22 the components are written as

$$\begin{aligned} v_{0x} &= v_0 cos(\theta) \\ v_{0y} &= v_0 sin(\theta) \end{aligned} \tag{4.16}$$

In this situation, the horizontal and vertical motion can be analyzed independent from each other since they do not have any effective component on each other. In other words, there is no acceleration in the horizontal direction and the horizontal velocity is constant. The vertical motion occurs due to the gravitation. Hence the components of the acceleration can be depicted as

$$\begin{aligned} a_x &= 0 \\ a_y &= -g \end{aligned} \tag{4.17}$$

As seen from the equation 4.17, the horizontal component of the velocity remains through the motion and the vertical component of the velocity increases linearly. Therefore at any time t, the components of the velocity might be written as

$$\begin{aligned} v_x &= v_0 cos(\theta) \\ v_y &= v_0 sin(\theta) - gt \end{aligned} \tag{4.18}$$

where the magnitude of the velocity is given by

FIGURE 4.23. Ball passes on the ground.

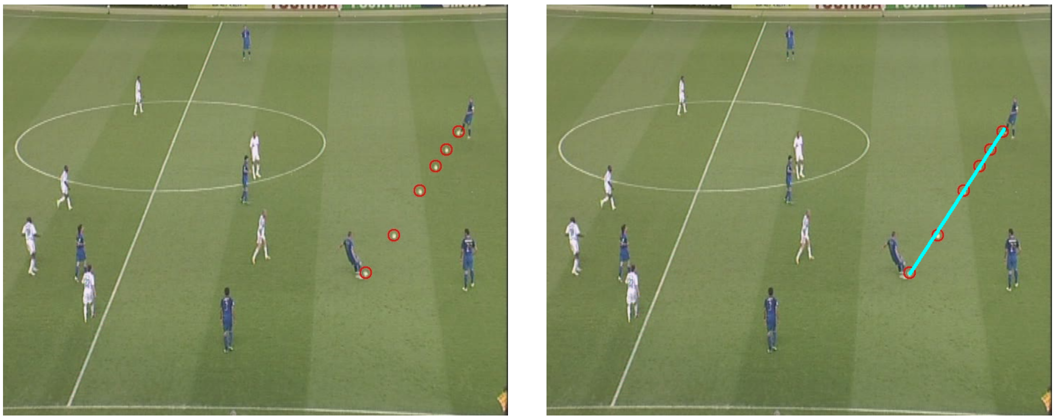


FIGURE 4.24. Linear characteristic of the ground pass.

$$v = \sqrt{v_x^2 + v_y^2} \tag{4.19}$$

Similarly the displacement at any time t can be written as

$$\begin{aligned} x &= v_0 t cos(\theta) \\ y &= v_0 t sin(\theta) - \frac{1}{2} g t^2 \end{aligned} \tag{4.20}$$

and the magnitude of the displacement is then

$$\Delta r = \sqrt{x^2 + y^2} \tag{4.21}$$

Using the equations 4.20 we can derive

$$y = tan(\theta).x - \frac{g}{2v_0^2 cos^2(\theta)}.x^2 \tag{4.22}$$

If we analyse the above equation 4.22 we find out that it is in the form of a parabola equation in the form of

$$y = ax^2 + bx + c \tag{4.23}$$

where a, b and c are constants.

Due to the gravitational force and the motion characteristics of the ball a is always negative in our experiments.

Geometrically, three nonlinear points are enough to define a parabola. Our main approach

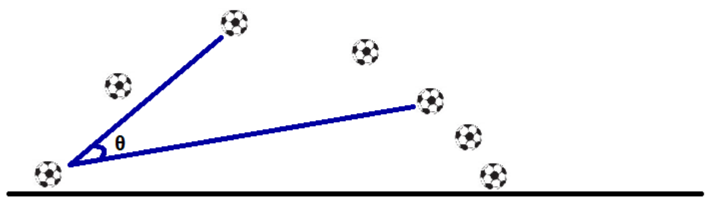


FIGURE 4.25. Five points example visualized how to detect the flying state.

is built on top of that mathematical principle. Even though three sequential points are theoretically enough to define a parabola, in our method, in order to be sure about the movement of the ball we used a window composed of 5 consecutive points. Using 5 points makes the algorithm more robust and the length of the window makes it more suitable to detect the parabolic trajectories. Using 3 points might led to missing the curved shape of the ball trajectory and thus 5 gave the best results on our experiments.

We have invented and developed two sperate methods in order to detect and ensure the rolling or flying ball states. The first method checks if the part of the track under analysis is in the form of a parabola using the trajectory mining methods. This is based on the detection of the angle of the ray from the first point on the path to the third point and the fifth point. A visualization of a scene is depicted in figure 4.25.

The flying ball states are mainly observed through the following actions.

- Possession to possession
- Possession to out of play
- Out of play to possession

In addition to trajectory mining method, we have also used the area between the shortest path and the actual path as another classification method. The appearances of the flying ball motions are depicted in figure 4.26. The method also helps us to detect the situations when the ball is kicked high in the air and is not visible on the screen and cluttered by the background of the spectators. The points selected for the analysis are also filtered out according to a predetermined distance level. Because if the user has the possession of the ball and makes movements with the ball this might lead to more false positives for the flying mode. While the ball is flying it travel a longer path than it is at the foot of a player.

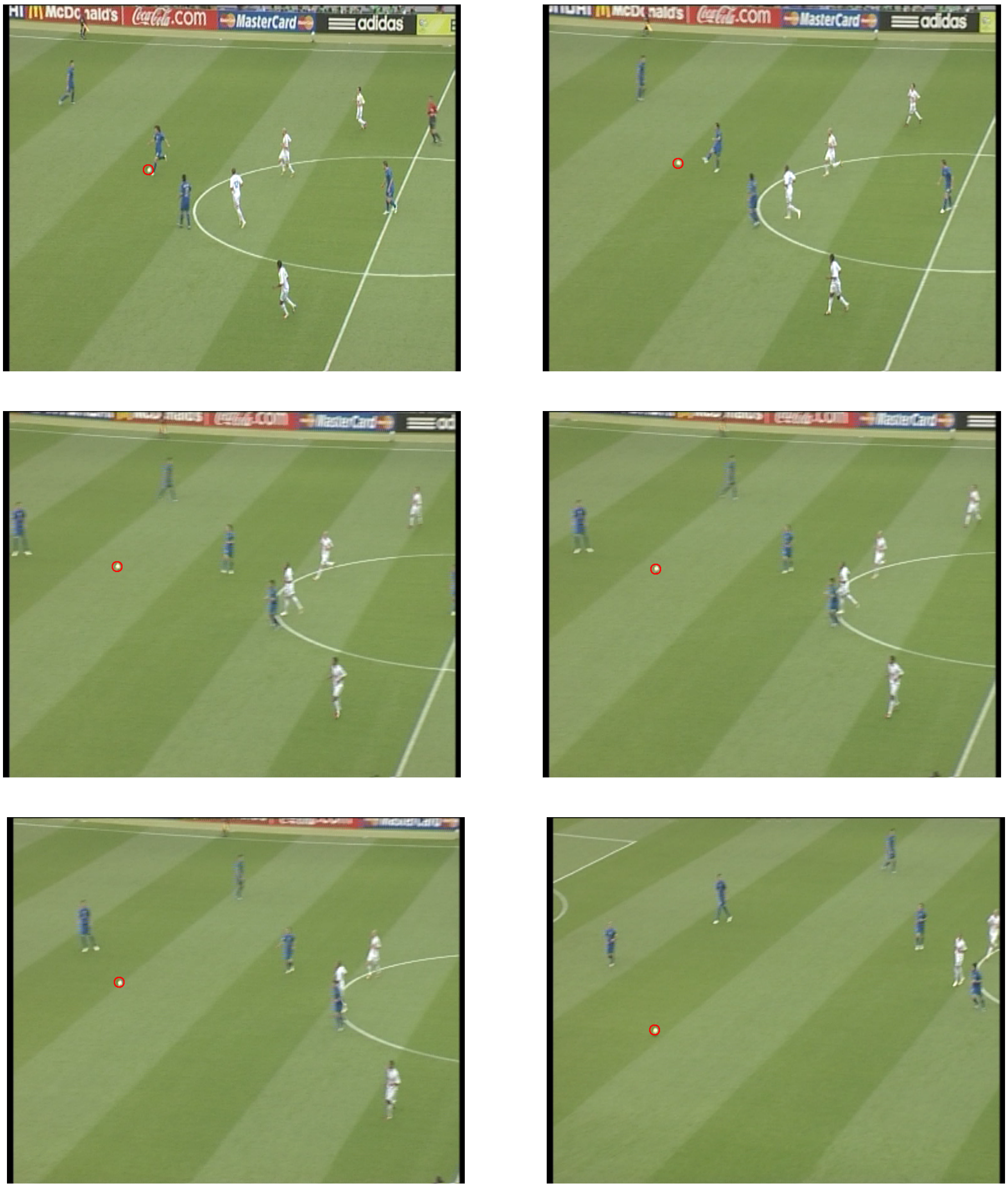


FIGURE 4.26. Ball passes in the air.

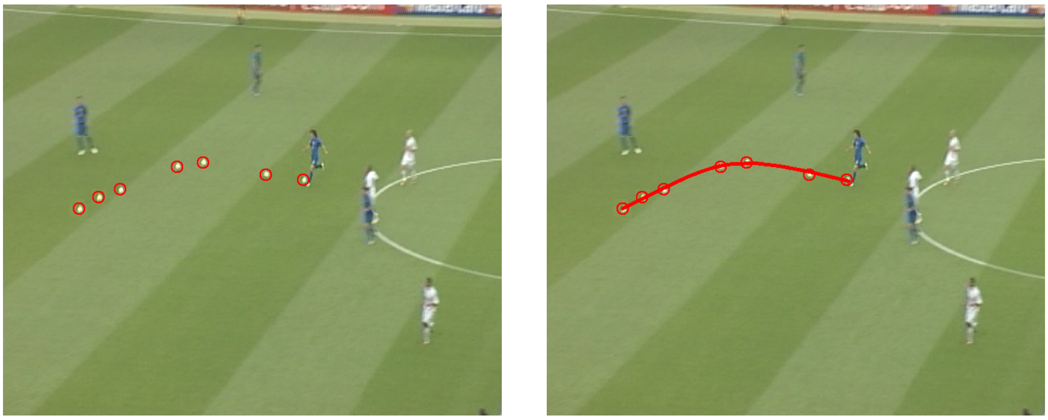


FIGURE 4.27. Nonlinear characteristic of flying ball.

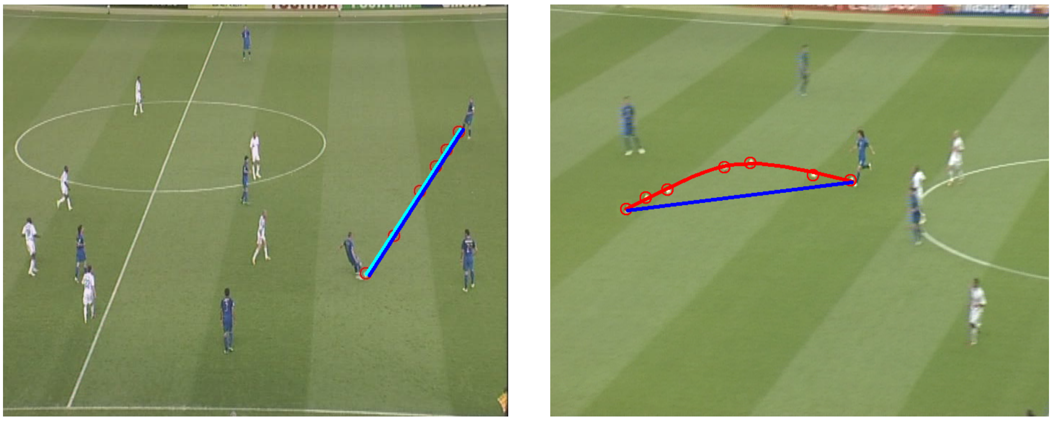


FIGURE 4.28. Flying and Rolling Ball States with shortest path indicated.

The method also helps us to detect the situations when the ball is kicked high in the air and is not visible on the screen and cluttered by the background of the spectators. An example of such a scenario is presented in the following sections.

4.9 Experiments

This section presents detailed results of the experiments conducted as a result of the thesis. Sports content extraction applications are more interesting and demanding since soccer is the most popular sports all over the world. Algorithms and methods developed throughout this dissertation are evaluated on TV broadcast videos from various international competitions. Therefore, main focus of the proposed algorithms focuses in the field of broadcast soccer videos. The invented algorithms have been tested on the FIFA 2006 World Cup final game between Italy and France and the FIFA World Cup 2010 games between England-USA and

Korea-Greece. Those videos include all possible actions that take place during a typical soccer game. The ground truth for the evaluations is generated manually.

However, even though the main goal of this study was to investigate the problem in sports video domain, especially soccer, the tracking methods developed might also be generalized to other sports domains with a similar structures. For instance, in field hockey ball tracking application, we present one of those generalizability approaches where the target is much more smaller and faster. From computer vision point of view, the complexity of the scenes and the recognition of the movements increases the need for the advanced machine vision algorithms.

The experiments have been carried out starting from the data acquisition phase to higher level abstraction. The input to the ball detection, tracking and ball action recognition consists of recorded videos broadcasted sports games. Those videos are acquired from various TV channel sources using a simple TV recorder interface. Using the recording hardware cards, games can be recorded on hard drives of the PCs for latter analysis. The recordings has been saved in 25 frames per second and $720x576$ resolution. Due to the broadcasting methods, we usually suffer from the interlacing problems as well as the low quality streams. Therefore the first step after data acquisition was to remove the interlacing within the frames.

4.9.1 Field Segmentation

As in many other field games, interesting part of the actions during the game always take place on the field. For this reason, the experimental results showing the filtering of the field of interest play a crucial role in the system performance. The first step in the analysis is the determination of the region of interest. We have used statistical Gaussian distance measures instead of linear measured proposed in the literature(see chapter 4.3). The results of the experiment is as shown in table 4.1.

Game	Camera	#Frames	False Positives	False Negatives	Accuracy
Italy - France	Broadcast	45000	400	0	0.991
Italy - France	Highangle	45000	44	0	0.999
England -USA	Broadcast	10000	127	0	0.987
Korea -Greece	Broadcast	10000	234	0	0.976

TABLE 4.1. Field Segmentation Results

4.9.2 Player Detection

Player detection method has been implemented using simple and color features of the players on the field [Jia Liu, Xiaofeng Tong, Wenlong Li, Tao Wang, Yimin Zhang, 2009, Nunez et al., 2008]. In this thesis, we used the player positions in order to extract higher level ball players interactions and ball actions. For this reason we have implemented a color and shape based method to extract the players. Our focus was to be able to segment the players around the ball. The following table 4.2 shows the results about the detection of the players interacting with the ball.

Game	Camera	#Frames	Detected	False Positives	False Negatives	Accuracy
Italy - France	Broadcast	45000	34204	0	10796	0.76
Italy - France	Highangle	45000	35555	0	9445	0.79
England -USA	Broadcast	10000	7168	0	2832	0.72
Korea -Greece	Broadcast	10000	6402	0	3598	0.64

TABLE 4.2. Player Detection Results

As seen on the table we usually had no false positives but had problems in detectiog the players close to the ball. This happened due to the camera zoom or occlusion factor which frequently happened throughout a typical soccer game. However, we are able to detect the players close to the ball with a good accuracy.

4.9.3 Ball Detection

The next step after the detection of the region of interest and the players was the detection of the ball. For this experiment we have used the same frames as in the field segmentation. Detection of the field is implemented using the shape and color features as explained in Chapter 4.6 The results are depicted in table 4.3

On the average the only detection of the long video sequence of broadcast video frames is experimented to be about $\%45.68$. In some similar works, the results are presented to be around $\%99.00$ [Pei et al., 2009]. However the detection algorithm is implemented on some hundred of selected frames on thsi videos. And it is not clearly stated that if all possible soccer events have been investigated. In our algorithm, we run the developed method on a typical ongoing broadcast soccer game video that has player occlusions, shots, occlusion by field

Game	Camera	#Frames	Detected	False Positives	False Negatives	Accuracy
Italy - France	Broadcast	45000	21353	21011	2636	0.47
England -USA	Broadcast	25000	12230	11245	1525	0.49
Korea - Greece	Broadcast	25000	10118	12400	2842	0.40

TABLE 4.3. Ball Detection Results

lines and so on. Additionally, the performance of the ball observation algorithm is improved by the novel tracking methods explained in the following section.

4.9.4 Ball Tracking

One of the main goals of this dissertation is to propose, develop and investigate a novel small and single target tracking algorithm with applications to broadcast soccer videos. The small and single target analyzed for the broadcast sports videos was the ball regarding its importance within the game. The tracking and trajectory analysis methods have been presented and explained in detail in the previous sections.

Major part of the experiments was aiming the algorithm development and implementation in order to track the ball and events happening around the ball 4.4.

Game	Camera	#Frames	Tracked	False Positives	False Negatives	Accuracy
Italy - France	Broadcast	45000	35112	2426	7462	0.78
England -USA	Broadcast	25000	17751	1760	5489	0.71
Korea - Greece	Broadcast	25000	16876	2625	5499	0.68

TABLE 4.4. Ball Tracking Results

Similar work has been presented for the ball tracking so far using various probabilistic tracking methods [Liang et al., 2005, Yu et al., 2003c,d] However the genrality and applicability of their algorithm is still a big question. [Liang et al., 2005] presented accuracy from %76.7 to %84.5 only on 650 frames of data. [Yu et al., 2003d] presented %100 of accuracy on 4400 frames of selected data and [Yu et al., 2003c] presented %96.65 of accuracy on 2000 frames. Similarly, [Kim and Kim, 2009, Liu et al., 2010] also presented their work on a limited amoun of data which is away from deeper evaluation of ball tracking.

