

# (A) Vision for 2050

## Context-Based Image Understanding for a Human-Robot Soccer Match

Udo Frese<sup>1,2</sup>, Tim Laue<sup>2</sup>, Oliver Birbach<sup>2</sup>,  
Jörg Kurlbaum<sup>1</sup>, and Thomas Röfer<sup>2</sup>

<sup>1</sup> Fachbereich 3 - Mathematik und Informatik, Universität Bremen,  
Postfach 330 440, 28334 Bremen, Germany

E-Mail: {ufrese,jkur}@informatik.uni-bremen.de

<sup>2</sup> Deutsches Forschungszentrum für Künstliche Intelligenz GmbH,  
Sichere Kognitive Systeme, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany

E-Mail: {Tim.Laue,Oliver.Birbach,Thomas.Roefer}@dfki.de

**Abstract.** We believe it is possible to create the visual subsystem needed for the RoboCup 2050 challenge – a soccer match between humans and robots – within the next decade. In this position paper, we argue, that the basic techniques are available, but the main challenge will be to achieve the necessary robustness. We propose to address this challenge through the use of probabilistically modeled context, so for instance a visually indistinct circle is accepted as the ball, if it fits well with the ball’s motion model and vice versa.

Our vision is accompanied by a sequence of (partially already conducted) experiments for its verification. In these experiments, a human soccer player carries a helmet with a camera and an inertial sensor and the vision system has to extract all information from that data, a humanoid robot would need to take the human’s place.

## 1 Introduction

Soon after establishing the RoboCup competition in 1997, the RoboCup Federation proclaimed an ambitious long term goal (depicted in Fig. 1).

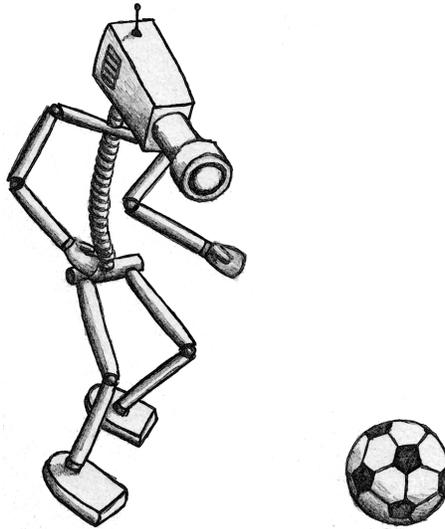
*“By mid-21st century, a team of fully autonomous humanoid robot soccer players shall win the soccer game, comply with the official rule of the FIFA, against the winner of the most recent World Cup.”*

*Hiroaki Kitano, Minoru Asada [2]*

Currently, RoboCup competitions take place every year. Within a defined set of different sub-competitions and leagues, incremental steps towards this big goal are made [3]. Although a rapid and remarkable progress has been observed

---

<sup>1</sup> This paper is an extended version of [1] presented at the celebration of the 60<sup>th</sup> birthday of Bernd Krieg-Brückner.



**Fig. 1.** Illustration of the RoboCup vision: An autonomous humanoid robot playing soccer.

during the first decade of these robot competitions, it is not obvious, if and how the final goal will be reached. There exist rough roadmaps, e.g. by [4], but in many research areas, huge gaps must be bridged within the next 40 years.

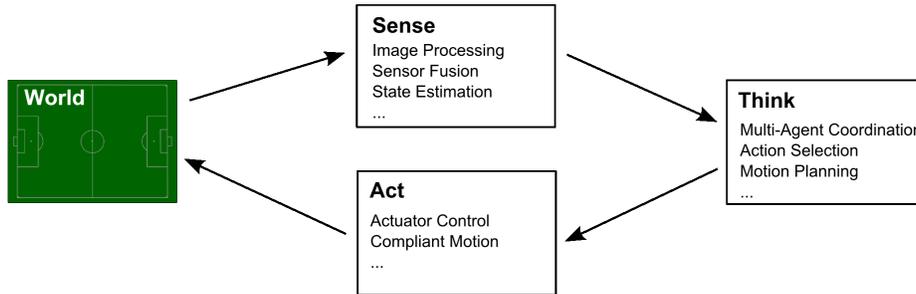
While this is obvious for several areas, e.g. actuator design and control, we claim that the situation is surprisingly positive for vision:

