ASPOGAMO: Automated Sports Game Analysis Models

Michael Beetz[1], Nicolai v. Hoyningen-Huene[1], Bernhard Kirchlechner[1], Suat Gedikli[1], Francisco Siles[1], Murat Durus[1], and Martin Lames[2]

[1] Intelligent Autonomous Systems Group,
Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany
http://aspogamo.cs.tum.edu
[2] Institut für Sportwissenschaften/Sportzentrum,
Universität Augsburg, Universitätsstr. 3, 86135 Augsburg, Germany

Abstract. We propose *automated sport game models* as a novel technical means for the analysis of team sport games. The basic idea is that automated sport game models are based on a conceptualization of key notions in such games and probabilistically derived from a set of previous games. In contrast to existing approaches, automated sport game models provide an analysis that is sensitive to their context and go beyond simple statistical aggregations allowing objective, transparent and meaningful concept definitions. Based on automatically gathered spatio-temporal data by a computer vision system, a model hierarchy is built bottom up, where context-sensitive concepts are instantiated by the application of machine learning techniques.
We describe the current state of implementation of the ASPOGAMO system including its computer vision subsystem that realizes the idea of automated sport game models. Their usage is exemplified with an analysis of the final of the soccer World Cup 2006.

Keywords: GAME ANALYSIS, SPORTS VIDEO PROCESSING, MODEL BUILDING

1 Introduction

Providing effective support for the interpretation and analysis of sport games requires the systematic and comprehensive observation of games and the abstraction of the observed behavior into informative models. Such models, usually created mentally by coaches and scouts, enable to infer the strengths and weaknesses of individual players and teams as well as of strategic planning and tactical decision making. Due to the human way they are subjective and limited to only small subsets of aspects at one time.

In the ASPOGAMO project, we investigate a new generation of sport game models that

– are based on players' positions, motion trajectories, and ball actions as their primitive building blocks;
– represent the interaction between ball actions, game situations, and the effects of ball actions and thereby allow for more comprehensive assessment of the games;
– use concepts such as scoring opportunity, being under pressure, and passing opportunities, classifying situations and interpreting the game events. These concepts are defined transparently and therefore constitute objective criteria for classification and assessment.
– can be acquired automatically by a camera-based observation system.

The benefit of this system is its possibility to model even cognitive abstractions based on an automatically gathered position data pool. The generated concepts are reproducible and revisable and hence objective. If highlevel definitions are modified or added, the change is transparently distributed to all other concepts. Although the system will never come close to human cognitive abilities, which are based on a much wider data base, automated models can help and accelerate human analysis of complex activities in sports games. They create new opportunities in sport science to get insights into the process instead of comparing the final results only after the end of the game.

In computer vision, video indexing and the extraction of statistics in major sports games have received a great deal of attention due to their controlled setting, e.g. a soccer field or tennis court with well separable actors and predefined rules for actions. Surveys of computer aided sport analysis are given in (Wang & Parameswaran, 2004b; Yu & Farin, 2005; Setterwall, 2003). The field can roughly be split into indexing and tracking methods. The first divide a predominantly broadcasted video into several labeled parts, while the latter estimate trajectories of moving objects like players or their equipment in recorded images.

In broadcasted material, the shots themself provide information about the content. Multiple cameras are usually employed to generate different game perspectives and view types such as player close-up, panoramic view or slow-motion replay. Broadcasters adhere to well established video production rules to select video feed for view consistency, making it easy for viewers to follow the game. For example, the sequence of views that tracks a game-service point in tennis usually begins with a close-up of the player preparing to serve, followed by a change to the panoramic court-view after the ball is served, until the break, when view changes to a close-up of the player, who won the point. A close relationship is maintained between the temporal view changes and the semantic game events. Syntactic structures of the game can therefore be reverse-engineered by recognizing these view type changes. Audio cues may also be used; since the crowd is usually quiet during play, the detectable sounds are ball hits, followed by loud eruption of cheers and applause when point is won. Similar structures can be observed in most major sports videos. Much research has been done exploiting these domain constraints by visual information (Ekin et al., 2003), audio clues (Rui et al., 2000; Lao et al., 2006a) and the multimodal integration of video, audio and chroma keying (Babaguchi et al., 2002; Han et al., 2002). Applications have been found almost in all major sports as tennis (Pingali et al., 1998), baseball (Rui et al., 2000; Han et al., 2002), basketball (Zhou et al., 2000), soccer (Ekin et al., 2003; Assfalg et al., 2002) and American football (Babaguchi et al., 2002; Li & Sezan, 2001).

On the other side, tracking methods focus on extracting the positions of the players and the ball or puck in ball games. Different segmentation and motion tracking techniques have been used including color based methods (Xu et al., 2004; Pingali et al., 1998; Duan et al., 2003; Liang et al., 2005), feature extraction (Jin, 1994; D'Orazio et al., 2002) and additional postprocessing (Yu et al., 2003) for single camera settings as well as some integrating information captured by multiple cameras (Pingali et al., 2000).

While most of them deliver raw trajectory data or simple statistics, Sudhir et al. (Sudhir et al., 1998) not only track tennis players while estimating the camera pose, but also map the spatial data to limited higher level classifications of the scene such as Baseline-Rallies, Passing-Shot, Serve-And-Volley and Net-Game using a lookup table. Similar work was done by Lao et al. (Lao et al., 2006b) integrating also audio cues. Wang and Parameswaran (Wang & Parameswaran, 2004a) classify tennis ball trajectories extracted from videos of calibrated cameras to 58 tactic patterns defined by sport experts using Bayesian inference.

Most of the methods so far reveal semantics in broadcasted material that has been already encoded by humans, or provide only low-level statistical informations based on trajectories as e.g. the covered distance containing far less semantics. Few enrich videos with semantics that are interesting for sport scientists and trainers beside limited classifications. The ASPOGAMO system tries to automatically gather a spatio-temporal information basis and build a rich context-dependent domain knowledge upon it. Transparent definitions can be stated in an objective way allowing the applicability to unseen games, their usage for indexing and may also reveal unknown insights into the structure of team sport games.

We describe the framework instantiated for the soccer domain by giving an overview of the complete system in the first section. It is splitted functionally into two main components: Section 3 details the automated observation system that extracts the position data of all soccer players and ball trajectories from multiple pan-tilt-zoom cameras. The automated models, built on these data, and the generation of high-level analysis are explained and exemplified for the final game of the Football World Cup in 2006 in Section 4. Finally, we draw our conclusions and give future prospects.

2 System Overview

The ASPOGAMO system extracts meaningful sports game models from video footage presenting an interface for analysis. An overview of the pipeline with the two main components of the ASPOGAMO system, the observation system and the automated hierarchical model is depicted in fig. 1.