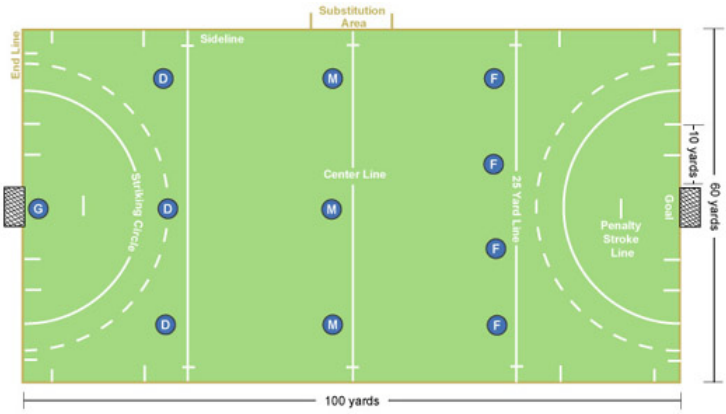


FIGURE 4.29. Diagram of a Hockey Field.

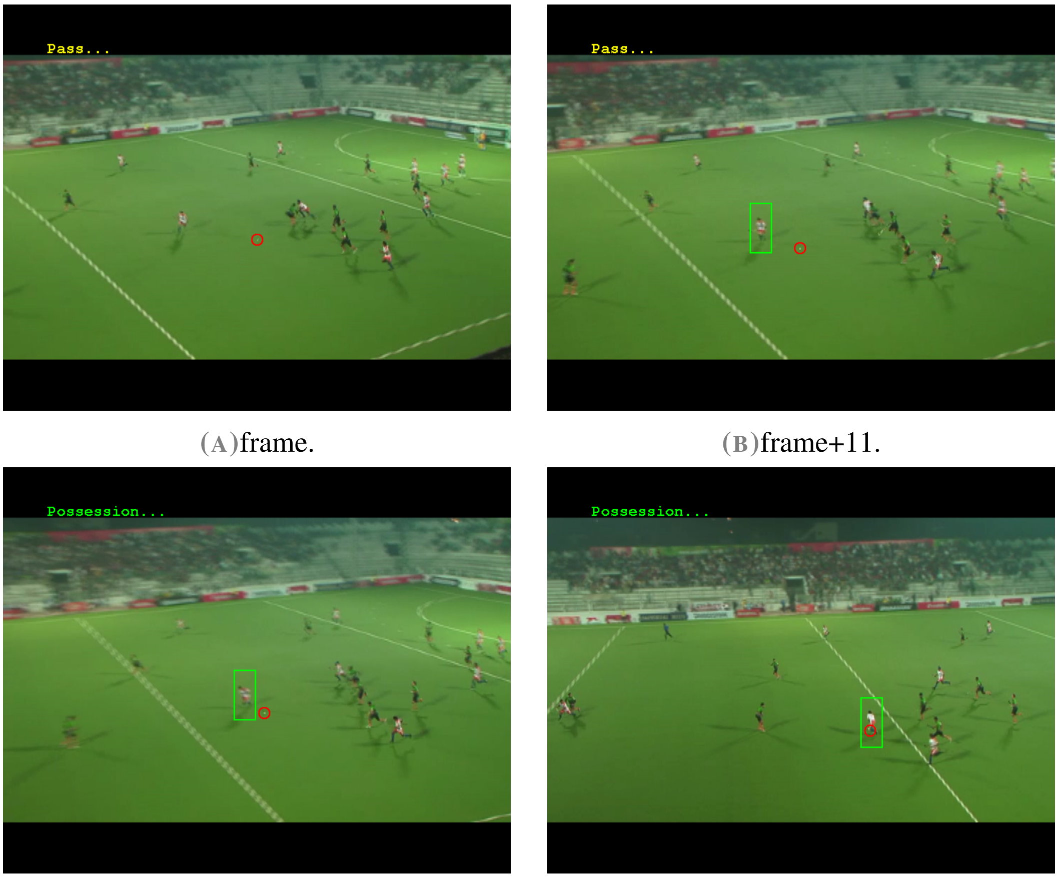
For proving the generality of the methods proposed, we have also investigated the performance of the proposed algorithms on hockey videos. Even though the main target of this research was to propose novel small single target tracking method for broadcast soccer videos, the field hockey has also similar structures as soccer.

Field hockey, commonly known as hockey, is a team sport which is played between two teams consisting of each from eleven players with hockey sticks and a ball. As a general structure it is similar to the soccer games. As in soccer the goal keepers are the only players who are allowed to touch the ball with any part of their body only applying within the shooting circle.

According to the rules, all players in the team that has possession of the ball are attackers, and those in the team without the ball are defenders. Through the game being played a team is always "attacking" for goal or "defending" the opposite goal. Different than the soccer games the game is controlled by two umpires each of whom responsible for half of the field. The control regions for each referee is defines roughly diagonal. The field model is as shown in figure 4.29.

Similar to soccer, the game is composed of two halves each consisting of 35 minutes which may sometimes vary on the regional leagues. The game is started from the center of the field as it is in soccer games. The game is restarted from the center of the field at the beginning of each half as well as after each goal.

Another difference in field hockey is in the characteristic of the actions by players. For instance, they can only play with the ball using the face of their sticks. Instead of the direct play of the ball in soccer games, a player can only move the ball using those special sticks. For instance, tackling is only permitted when the opponent does not make contact with the attacker who has the possession of the ball or his stick before playing with the ball. This is similar to



(A)frame. (B)frame+11.

(C)frame+20. (D)frame+50.

FIGURE 4.30. Hockey Ball Tracking Frames.

tackling in soccer games where the players use their feet instead of sticks. In field hockey, ball contact with the feet is allowed as long as the player does not become advantageous from the contact. When the ball crosses over the sidelines, it is returned to play with a sideline hit by one of the team members from the opponent team of the player who touched the ball last. The ball is returned to the game with a sideline hit similar to the throw in in soccer games.

In general, game characteristics regarding the contact with the ball as well as passing and possession situations is similar. But more differently the target, i.e. the ball, to be tracked in hockey is much smaller than the soccer ball. Therefore it makes the detection and tracking a more difficult task to be carried on. Results from the hockey ball tracking is depicted in figure 4.30.

Based on all the similar structure defined above, in order to strengthen the generality of the proposed technique for other broadcasted videos, we have performed evaluations on a broadcast hockey video game between Mumbai Marines and Bhopal Badshahs. The game is the fifth game of the Bridgestone World Series Hockey 2012. More details about the experiment

Game	Camera	#Frames	Tracked	False Positives	False Negatives	Accuracy
Mumbai Marines - Bhopal Badshahs	Broadcast	5000	2916	540	1544	58.2
Mumbai Marines - Bhopal Badshahs	Broadcast	3000	1649	201	1150	51.1

TABLE 4.5. Hockey Ball Tracking Results

is depicted in table 4.5. Even though the target was much smaller and faster we observed promising results on that type of data. Up to the best of our knowledge this work is the first to be evaluated on field hockey data.

4.9.5 Ball Trajectory Flying / Rolling

Another novel approach investigated through the thesis was the classification of flying and rolling ball states from broadcast soccer videos. Up to the best of our knowledge, such a state classification has not been studied before. We have proposed novel methods in order to extract the ball states automatically as rolling or flying. The results of ball trajectory mining experiments are depicted in table 4.6.

Game	Camera	#Frames	Correct Classifications	False Classifications	Accuracy
Italy - France	Broadcast	8000	5554	2446	0.69
England -USA	Broadcast	5000	3558	1442	0.71
Korea - Greece	Broadcast	5000	3371	1629	0.67

TABLE 4.6. Ball Trajectory Mining Results

The invented motion cues were then able to extract the ball states with a good accuracy.

4.10 Conclusion

In this chapter, we introduced the ball detection and tracking methods developed through the thesis. The detection and tracking of the ball is not straightforward and it requires preprocessing of the frame in order to extract the related content. The first step in the determination of the field of interest was the segmentation of the field region where many actions take place. Then we focused on the detection of the player which will then be used to extract higher level ball actions in Chapter 5. The main part of the section was related to the development of a ball detection and tracking system. An efficient ball detection algorithm is proposed and explained. The proposed algorithm not only run the detection of the ball using shape and color features but also improves the detection by using smart detection methods. However only the detection was not accurate enoguh to determine the ball localisations. Therefore a novel method of particle filter is proposed and developed for ball tracking. The development of the filter is explained and related experimental results are provided. The experiments proves that the method performs more accurate than the state of the art methods explained in the literature. Following the ball tracking, a new method of flying and rolling ball classification algorithm is presented. The evaluations and the results are given along with the experimental conclusions.

CHAPTER 5

Ball Action Recognition

The ball is the main component of a soccer game, and therefore actions happening around the ball have much higher importance than other actions throughout the game. Ball states have direct effect on the decision of the referee and the way rules are related to the position and the state of the ball. Most state of the art approaches have been focusing on player or human movements for action recognition [Kjellstrom et al., 2010, Li and Ibrahim Sezan, Mendoza and Blanca, 2007, Needham, 2003]. In contribution to existing approaches for sports games, especially soccer, we employ the ball information in order to extract actions during the game. The proposed approach enriches the action cues and improves the representation of a soccer game. Other than other state of the art methods [Kjellstrom et al., 2010, Li and Ibrahim Sezan, Mendoza and Blanca, 2007, Needham, 2003], we define similar actions based on more effective data cues.

In this chapter, we propose and investigate higher level sports game actions based on ball trajectories and player positions. We investigate the semantic analysis of the game by the association of the ball and player motion parameters. In other words, those features allow us to extract the interactions between the ball and the players. Since the ball is a more active element of the game that is always in the center of subject of interest throughout the game, we can more effectively analyze the content of the game.

In this chapter, we also discuss a method of detecting important scenes based on the ball kinematics and try to deduce the current state of the game using those motion cues.

In addition to the actions happening around the ball, we also investigate the ball player interaction scenarios in order to predict the actions happening throughout the game. By using the positional information of the ball and the detected players, higher level events are extracted at abstraction level. Those events have been defined in three separate classes namely, $Pass$, $Possession$ and $OutofField$. In our method, we derive metrics for those definition models of the ball states based on ball position, ball velocity and player localizations.

The ball velocity is one of the important factors in defining the ball states, especially during occlusions. For this reason, we propose a ball speed estimation model based on ball position on the tracks. This speed estimation is built on the motion model explained in chapter 4.7. As the real world coordinate conversion is straightforward using the camera parameters from the field calibration which is obtained via the approach of [Gedikli, 2008], we only focus on the interpretation of the speed of the ball in this thesis.

We evaluated the mentioned algorithms on the final game of FIFA World Cup 2006, two games from FIFA World Cup 2006 and broadcast hockey video game between Mumbai Marines and Bhopal Badshahs of the Bridgestone World Series Hockey 2012. Throughout the experiments we observed promising results and improved accuracy over current state of the art approaches. Using the ball features as the main point of the approach has allowed us to precisely recognize actions during the game. Experimental details and results are presented in chapter 5.6.

5.1 Markov Model

The Markov model and Markov processes are statictical methods that have been introduced in the 1960s [Baum and Petrie, 1966, Baum et al., 1970b]. The Markov models have a very strong application area regarding their mathematical models and practical applicability. The simplest Markov model is the Markov chain which is used to model the state of a system with one random variable that changes during time. In other words, it represents a random memoryless system in which the next state depends only on the current state and the transition probabilities.

Formally speaking, a Markov chain is a sequence of random variables $X_1, X_2, X_3, \ldots$ where the next state is only dependent on the current state and the state transitions.

$$P[X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = P[X_{n+1} | X_n = x_n] \tag{5.1}$$

In the above equation, values of X_i form a countable set, S is called the state space of the chain. The most important characteristic of the Markov chains is that the current state, future states and past states are independent of each other. The probability of going from state i to state j in n time steps is

$$a_{ij}^{(n)} = P[X_n = S_j | X_0 = S_i] \tag{5.2}$$

and the single-step transition is

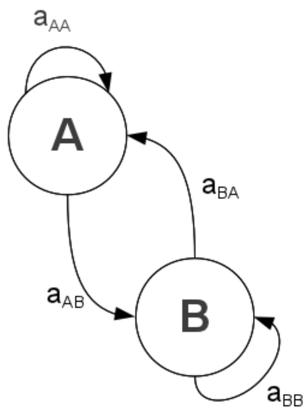


FIGURE 5.1. Simple two-state Markov Chain.

$$a_{ij} = P[X_1 = S_j | X_0 = S_i] \tag{5.3}$$

Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states. A simple graph for a two-state Markov chain is depicted in figure 5.1

Another variation of Markov chains is the homogeneous or time-homogeneous Markov chain. In time-homogeneous Markov chains the system state stays exactly in one of the defined states for every time period. At the end of each determined period, the state moves to another state or stays on the same state for another period. In time-homogeneous Markov chains the total probability of movement from a state must be equal to one. Here we should also note that the movement from a state to itself does not count as movement. Another interesting property of homogeneous Markov chains is that the transition probabilities do not change over time. This is a requirement for homogeneous Markov chains [Baum and Petrie, 1966, Baum et al., 1970b]. For a time-homogeneous Markov chain the transition probabilities of n step is represented as

$$a_{ij}^{(n)} = P[X_{k+n} = S_j | X_k = S_i] \tag{5.4}$$

where the one step transition is given by

$$a_{ij} = P[X_{k+1} = S_j | X_k = S_i] \tag{5.5}$$

The above n step probabilities satisfy the Chapman-Kolmogorov equation,

$$a_{ij}^{(n)} = \sum_{r \in S} a_{ir}^{(k)} a_{rj}^{(n-k)} \qquad 0 < k < n \tag{5.6}$$

So far the above equations describe a Markov chain of first order. In other words the current state is only dependent on the last step. However in some situations the current step depends not only on the last but more past steps. In such cases, a Markov chain of higher order is proposed [Begleiter et al., 2004, Rissanen, 1983]. For example, a Markov chain of order m represents a Markov chain with a finite memory as

$$\begin{aligned} &P[X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1] = \\ &P[X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m}] \qquad n > m \end{aligned} \tag{5.7}$$

In other words, the equation 5.7 represents a Markov chain with memory m, in which the current state depends on the past m states. In higher order Markov chains, we can also form a chain keeping the Markov property by taking as the state space the ordered m values from the state space as

$$Y_n = (X_n, X_{n-1}, \ldots, X_{n-m+1}) \tag{5.8}$$

The above equations hold for the discrete-time Markov chains. Generally speaking, the Markov property is defined as, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t, t \in T)$ where T is a totally ordered indexed set and S being a measurable space, the S valued stochastic process $X = (X_t, t \in T)$ adapted to the filtration $(\mathcal{F}_t)$ depicted as [Durrett, 1999],

$$\mathbb{P}[X_t \in A | \mathcal{F}_s] = \mathbb{P}[X_t \in A | X_s] \tag{5.9}$$

For example for autonomous systems where the state is fully observable, Markov chains are used. And when the system is controlled and fully observable a Markov decision process is applied. In cases where the system state is partially observable and the system is autonomous Hidden Markov models (HMMs) are used and when the system is controlled and partially observable Markov decision process approach is used. In the ball action recognition system we have an autonomous system model where the ball states are partially observable. For that reason we have implemented a hidden Markov model approach for ball action recognition in broadcast soccer videos. A detailed explanation of the Hidden Markov model approach is presented in section 5.1.1.

5.1.1 Hidden Markov Model (HMM)

Each process in real world usually generates an output and this output is represented as a signal. The signals can be found in various natural characteristics such as discrete or continuous forms. The HMM based methods in broadcast soccer video analysis proposed high level actions based on simple scene content like dominant color and motion intensity [Li and Ibrahim Sezan, Xie and Chang, 2002]. Those cues are employed in order to extract meaningful contents, play, non-play, break etc., from the game video data.

One of the most interesting areas of information science is to analyze this signals in order to construct models to represent the signals in terms of mathematical models. A hidden Markov model (HMM) is a type of statistical method in which the system is assumed to have unobserved states [Rabiner, 1989]. Hidden Markov Models relate the states to each other over adjacent time steps. In a hidden Markov model the hidden variables are controlled through a Markov process. In such manner, an HMM is a form of dynamic Bayesian network [Baum et al., 1970a,a,b]. In HMMs the observation of the states are not always available, therefore the state is considered to be hidden.