*Within the next decade, it will be possible to develop a vision system that is able to provide all environmental information necessary to play soccer on a human level.*

Annual RoboCup competitions are always bound to strict rule sets (defined for the state of the art of the competing robots) and demand competitive robot teams. Thus only incremental progress adapting to actual rule changes (which continuously raise the level of complexity) is fostered. By developing the aforementioned vision system independently of these competitions, we hope to set a new landmark which could guide the incremental development.

Because a *real* human level soccer robot will not be available for a long time, our vision is accompanied by a (partially already conducted) set of experiments that verify our claim without needing a robot.

This paper is organized as follows: Section 2 roughly identifies the challenges for playing robot soccer and compares them to the state of the art in robotics. In Sect. 3 we explain, why the basic techniques for the vision system are available. We argue, why the remaining challenge is robustness, for which we present our idea of a solution in Sect. 4. Finally, a sequence of experiments to verify our claim is described in Sect. 5.



**Fig. 2.** The Sense-Think-Act cycle roughly depicting major tasks for playing soccer with a humanoid robot.

## 2 Challenges for Playing Soccer

The global task of playing soccer consists of several different, interdependent challenges. We roughly categorize them according to the Sense-Think-Act cycle (see Fig.2). This should be considered as a possible architecture for illustration. In the following, the challenges are described in reverted order but with decreasing degree of difficulty.

### 2.1 Challenges for Actuation

The hugest obvious gap may be observed in the field of actuation. Nowadays, the probably most advanced humanoid robot, Honda’s ASIMO, is capable of running at a top speed of six kilometers per hour [5]. This is an impressive result, but still more than five times slower than the top speed of a human soccer player. A similar gap regarding kicking velocity has been pointed out by [6, 7]. They showed that a state-of-the-art robot arm (with a configuration comparable to a human leg) is six times slower than required to accelerate a standard soccer ball to an adequate velocity. It is still an open issue whether today’s motor technology could be developed further on enough, or if more efficient actuators, e.g. artificial muscles, will be needed. Since soccer is a contact sport leading to physical human-robot interaction [6], not only direct position control but also approaches for compliant motion, such as impedance control, need to be taken into account. As pointed out by [7], joint elasticity appears to be a crucial aspect for velocity as well as for safety.

Additionally, the problems of energy efficiency and power supply need to be solved. The ASIMO robot for example is, according to [5], capable of walking (with a speed of less than three kilometers per hour) for 40 minutes.

### 2.2 Challenges for Thinking

In this area, two different levels may be distinguished: motion planning and high-level multi-agent coordination. The latter is a research topic in the RoboCup

Soccer Simulation League since a while and has reached a remarkable level. Dealing with the offside rule as well as playing one-two passes are standard behaviors, complex group tasks as playing keep-away soccer serve as a test bed for learning algorithms [8]. This area could be considered to be already quite close to human capabilities.

On the other hand, when playing with real humanoid robots, sophisticated methods for motion planning are needed. The current research frontier on humanoid motion control is balancing and dynamic foot placement for walking robots. Algorithms for full-body motion planning exist [9], but are subject to restrictions that make them inapplicable to tasks as playing soccer.

Here is a big gap to human level soccer. As an example consider volley-kicking. The player has to hit the ball exactly at the right time, position, and velocity, with a motion compatible to the step pattern, allowing balancing, and considering opponents. Last but not least, all this must happen in real-time.