Image sequences captured by one or more pan-tilt-zoom cameras – possibly including also broadcasted material – form the input of the system. The observation system bundles state-of-the-art computer vision techniques to estimate camera perspectives as well as to segment and to track all soccer players and the ball in real-time exploiting domain knowledge to make the problem tractable. For example, the segmentation relies heavily on characteristics of the inspected domain as e.g. the appearance of the playing field and is implemented for soccer only, so far. The automatically gathered and possibly revised trajectories are fed by the observation system to the knowledge base. They build the base of all models assuming that this information is sufficient to completely describe a game. The knowledge base is organized as an ontology and provides hierarchical models of games. Contextual concept definitions are grounded transparently in the data by datamining techniques supposing valid declarations and adequate data. The model framework is not specific for soccer but can incorporate other sports games as well, given an adjusted conceptualization. Users of the ASPOGAMO system, which are sport scientists, the media and the viewers, can query the knowledge base in terms of the model and receive answers visualized in appropriate ways. Visualizations can be abstract images or videos but also augmented video parts of the original material. All the data needed for augmentation has been already gathered by the observation system. The automated sports game models provide in conjunction with their groundings a rich source of analysis.

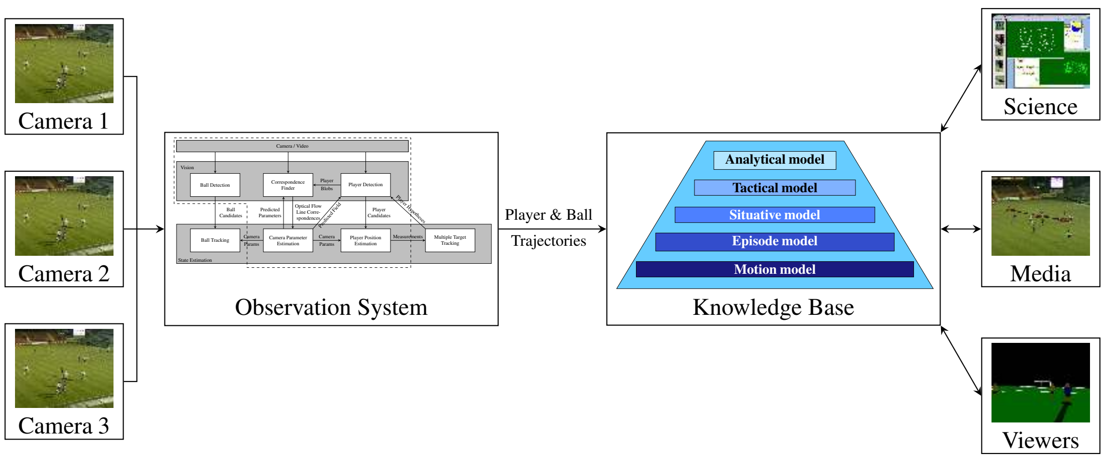


Fig. 1. Overview of the ASPOGAMO system.

3 Observation System

The observation system extracts player and ball motion data from video footage (Beetz et al., 2007; Gedikli et al., 2007; Beetz et al., 2006). For the acquisition of trajectories it is necessary to solve a complex probabilistic estimation problem consisting of subproblems interacting in subtle ways. One of these problems is the estimation of the camera parameters, such as the position of the camera related to the field as well as its pointing direction and zoom factor. Also, the detection of players in the video frames and of team affiliations need to be considered. Finally, the tracking of all players disambiguating them through occlusions needs to be solved by fusing the position estimates from different cameras and integration over time. To deal with all these problems, the observation system consists of two basic components, named the Vision module and the State Estimation module, as shown in Figure 2. The Vision module is in charge of the image processing and computer vision tasks, while the State Estimation module is used to estimate camera parameters and the position of each player.

Based on this functional description, the path followed by the data flow is as follows: The *Vision* module using the *Correspondence Finder* and *Player Detection* submodules, performs the estimation of the camera parameters, and, at the same time, the localization of the players for every image in the video stream. The *Correspondence Finder* uses the predicted camera parameters and player localizations to find line correspondences between the current and previous video frames, avoiding the unwanted effects of possible occlusions. These line correspondences are sent to the *Camera Parameter Estimation,* where a prediction of the camera parameters for the next frame is made. The *Player Detection* combines the predicted camera parameters, the player hypotheses, the results of a local spatial variance filter and a color template matching to deliver the player position measurements to the *Player Position Estimation*. The player positions and their uncertainties are projected to world coordinates. This information is used by the *Multiple Target Tracking* module to combine and associate the individual positions over time to consistent team configurations, and to predict a set of hypotheses required in the next frame.

The dashed box in fig. 2 indicates the parts of the observation system that are cloned for each camera.

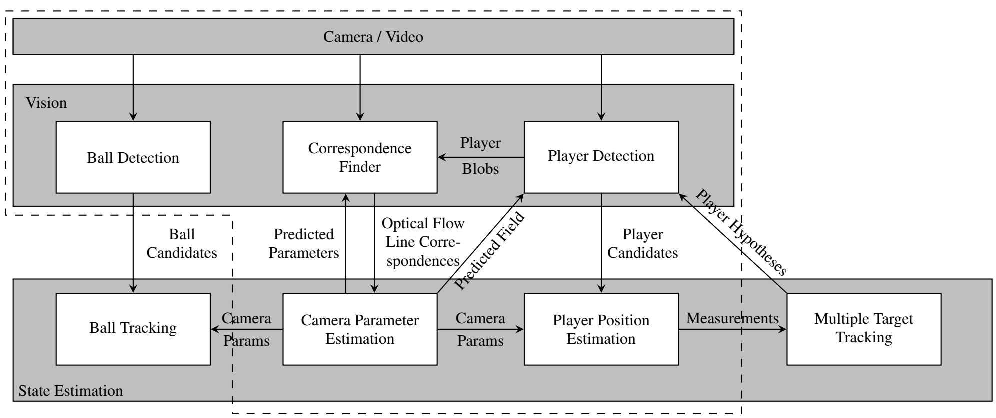


Fig. 2. The observation system of ASPOGAMO.

3.1 Camera Parameter Estimation

The only sensors used by the ASPOGAMO system are rotating and zooming cameras. Since all measurements are done in the images taken by these cameras, the mapping characteristics of these cameras are required to transform measurements from image coordinates into measurements in real world coordinates. This mapping is mathematically described by the pinhole camera model with one radial distortion coefficient. This model has twelve free parameters, which have to be estimated by the *Camera Parameter Estimation* module.

Estimating all parameters for every frame is unreliable, inefficient and unnecessary, as most of these parameters stay constant during the whole game. Since the cameras are usually mounted on tripods around the field, they are fixed in their position, changing only their orientation and zoom factor to track the game. Considering these kind of camera configurations and making some reasonable assumptions, the camera parameters are split into two sets: the constant parameters, which stay constant during the whole game, e.g. the position of the camera, and the dynamic parameters, which have to be estimated successively for each frame (tilt α, pan γ and focal length f). This allows ASPOGAMO to estimate the constant parameters accurately from multiple views beforehand, while increasing the efficiency and robustness of the estimation of the dynamic parameters during the game.