Before we start to explain the details of our method, we present the elements of an HMM for the reader for easier understanding of the derivation.

$$\begin{aligned}
N \Rightarrow \quad & \text{number of states in the model.} \\
M \Rightarrow \quad & \text{number of distinct observation symbols per state.} \\
& V = \{v_1, v_2, \ldots, v_M\} \\
A = a_{ij} \Rightarrow \quad & \text{the state transition probability distribution} \\
& a_{ij} = P[X_{k+1} = S_j | X_k = S_i] \qquad 1 \leq i, j \leq N \\
B = \{b_j(n)\} \Rightarrow \quad & \text{the observation symbol probability distribution in state j} \\
& b_j(n) = P[v_k \quad at \quad t | X_k = S_j] \qquad 1 \leq j \leq N \quad 1 \leq k \leq M \\
\pi = \{\pi_i\} \Rightarrow \quad & \text{the initial state distribution} \\
& \pi_i = P[X_1 = S_i] \qquad 1 \leq i \leq N
\end{aligned} \tag{5.10}$$

As explained in section 5.1 the system considered is described at any time being in one of N distinct states, X_1, X_2,... X_N, where $X_i \in S$ and $1 \leq i \leq N$. State changes occur at $k = 1, 2, \ldots N$ and the actual state at time k is given by X_k. State transition probabilities a_{ij} are defined in the form

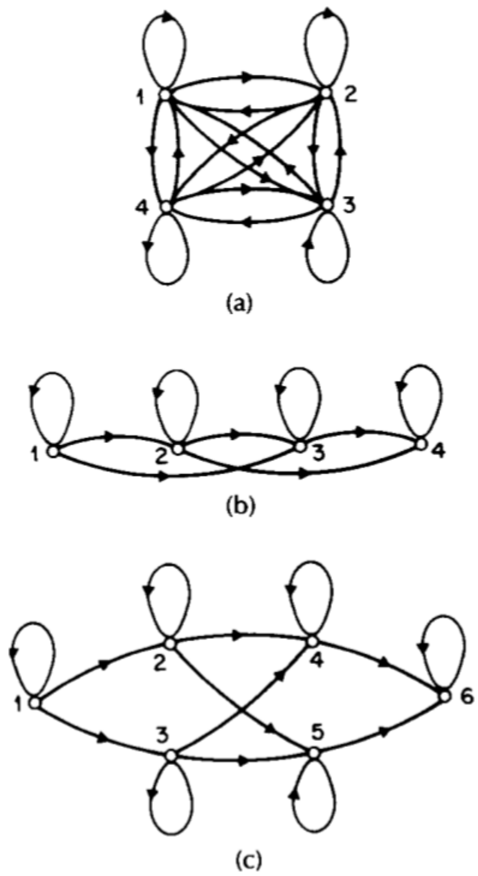


FIGURE 5.2. Illustration of 3 distinct types of HMMs. (a) A 4-state ergodic model. (b) A 4-state left-right model (c) A 6-state parallel path left-right model [Rabiner, 1989].

$$a_{ij} = P[X_k = S_j | X_{k-1} = S_i] \quad 1 \leq i, j \leq N \tag{5.11}$$

where

$$\begin{aligned} a_{ij} &\geq 0 \\ \sum_{j=1}^{N} a_{ij} &= 1 \end{aligned} \tag{5.12}$$

to obey standard stochastic constraints, i.e. the total probability theorem. In the observable Markov models each state corresponds to a physical observable event, however in the Hidden Markov model the problem deals with a Markov chain in which some states are only partially observable.

Given the state model is known, the probability that the system will follow a desired order for the required model for d time steps is evaluated as the probability of the observation sequence depicted as

$$O = \{S_1, S_2, S_3, \ldots, S_d, S_{d+1}\}, \tag{5.13}$$

where the model is

$$\begin{aligned} P(O|Model) &= P[S_1, S_2, S_3, ..., S_d, S_{d+1}|Model] \\ &= P[S_1] \cdot P[S_2|S_1] \cdot P[S_3|S_2] \cdot \ldots \cdot P[S_d|S_{d-1}] \\ &= \pi_1 \cdot a_{12} \cdot a_{23} \ldots a_{(d)(d+1)} \end{aligned} \tag{5.14}$$

where π_1 is the probability density function of duration d at the same state.

Markov decision process explains a Markov chain in which the state transitions depend on the current state and an action vector that is applied to the system. In some cases the states might not be fully observable, in that case we apply the Partially Observable Markov decision process (POMDP) which is a NP complete problem. POMDP has many application areas especially in robotics and action recognition.

There are three main problems researched in Hidden Markov Models. The first problem is the simplest problem of the HMMs which is the direct calculation of the probability of the observation $P(O|\lambda)$ sequence given the observation sequence $O = O_1 O_2 \ldots O_T$ and a model $\lambda = (A, B, \pi)$. This approach is usually used in chosing an appropriate model for a process.

The second problem deals with the optimization of the model parameters in order to describe how a given observation sequence comes about. In the training phase, an observation sequence is chosen in order to optimize the model parameters to optimally adapt model parameters to the observed data. The determination of the training sequence is the crucial task in this problem which optimally adapt model parameters to the observation data. In other words, the second problem deals with adjusting the model parameters $\lambda = (A, B, \pi)$ to maximize the probability of the observation sequence i.e. $P(O|\lambda)$.

The last problem is related to our approach and attempts to uncover the hidden part of the ball state sequence in order to find the correct state sequence. In our case this is the correction of the ball states namely, pass, possession and out of play states. We have proposed rule based HMM optimization methods for the case of ball action recognition. Up to the best of our knowledge this rule based approach for the summarization of broadcast videos with the aim of ball action recognition is a novel method in extracting the events. The method is explained in detail in section 5.2.

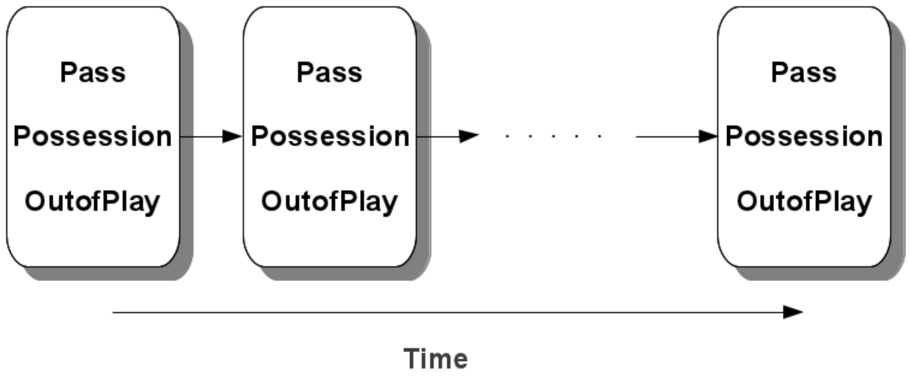


FIGURE 5.3. General Duration Synthesis for Ball Actions.

5.2 Player Ball Interaction

The extraction of ball action states and the interpretation of motion classes for broadcast soccer games are studied in this part of the thesis. We investigate game characteristics and conclude ball player interactions using the novel features for the automatic interpretation of soccer games. We propose an HMM based approach to the problem and investigated the results on broadcast soccer videos.

The ball is always either in control or in action of the players during the game. Player and ball interaction has been studied so far for tracking applications or manually annotated team pass or possession information. However the automatic ball state recognition with focus on autonomous player ball interaction is not studied in detail. In this thesis, we have defined three distinct states for the ball actions,

- $Possession$
- $Pass$
- $Outofplay$

The simple duration synthesis for a ball action model is depicted in figure 5.3.

Technically, $Possession$ means that the ball is under the control of a player. According to our algorithm, it means that the ball is in a static state or stays within a distance to one of the detected player's reach area (see Fig. 5.4). In other words, the tracked ball position is in such a state that it might be associated with a player position for a definite time of duration. The reach area is defined as the half of the height of an average player. The center of the movement is always at the foot of the players. Therefore the center of the player reach region is centered at the foot of the player in figure 5.4.

$Pass$ is defined as the state where the ball cannot be associated with any of the players but still found on the field. The passing state is the most common state of the ball during the game. It happens when the ball is on the field and not in the control of a player. In other words, if the ball is on the field but out of the reach area of the players, then it is classified as in the $Pass$ state.

The $OutofPlay$ state happens when the ball is not found on the field during the game. However the ball sometimes might be in the flying state and still in the game. Normally during such situations the players also stop moving and the camera motion is zoomed out to get a broader view of the scene. In such situations, when the zoom factor of the camera is changed radically we reinitialize the system since the tracker and player detection module also need to be reinitialized because of drastic zoom factor changes of the camera.

For the definition of the mentioned ball action states above, we have used the distance and the velocity of the ball. the ball velocity is computed by using the ball positions over the frames assuming a 25 frames per second. $Possession$ happens simply when the ball is found within a distance where a player can take control of the ball. The player can take the control of the ball when the ball is not moving at a fast speed at his reach region. If the ball is found to be in the reach area of a player but still moving at a high speed then it is still classified as in the $Pass$ state.

The distance a player can reach at a moment is estimated dynamically using the player size. We define the reachable position as the half of the detected player height centered at their feet. This is set at the beginning of a game as a result of the detections once for a whole game. The range for the possession is computed and associated with each detected player around the ball. An example is shown in figure 5.4.

As long as the ball is found on the field, it is either in $Possession$ or $Pass$. When the ball is flying over a player, it might falsely be regarded as in the possession of that player by using only the positional information. Therefore if the ball is still in motion in the reachable area of a player, then it is regarded as in $Pass$ state. In summary,

- The ball is $OutofPlay$ when the ball position is not found on the field.
- The ball is in $Possesion$ when it is within the reachable position of a player on the field and has a negligible velocity.
- The ball is is $Pass$ mode when it is not within the reachable position of a player or at a higher speed and on the field.

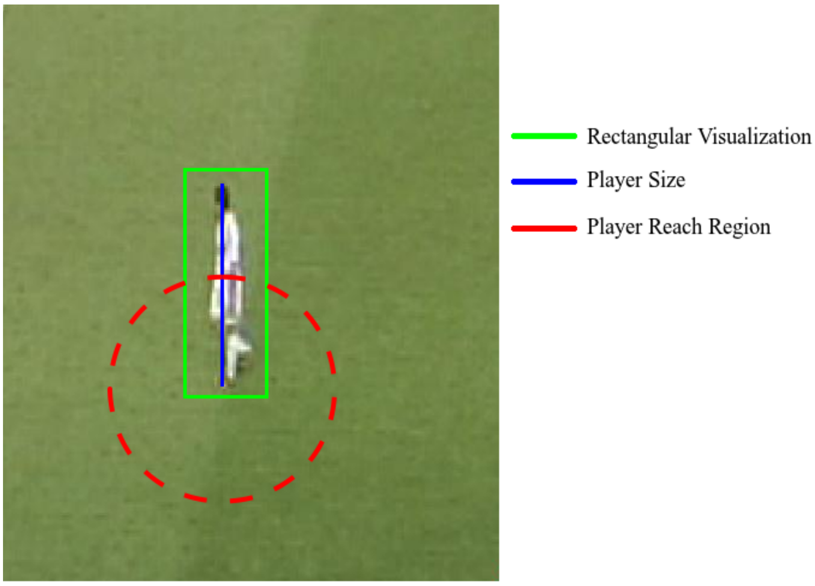


FIGURE 5.4. Dynamic player possesion reach region.

5.3 Ball Action Modeling

Ball actions defined in section 5.2 cannot always be clearly observed by using the defined metrics. Therefore, we proposed and developed a Hidden Markov Model approach in order to recognize the ball actions during the game. Actions are analyzed in three different classes namely $Pass$, $Possession$ and $OutofPlay$ as explained in the previous section.

For the first part of our analysis, we gather the ball states extracted using the ball kinematics. However we cannot always get accurate observations due to the complexity of the scene. Therefore detected ball states then needs to be filtered in order to achieve more accurate results and to uncover the hidden states.

Ball state sequences include the consecutive ball states detected using the player positions associated with the ball kinematics. An example of a state sequence is depicted in figure 5.3. In each timestamp the ball is in either of the three states depicted in figure 5.3.

As the soccer game is played in a controlled environment and setting belonging to the rules, we are then able to build a Markov model according to the expected motion actions. The transitions between the states can be also explained as

- $Pass$ to $Possession$ -> the ball is acquired by another player. This is one of the most common actions during the game where the players try to move the ball to the forward players. After the passing state the ball comes in the reachable area of a player and is possessed by one of the players.
- $Pass$ to $Pass$ -> the ball is still in passing state between players, the next state will be

most likely possession. That state happens mostly during long passes when the distance between the players are larger.

- $Pass$ to $OutofPlay$ -> fail pass, the ball state needs to be reinitialized manually with possession and pass. That happens when the player in possesion is under pressure and kicks the ball towards out of the field.

- $Possession$ to $Possession$ -> the ball is still in possession of a player. The player moves with the ball in order to find a better situation of pass or shoot in order to score for his team.

- $Possession$ to $Pass$ -> the ball is passed from the previous player. In this state sequence, the player transfers the ball to one of his team mates in order to support the attack.

- $Possession$ to $OutofPlay$ -> does not often happen, but can happen when a player is under pressure.

- $OutofPlay$ to $Possession$ -> the current player restarts the game. The track and the ball action state needs to be reinitialized with the associated player who will start the game.

- $OutofPlay$ to $Pass$ -> does not happen since after the ball is outside of the field, the game needs to be restarted with the throw of a ball with the possession of a player.

- $OutofPlay$ to $OutofPlay$ -> the ball state is still out of the field. This happens when the ball cannot be detected within the field regions.

5.3.1 HMM-Based Ball Action Recognition

Ball positions might not be always clearly observable due to the observations or the state extraction measures. This section describes an HMM-based ball action recognition method in order to extract hidden ball states. The major problem investigated here is the generation of optimal ball action states from the HMM based on a decision tree defined by the measures in the previous section. In other words, given an HMM λ of length T of a action parameter sequence to be generated, the problem for generating action states from the HMM is to obtain a observation sequence of $O = O_1 O_2 \ldots O_T$ which maximizes $P(O|Model) = P(O|\lambda, T)$ with respect to O,

$$\begin{aligned} O^* &= \arg\max_O P(O|\lambda, T) \\ &= \arg\max_O \sum_{all\ q} P(O, q|\lambda, T) \end{aligned} \tag{5.15}$$

where q stands for ball state sequences. Since there is no known method to analytically obtain the ball action state sequence which maximizes $P(O|\lambda, T)$ in a closed form,this problem is approximated by using the most likely state sequence in the same manner as the Viterbi algorithm, i.e.,

$$\begin{aligned} O^* &= \arg\max_O P(O|\lambda, T) \\ &= \arg\max_O \sum_{all\ q} P(O, q|\lambda, T) \\ &\simeq \arg\max_O \max_q P(O, q|\lambda, T) \end{aligned} \tag{5.16}$$

Using Bayes' theorem, the joint probability of O and q can be simply written as

$$\begin{aligned} O^* &\simeq \arg\max_O \max_q (P(O, q|\lambda, T)) \\ &= \arg\max_O \max_q (P(O|q, \lambda, T)P(q|\lambda, T)) \end{aligned} \tag{5.17}$$

Then the problem is divided into two smaller optimization problems. In summary, optimization of the observation sequence O given the HMM λ and the length T can be approached by two optimization problems as

$$q^* = \arg\max_q P(q|\lambda, T) \tag{5.18}$$

$$O^* = \arg\max_O P(O|q*, \lambda, T) \tag{5.19}$$

As seen in equation 5.19, the optimum solution for the observation sequence O^* is given by the optimum state sequence. Therefore, we firstly focus on the solution of the optimization problem of the ball action state sequence q and then propose our method for the ball observation states O.

Given the HMM model λ of length T, we can then calculate $P(q|\lambda, T)$ similar to equation

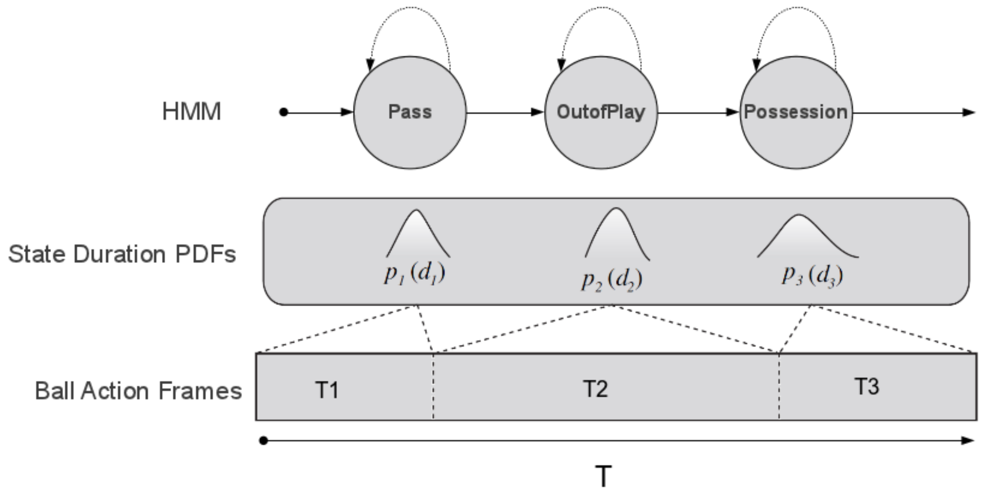


FIGURE 5.5. Example of a ball state duration synthesis.

5.14

$$P(q|\lambda, T) = \prod_{t=1}^{T} a_{q_t-1q_t} \tag{5.20}$$

where $a_{q0q1=\pi_{q1}}$.

The solution to the optimization problem can be completely determined if the value of $P(q|\lambda, T)$ can be obtained for every possible sequence. However this solution is impractical since too many combinations exist for q. If the state duration is controlled by self-transition probability, the state duration probability density associated with state i follows a geometrical distribution

$$p_i(d) = (a_{ii})^{d-1}(1 - a_{ii}) \tag{5.21}$$

where $p_i(d)$ represents probability of d consecutive observations in state i, and a_{ii} is self-transition probability associated with state i. The mentioned exponential state duration probability density is inappropriate for controlling the state duration. Therefore the state duration distributions for ball actions are modeled by Gaussian probability density functions as shown in figure 5.5.

In our model, the HMM is a left to right model with no skips since the ball should be at any of the defined states throughout the game. Then based on this reality, we can characterize the probability of the state sequence $q = (q_1, q_2, \ldots, q_T)$ as explicit state duration distributions. If $p_k(d_k)$ is the probability of being d_k frames at state k, then the probability of the state sequence q can be written as

$$P(q|\lambda, T) = \prod_{k=1}^{K} p_k(d_k) \tag{5.22}$$

where K is the total number of ball states visited during the all analyzed T frames. In a mathematical expression,

$$\sum_{k=1}^{K} d_{q_k} = T \tag{5.23}$$

We employed a single Gaussian pdf for modeling the state duration probability density because of the simplcity of the represention and validity. The state duration pdf is expressed as

$$p_k(d_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} exp(-\frac{(d_k - m_k)^2}{2\sigma_k^2}) \tag{5.24}$$

The optimized state sequence q^* is then obtained by using the Lagrange multipliers method of equation 5.22 as

$$d_k = m_k + \rho \cdot \sigma_k^2,\ 1 \leq k \leq K, \tag{5.25}$$

$$\rho = \frac{(T - \sum_{k=1}^{K} m_k)}{\sum_{k=1}^{K} \sigma_k^2} \tag{5.26}$$

where m_k and σ_k are the mean and variance of the duration distribution of state k, respectively as seen in figure 5.5. In our method ρ is assumed to be zero since the ball actions happen at each frame at a constant rate. The hidden states through an state are uncovered by using the state distribution pdfs.