### 2.3 Challenges for Sensing

According to [2], it is evident that the robots' sensorial capabilities should resemble the human ones. Thus, we could assume to deal with data from cameras and inertial sensors emulating the human eyes and vestibular system. The required information are estimates of the own position and the positions of the ball and of other players. In case of tackles or dribbling, the latter will be needed to be recognized in more detail (e.g. the positions of the feet and limbs).

Current solutions for these tasks and our idea how to bridge the remaining gap are presented in the following section.

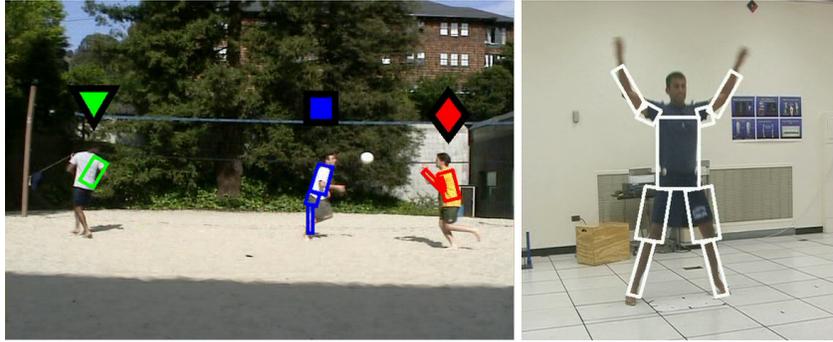
## 3 The Vision System

Our main thesis is that the “sense” part of the RoboCup 2050 challenge can be realized within a decade starting from the current state of the art in computer vision. This is remarkable, since the “act” and “think” parts are apparently light-years away from reaching human level performance and for computer vision in general, this is also true. The reason, why we believe such a vision system can be realized, is that unlike a household robot for instance, a soccer robot faces a rather structured environment.

### 3.1 State of the Art

The objects relevant in a soccer match are the ball, the goals, the line markings, and, of course, the players as well as the referees. Ball, goal, and line markings are geometrical features, i. e. circles and lines. There is a large number of algorithms for detecting them in images, from the classical Hough transform [10] up to a range of more elaborate methods [11].

Recognizing other players is more challenging. It is particularly difficult because we will probably need not only the general position but the detailed state



**Fig. 3.** Tracking sportsmen and their limbs: Three persons which are running and passing a ball (left image), one person performing jumping jacks (right image). Both images are results of [13] and are used with permission of Deva Ramanan.

of motion for close range tackling and to infer the player's action for tactical purposes. Fortunately, people tracking is an important topic in computer vision with a large body of literature [12]. An approach which is capable of tracking the body as well as the limbs of sportsmen (see Fig. 3) has been presented by [13].

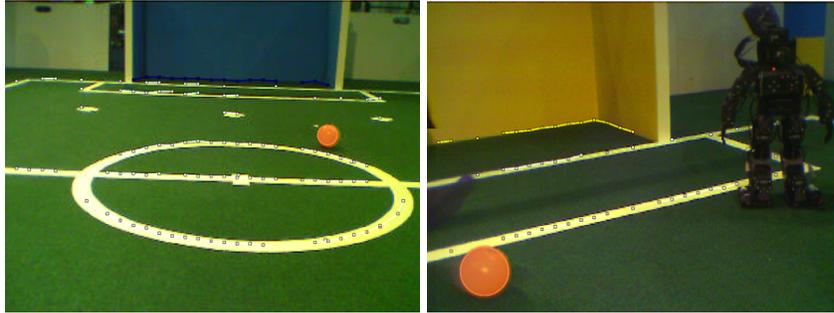
Furthermore, soccer scenes are lightly colored with green lawn and the players wearing colored clothes of high contrast. In the RoboCup competition, this idea is taken to an extreme, where most teams rely on color segmentation on a pixel-per-pixel basis as their primary vision engine (see Fig. 4). This will not be possible for real-world soccer, mainly due to changing lighting conditions. Still, color can provide a valuable additional cue, at least when looking below the horizon, where objects are in front of green lawn.