Continuous Camera Calibration ASPOGAMO uses model-based localization for its estimation of the dynamic parameters, where the model describes the appearance and geometry of real world features on the field. The basic idea is to determine the parameters of the camera model that lead to the best fit between the image and the projected field model (see Figure 3). The quality of an estimate is determined from correspondences between

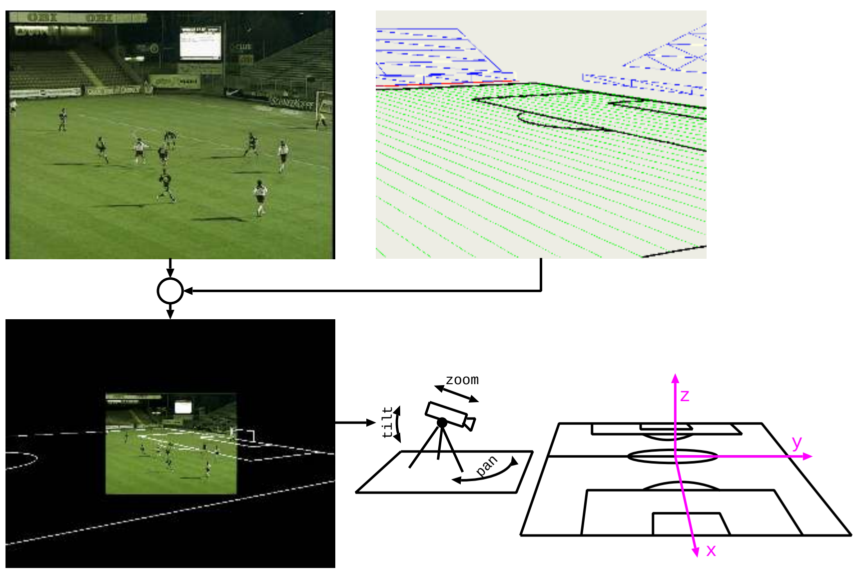


Fig. 3. Model based estimation of dynamic camera parameters.

model points and image points.

First, predicted camera parameters are used to project the field model onto the image (Figure 4 left). Then, correspondences between image points and model points are searched perpendicular to the corresponding model line (Figure 4 right). Finally, the optimal parameters are determined by minimizing the distances between model points and corresponding image points.

Fig. 4. Projected fieldmodel using predicted parameters (left), finding correspondences along the search lines (right).

Since the estimation is done successively for each frame, it can be described as a tracking problem and is solved using the *Iterated Extended Kalman Filter* (IEKF) (Bar-Shalom & Fortmann, 1988). IEKF provides also the uncertainty of the estimated parameters as a covariance matrix, which is used to determine the search window for the next time step as well as to determine the uncertainties for the player positions.

Advanced image processing and probabilistic estimation techniques enable ASPOGAMO to reliably track the camera parameters even when confronted with inhomogeneous lighting conditions, image compression artefacts, fast camera motion, missing line features and noisy and sometimes wrong observations. These techniques are briefly described below.

To deal with blurry lines and edges, ASPOGAMO uses probabilistic color classes and characterizes lines as well as edges by color transitions. Image points corresponding to given model points are found by searching the color distribution along the search line that best fits the expected color transition. Additionally, the variances of the respective correspondences are heuristically estimated from the match.

Most outliers (wrong measurements) are caused by occlusions of field lines by players. Therefore, all correspondences lying within player regions are removed in a first step (See Figure 4 right).

Furthermore, ASPOGAMO uses a robust optimization method to suppress the impact of the remaining outliers to the final estimation. This is done by integrating the robust *M-Estimator* (Huber, 1981) into the MAP estimation of the Kalman Filter.

To obtain reliable estimates for sequences without sufficient field lines, ASPOGAMO uses the optical flow information between subsequent images to predict the camera parameters in the time update stage of the Kalman Filter. In a first step, the relative change in the parameters are obtained in a closed form solution from the homography, which itself is robustly determined by RANSAC (Fischler & Bolles, 1981), between both images. Then, this solution is used as an initial point for a non linear least squares estimation, where again the M-Estimators are used to suppress outliers in the measurements.

A monitoring process, using several low level features which complement each other, is used to detect when the estimation fails. In such a case, the estimation process is automatically reinitialized using detected field lines in the image. This however is only possible for images with a sufficient number of visible field lines (e.g. goal area).

3.2 Player Detection and Localization

In order to estimate the real player positions, the positions of the players in image coordinates are required along with the camera parameters. Every image in the sequence is processed in three steps: segmentation of regions, which possibly contain players, called blobs, the player localization inside these blobs and the mapping of the player positions in the image to real world coordinates on the field.

The blob segmentation exploits the homogeneity of the green grass field. The spatial variance of the intensity image is thresholded and the regions showing a high variance are assumed to contain the searched players. The threshold to classify the variance image into player blobs and field regions is selected adaptively based on a Maximum Likelihood approach.

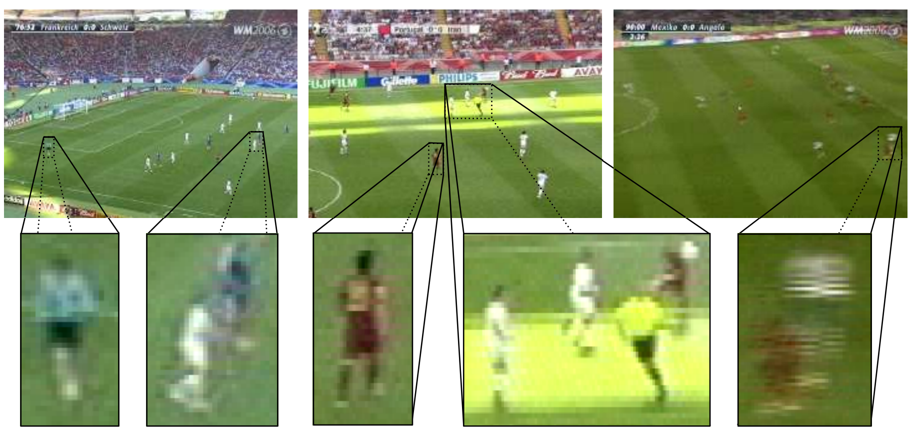


Fig. 5. Hard scenes for the player segmentation: bad resolution due to the small player sizes and player occlusions (left), illumination changes due to strong shadows (middle), washed-out of the pixels due to camera motion blur (right).

The next step is to robustly locate the players inside the blobs. A color template matching procedure is applied, where color templates are generated for every different player type. The color template is formed by three sections, where the upper section represents the shirt, the middle the shorts and the bottom the socks of the player. Every section of the template is associated to one or more color classes, which are modeled as Gaussian distributions in the color space. To generate these color classes, some regions containing the most representative colors are selected beforehand from the input video. The system can also handle cases where one team is completely dressed in a uniform color or both teams share some colors between them. If occlusions between players appear, the template matching is supported by the application of geometrical constraints. This geometrical constraints incorporate the expected player size and their shape.

Finally, the image positions of the players are projected to the field using the already estimated camera parameters. Additionally, the covariance matrix for each player position is propagated from the covariance matrices of the camera parameters and the template matching.