The classification of the observations are accomplished through a decision tree based method proposed on the rules proposed above. The decision tree for ball action recognition is depicted in figure 5.6.

The action states are determined based on the rules defined through the tree-based approach depicted in figure 5.6. However ball kinematics or player positions might not be always available or measurements might be misleading. The hidden states are determined based on the optimization of the state sequences according to the duration pdfs proposed above. Example sequences of ball action recognition algorithm are depicted in figure 5.7.

In the typical HMM method the state is extracted according to the last state and the state transition probability. Since we cannot determine a static transition probabilities for transitions

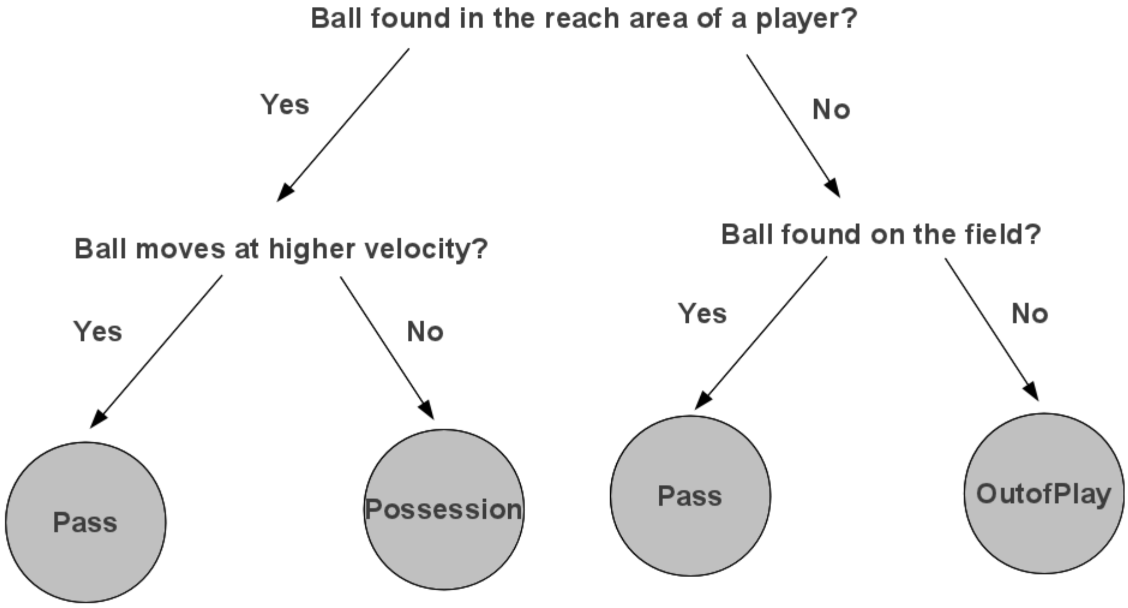
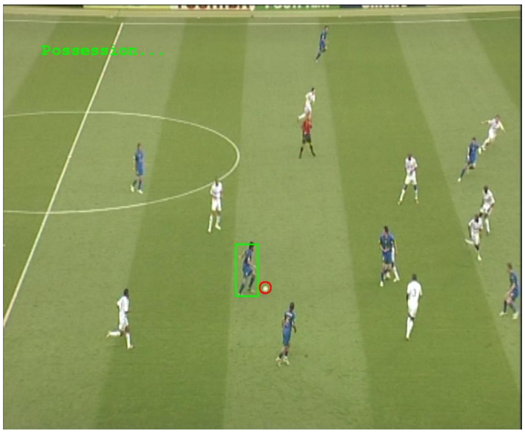


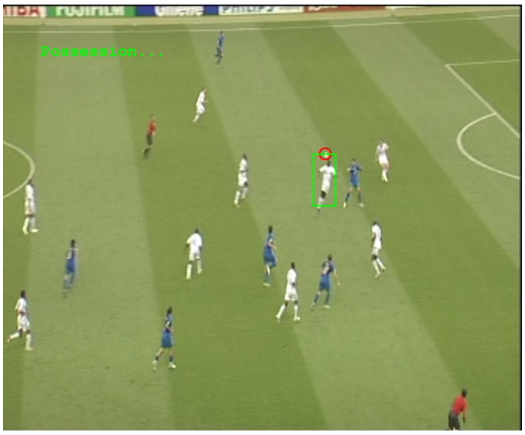
FIGURE 5.6. Ball action decision tree.

due to the characteristics of the game, we proposed and implemented a tree based HMM approach. Here we belong to the general characteristics of the HMMs regarding that the next state depends on the current state and the transition probabilities. But here we derive the state transition probabilities using the last and the next steps in order to extract the hidden states.

As seen in the figures depicted in figure 5.7, the proposed algorithm has showed a well accuracy on the classification of the ball action states.



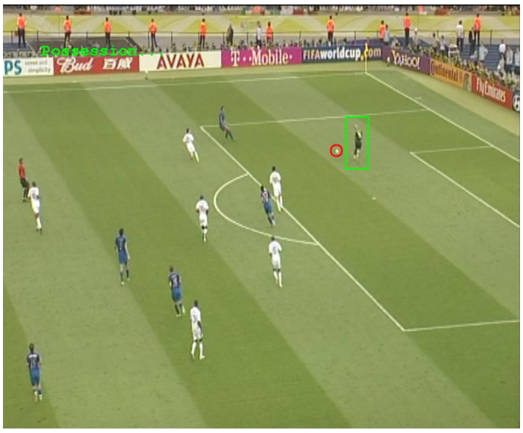
(A) Action:Possession



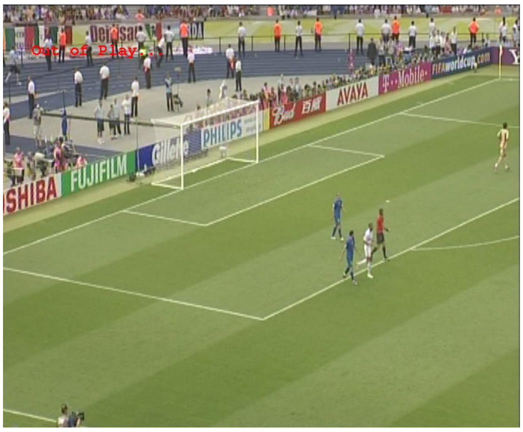
(B) Action:Possession

(C) Action:Pass

(D) Action:Pass



(E) Action:Possession



(F) Action:OutofPlay

FIGURE 5.7. Ball action recognition examples.

5.4 Detection of Important Scenes

Soccer is always played in a controlled environment with the official and global predefined rules. The general actions happening throughout the game can be estimated generally in most cases. [Xu et al., 2001] and [Xie et al., 2004] segmented the soccer game into play and break classes by respective visual patterns such as shot type, camera motion during play, break and state transitions. Some methods focused on the extraction of the excitement based on the observation of the proposed low level features [Hanjalic, 2005]. In [Tjondronegoro et al., 2004] employed the whistle sounds, crowd excitement and text boxes to summarize sports video content.

However significant actions happening throughout the game always take place around the ball. Therefore, ball detection and tracking provide us with the more important key features in order to build the semantic analysis of the game. In this thesis, we have proposed and developed methods in order to detect important scenes based on the ball state information. The method is then confirmed automatically using the scene calssification part of the camera parameter estimation system of [Gedikli, 2008].

Since the ball is the main element of the game, in normal conditions, it is always in the center of the broadcast camera's field of view. Additionally, in many situations camera motion is adjusted according to the ball motion. However, camera movements does not provide much detailed information about the ball motion and the scene content. In [Thomas, 2006] the authors tried to estimate ball motions using a motion sensor mounted on the camera, but obviously it does not provide reliable and extensive information about the game. We approach the problem in another way by extracting the scene content and the ball localizations. Our method might also be applied to feed the sensor mounted on the camera and move the camera according to the estimated ball trajectories.

The first feature we have investigated depend on events where the ball is not seen at all and almost all of the frames contain pixel values from the grass and the players. This case occurs when a player is injured or the referee had a decision to restart the game from a different position. In such cases the situation is also confirmed with and increased camera zoom parameter. Example scene is depicted in 5.8.

Another interesting scenario that we investigated was the detection of the ball actions close to the goal areas. Those scenes include to most interesting and exciting parts of the game. The main aim of a typical soccer game is to score a goal against the opponent team. In other words, the goal of every game tactic is then to approach a new score is by carrying the ball close to the opponents goal area in order to increase the probability of scoring or to defend

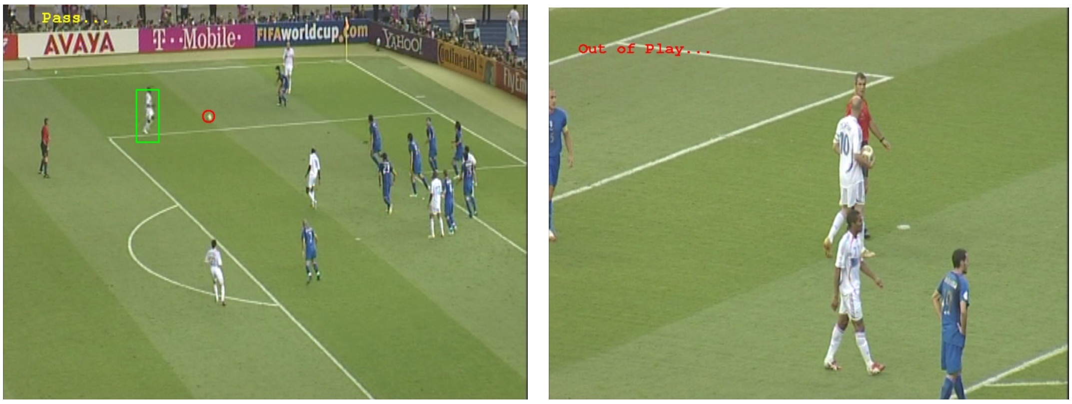


FIGURE 5.8. Camera zoom on critical situation and after player injury.

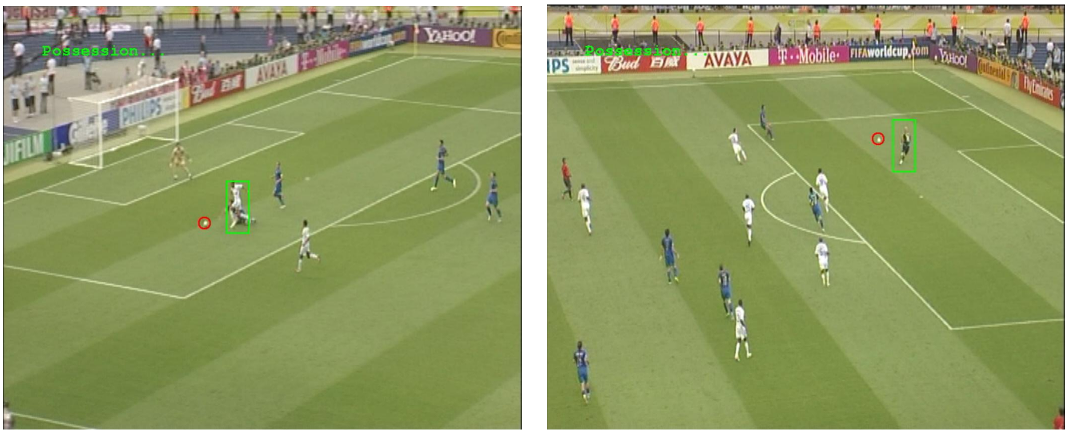


FIGURE 5.9. Example of important scenes detected close to the goal areas.

in front of the own teams goal to from a scoring opportunity. Those actions always happen around the goal area in case of scoring or defending. Thus scenes where the ball is close to the goal areas provide us critical positions where the analysts or the audience should pay more attention on. The features that we used for this scene detection and classification are the ball occurrence and the field detection. For this reason we first calibrate the camera in order to calculate the distance and proposed a critical level measure to detect the important scenes. An example scenario is depicted in figure 5.9.

In this section we developed two methods in order to gather general scene information using image processing methods. The first method is already implemented in the literature by the studies given above [Hanjalic, 2005, Tjondronegoro et al., 2004, Xie et al., 2004, Xu et al., 2001]. For the important secene classification the novelty we proposed in addition to the literature is the recognition of the important scenes based on ball kinematics. As the ball is played close to the goal area the audience gets more excited and interested in the game.

5.5 Actions Happening in Neighborhood of the Ball

In a soccer game, all the players move as a team based on a tactic of their team. Each movement of the players aims to belong to this tactic in order to have more control of the ball. In this section, we present an novel feature that we found out during the experiments.

The theory to be implemented here based on the correlation of the position of the ball and the players. Two teams change their positions according to the motion of the ball. In this part we observed the center of the detected players and it is relation to the position of the ball. Our experimental results proved us the correlation between the middle point of the detected players and the ball position.

An example graph showing the change of the distance between the ball position and the middle point of the players is depicted in figure 5.10

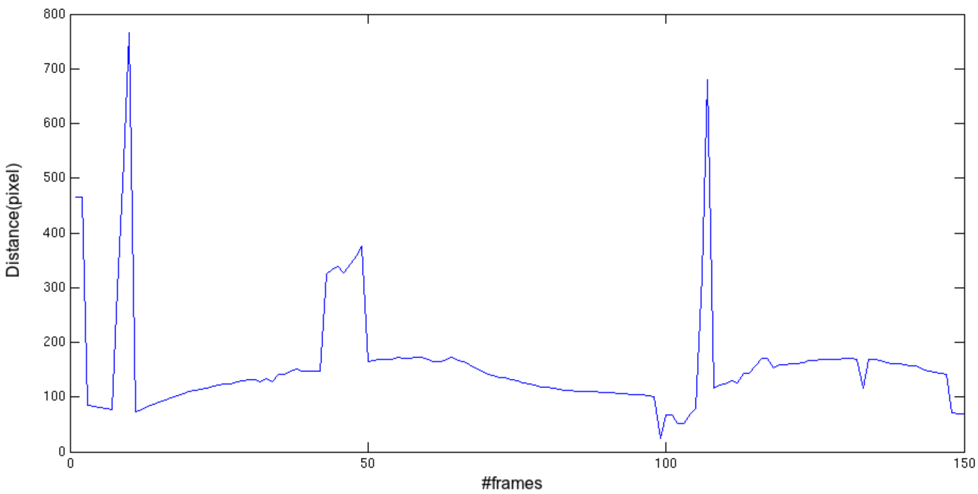


FIGURE 5.10. Distance to the middle point of the detected players.

The graph in figure 5.10 also confirms our hypothesis so that ball is the main component of a soccer game and the players take actions according to the ball. Since each moving player aims to get the possession of the ball, they take an action towards the ball which is balanced by the opponent team. This balanced action can be observed in the graph depicted in figure 5.10. The outlier areas occured because of the ball tracking errors due to many player occlusions.

5.6 Experiments

Ball action recognition using novel motion cues is one of the major novelties proposed in this thesis. Ball actions and its interaction with the players are inevitable for an intelligent automated soccer analysis system. In this part of the thesis, for the recognition of the ball actions, we did our evaluations on soccer and hockey videos as mentioned. However we focused on the applications in broadcasted soccer videos for important scene classification since the main goal of the thesis is the analysis of soccer videos. We tried to cover all the possible cases where the ball is in active play and in action for our dataset. Ball play including the possesion and pass constitutes the most part of a typical soccer game.

5.6.1 Ball Action Recognition

Ball tracks and positions provide a very important amount of valuable information for further level analysis. In this chapter, the fusion of the player positions and the ball tracks are investigated and novel motion cues are proposed. The ball action classes are defined as $Pass$, $Possession$ or $OutofPlay$. The definition of the actions are evaluated as it is explained in section 5.2. The experiments on the presented dataset showed promising results. Table 5.1 shows the detailed information about the experiments. The frames and scenes for the analysis chosen as the general cases that happen mostly in a typical soccer game. However we also included the scenes where the ball is heavily occluded and attacked by more players as explained in previous sections. In that manner, our dataset represents the general characteristics of a typical soccer game.

Game	Camera	#Frames	Correct Classifications	False Classifications	Accuracy
Italy - France	Broadcast	10000	8485	1515	0.85
Italy - France	Highangle	10000	8946	1054	0.89
England -USA	Broadcast	10000	8283	1717	0.83

TABLE 5.1. Ball Action Recognition Results

Same structure has been implemented on the hockey ball and player positions. The results were not as accurate as in the soccer game sine the number of the hidden states are much more than the hidden states in soccer videos. The results depicted in figure 5.2 also proves that reality.

5.6.2 Important Scene Detection

Another interesting work that has been investigated through the thesis was the extraction of the important scenes using the visual content included within the frames. Up to now most of

Game	Camera	#Frames	Correct Classifications	False Classifications	Accuracy
Mumbai Marines - Bhopal Badshahs	Broadcast	5000	3118	1882	62.4
Mumbai Marines - Bhopal Badshahs	Broadcast	3000	1683	1317	56.1

TABLE 5.2. Hockey Ball Action Recognition Results

the work has been focusing on the audio cues in order to detect exciting scenes. Using the methods explained in Chapter 5.4 the results are depicted in table 5.3.