The background above the horizon, including the stadium and the audience is of course also visible and unfortunately rather undefined. However, if it is relevant for the soccer robot at all, then not for recognition, but only in the sense of a general landmark. For this purpose there are nowadays well working techniques, such as the Scale Invariant Feature Transform (SIFT) [14].

Overall, understanding a soccer scene from the player's perspective seems much easier than for instance understanding an arbitrary household, traffic, or outdoor scene. Indeed there are already half-automatic systems in the related area of TV soccer scene analysis, for example the ASPOGAMO system by [15, 16], proving that soccer scene understanding in general is on the edge of being functional.

### 3.2 Open Problems

So, is a vision system for the RoboCup 2050 challenge an easy task? We believe it is not. It is surprisingly a realistic task but well beyond the current state of the art. The first problem is that the camera is moving along with the head of the humanoid soccer robot. To predict a flying ball, the orientation of the



**Fig. 4.** Two images taken and processed by a humanoid soccer robot on a standard RoboCup field [17]. The colored dots and lines indicate perceived parts of goals and field lines; the recognized ball is bordered by an orange circle.

camera must be known very precisely. It seems unrealistic that the necessary precision can be obtained from the robot’s forward kinematic, since unlike an industrial robot, a humanoid robot is not fixed to the ground. So our solution is to integrate an inertial sensor with the camera and fuse the complementary measurements of both sensors in a probabilistic least-square framework.

The second problem is the player’s perspective. It is much more difficult than the overview perspective used in TV soccer scene analysis. In the TV perspective the scale of an object in the image varies by a factor of about 3 [15, Fig. 5] whereas in the player’s perspective it can vary by a factor of 250 assuming the distance to an object ranging from 0.5m to 125m. Hence, for instance the people detection algorithm must handle both extreme cases, a person only the size of a few pixels, where an arm or a leg maybe thinner than a single pixel and a person much larger than the camera’s field of view, only partially visible. Furthermore, in an image from the player’s perspective, other players will extend beyond the green lawn of the field into the general background. Hence it is not possible to search for non-green blobs as an easy first processing step. This can also happen for a flying ball, which is then particularly difficult to detect.

However, the third and most severe problem is, that from our experience, most of the academic computer vision systems perform on the level of lab demonstrators requiring nicely setup scenes and lighting conditions and usually considerable parameter tweaking. So, to summarize, for the vision part of the RoboCup 2050 challenge, we do not need a new level of functionality as for many other grand challenges, but we need a new level of robustness.

## 4 Robustness Through Context

We propose to address the question of robustness by utilizing probabilistically modeled context information, formulating the overall scene understanding and prediction problem as a global likelihood optimization task. Using contextual

associations is not entirely new [18, 19] and recent progress has shown that exploiting context leads to promisingly robust results. Recent work worth noting includes modeling of object to object relations as done by [20]. Here, the authors improved the object categorization accuracy by integrating a post-processing step that maximizes the object categorization agreement based on contextual relevance. More generally, [21] proposed a hierarchical framework allowing interaction between objects and regions within the image plane.

Leaving the image plane and therefore 2D models behind, [22] propose a mechanism for estimating rough 3D scene geometry from a single image which is then used as an additional input to the object detection methods. Therefore, detections which fail to fit into the estimated scene geometry at the detected position are rejected from the global scene interpretation (e. g. pedestrians who do not stand on the ground plane). This idea is taken one step further by [23] and [24] where the relationships between objects and scene geometry is modeled, allowing suitable recognition of objects for a mobile robot application.

These promising results show that exploiting context within the image and the scene is particularly well suited to the task of creating a vision system for RoboCup 2050 and also that this task is well suited to study this methodology.