The player recognition is a very challenging task that is effectively tackled by the ASPOGAMO system. Our robust and accurate segmentation of the players accomplishes also small size of the players due to low image resolution, inherent player occlusions, overlapping color classes and blurry images caused by fast camera motion. To get a glimpse of the variety, some of the mentioned challenges are depicted in Figure 5.

3.3 Player Tracking

The detection of players is performed locally in every frame for each camera. The task of the *Tracker* module is to form consistent trajectories for all visible players independent from the current view. As there is only one real player configuration at a time, the informations gathered in each single camera view have to be fused. In addition, an integration over time contributes to consistency and therefore results in a better approximation.

We use a Rao-Blackwellized Resampling particle filter (RBRPF) for multiple target tracking of each single player identity (von Hoyningen-Huene & Beetz, 2009). This particle filter can handle similar appearances of players of the same team, frequent occlusions and false alarms by focusing on the data association problem of multiple target tracking. The RBRPF approximates the posterior over all complete player configurations. Single positions are represented as Gaussians forming a multi-modal posterior to take the uncertainty in the data association of players and detected measurements into account. New states are sampled with an importance density over the possible data associations and an analytical solution to find the optimal Gaussian for the given association and predicted state. Kalman filters deliver theoretically optimal sample distributions and its use refers to the Rao-Blackwellization of the filter. The sampling is based on the approach by (Särkkä et al., 2004; Särkkä et al., 2007), which presupposes either the independence of associations or the knowledge and fast computability of the dependency to make the computation of association probabilities tractable. We relax this assumption by sorting the associations of measurements of a sweep randomly. Only few associations are highly probable for each predicted state and, therefore, the algorithm can make use of memoization to increase performance. A constant velocity model approximates the dynamics of the athletes for prediction; the team affiliation is represented as a simple appearance model, which influences the sampling of associations. We assume clutter or false alarms to be distributed uniformly over the measurement area. Sampled states are weighted according to their fit to the measurements and dynamics and the posterior probability of their state of the former time step. Resampling replicates particles with a high probability by sampling more associations for them. The discreteness of associations allows the subsumption of equal states, reduces the computation time for unambiguous player configurations and offers real-time performance in the first place.

The fusion of measurements of different camera perspectives reduces the probability of occlusions and increases the accuracy of the tracking result. It is done by incorporating every measurement sweep sequentially, which is computationally equal to the parallel case. The player detections of each camera produced asynchronously by several independent modules are synchronized to feed the *Tracker* with the correct sequence of measurement sweeps.

3.4 Ball Tracking

In any ball game like soccer, the ball is invariably the focus of attention. Images showing the ball paths have become indispensable for analyzing scenes since they tell about the tactical situation on the field. In ASPOGAMO the tracking of the ball is implemented in every frame for each camera following the particle filter framework (Arulampalam et al., 2002). Particles stand for hypothetical ball states including position, velocity and acceleration in real world coordinates. Every frame, their current state is updated according to the ball dynamics supposing a constant acceleration motion model. New particles are resampled according to their weights of the previous step. Using the known camera parameters, their position can be projected into each monocular camera image. Summation of color template matching in the original image and shape constraints in the variance image determine the new weight of the sampled particle. Since the ball is small in size and always moving, there always exist a motion blur in the color image while detecting the ball. To avoid this problem, the spatial variance of the intensity image is examined. If the ball is occluded by a player, its weight is set to a threshold value that is lower than the usual one exhibited by color and shape template matching. Of course, two cameras are needed at minimum to determine the 3D trajectory of the ball.

4 Model for Football Games

The outcome of the observation system is a set of ball and player trajectories, which has limited ability for direct inspection because already for a short time period the data set gets too complex to be displayed in a single picture (see Figure 6). Therefore, the raw data have to be reduced or abstracted according to the analysis tasks. A model denotes such a set of abstractions and transformations of data.

4.1 The Model Hierarchy

We use a hierarchical structure for our model (Beetz et al., 2004c; Beetz et al., 2004b; Beetz et al., 2004a; Beetz et al., 2005) (see Figure 7) allowing for a bottom up development of model levels. Each level builds concepts upon lower level concepts providing a big range of different model abstraction levels.

The most basic layer of the model, the motion model (Beetz et al., 2004b), represents the positions and motions of the players and the ball. This level can be generated automatically from the position data created by the observation system. The data is stored in

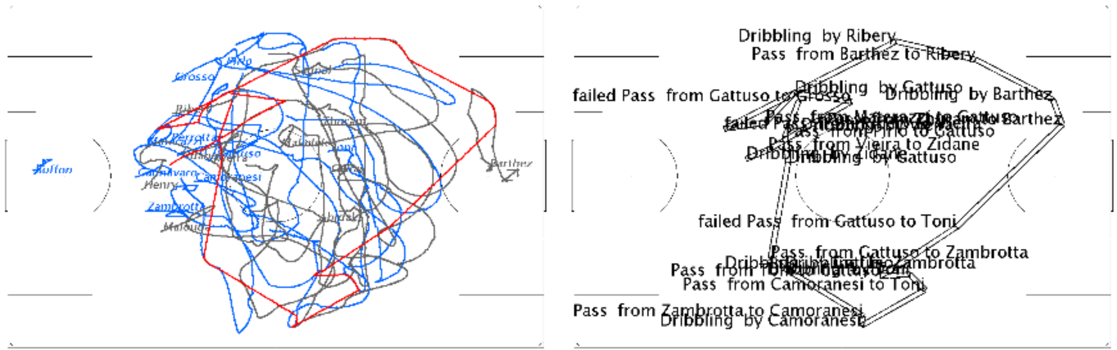


Fig. 6. Complexity of trajectories for all players and the ball with corresponding ball actions belonging to only 34 seconds within the World Cup final 2006.

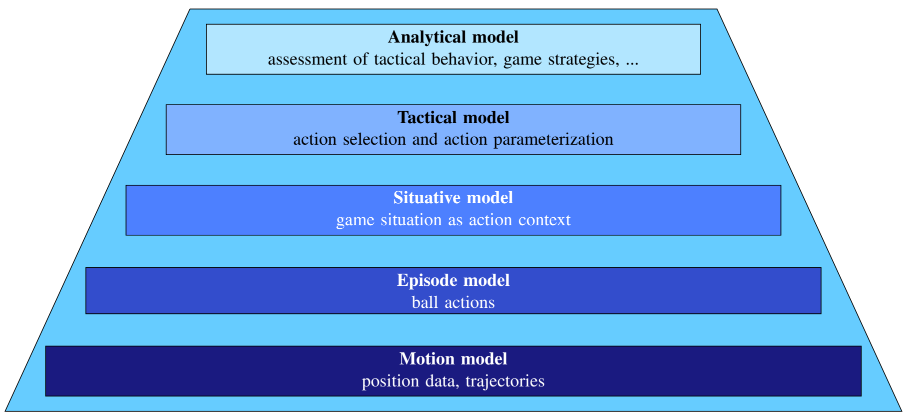


Fig. 7. ASPOGAMO model hierarchy.

a compact and structured way to ease subsequent interpretation and analysis of position information. To achieve this, the continuous motions are segmented into curve segments that can be approximated with a specified accuracy by simple curve functions. The position and motion data do not only build the basis for subsequent building of higher level concepts but already allow for interesting analysis, e.g. static or dynamic heat maps representing the distributions of player positions or velocity profiles of players giving hints on the physical strain of players.