Game	Camera	#Frames	Important Situation Detected	Accuracy
Italy - France	Broadcast	5000	4951	0.99
England -USA	Broadcast	5000	4890	0.98
Korea - Greece	Broadcast	5000	4894	0.98

TABLE 5.3. Important Scene Detection Results

Tables presented above provide the reader a deeper understanding of the experimental results. Our work focuses on the soccer game videos regarding the complexity and the challenges of the problem from the computer vision perspective. Through the experiments many challenging problems are investigated as presented in the previous chapters. The novelities of our work mainly depends on the tracking and action extraction of the ball object in a game. Based on above results and current literature we can conclude that our work has been contributed to the computer science regarding the novel approach to the tracking problem for broadcast soccer videos, feature extraction and action recognition.

5.7 Conclusion

In this chapter, we investigated higher level analyis in ball team games. The motion cues based on ball and player localizations are proposed and evaluated. We built an action recognition module based on the ball dynamics. In our method, we derived metrics for those definition models of the ball states based on ball position, ball velocity and player localizations. The new measures defined for ball action recognition showed a very good performance in the extraction of the high level events throughout the game. The ball player interaction scenarios are studied and presented extensively.

Ball velocity is one of the important factors in defining the ball states, especially during occlusions. For this reason, we developed a ball speed estimation model based on ball position on the tracks. This speed estimation is built on the motion model proposed.

We evaluated the mentioned algorithms on the final game of FIFA World Cup 2006 and two game from FIFA World Cup 2010 and broadcast hockey video game between Mumbai Marines and Bhopal Badshahs of Bridgestone World Series Hockey 2012. Throughout the experiments we observed promising results and improved accuracy than the current state of the art approaches. Using the ball features as the major point of the approach has let us be able to precisely recognize the actions during the game.

CHAPTER 6

Web Based Sports Monitoring

Advances in the web based video analysis have been increasing the last decade with the development of computational performance and the bandwidth of the communication channels. Web based sports analysis constitutes one of the most important fields in web based video analysis. The audience has an increasing interest in the analysis of their team as much as the trainers. They would like to see the automatic analysis of their teams throughout the games to highlight and gather more detailed information. Installation procedures and portability is one of the main effects in the usage of a program, especially for inexperienced users. Therefore, we have proposed and developed a web based Ball Observation System in this thesis in order to provide a platform independent system available to more users. The system mostly aims for the analysis of broadcast soccer video data. This application aims to gather experiences from various users like coaches, players and fans to improve the architecture and generality of the system. Proposed web system has also been designed in a more user friendly way than the ASPOGAMO Ball Observation System interface in order to make it available as many users as possible.

6.1 Related Work

Although many manual and automatic sports analysis tools have been proposed in the literature, the number of the web based systems are still limited. Most of the existing approaches focus on the graphical representations based on the manually annotated data rather than running the automated analysis on a web based architecture. In [Snoek and Worring, 2003], authors has presented a so called Time Interval Maximum Entropy (TIME) framework for the web based multimodal indexing of events in soccer videos. [Snoek and Worring, 2003] propose efficient solutions on the representation, inclusion of contextual information and synchronization and fusion of the information from various sources.

Tango Online system of Match Analysis Inc. also provides a web based soccer analysis system [Match Analysis Inc., 2013]. The system works basically in four steps, uploading of the data to match analysis system, manual interaction of the game analysts for the localization and classification tasks, saving the processed data to the database and interpretation of the results and visualization to the players. But as stated, many of the processes are done by the manual work of trained professionals. It does not provide automatic analysis of the game like the ball observation system presented in this thesis.

Another web based game analysis system is the ESPOR system of espor Inc. ESPOR system also provides the users statistics about the sports games [espor Inc., 2013]. Different than the Tango online system mentioned above, they provide statistics for five different sports including the soccer. The system is said to be available in five different sports namely, soccer, basketball, lacrosse, volleyball, ice hockey. However, the system does not provide further automatic analysis and it does not do more than a online video editing and annotation software. The salaried professionals annotate the actions in required videos in different levels based on the price of the product. As a user friendly online video editing tool it is easy to use but as mentioned it does not include the automated tracking and action recognition systems presented in this thesis.

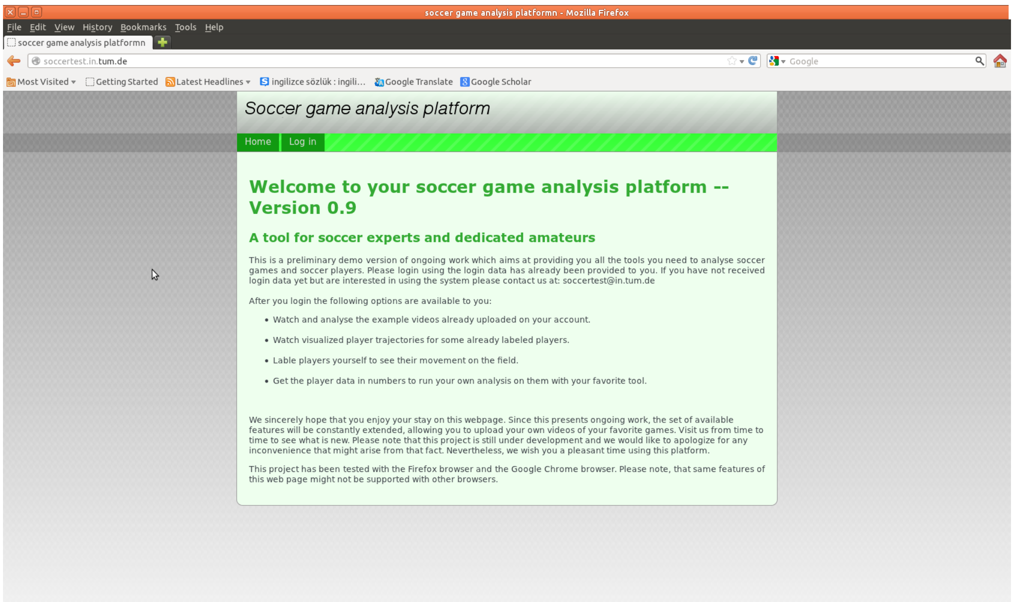


FIGURE 6.1. Web System Entrance.

6.2 System Architecture

As presented in section 6.1, the existing web based sports analysis systems depend on the manual annotation of video after uploading it to the web site. Those implementations provides the user only a nice representation of the common positions but not real automated analysis. [Bigontina, 2011] has proposed a web system architecture for the player detection of ASPOGAMO in his work. In this chapter, we present the web based Ball Observation System and its integration to the existing ASPOGAMO web system.

The entrance to the proposed web system is depicted in figure 6.1.

Before we continue to the details of the system developed, we would like the present the reader state of the art technologies in web based analysis tools design. While providing the reader the necessary knowledge for the rest of the chapter, we also explain the reasons stated for the usage of the mentioned systems.

6.2.1 Web Technologies

Web technologies have been emerging rapidly in the last decade. In this chapter, the integration and embedding of the Ball Observation System into a web application tool is investigated. Such a system is designed to allow the video analysts, trainers and audience to upload videos and run the tracking system for their videos from anywhere. As a result of this integration, the analysis system becomes platform and location independent. In the following section details of the proposed system is explained along with the already existing ASPOGAMO Web analysis tool.

6.2.2 Web 2.0

Web 2.0 is the last generation of stable web technologies which was firstly presented by Darcy Dinucci. Nowadays, the last generation of stable web technologies reached is Web 2.0 which is firstly proposed by Darcy Dinucci and she explained in her article Fragmented Future [DiNucci, 1999]. *"The Web we know now, which loads into a browser window in essentially static screenfuls, is only an embryo of the Web to come. The first glimmerings of Web 2.0 are beginning to appear, and we are just starting to see how that embryo might develop. The Web will be understood not as screenfuls of text and graphics but as a transport mechanism, the ether through which interactivity happens. It will [...] appear on your computer screen, [...] on your TV set [...] your car dashboard [...] your cell phone [...] hand-held game machines [...] maybe even your microwave oven."*

As she stated, the main feature of the Web 2.0 is the interaction of the users rather than just retrieving information. Web 2.0 allows the users to participate and contribute to the application that they use. [Best, 2006] stated the major highlights of the Web 2.0 is the rich user experience, user participation, dynamic content, metadata, web standards and scalability.

Key features of Web 2.0 presented in [WebAppRater, 2013] as,

- Free classification of information
- Rich user experience
- User contribution
- Long tail
- User Participation
- Basic Trust
- Dispersion

Regarding the Web 1.0, Web 2.0 has brought novelties in Rich Internet Application (RIA), Web-oriented architecture as well as social web. RIA improves the graphical visualizations and the usability of the web pages. RIA usually refers to client side technologies used in Web 2.0 including Ajax and JavaScript. Javascript is used to upload and download new data from the web server without reloading the whole page use by Ajax programmmers. Examples of Web-oriented architecture applications include RSS (Rich Site Summary) and Web services. Web 1.0 was composed of static web pages but in Web 2.0 third party plugins like Adobe Flash or Javascript dynamic animations and more aesthetic layouts can be developed. The social web made the interaction of users more effective by allowing the users to interact with the webpage while giving them the availability to send and receive data from the server asynchronously. In Web 1.0 the user had to wait for the data to be reloaded before they can do anything on the webpage. This routine always made the whole interaction slower and made the user wait for a page to complete the reload. Simply the mentioned asynchronous data exchange increases the performance of the website and the requests are completed quicker than independent of queueing of the bytes. The data types fetched is usually formatted in XML (Extensible Markup Language) and JSON (Javascript Object Notation). Those data types might then be used like in desktop applications by using the DOM (Document Object Model).

On the server side, Web 2.0 uses the same technology as Web 1.0 Languages. Web 2.0 mostly changed the way the data is exchanged and the novelties and efficiencies occur in that field. In other words the Web 2.0 forms a more participatory web belonging to the same server side technology as Web 1.0.

Web 3.0 has been under development recently but it is not as stable as Web 2.0 yet. Briefly, the Web 3.0 focuses on more semantic interpretation and personalization. This means that the computers will be making logical decisions and generating information with the humans. In summary, the users in Web 1.0 was in passive mode and the communication was synchronous, in Web 2.0 users were able to interact with the computer and the communication was asynchronous and Web 3.0 makes the computer make decisions based on the semantics and produce information for the user in addition to Web 2.0. But as mentioned, Web 3.0 is not as mature as Web 2.0 yet.

6.2.3 HTTP Protocol

The asynchronous data communication presented in section 6.2.1 is accomplished through the Hypertext Transfer Protocol (HTTP). Here the hypertext means the multi-linear set of objects building a network by using hyperlinks between the nodes.

If a client wants to view a website his/her user agent (web browsers, web crawlers, etc.) has to download the necessary contents from the server. The format of this data transfer was first defined in HTTP/0.9 but was improved and than defined by the World Wide Web Consortium (W3C) in 1990 as HTTP/1.0 by Request of Comment (RFC) 1945 [Berners-Lee, 1996]. The current version is HTTP/1.1 and was specified in RFC 2616 in 1999 [Fielding et al., 1999].

The structure of HTTP is a request/response protocol which is implemented in ASPOGAMO Web-Analysis tool as depicted in figure 6.2.

On the server side of the HTTP requests a web server is responsible for listening to the specified port and handling of the incoming data. The HTTP server parses the received requests and provides the required data for the clients. The implementation of the server side also uses the HTTP to provide the data the requested by the user.

Currently the widely used web servers are running Apache systems [Netcraft, 2013]. Apache is a robust widely used and robust server which shows less rbost performance in very large number of connections. Regarding those main specifications, the Apache web server interface is implemented knowing that the system will not be used for very large number of users at the beginning and system will be running very stable on the server side.

On the client side many standard tools exist like JavaScript, Ajax and Flash. The basic HTML is very difficult to use for frameworks that include the usage of the videos like in our case. HTML5 is capable of displaying the videos but when it comes to the processing of the video it does not provide enough functionality. Because the Ball Observation system

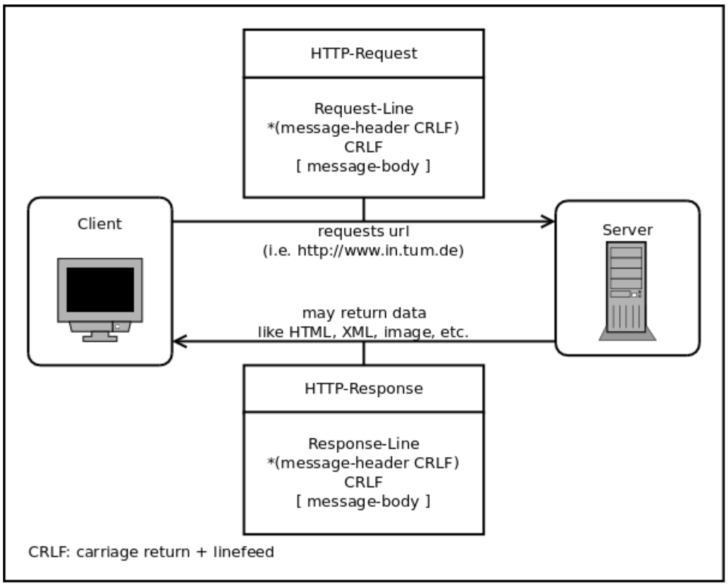


FIGURE 6.2. Diagram of the HTTP request/response protocol [Bigontina, 2011].

requires the editing of the video to cut the corresponding scene as well as the generation of the template. Additionally the correction of the tracks is required. At the first look flash seems to be the ideal solution regarding the display of the videos but the disadvantage of flash is that it only preserves a rectangular area for processing and it is not suitable for applications where the semi manual or manual processing of the video required. For this reasons, it is not suitable for the architecture proposed in ASPOGAMO web system.

The most important specification of our system on the client side is that the system quickly responses to user inputs and provides the requested data. This has been solved by implementing Ajax and XMLHttpRequests. The data transmission is done through JSON (JavaScript Object Notation) regarding its readability and efficiency.

Additonally a database is also essential for the web analysis system. For this purpose, MySQL is used for storing that data regarding ease and the widely offered sources.

To sum up, on the server side the Apache web server is used with Python programming, on the client side JavaScript is used for DOM manipulations and asynchronous requests with MySQL as the database.

6.2.4 Implementation Model

The website developed builds an interface for the Ball Observation System of Aspogamo project. The web system is designed using Python programming language and the Ball Observtaion system is adapted to the existing system using an interface between $C++$ and

FIGURE 6.3. User Interface of Web Based Ball observation system.

Python modules. Basically it allows the users to upload videos, provide game information and perform the analysis on their videos. The website is designed in three separate areas of access, namely public area, user area and administration area.

The public area is available to all users on the web and provides general information about the project. The public part of the web system is as shown in 6.1.

The users need an account to get access to the user area for analyzing their own soccer videos. In the first step, they are asked to create an account that would be associated with their videos. This is required to enter the user area as in figure 6.3.

Then user should then use this account to log in and upload videos to the database. After the user logged in, he has access to the analysis part of the web system that allows him to process his own data.

After the videos are uploaded to the database they are firstly converted to one of the already defined formats. Then the video becomes ready for the analysis and an analysis screen is opened in front of the user. There the user can run the ball tracker and also manually edit the tracks if it is needed. An image of the web based ball observation system is as shown in figure 6.4.

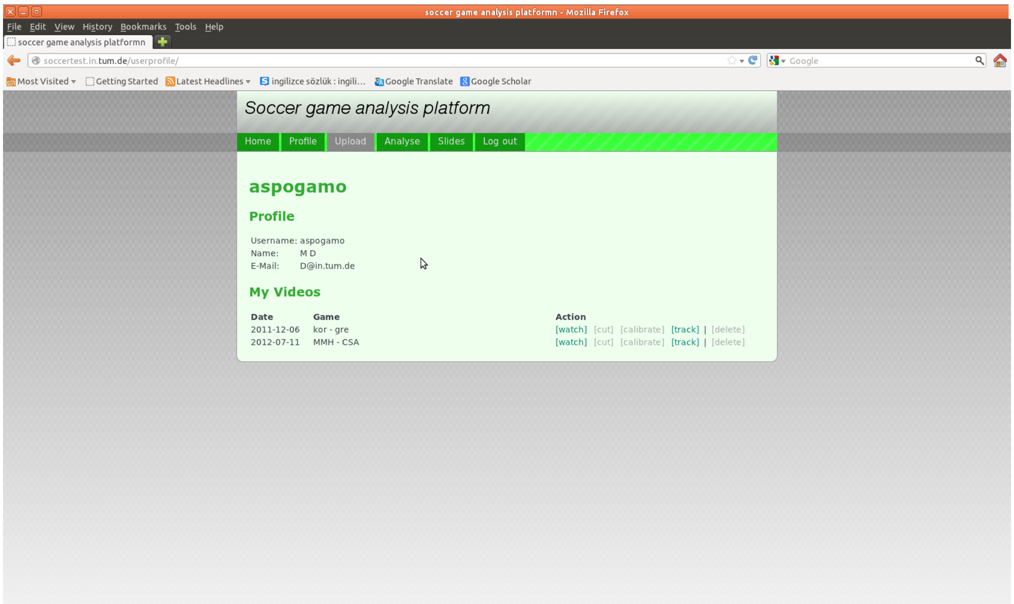


FIGURE 6.4. Main User Interface of the Ball Observation System Analysis.

The screen for ball tracking is as shown in figure 6.5.

FIGURE 6.5. Web Based Ball Observation System.

6.3 Conclusion

In this chapter, we presented the web based Ball Observation System interface. Web based sports analysis systems is an emerging application area with platform independent easy to use developments. Installation procedures and portability is one of the main effects in the usage of a program, especially for inexperienced users. The audience has an increasing interest in the analysis of their team as much as the trainers. In the web based Ball Observation System, we provide the users an access to the ball tracker algorithm developed throughout the thesis. The interface has been explained and possible application areas are discussed.