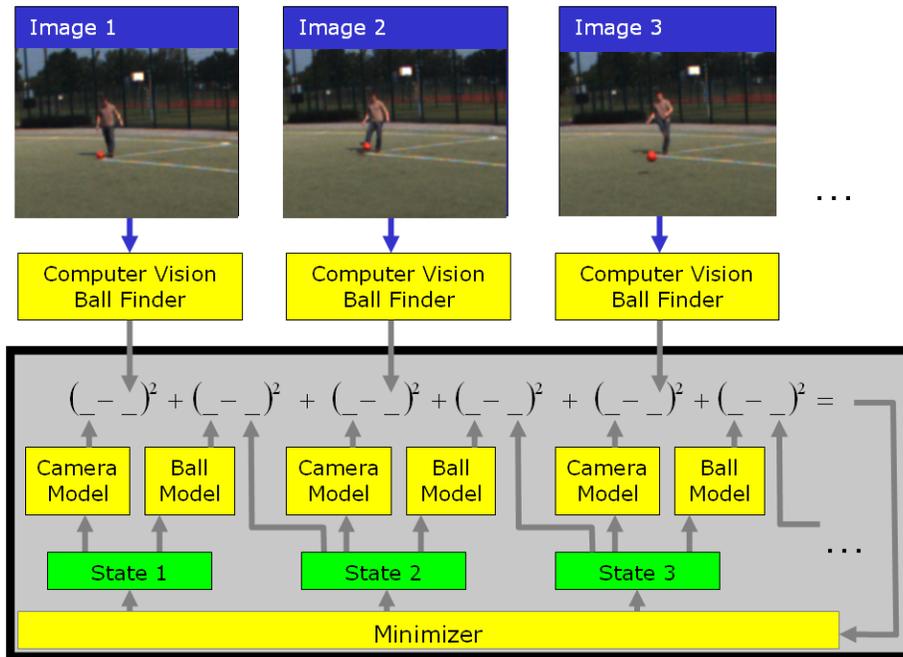
#### 4.1 Data-Driven Bottom-Up Processing

Most current vision systems use a data-driven bottom-up approach [25, 26, 16, as examples]. Usually, low level features are extracted from the image and then aggregated through several stages to high level information. Each stage may incorporate background knowledge at its own level but does not take information from higher levels into account. It simply takes some input from the previous lower level and passes the result of the computation to the next higher level.

As an example, a classical Hough transform starts by classifying pixels as edge or not by thresholding the result, for instance, of a Sobel filter. Similarly, the system by [16] starts by classifying pixels as lawn or not on the basis of their color. This is a hard decision taken on the lowest level without any higher level knowledge, such as the fact that we are looking for a ball or the ball's motion model. Such a pixel-wise classification can be very ambiguous. Often we could, for instance, classify a borderline pixel correctly as belonging to the ball, although it looks rather greenish, if we considered the context of the ball or its motion model. However, in conventional vision systems, on the low level this knowledge does not exist and on the higher level, the fact, that this pixel was borderline in the classification, is lost due to committing to a hard decision on the lower level.

To make this idea more concrete, we will now describe a ball tracking system in the conventional data-driven design (Fig. 5). We will later discuss how it would look like in the context based design we propose in this paper.

In a data-driven approach, a *ball finder* algorithm is executed on every image yielding the 2D image circle (position, radius) corresponding to the ball (or several hypotheses thereof). These circles are passed to a probabilistic estimator, such as Gaussian-Maximum-Likelihood that finds the mostly likely states, i. e.



**Fig. 5.** Information flow in a ball tracker based on the conventional data-driven paradigm (see text). The ball finder is not part of the optimization loop (gray box).

ball positions and velocities. Such an estimator is based on model likelihoods, i. e. functions with uncertainty, that specify the following:

1. A *camera model* defines when the ball was at a certain position, where it would be in the image. This links 3D positions to 2D circles.
2. A *ball motion model* defines when the ball was at a certain position having a certain velocity, where it would be and which velocity it would have after one timestep. This links 3D positions and velocities over time.

The estimator then finds the most likely states by minimizing a sum of squared differences, which indicates how likely an assumed sequence of states is. It consists of 1. the squared differences between where the ball has been observed by the ball finder and where the ball model would expect it to be observed if the assumed state sequence was true and 2. the squared difference between the states predicted by the ball model and the assumed corresponding states.