At the second level of our model hierarchy, ball actions are represented. From the motions of the players and the ball, the system identifies ball contacts as accelerations of the ball caused by a contact with a player. The ball contacts are classified into different categories like pass or shot with sub-categories such as a center or a through ball. This model allows for an analysis of actions, their frequencies and their a priori success rate. Possible analysis include passing dyad or pass success rates of players.

The concept “situation” can be viewed as a snapshot of the game. On the situative level, all features such as player and ball positions as well as their interactions are represented at a specific time stamp. Considering the spatial constellation of passing opportunities, properties such as whether or not a player is being marked or is being attacked are modeled on this layer.

Based on the situation concept, tactical set-ups can be defined on the tactical layer. These include typical action selection criteria of players, action parameterizations or tactical activities.

Finally the analytical level is comprised of assessment models for situations and tactical behaviors as well as the identification of typical attack and defense systems.

The concepts and the relationships between them are stated in form of an ontology allowing an automatic reasoning about them. The ontology has a human readable as well as a parsable interface and constitutes the domain knowledge as well as facts of the seen games.

4.2 Visualizations

An abstraction of data can also be done visually, exploiting the advanced visual cognitive abilities of humans. Position information about the players can, for example, be abstracted into heat maps (Figure 8) that represent the probability distribution over player positions on the field. The tactical lineup and positional structure inside a team is made visible by depicting a 11x11 grid which is colored according to the most frequent player label for relative player indices.

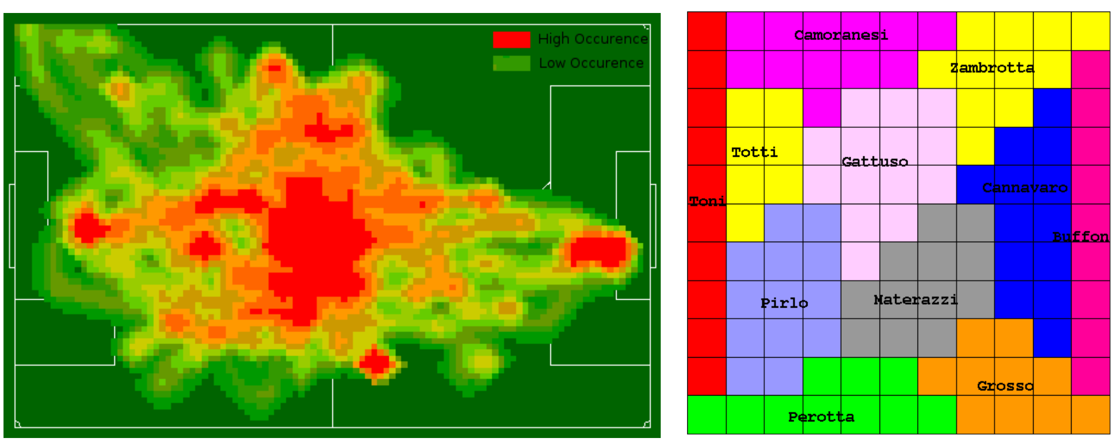


Fig. 8. Heat map visualization for Zidane and a grid determining the tactical lineup of the Italian team.

While these visualizations are a big achievement themselves, they are limited to a specific context and new ones are not easily developed without creativity. The visualizations rely on a specific relationship between the data that they depict. The ASPOGAMO system incorporates all data in an ontology providing a convenient way to describe the context needed for a specific visualization. The applicability to specific data subsets or transformations can be deduced automatically by their ontological tags.

4.3 Model Acquisition

The central part of our analysis system is the acquisition of an informative model of the observed football game (Beetz et al., 2004a) by applying statistical learning to a number of games resulting in a model of football games in general. This model provides a high abstraction level to view the game as a whole but also allows to perform detailed analysis for special aspects of the game. The system adapts the model automatically, when data for a new game are available.

Our model is divided into two parts: A static part that is not dependent on a specific game or class of games, and a dynamic part that has to be adapted for the context it is used in.

The static part contains concrete definitions of the individual model layers that are independent of a specific game or game class. This part can be specified once for all football games and applied to all games in the same way. It contains general rules and classifications like the offside law, what a successful shot is or the number of allowed substitutions.

The dynamic part consists of concepts that relate to the same definition but different specifications. For example the concept of scoring opportunity has a well defined meaning, but depends on the quality of the involved teams. A given situation would be a scoring opportunity e.g. for a World Cup player, because he would always score, but could not be classified as such for another one (e.g. a junior player). So some (unfortunately most) parts of the model have to be specific to the game, to the teams or to the players and therefore have to be specified relative to their context. We solve this problem by enabling our system to learn the dynamic part from observed games automatically by machine learning techniques given the abstract meaning that is consistent through all contexts (von Hoyningen-Huene et al., 2007).

As an example we examine the episode model of our model hierarchy. We can easily specify statically, that a successful pass denotes a ball contact of one player of a team followed by a ball contact of another player of the same team. For unsuccessful passes however, the definition is very hard to specify in a general way. But we can state that the characteristics of unsuccessful passes should be similar to passes regarding the velocity of the ball, the direction, the ball was played in, or the positions of team mates and of opponents. This definition holds for all games even if the attributes are highly correlated with the abilities of the players and, therefore, depends on the league or competition, in which the specific game took place. Transforming this static definition into a datamining task, the system can learn rules for each league that specify which ball actions should be considered as passes and which ones rather as shots. Taking the known set of successful passes, shots and dribblings as training data, a decision tree (Quinlan, 1993) or respectively a regression tree (Witten & Frank, 2005) is learned automatically for each binary classification (pass or no pass, shot or no shot, dribbling or no dribbling). The tree is split into several rules by logical disjunction of the nodes on all possible paths beginning at the root and ending at a leaf. The abstraction comes into play by the pruning of the tree that is a part of the learning algorithm to avoid overfitting. The rules are transformed to descriptions in the ontology and in this way they are integrated into the knowledge base. The description (consisting of the rules) as well as the concept itself (referring to the classification of instances) can be inspected transparently.

There exists also a second class of definitions that contain a dynamic part. Most of the continuous attributes of actions in team sports are discretized into classes like slow, normal and fast or short to far. This is usually done by simple thresholding at predefined ranges. If we look closely to these kind of concepts, the sensitiveness to their context becomes evident. The velocity of a fast sprint in a international competition obviously differs from values for a burst of speed in the minor league. Still, there is a common definition, that partitions the usual velocity range in a predefined number of parts, naming the part with the highest velocities as fast. Partitioning is achieved by datamining techniques called clustering (MacKay, 2003) which iteratively find a locally optimal subdivision. Also a smooth fragmentation can be achieved automatically by using probabilistic assignment to clusters obtained by fuzzy clustering (MacKay, 2003). The partition and the naming of each part is again transformed into descriptions in the ontology to provide a seamless interface.