They would like to see the automatic analysis of their teams throughout the games to highlight and gather more detailed information. Feedback to the system from various users like coaches, players and fans will improve the architecture and generality of the system. Proposed web system has also been designed in a more user friendly way than the ASPOGAMO Ball Observation System interface in order to make it available as many users as possible.

CHAPTER 7

Conclusion and Perspectives

In this chapter, we present the reader a general overview of what is presented throughout the thesis and conclude our work with a discussion about the methods developed. Following the discussion about the outcomes and the novelties, we then describe the possible future perpectives that can be developed on our work.

7.1 Summary

The thesis is written in a way to give the reader a complete understanding of the scenario proposed and investigated. However the chapters can also be read independently easily and provide the reader different perpectives to the work presented. We can brielfy describe the thesis to introductory , single and small object tracking theory and applications, action recognition, web developments and conclusions. The introductory chapters described the general outlook idea investigated and described the challenging problems solved. In this chapters, we provided a very detailed survey about the problems and solutions in the scientific literature. Following the definition and the state of the problem in the literature, we explained the theoretical background of the problem with related proof of the concepts. We then described our technique by reasoning on the developments and the advantages. The details of the proposed detection and tracking algorithms are explained in detail and results of the experiments are depicted. Following the low level detection and tracking, the next chapter provided the reader more insights into the ball action measures and recognition algorithm developed. In the last part of the thesis, an architecture of the web based Ball Observation System is explained and presented.

7.2 Conclusion

In this thesis, tracking of the ball in sports videos are presented as well as the detection of the players for extraction of ball actions. The proposed approach for tracking includes novel small and single target tracking methods in a controlled environment. The controlled environment means that the area and the way of interaction of the target with its environment are predefined up to a limit. As an application we chose the sports video domain and used the broadcast soccer videos and field hockey videos in order to realize and test our approach. The cues proposed and fusion of the target kinematics with the nature of the game have improved the accuracy of the results of the tracking approach. The development of the particle filter that is designed for tracking small and single targets in soccer videos was one of the major contributions of this thesis. The deeper analysis of the trajectory has been also used to classify between the flying and rolling states of the ball in 2D image plane. The features invented have showed promising results that also proved the accuracy of our method. In addition, the evaluations on the hockey dataset have also proved the generality of the method developed while tracking the ball.

In addition to the soccer ball tracking approach developed, we have also proposed new measures for the recognition of actions and designed a system in order to extract the higher level movements of the target. As the ball being the main component is decision making, we proposed measures to detect the important scenes as well as the actions happening around the ball. As ball plays the key role in the game, the decision taking the ball state as the center made the recognition of actions around the ball feasible to recognize. Analysis on the ball and player interaction presents the results and the feasibility of the automatic ball and player action recognition system. The ball player interaction method has solved the action classification by using the novel features used to define possession, pass and out of play actions. Such features are mainly dependent on the ball characteristics and kinematics that had been tracked through the game. The most difficult problems were the tracking of the ball and the association of the hypotheses with the true target. The accurate tracking has been achieved using the smart resampling algorithm proposed and the invariant and reliable features for the classification of the hypotheses and actions. The rolling and flying ball state classification has been proposed on the characteristic of the trajectory. Up to the best of our knowledge our work presents the first results on such state classification of a rolling or flying object in 2D based on the trajectory and the motion cues defined. It is based on simple features but provided very reliable and accurate results during our experiments.

Ball action recognition represents higher level events are determined automatically based on the ball kinematics and the positional information of the detected players. In this thesis, our aim was to robustly track the ball during the game and associate it with the player position information gathered. The tracking results were encouraging regarding the other works have been done in this field. The system is also capable of extracting more high level information based ball player interactions. Ball action recognition has presented the classification of the scenes in typical game based on the ball states. Here the nature of the game and the rules are used to propose decisions. The classification is based on the association of the information from the ball tracks and the player positions. The association of such information had been built in order to define the actions happening during the game.

In addition to the algorithms developed for tracking and action recognition, we have also proposed a web based Ball Observation System system for the usage of the tracker for web users. The architecture and the design processes have been discussed and explained in detail. Such web applications allow our system to get the feedback from various type of users in order to improve the algorithms developed and proves the generality of the system performance.

7.3 Future Directions

The detection and tracking algorithms have been evaluated in long image sequences and provided robust and accurate results. Since ball is the major element of the game and actions take around it, the camera always try to focus on the ball. Therefore the results of the tracking can be used in any other means than software development. For instance the ball localisations can ben fed to a microcontroller unit to manage the broadcast camera movements throughout the game. Other than the camera motion management, camera selection might also be accomplished according to the content in multi-camera configurations.

The importance of the data can also be used to design other software systems. One of the future developments might be the integration of a speech generation system with the already developed ball detecting and tracking system. Such integration might be implemented in order to develop an automatic artificial sports game reporter.

APPENDIX A

Screenshots

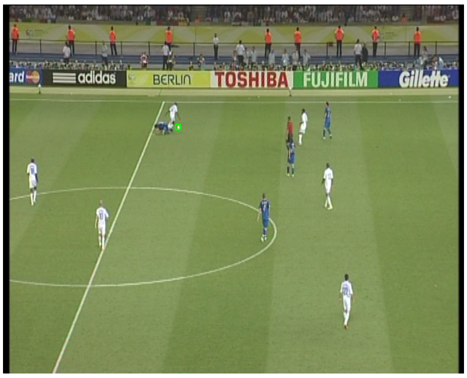
A.1 Ball Detection

(A.1.1)frame.

(A.1.2)frame+45.

(A.1.3)frame+75.

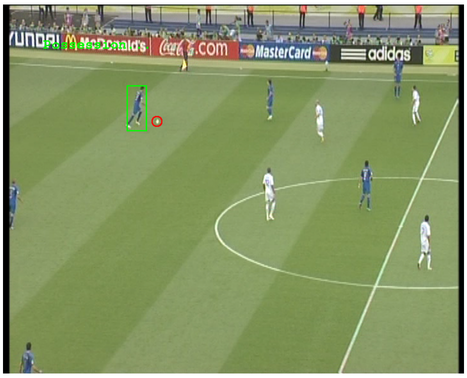
(A.1.4)frame+110.



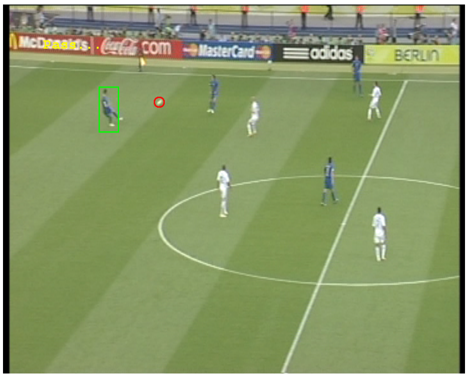
(A.1.5)frame+173.

(A.1.6)frame+181.

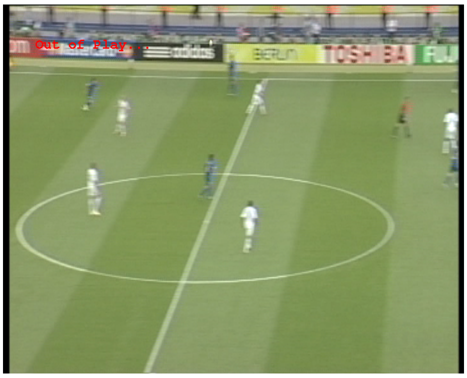
A.2 Ball Tracking



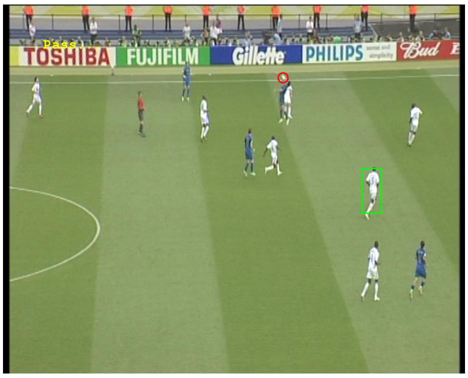
(A.1.1)frame.



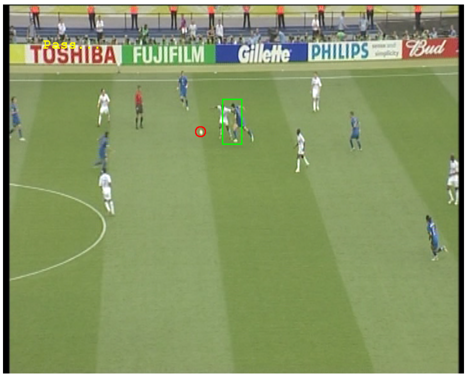
(A.1.2)frame+17.



(A.1.3)frame+31.

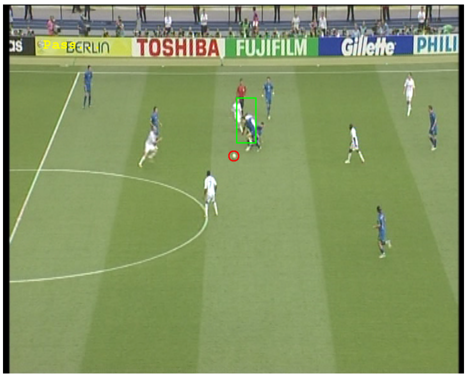


(A.1.4)frame+64.

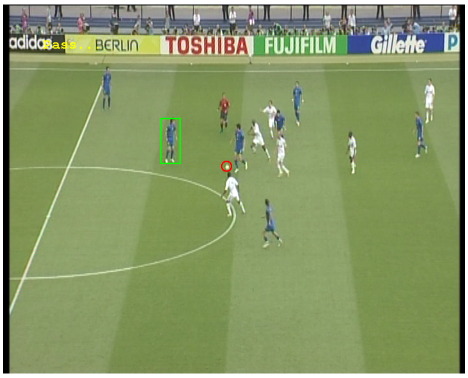


(A.1.5)frame+121.

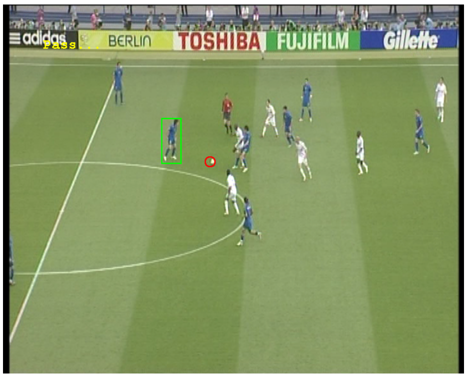
(A.1.6)frame+127.



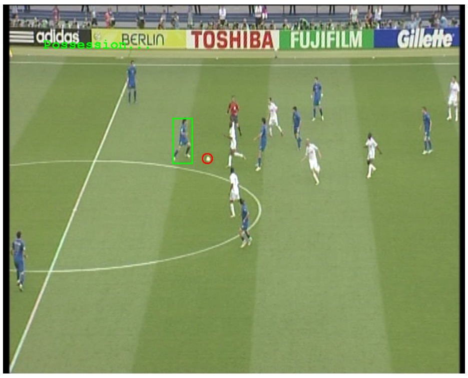
(A.1.7)frame+163.



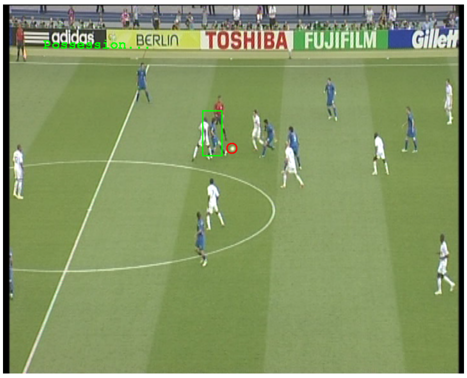
(A.1.8)frame+207.



(A.1.9)frame+218.

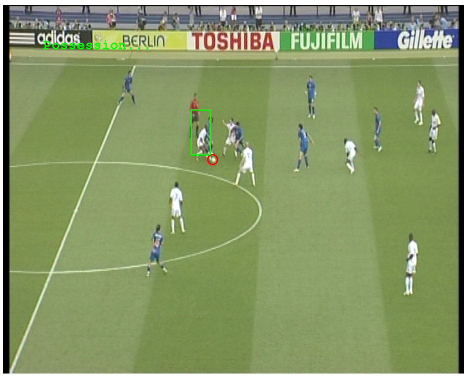


(A.1.10)frame+224.

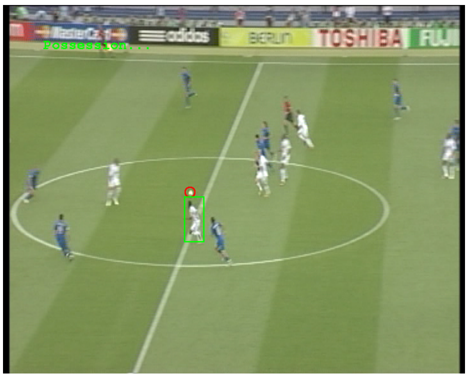


(A.1.11)frame+250.

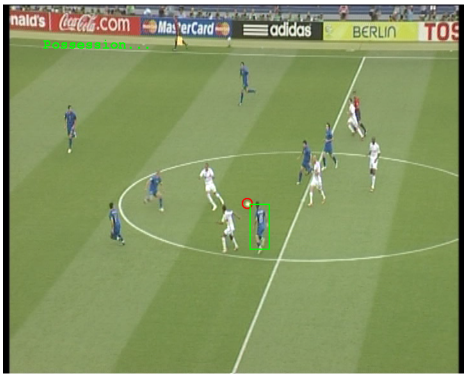
(A.1.12)frame+253.



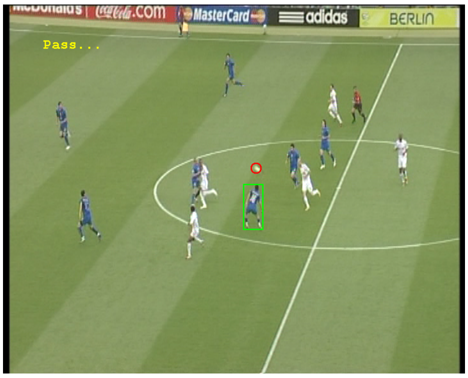
(A.1.13)frame+267.



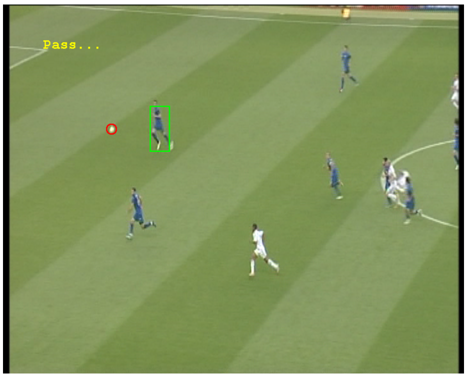
(A.1.14)frame+304.



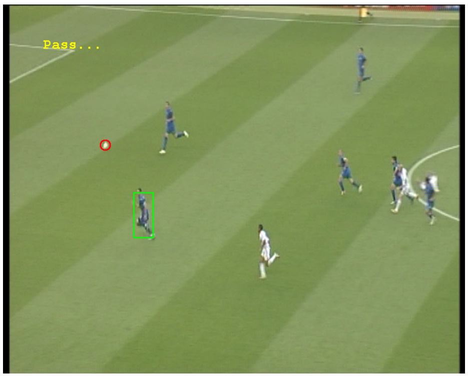
(A.1.15)frame+318.



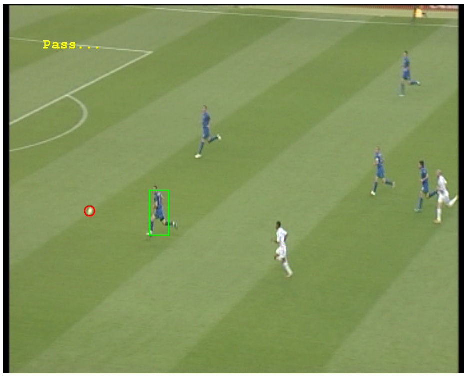
(A.1.16)frame+333.



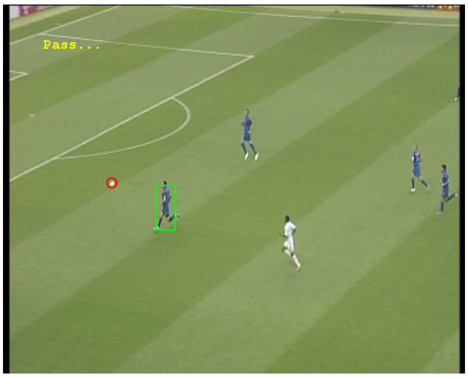
(A.1.17)frame+365.



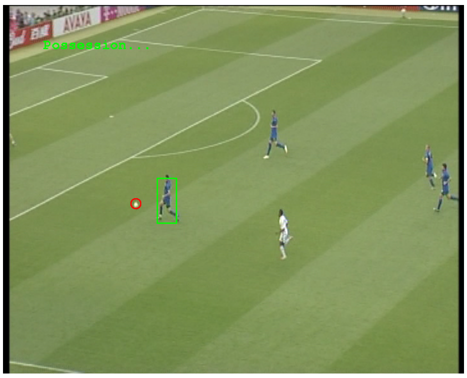
(A.1.18)frame+368.



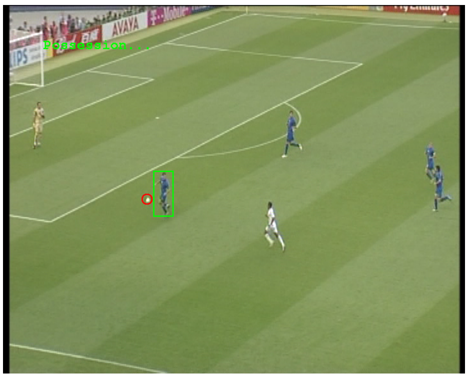
(A.1.19)frame+377.