The models 1. and 2. are Gaussian likelihoods, i. e. formulas with uncertainty, so there is a wealth of algorithms for handling this kind of problems. In particular, figure 5 is a conceptual description. Most actual implementations do not minimize over the whole state sequence in every step but instead use an incremental estimator, such as the Extended or Unscented Kalman Filter [27–29]. However, this changes only the computation not the result. The key point of this

approach is that the ball finder, more general the computer vision, is outside the optimization loop and there is no information flow from the estimator back to the ball finder. Hence the ball finder cannot take the high level context of the ball’s motion into account when deciding on the low level, whether something is accepted as a ball or not.

## 4.2 Global Likelihood Optimization

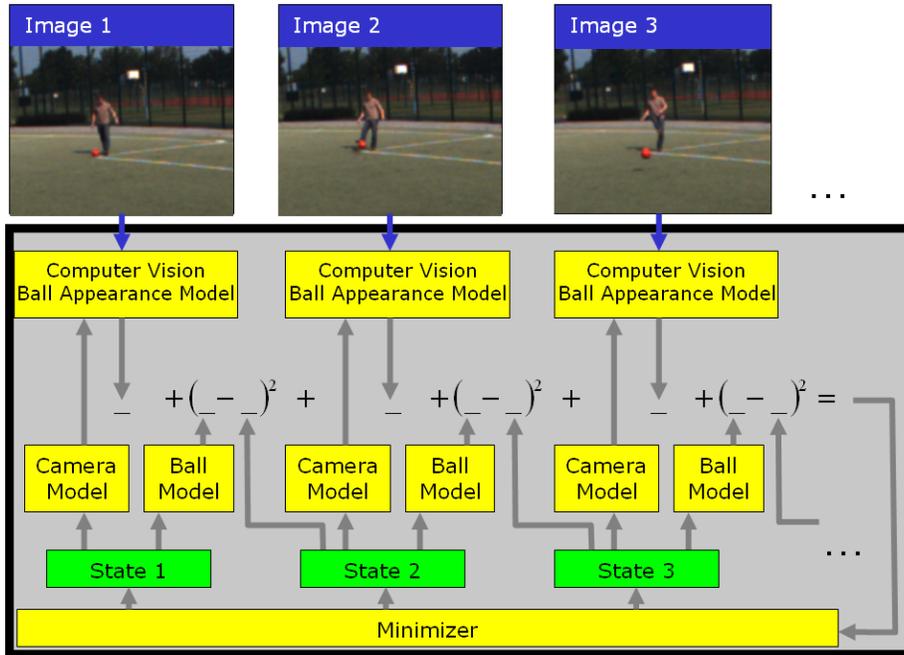
We believe that much of the brittleness of current vision systems originates from this way of committing to hard decisions on an early level of processing. So our approach for increased robustness is an overall likelihood optimization.

The problem is to understand an image sequence, i. e. estimating over time. Indeed, successive images are linked by a motion model and this provides most of the context we want to build upon. However, we propose not to use incremental filters, such as EKF, but to look back into the raw images of the last few seconds at least. This approach has surprising advantages. Imagine the ball is kicked, but during the first 100ms there is too little contrast to the background so it is not detected. Now when it is detected, there is new information on where the ball has been before from the ball’s motion model. The old images are still in memory and tracking the ball back in time is much less ambiguous than finding the ball without context and will probably succeed. Paradoxically, once the system has detected the ball, it has already observed it for 100ms. The first useful prediction is not delayed at all, because prior to that the ball must have been observed for some time anyway. In the example above, we would realize this (Fig. 6) by replacing the ball finder by

3. a *ball appearance model* that indicates for a given circle in the image plane how much it looks like a ball. This links the 2D image circle to the actual image.