All these models, the static and dynamic ones, are accessible to the system due to their integration into the ontology. Their definition and semantics can be inspected and analyzed by the user resulting in more alternatives to analyze a football game and attaining more objectivity from the transparency of the models. For example, the user can retrieve the scoring opportunities of the teams and then analyze how they might arise by interpreting the generated rules for the concept. A concept for all situations, in which the ball was lost, could be stated and from the resulting rules, some reasons for the failed ball action may be derived. Clustering the situations, in which the ball was lost, can help in gaining insights, how the opponent team forced the loss of ball possession, etc.

4.4 Example of Analysis

In our soccer ontology each pass is assigned to a set of attributes including its length (short, middle, long), speed (fast, middle, slow), direction (forward, cross, back), and risk (risky, safe). Although their definitions contain already dynamic parts, they build the base of further analysis. We present three different kind of analyses inspecting the passes in the final game (Italy vs. France) of the FIFA soccer world championship 2006.

Defensive pass classes Instead of examining statistics over all attributes and obtaining a multi-dimensional distribution, the analyzer can ask, “What are the four most typical classes of passes played by defenders?”. Therefore, he introduces new concepts to the system that partition all passes of defenders of one team and are named $p_1^I, \ldots, p_4^I$ for Italy and $p_1^F, \ldots, p_4^F$ for France. The subfigures of fig. 9 visualize these new concept. The four classes for the Italian defenders are the long cross passes, the medium length forward passes, the very long forward passes and the fast forward passes. For the French team, on the other hand, these passes are the long or fast passes not played back, the passes not played long with medium velocity, the slow passes and the fast or medium velocity back passes.

Comparing these two sets of classes, one can see, that, while the Italian team prefers cross passes in the backfield (a) and tries to play long passes to the right wing player, the French team aims at passing to the outside (f) and often plays the ball back to the goalkeeper if there are no suitable pass receivers (h). One can also see that different types of passes are spatially local.

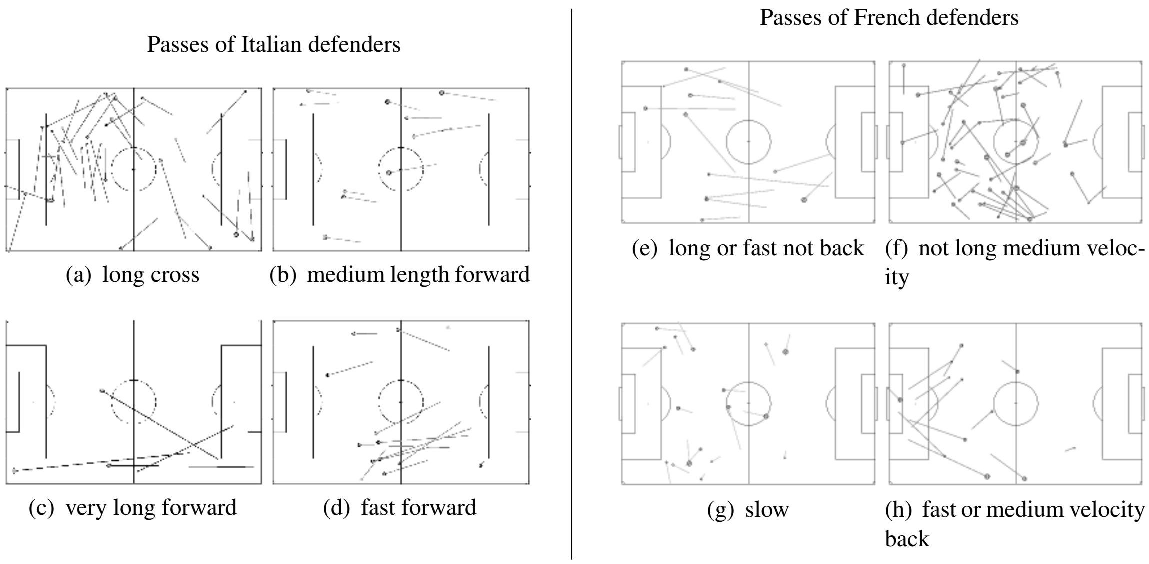


Fig. 9. Typical pass types of defensive players of the Italian (a)–(d) and French (e)–(h) team during the World Cup Final 2006. One can see that the passing game of both teams can be characterized using different classes of passes that are typical for the respective teams.

We can also include positional and situational features as well as the roles of the playing and receiving players. With this information, one can analyze the individual classes of the whole Italian team in greater detail, automatically creating descriptions of the classes.

Descriptions of subsets of passes An example of a more abstract analysis examines the typical passing behavior of a team. We will now describe how this is performed and how it relates the different layers of the model. First, a specific subclass of passes is selected from the episode model by the user, for example, all passes of the Italian team. Next, these passes are partitioned on the situation model layer. This is achieved by an automatic clustering of spatial and situational features, that are computed for the start and end situations of the passes. The automatic clustering delivers regions that need not be specified beforehand but are generated adaptively for existing structure in the data. In our example five clusters were found automatically and can be inspected in Figure 10.

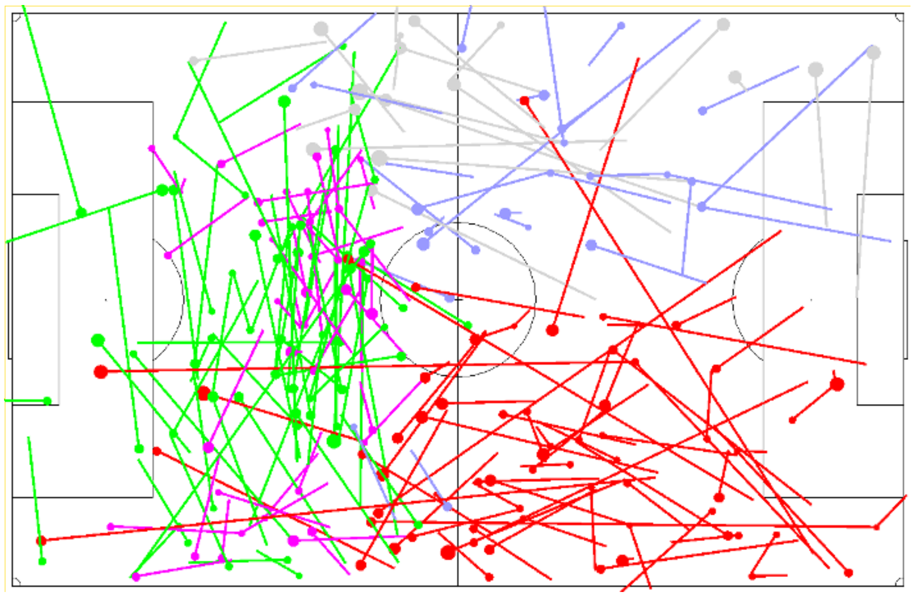
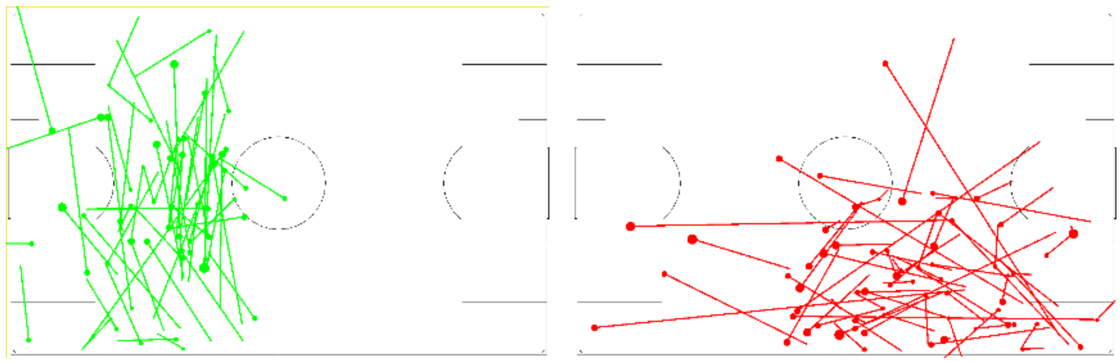