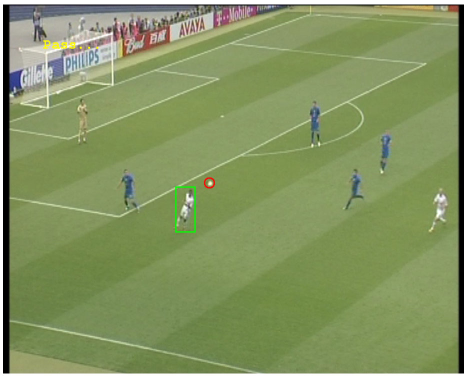
(A.1.20)frame+392.



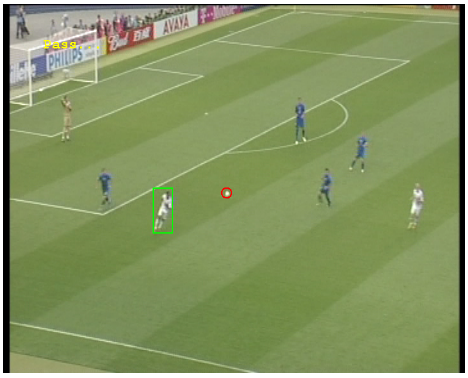
(A.1.21)frame+405.



(A.1.22)frame+415.

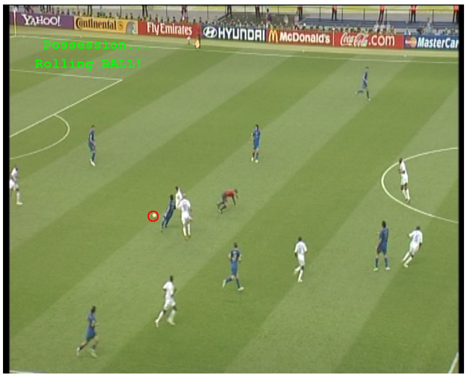


(A.1.23)frame+454.

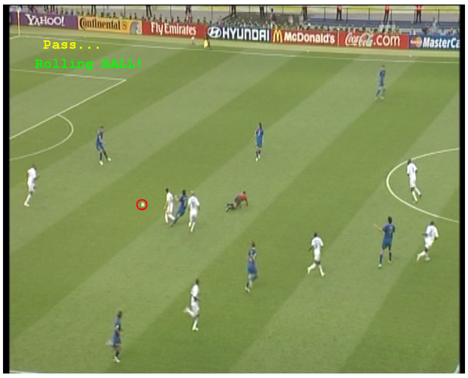


(A.1.24)frame+459.

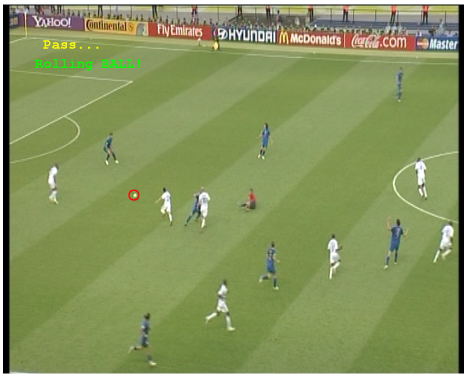
A.3 Ball Flying/Rolling



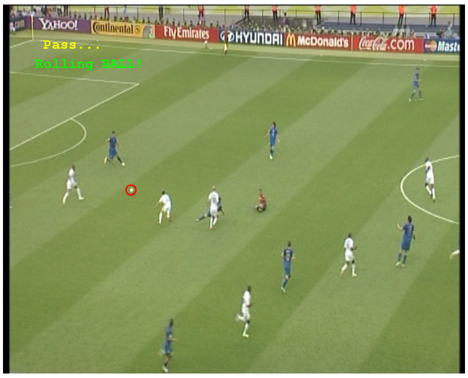
(A.1.1)frame.



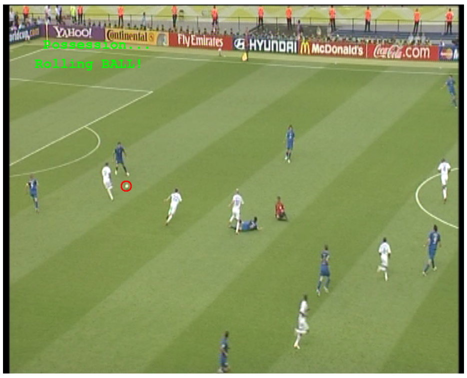
(A.1.2)frame+5.



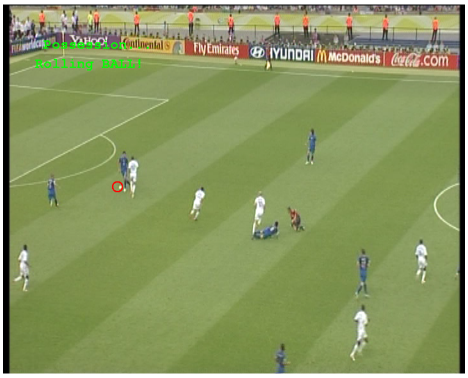
(A.1.3)frame+10.



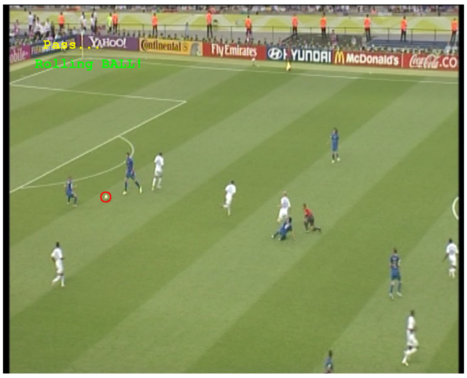
(A.1.4)frame+14.



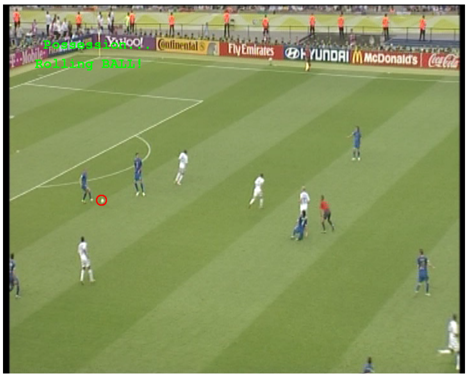
(A.1.5)frame+24.



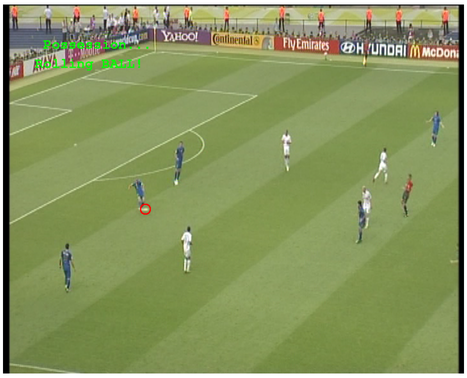
(A.1.6)frame+34.



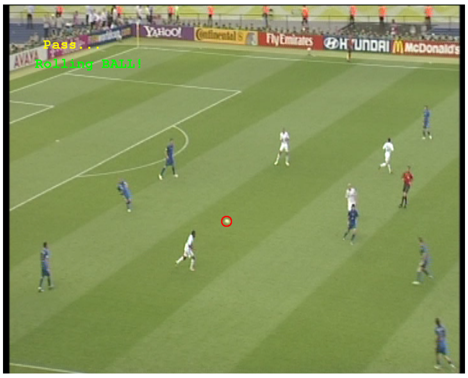
(A.1.7)frame+42.



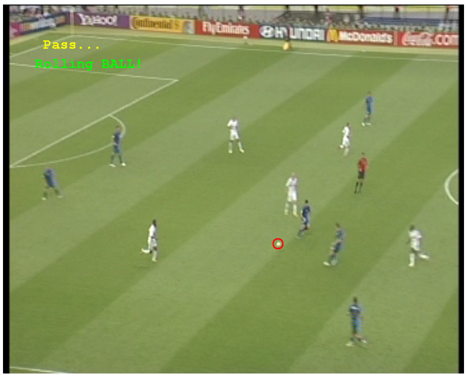
(A.1.8)frame+48.



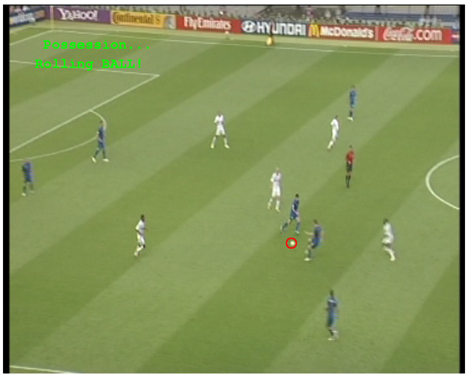
(A.1.9)frame+79.



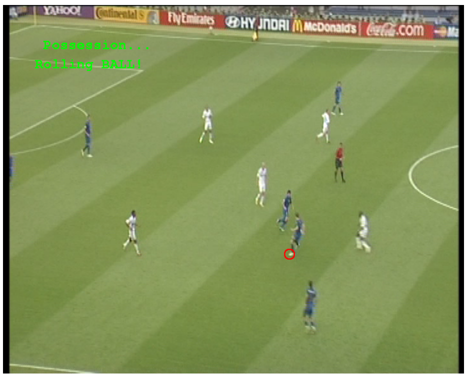
(A.1.10)frame+88.



(A.1.11)frame+98.



(A.1.12)frame+101.



(A.1.13)frame+104.

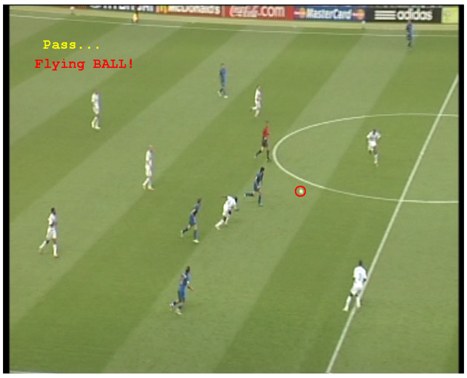
(A.1.14)frame+109.

(A.1.15)frame+113.

(A.1.16)frame+117.

(A.1.17)frame+121.

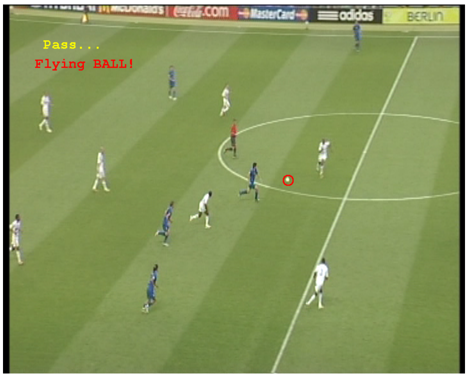
(A.1.18)frame+124.



(A.1.19)frame+130.

(A.1.20)frame+133.

(A.1.21)frame+137.



(A.1.22)frame+140.

Bibliography

A T Alouani, T R Rice, and N Auger. On the generalized nearest neighbor filter. *1990 Proceedings The TwentySecond Southeastern Symposium on System Theory*, 1990.

Vasilios Armatas, Athanasios Yiannakos, Dionisis Ampatis, and Paraskevi Sileloglou. Analysis of the successful counter-attacks in high-standard soccer games. *Inquiries in Sport and Physical Education*, 3(2):187–195, 2005.

M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

N. Babaguchi and R. Jain. Event detection from continuous media. *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, 2:1209–1212, 1998.

N. Babaguchi, Y. Kawai, and T. Kitahashi. Event based indexing of broadcasted sports video by intermodal collaboration. *IEEE Transactions on Multimedia*, 4(1):68–75, mar 2002.

J. Baker. Reducing bias and inefficiency in the selection algorithm. 1987.

L Ballan, M Bertini, A Del Bimbo, and A Jain. Automatic trademark detection and recognition in sport videos. *2008 IEEE International Conference on Multimedia and Expo*, pages 901–904, 2008.

Y Bar-Shalom and E Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460, 1975.

Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. 1987.

Sian Barris and Chris Button. A review of vision-based motion analysis in sport. *Sports Medicine*, 38(12):1025–43, jan 2008.

Ricardo M L Barros, Milton S Misuta, Rafael P Menezes, Pascual J Figueroa, A Felipe, Sergio A Cunha, Ricardo Anido, and Neucimar J Leite. Analysis of the distances covered by first division Brazilian soccer players obtained with an automatic tracking method. *Instrumentation*, 6(March):233–242, 2007.

L E Baum, T Petrie, G Soules, and N Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970a.

LE Baum and T Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 1966.

LE Baum, T Petrie, G Soules, and N Weiss. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The annals of mathematical statistics*, 1970b.

BBC News. London 2012: BBC gives first sight of Olympic titles. *http://www.bbc.co.uk/news/entertainment-arts-18647587*, 2012.

M. Beetz, B. Kirchlechner, and M. Lames. Computerized Real-Time Analysis of Football Games. *IEEE Pervasive Computing*, 4(3):33–39, jul 2005.

Michael Beetz, Thomas Stammeier, and Sven Flossmann. Motion and episode models for (simulated) football games: Acquisition, representation, and use. *3rd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1368–1369, 2004.

Michael Beetz, Suat Gedikli, and Jan Bandouch. Visually tracking football games based on TV broadcasts. *20th international joint conference on Artifical intelligence.*, pages 2066–2071, 2007.

Michael Beetz, Nicolai Hoyningen-huene, Bernhard Kirchlechner, Suat Gedikli, Francisco Siles, Murat Durus, and Martin Lames. ASPOGAMO : Automated Sports Game Analysis Models. *International Journal of Computer Science in Sport*, 8 (1), 2009.

Ron Begleiter, R El-Yaniv, and G Yona. On Prediction Using Variable Order Markov Models. *J. Artif. Intell. Res. (JAIR)*, 22:385–421, 2004.

Tim Berners-Lee. The Original HTTP. *http://www.w3.org/DesignIssues/OldDocs.html*, 1996.

D Best. Web 2.0: Next Big Thing or Next Big Internet Bubble? *Lecture Web Information System Technische Universiteit Eindhoven*, 2006.

Andreas Bigontina. Implementation of a Web Interface for Automated Football Analysis. *Bachelor Thesis*, 2011.

Samuel Blackman and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 70 edition.

R.S. Bucy and P.D. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. John Wiley & Sons, 1968.

Cairos Technologies AG. Vis.Track. *http://www.cairos.com/unternehmen/vistrack.php*, 2012.

J F Canny. *Finding Edges and Lines in Images*. PhD thesis, MIT, 1983.

George Casella and Christian P Robert. Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

Eric Hern X B Ndez Castillo, Ziziliz Zamudio Beltr X B N, and Juan Manuel Ibarra Zannatha. Soccer Ball Speed Estimation Using Optical Flow for Humanoid Soccer Player. *IEEE Electronics Robotics and Automotive Mechanics Conference*, pages 178–183, 2011.

Rick Cavallaro. The FoxTrax hockey puck tracking system. *Computer Graphics and Applications, IEEE*, (April):6–12, 1997.

Z Cernekova, I Pitas, and C Nikou. Information theory-based shot cut/fade detection and video summarization. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(1): 82–91, 2006.

Hua-Tsung Chen, Chien-Li Chou, Wen-Jiin Tsai, and Suh-Yin Lee. 3D ball trajectory reconstruction from single-camera sports video for free viewpoint virtual replay. *2011 Visual Communications and Image Processing VCIP*, pages 1–4, 2011.

Rong Chen and Jun S. Liu. Mixture Kalman Filters. *Journal of the Royal Statistical Society.*, 62:493–508, 2000.

Sunghoon Choi, Yongduek Seo, Hyunwoo Kim, and Ki-sang Hong. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. *Science And Technology*, pages 1–15, 1997.

Tom Reilly Christopher Carling, Mark Williams. *Handbook Of Soccer Match Analysis: A Systematic Approach To Improving Performance*. Routledge, 2006.

F Coldefy and P Bouthemy. Unsupervised soccer video abstraction based on pitch, dominant color and camera motion analysis. *12th annual ACM international conference on Multimedia.*, pages 268–271.

S B Colegrove and S J Davey. On using nearest neighbours with the probabilistic data association filter, 2000.

T Cover and P Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

Willie D. Jones. Will soccer-ball tracking ever reach its goal? *IEEE SPECTRUM*, page 20, sep 2011.

Dartfish. DartTrainer. *http://www.dartfish.com/en/software/index.htm*, 2012.

D DiNucci. Fragmented Future. 1999.

Paul Dirac. *Principles of quantum mechanics (4th ed.)*. Oxford at the Clarendon Press, 1958.

T D'Orazio and M Leo. A review of vision-based systems for soccer video analysis. *Pattern Recognition*, 43(8):2911–2926, 2010.

Arnaud Doucet. On Sequential Simulation-Based Methods for Bayesian Filtering. *Signal Processing*, (CUED/F-INFENG/TR.310):1–26, 1998.

Ling-Yu Duan, Min Xu, Tat-Seng Chua, Qi Tian, and Chang-Sheng Xu. A mid-level representation framework for semantic sports video analysis. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03*, page 33, 2003.

Rick Durrett. *Essentials of Stochastic Processes*. 1999.

Ahmet Ekin, a Murat Tekalp, and Rajiv Mehrotra. Automatic soccer video analysis and summarization. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 12(7):796–807, jan 2003.

Elite Sports Analysis. Focus X2. *http://www.elitesportsanalysis.com/products.htm*, 2012.

espor Inc. Pro Soccer. *http://www.espor.com/sports/soccer/*, 2013.