In the data-driven approach the ball finder determines where the ball is and the camera model computes where the ball should be according to the state. Then both are compared with the difference indicating the likelihood of the state. In the global likelihood optimization approach, instead the camera model computes where the ball should be and then the ball appearance model computes how “ball-like” the image looks there. The difference is, that now the ball appearance model, i. e. the computer vision, is inside the optimization loop. Now the lower layer does not have to commit early to a single image circle, but instead it gradually assesses different circles in the image as requested by the minimization algorithm. In this approach, an indistinct ball would get a lower likelihood in 3. but this could be compensated by 1. and 2. if it fits well to the context of a flying ball. This mechanism allows the now implicit ball finder to utilize the context provided by the ball motion and implement the behavior described above.

As a final remark, the provided functionality greatly resembles the priming mechanism known from psychology [30]. This makes us even more believe that the approach of a global likelihood optimization directly in the images is an elegant way to greatly increase robustness of computer vision algorithms.

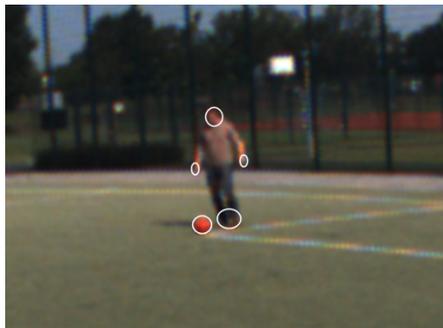


**Fig. 6.** Information flow in a ball tracker based on the global likelihood optimization paradigm (see text). Here, computer vision, i.e. the ball finder, is part of the optimization loop (gray box).

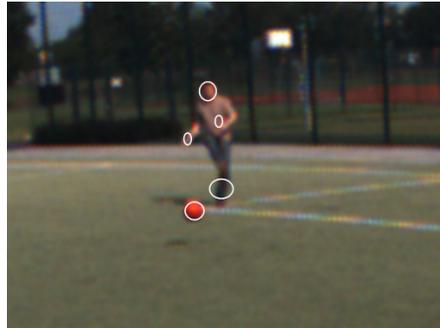
### 4.3 Exploiting Semantic Context

Doubtless, if the goal is to track a flying ball, the ball’s motion model provides the most valuable context. However, there is other, more semantic information that could be incorporated to increase robustness. Figure 7 illustrates how a broader context could help distinguishing the real ball from false responses of the computer vision. We have constructed this example to illustrate what we believe could happen. It is not based on real computation.

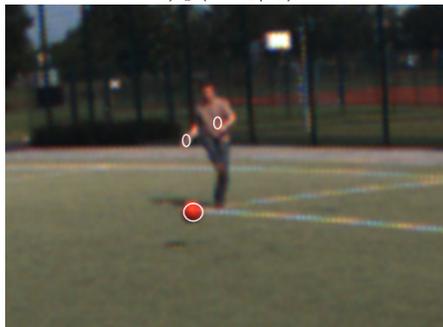
In a) only the first image  $I_1$  is considered ( $p(\text{ball}_1|I_1)$ ). The computer vision gives varying responses for different image circles. After applying the necessary threshold needed to give a final result, there are five possible ball locations in the example, namely the true ball and the player’s head, hands, and both feet together. The same happens in b), where in the example the 9th image  $I_9$  is considered alone ( $p(\text{ball}_9|I_9)$ ). Now in c) all images  $I_{1..9}$  are considered together ( $p(\text{ball}_9|I_{1..9}, \text{ball model})$ ). The link between the images is provided by the ball motion model as described in the previous section. In the example, the player’s head and feet could be ruled out, because their motion does not correspond to the ball motion model. However, in this example, the two hands remain valid hypotheses, because their side-swinging motion would actual correspond to a distant flying ball. In d), the additional con-