Fig. 10. Clusters for the Italian team's passes.

In the next step of the analysis, the system automatically creates several alternative descriptions for the partitions previously found. This is done as follows: first, different non-overlapping feature sets are selected. Decision trees that discriminate between an individual cluster and the rest of the specified passes are learned. All paths to nodes labeled with the same cluster are combined by con- and disjunctions, transforming the decision tree to a rule. This rule constitutes a description of the cluster in terms of the given features. By the use of spatial features only, we get rules that discriminate the clusters spatially and, therefore, form a spatial description of each cluster. In our example, extracted with the x coordinate x_s of the starting and the y coordinate y_e of the final situation of a pass as features, cluster 3 can be described as all passes fulfilling $x_s > -8.52m \wedge y_e > -0.3m$, the passes from the French half of the field (reaching 8.5m into the Italian one) to the right side (see Figure 11b) with an accuracy of more than 95%.

These descriptions are sorted by a quality measure derived from the complexity of the rule as well as the accuracy of the description. The accuracy is computed as the ratio of the number of entities that belong to both, the cluster and the describing rule, and the number of entities that belong to at least one of them. Descriptions with high quality measure are presented as alias to the clusters, e.g. the cluster depicted in Figure 11a is described as long cross passes.

Comparing different descriptions A possible subset of the discarded descriptions still contains interesting information. Short rules that represent sub- or supersets of the cluster are presented as asymmetric descriptions to the user, too. In our running example, the rule $x_e > -3.87m \wedge y_s > -5.72m$ (the passes to the front from the right side of the field)



(a) long cross passes

(b) passes from the right to the front

Fig. 11. Two classes of typical passes of the whole Italian team during the first half of the World Cup final 2006.

defines a subset of the cluster depicted in Figure 11b, again with about 95% accuracy. So, we have the following two simple descriptions that are strongly correlated:

- Firstly the cluster is described as the passes from the right side of the field to the opponent's (front) half [3] and
- secondly it contains all passes from the front half to the right side of the field.

From these descriptions, the system creates the following rule comparing the different descriptions:

$$xe > -3.87m \wedge ys > -5.72m \Longrightarrow ye > -0.30m \wedge xs > -8.52m$$

In other words, the passes from the right side of the field to the front are played from the front to the right.

The consequences of this rule and the propositions following from it are not immediately obvious, because the start and end positions of the passes are intermixed. For an easier interpretation, we group them together to form two rules that state that

- passes from the left front area of the field are not played to the right side of the field and
- passes to the right back area of the field are not played from the front of the field.

Whereas the second rule is a rule that is fairly usual for football games, the first one, the lack of cross passes from the front left side of the field to the right side, is again a specialty of the Italian play.

The two clusters examined in this example make up about one third of all passes played by the Italian team each. So, these two clusters, the extracted rules and their interpretation should give a fairly good insight into the passing behavior and its characterization.

[3] The area is not exactly one half of the field but a partition close to it created by the system from the data

5 Conclusions

In this article we have presented ASPOGAMO, a new generation of sports game analysis models. Trajectories of the ball and the players extracted from video by our state-of-the-art camera-based observation subsystem build the base of these models. Semantics of higher levels of abstraction are automatically grounded by adapting static concept definitions to the appropriate context of the observed game using statistical learning methods. The automatic and transparent acquisition of high-level facts provides objectiveness and comparability. Games can be analyzed in an explorative and objective way, gathering meaningful sub-categories of concepts and propositions specific for the game that is inspected. ASPOGAMO will provide new opportunities for sport scientists to analyze sports games, support scouts in inferring the strengths and weaknesses of the players and help coaches in the strategical planning and tactical decision making processes. To demonstrate the possibilities of automated sports game models, we provided some exemplar of analysis for the final of the soccer world championship 2006.

For further improvements we will investigate the way of defining new concepts for sport scientists, which can be stated clearly and naturally, but still ensures objectivity and the ability for automated processing. We will examine the scalability of our approach to huge data sets providing a richer source for analysis, but also making high demands on memory and performance requirements of the used algorithms.

References

[Arulampalam et al., 2002]Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

[Assfalg et al., 2002]Assfalg, J., Bertini, M., Colombo, C., & Bimbo, A. D. (2002). Semantic Annotation of Sports Videos. *IEEE MultiMedia*, *9*(2), 52–60.

[Babaguchi et al., 2002]Babaguchi, N., Kawai, Y., & Kitahashi, T. (2002). Event Based Video Indexing by Intermodal Collaboration. *IEEE Transactions on Multimedia*, *9*(2), 68–75.

[Bar-Shalom & Fortmann, 1988]Bar-Shalom, Y. & Fortmann, T. (1988). Tracking and data association. Academic Press.

[Beetz et al., 2007]Beetz, M., Gedikli, S., Bandouch, J., Kirchlechner, B., von Hoyningen-Huene, N., & Perzylo, A. (2007). Visually tracking football games based on tv broadcasts. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, (pp. 2066–2071).

[Beetz et al., 2006]Beetz, M., Bandouch, J., Gedikli, S., von Hoyningen-Huene, N., Kirchlechner, B., & Maldonado, A. (2006). Camera-based observation of football games for analyzing multi-agent activities. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, (pp. 42–49).

[Beetz et al., 2005]Beetz, M., Kirchlechner, B., & Lames, M. (2005). Computerized real-time analysis of football games. *IEEE Pervasive Computing*, *4*(3), 33–39.