Extron Electronics. Sports Stadium. *http://www.extron.com/company/article.aspx?id=ssfa*, pages 30–31, 2012.

D. Farin. Current and Emerging Topics in Sports Video Processing. *2005 IEEE International Conference on Multimedia and Expo*, pages 526–529, 2005.

A Farina and S Pardini. Track-While-Scan Algorithm in a Clutter Environment. *IEEE Transactions On Aerospace And Electronic Systems*, AES-14(5), 1978.

R Fielding, J Gettys, J Mogul, and H Frystyk. Hypertext transfer protocol–HTTP/1.1. pages 1–114, 1999.

FIFA. Football Stadiums. 2007.

FIFA. FIFA Football Tests. *http://footballs.fifa.com/Football-Tests*, 2013.

Kazuhiro Fukui and Osamu Yamaguchi. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan*, 29(6):49–58, jun 1998.

Suat Gedikli. *Continual and Robust Estimation of Camera Parameters in Broadcasted Sport Games*. PhD thesis, 2008.

Nicolas Gengembre and Patrick Pérez. Probabilistic Color-Based Multi-Object Tracking with Application to Team Sports. Technical Report RR-6555, INRIA, 2008.

DN Ghista and X Liu. Analysis of Spinning Ball Trajectories of Soccer Kicks and Basketball Throws. *Applied Biomedical Engineering Mechanics*, pages 425–444, 2008.

John Eric Goff and Matt J. Carré. Investigations into soccer aerodynamics via trajectory analysis and dust experiments. *Procedia Engineering*, 34:158–163, jan 2012.

A Hanjalic. Adaptive extraction of highlights from a sport video based on excitement modeling. *IEEE Transactions on Multimedia*, 7(6):1114–1122, 2005.

M Hansen, P Anandan, K Dana, G Van Der Wal, and P Burt. Real-time scene stabilization and mosaic construction. *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, pages 54–62, 1994.

Sadatsugu Hashimoto and Shinji Ozawa. A system for automatic judgment of offsides in soccer games. *2006 IEEE International Conference on Multimedia and Expo*, pages 1889–1892, jul 2006.

Shoji Hirano and Shusaku Tsumoto. Finding interesting pass patterns from soccer game records. *Knowledge Discovery in Databases: PKDD 2004*, pages 209–218, 2004.

Andreas Hofhauser, C Steger, and N Navab. Harmonic deformation model for edge based template matching. *Third International Conference on Computer Vision Theory and Applications*, 2008.

Shaun Holthouse and Igor van de Griendt. Tracking Balls in Sports. *United States Patent Application*, page 7, 2009.

J Hossein-Khani, H Soltanian-Zadeh, M Kamarei, and O Staadt. Ball Detection with the Aim of Corner Event Detection in Soccer Video. *2011 IEEE Ninth International Symposium on Parallel and Distributed Processing with Applications Workshops*, pages 147–152, 2011.

Nicolai Hoyningen-Huene. *Real-time Tracking of Player Identities in Team Sports*. PhD thesis, TUM, 2011.

Yu Huang, Joan Llach, and Sitaram Bhagavathy. Players and ball detection in soccer videos based on color segmentation and shape analysis. *Multimedia Content Analysis and Mining*, pages 416–425, 2007.

inmotiotec GmbH. inmotec. *http://www.abatec-ag.com/?id=55*, 2012.

S Iwase and H Saito. Parallel tracking of all soccer players by integrating detected positions in multiple view images. *Proceedings of the 17th International Conference on Pattern Recognition 2004 ICPR 2004*, 4(C):751–754 Vol.4, 2004.

A.G. Jaffer and Y. Bar-Shalom. On optimal tracking in multiple-target environments. *3rd Symposium on Nonlinear Estimation, San Diego, CA*.

A H Jazwinski. Stochastic processes and filtering theory. *IEEE Transactions on Automatic Control*, 17(5):752–753, 1972.

Hongqi Wang Jia Liu, Xiaofeng Tong, Wenlong Li, Tao Wang, Yimin Zhang. Automatic player detection, labeling and tracking in broadcast soccer video. *Pattern Recognition Letters*, 30(2):103–113, 2009.

J Jokovic and D Orevic. Scene cut detection in video by using combination of spatial-temporal video characteristics. *2009 9th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services*, 1:479–482, 2009.

SJ Julier and JK Uhlmann. A New Extension of the Kalman Filter to Nonlinear Systems. *Int Symp AerospaceDefense Sensing Simul and Controls*, 3(2):26, 1997.

BM Junior and Ricardo De Oliveira Anido. Distributed real-time soccer tracking. *ACM Second international workshop on Video surveillance & sensor networks*, pages 97–103, 2004.

R E Kalman. A New Approach to Linear Filtering and Prediction Problems 1. *Journal Of Basic Engineering*, 82(Series D):35–45, 1960.

R E Kalman and R S Bucy. New results in linear filtering and prediction theory. *Journal Of Basic Engineering*, 83(1):95–108, 1961.

Hirokatsu Kataoka and Yoshimitsu Aoki. Football players and ball trajectories projection from single camera's image. *2011 17th KoreaJapan Joint Workshop on Frontiers of Computer Vision FCV*, pages 1–4, 2011.

Hirokatsu Kataoka, Kiyoshi Hashimoto, and Yoshimitsu Aoki. Player position estimation by monocular camera for soccer video analysis, 2011.

Hyunwoo Kim Hyunwoo Kim and Ki Sang Hong Ki Sang Hong. Soccer video mosaicing using self-calibration and line tracking. *Proceedings 15th International Conference on Pattern Recognition*, 1:592–595, 2000.

Jong-Yun Kim Jong-Yun Kim and Tae-Yong Kim Tae-Yong Kim. Soccer Ball Tracking Using Dynamic Kalman Filter with Velocity Control. *Sixth International Conference on Computer Graphics Imaging and Visualization*, 0:367–374, 2009.

Taeone Kim, Yongduek Seo, and KS Hong. Physics-based 3d position analysis of a soccer ball from monocular image sequences. *ICCV '98 Proceedings of the Sixth International Conference on Computer Vision*, pages 721–726, 1998.

Genshiro Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal Of Computational And Graphical Statistics*, 5(1):1–25, 1996.

H Kjellstrom, D Kragic, and MJ Black. Tracking people interacting with objects. *Computer Vision and Pattern Recognition*, 2010.

C Kwok, D Fox, and M Meila. Real-time particle filters. *Proceedings of the IEEE*, 92(3): 469–484, 2004.

M Leo, N Mosca, and P Spagnolo. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. *Second ACM/IEEE International Conference on Distributed Smart Cameras*, pages 1–10, 2008a.

M Leo, N Mosca, P Spagnolo, P L Mazzeo, and A Distante. Real-time multi-view event detection in soccer games. *Second ACM/IEEE International Conference on Distributed Smart Cameras, 2008.*, pages 1–10, 2008b.

M Leo, N Mosca, P Spagnolo, P L Mazzeo, and A Distante. Real-time multi-view event detection in soccer games. *2008 Second ACMIEEE International Conference on Distributed Smart Cameras*, pages 1–10, 2008c.

B. Li and M. Ibrahim Sezan. Event detection and summarization in sports video. *Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001)*, pages 132–138.

X R Li. The PDF of nearest neighbor measurement and a probabilistic nearest neighbor filter for tracking in clutter. *Proceedings of 32nd IEEE Conference on Decision and Control*, 1993.

X R Li and Y Bar-Shalom. Stability evaluation and track life of the PDAF for tracking in clutter. *IEEE Transactions on Automatic Control*, 36(5), 1991.

Y Li, A Dore, and J Orwell. Evaluating the performance of systems for tracking football players and ball. *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005.

Ying Li Ying Li, Guizhong Liu Guizhong Liu, and Xueming Qian Xueming Qian. Ball and Field Line Detection for Placed Kick Refinement. *2009 WRI Global Congress on Intelligent Systems*, 4:404–407, 2009.

Dawei Liang, Yang Liu, Qingming Huang, and Wen Gao. A scheme for ball detection and tracking in broadcast soccer video. *Advances in Multimedia Information Processing*, pages 864–875, 2005.

Chung-chi Lin Chung-chi Lin, Ming-hwa Sheu Ming-hwa Sheu, Huann-keng Chiang Huann-keng Chiang, Chishyan Liaw Chishyan Liaw, and Ming-che Chen Ming-che Chen. Film-to-Video Conversion with Scene Cut Detection. *First International Conference on Innovative Computing Information and Control Volume I ICICIC06*, 1:285–289, 2006.

Jun S Liu and Rong Chen. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93(443):1032, 1998.

S X Liu, Lijun Jiang Lijun Jiang, J Garner, and S Vermette. Video based soccer ball tracking. *Image Analysis amp Interpretation SSIAI 2010 IEEE Southwest Symposium on*, pages 53–56, 2010.

Match Analysis Inc. Tango Online. *http://matchanalysis.com/tangoonline.htm*, 2013.

M Ángeles Mendoza and N Pérez De La Blanca. HMM-Based Action Recognition Using Contour Histograms. *3rd Iberian conference on Pattern Recognition and Image Analysis, Part I*, pages 394–401, 2007.

Yu Ming Yu Ming, Cui Guodong Cui Guodong, and Qi Lichao Qi Lichao. Player Detection Algorithm Based on Gaussian Mixture Models Background Modeling. *2009 Second International Conference on Intelligent Networks and Intelligent Systems*, 2009.

Jun Miura and Takumi Shimawaki. *Computer Vision and Image Understanding Archive*.

C Morimoto and R Chellappa. Fast 3D stabilization and mosaic construction. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 00: 660–665, 1997.

CJ Needham. *Tracking and modelling of team game interactions*. PhD thesis, 2003.

Netcraft. Web Server Survey. *http://news.netcraft.com/archives/category/web-server-survey/*, 2013.

Zhenxing Niu, Q Tian, and X Gao. Real-world trajectory extraction for attack pattern analysis in soccer video. *Electronic Engineering*, pages 635–638, 2010.

J R Nunez, J Facon, and A De Souza Brito. Soccer video segmentation: Referee and player detection. *15th International Conference on Systems Signals and Image Processing*, 1(55): 279–282, 2008.

Henry Orenstein and James J. Maune. Apparatus for detecting moving ball. *United States Patent Application*, 1999.

Kyoung-Jin Park Kyoung-Jin Park and A Yilmaz. Social Network Approach to Analysis of Soccer Game. *Pattern Recognition ICPR 2010 20th International Conference on*, 0:3935–3938, 2010.

Chaoke Pei, Shuyuan Yang, Li Gao, and Weipeng Ma. A Real Time Ball Detection Framework For Soccer Video. pages 1–4. IEEE, 2009.

Gopal Pingali, Agata Opalach, and Yves Jean. Ball tracking and virtual replays for innovative tennis broadcasts. *15th International Conference on Pattern Recognition.*, 2000.

S V Porter, M Mirmehdi, and B T Thomas. Video cut detection using frequency domain correlation. *Proceedings 15th International Conference on Pattern Recognition ICPR2000*, 3, 2000.

G G Lakshmi Priya and S Domnic. Video cut detection using Hilbert transform and GLCM. *International Conference on Recent Trends in Information Technology ICRTIT*, pages 749–753, 2011.

Prozone Sports Ltd. Prozone3. *http://www.prozonesports.com/product-prozone3.html*, 2012.

LR Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989.

Jinchang Ren, James Orwell, Graeme A Jones, and Ming Xu. Tracking the soccer ball using multiple fixed cameras. *Computer Vision and Image Understanding*, 113(5):633–642, 2009.

J Rissanen. A Universal Data Compression System. *IEEE Transactions on Information Theory.*, 1983.

K Sato and J K Aggarwal. Tracking soccer players using broadcast TV images, 2005.

Scanball. Scanfoot. *http://www.scanball.com/english/*, 2012.

T Shimawaki, T Sakiyama, J Miura, and Y Shirai. Estimation of ball route under overlapping with players and lines in soccer video image sequence. *18th International Conference on Pattern Recognition*, 1(c):359–362, 2006.

R Singer and R Sea. New results in optimizing surveillance system tracking and data correlation performance in dense multitarget environments. *IEEE Transactions on Automatic Control*, 18(6):571–582, 1973.

R Singer, R Sea, and K Housewright. Derivation and evaluation of improved tracking filter for use in dense multitarget environments. *IEEE Transactions on Information Theory*, 20 (4):423–432, 1974.

P Smith and G Buechler. A branching algorithm for discriminating and tracking multiple objects. *IEEE Transactions on Automatic Control*, 20(1):101–104, 1975.

CGM Snoek and M Worring. Time interval maximum entropy based event indexing in soccer video. *Multimedia and Expo, 2003. ICME'03.*, 2003.

Samuel F De Sousa, Arnaldo De A Araujo, and David Menotti. An overview of automatic event detection in soccer matches. *2011 IEEE Workshop on Applications of Computer Vision WACV*, pages 31–38, 2011.

SportsUniversal. Amisco. *http://www.sport-universal.com/technology/technology.html*, 2012.

Carsten Steger. Occlusion, clutter, and illumination invariant object recognition. *International Archives of Photogrammetry and Remote Sensing.*, XXXIV, part 3A:345–350, 2002.

Symah Vision. Epsis. *http://www.epsis.com*, 2012.

A Tanwer and P S Reel. Effects of threshold of hard cut based technique for advertisement detection in TV video streams. *Students Technology Symposium TechSym 2010 IEEE*, pages 211–216, 2010.

GA Thomas. Real-time camera pose estimation for augmenting sports scenes. *3rd European Conference on Visual Media Production.*, pages 10–19, 2006.

D Tjondronegoro, Yi-Ping Phoebe Chen Yi-Ping Phoebe Chen, and Binh Pham Binh Pham. Highlights for more complete sports video summarization. *IEEE Multimedia*, 11(4):22–37, 2004.

Touch-Line Data Systems Ltd. SoccerScout. *http://www.touch-line.co.uk/*, 2012.

O Urhan, M K Gullu, and S Erturk. Modified phase-correlation based robust hard-cut detection with application to archive film. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):753–770, 2006.

O Utsumi, K Miura, I Ide, S Sakai, and H Tanaka. An object detection method for describing soccer games from video. *Proceedings IEEE International Conference on Multimedia and Expo*, 1, 2002.

Kongwah Wan, Xin Yan, Xinguo Yu, and Changsheng Xu. Real-time goal-mouth detection in MPEG soccer video. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03*, page 311, 2003.

Kongwah Wan, Jinjun Wang, Changsheng Xu, and Qi Tian. Automatic sports highlights extraction with content augmentation. *Advances in Multimedia Information Processing*, 2005.

H Wang and Z Li. Tracking a rigid object in 3D from a single camera. *Automatic Inspection and Novel Instrumentation*, 3185:78–89 170, 1997.

WebAppRater. 7 key features of web 2.0. *http://webapprater.com/general/7-key-features-of-web-2-0.html*, 2013.

Patrick Wong and Laurence Dooley. High-motion table tennis ball tracking for umpiring applications. *Signal Processing ICSP 2010 IEEE 10th International Conference on*, pages 2460–2463, 2010.

Lexing Xie and SF Chang. Structure analysis of soccer video with hidden Markov models. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (1), 2002.

Lexing Xie, Peng Xu, SF Chang, Ajay Divakaran, and Huifang Sun. Structure analysis of soccer video with domain knowledge and hidden Markov models. *Pattern Recognition Letters*, (June 2003), 2004.

Ziyou Xiong Ziyou Xiong, R Radhakrishnan, A Divakaran, and T S Huang. Audio events detection based highlights extraction from baseball, golf and soccer games in a unified framework. *IEEE International Conference on Acoustics Speech and Signal Processing 2003 Proceedings ICASSP 03*, 3:401–404, 2003.

Peng Xu, Lexing Xie, SF Chang, Ajay Divakaran, Anthony Vetro, and Huifang Sun. Algorithms and system for segmentation and structure analysis in soccer video. *Proc. ICME*, 2001.

Zh Xu and Pendei Shi. Segmentation of players and team discrimination in soccer videos. *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology.*, 2005a.

Zhifei Xu Zhifei Xu and Pengfei Shi Pengfei Shi. Segmentation of players and team discrimination in soccer videos. *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology 2005*, 2005b.

K. Birmiwal Y. Bar-Shalom. Consistency and Robustness Evaluation of the PDAF for Target Tracking in a Cluttered Environment. *Automatica*, pages Vol. 19, 431–437.

Xinguo Yu, Hon Wai Leong, Joo-Hwee Lim, Qi Tian, and Zhenyan Jiang. Team possession analysis for broadcast soccer video based on ball trajectory. *Fourth International Conference on Information Communications and Signal Processing 2003 and the Fourth Pacific Rim Conference on Multimedia Proceedings of the 2003 Joint*, 3:1811–1815, 2003a.

Xinguo Yu, Qi Tian, and Kong Wah Wan. A novel ball detection framework for real soccer video. *2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698)*, 2:II–265, 2003b.

Xinguo Yu, Changshen Xu, Qi Tian, and HW Leong. A ball tracking framework for broadcast soccer video. *2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698)*, pages II–273, 2003c.

Xinguo Yu, Changsheng Xu, Hon Wai Leong, Qi Tian, Qing Tang, and Kong Wah Wan. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03*, pp:11, 2003d.

Guangyu Zhu, Qingming Huang, Changsheng Xu, Yong Rui, and Shuqiang Jiang. Trajectory based event tactics analysis in broadcast sports video. *Proceedings of the 15th international conference on Multimedia*, pages 58–67, 2007.

Guangyu Zhu, Changsheng Xu, Qingming Huang, Yong Rui, Shuqiang Jiang, Wen Gao, and Hongxun Yao. Event Tactic Analysis Based on Broadcast Sports Video. *IEEE Transactions on Multimedia*, 11(1):49–67, jan 2009.