[Beetz et al., 2004a]Beetz, M., Fischer, F., Flossmann, S., Kirchlechner, B., Unseld, A., & Holzer, C. (2004a). Watching football with the eyes of experts: Integrated intelligent systems for the automatic analysis of (simulated) football games. In *5th Annual Conference dvs-Section Computer Science in Sport*.

[Beetz et al., 2004b]Beetz, M., Flossmann, S., & Stammeier, T. (2004b). Motion and episode models for (simulated) football games: Acquisition, representation, and use. In *3rd International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS)*, (pp. 1370–1371).

[Beetz et al., 2004c]Beetz, M., Kirchlechner, B., & Fischer, F. (2004c). Interpretation and processing of position data for the empirical study of the behavior of simulation league robocup teams. In *KI 2004 Workshop*.

[D'Orazio et al., 2002]D'Orazio, T., Ancona, N., Cicirelli, G., & Nitti, M. (2002). A ball detection algorithm for real soccer image sequences. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1*, (pp. 10210)., Washington, DC, USA. IEEE Computer Society.

[Duan et al., 2003]Duan, L.-Y., Xu, M., Chua, T.-S., Tian, Q., & Xu, C.-S. (2003). A mid-level representation framework for semantic sports video analysis. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, (pp. 33–44)., New York, NY, USA. ACM.

[Ekin et al., 2003]Ekin, A., Tekalp, A. M., & Mehrotra, R. (2003). Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, *12*(7), 796–807.

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

[Gedikli et al., 2007]Gedikli, S., Bandouch, J., von Hoyningen-Huene, N., Kirchlechner, B., & Beetz, M. (2007). An adaptive vision system for tracking soccer players from variable camera settings. In *Proceedings of the 5th International Conference on Computer Vision Systems (ICVS)*.

[Han et al., 2002]Han, M., Hua, W., Xu, W., & Gong, Y. (2002). An integrated baseball digest system using maximum entropy method. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, (pp. 347–350)., New York, NY, USA. ACM.

[von Hoyningen-Huene & Beetz, 2009]von Hoyningen-Huene, N. & Beetz, M. (2009). Rao-Blackwellized Resampling Particle Filter for Real-Time Player Tracking in Sports. In Ranchordas, A. & Araujo, H. (Eds.), *Fourth International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 1, (pp. 464–470)., Lisboa, Portugal. INSTICC, INSTICC press.

[von Hoyningen-Huene et al., 2007]von Hoyningen-Huene, N., Kirchlechner, B., & Beetz, M. (2007). GrAM: Reasoning with grounded action models by combining knowledge representation and data mining. In Rome, E., Hertzberg, J., & Dorffner, G. (Eds.), *Towards Affordance-based Robot Control*, LNAI 4760, (pp. 47–62). Springer.

[Huber, 1981]Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons, N.Y.

[Jin, 1994]Jin, J. S. (1994). Computational simulation of depth perception in the mammalian visual system. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, (pp. 451).

[Lao et al., 2006a]Lao, W., Han, J., & de With, P. H. N. (2006a). Automatic sports video analysis using audio clues and context knowledge. In *EuroIMSA*, (pp. 198–202).

[Lao et al., 2006b]Lao, W., Han, J., & de With, P. H. N. (2006b). Ball-path inference based on a combination of audio and video clues in tennis video sequences. In *IEEE Benelux/DSP Valley Signal Processing Symposium*, volume 1, (pp.–).

[Li & Sezan, 2001]Li, B. & Sezan, M. I. (2001). Event detection and summarization in sports video. In *CBAIVL '01: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01)*, (pp. 132)., Washington, DC, USA. IEEE Computer Society.

[Liang et al., 2005]Liang, D., Liu, Y., Huang, Q., & Gao, W. (2005). A scheme for ball detection and tracking in broadcast soccer video. In *Advances in Multimedia Information Processing - PCM 2005, 6th Pacific-Rim Conference on Multimedia, Jeju Island, Korea, November 13-16, 2005, Proceedings, Part I*, (pp. 864–875).

[MacKay, 2003]MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

[Pingali et al., 2000]Pingali, G., Opalach, A., & Jean, Y. (2000). Ball tracking and virtual replays for innovative tennis broadcasts. In *ICPR '00: Proceedings of the International Conference on Pattern Recognition*, (pp. 4152)., Washington, DC, USA. IEEE Computer Society.

[Pingali et al., 1998]Pingali, G. S., Jean, Y., & Carlbom, I. (1998). Real-time tracking for enhanced tennis broadcasts. In *CVPR*, (pp. 260–265).

[Quinlan, 1993]Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.

[Rui et al., 2000]Rui, Y., Gupta, A., & Acero, A. (2000). Automatically extracting highlights for TV baseball programs. In *ACM Multimedia*, (pp. 105–115).

[Särkkä et al., 2007]Särkkä, S., Vehtari, A., & Lampinen, J. (2007). Rao-blackwellized particle filter for multiple target tracking. *Information Fusion Journal*, *8*(1), 2–15.

[Särkkä et al., 2004]Särkkä, S., Vehtari, A., & Lampinen, J. (2004). Rao-blackwellized monte carlo data association for multiple target tracking. In *Proc. of Intl Conf. on Information Fusion*, volume 7, (pp. 583–590)., Stockholm.

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

[Sudhir et al., 1998]Sudhir, G., Lee, J. C. M., & Jain, A. K. (1998). Automatic classification of tennis video for high-level content-based retrieval. In *CAIVD '98: Proceedings of the 1998 International Workshop on Content-Based Access of Image and Video Databases (CAIVD '98)*, (p. 81)., Washington, DC, USA. IEEE Computer Society.

[Wang & Parameswaran, 2004a]Wang, J. & Parameswaran, N. (2004a). Detecting tactics patterns for archiving tennis video clips. In *IEEE Sixth International Symposium on Multimedia Software Engineering*, (pp. 186–192).

[Wang & Parameswaran, 2004b]Wang, J. R. & Parameswaran, N. (2004b). Survey of sports video analysis: research issues and applications. In *VIP '05: Proceedings of the Pan-Sydney area workshop on Visual information processing*, (pp. 87–90)., Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Witten & Frank, 2005]Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques* (2^{nd} ed.). San Francisco: Morgan Kaufmann.

[Xu et al., 2004]Xu, R. Y. D., Allen, J. G., & Jin, J. S. (2004). Robust real-time tracking of non-rigid objects. In *VIP '05: Proceedings of the Pan-Sydney area workshop on Visual information processing*, (pp. 95–98)., Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Yu & Farin, 2005]Yu, X. & Farin, D. (2005). Current and emerging topics in sports video processing. In *IEEE International Conference on Multimedia and Expo (ICME)*, (pp. 526–529).

[Yu et al., 2003]Yu, X., Xu, C., Leong, H. W., Tian, Q., Tang, Q., & Wan, K. (2003). Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In *ACM Multimedia*, (pp. 11–20).

[Zhou et al., 2000]Zhou, W., Vellaikal, A., & Kuo, C. C. J. (2000). Rule-based video classification system for basketball video indexing. In *ACM Multimedia Workshops*, (pp. 213–216).