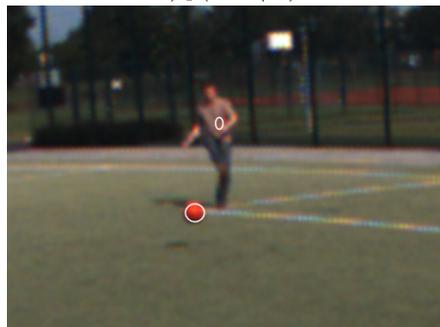
a)  $p(\text{ball}_1|I_1)$



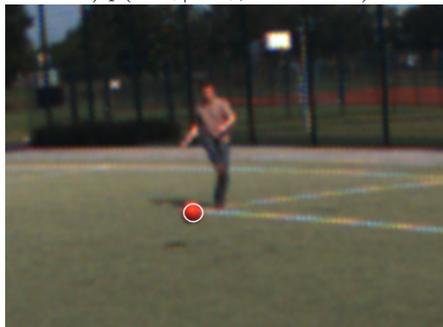
b)  $p(\text{ball}_9|I_9)$



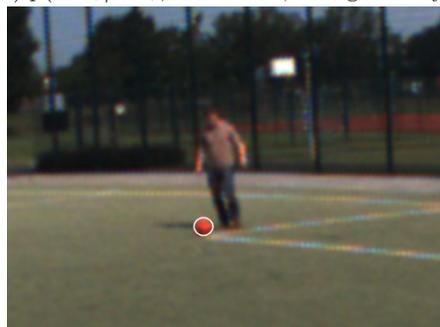
c)  $p(\text{ball}_9|I_{1...9}, \text{ball model})$



d)  $p(\text{ball}_9|I_{1...9}, \text{ball model, field geometry})$



e)  $p(\text{ball}_9|I_{1...9}, \text{ball model, field geometry, game semantics})$



f)  $p(\text{ball}_1|I_{1...9}, \text{ball model, field geometry, game semantics})$

**Fig. 7.** An illustration, how different types of context information could help to distinguish the real ball from things that look like a ball to the computer vision (see text).

text of the field geometry is added ( $p(\text{ball}_9|I_{1\dots 9}, \text{ball model, field geometry})$ ). This context makes the player’s right hand being a ball very unlikely, because from the image position and radius of the circle, the ball would be outside of the field. In e), additional information about game semantics is added ( $p(\text{ball}_9|I_{1\dots 9}, \text{ball model, field geometry, game semantics})$ ). From the game situation, the system expects the player to kick the ball towards the camera. A ball flying sideward as the falsely detected left hand is implausible with this context and can be ruled out. Of course, as explained above, once the system has found a unique ball in image 9, it already knows backwards, via the ball motion model, which was the correct ball in image 1. This situation is shown in f) ( $p(\text{ball}_1|I_{1\dots 9}, \text{ball model, field geometry, game semantics})$ ). So the system has already observed the ball for 9 images and can immediately provide a prediction.

How could this example be realized? First of all, using game semantics requires to detect the players, lines and goals to have information about the game situation. But how could the semantics of soccer be modeled? The classical AI approach would be to use logic, for instance ontologies to describe a soccer match. However, typical ontologies, such as OWL [31] express crisp knowledge in some sort of logic and are difficult to combine with the probabilistic and vague knowledge discussed in the previous section.

While in the long run modeling a rich and complex semantics, as ontologies provide, is desirable, probably the first step would be to start with a shallow semantics that is more easily integrated with probabilistic motion and perception models. A candidate would be to run a Hidden Markov Model (HMM) [32] for each player to model the different actions (kicking, running, tackling, etc.) a player may take, maybe subdivided into different phases. Then one could define how each phase would affect the players motion and the motion of the ball. The advantage is that easy algorithms for inference in HMM exist and even more important that HMM can be well combined with continuous probabilistic estimation by the Interacting Multiple Model Filter (IMM) [33, §11.6.6].

Overall, it is obvious that modeling the player’s behavior is essential for understanding a soccer match as a whole and is helpful even for tracking the ball. However, it is still rather unclear how the required background knowledge can be modeled in a way that can effectively help a computer vision system.

## 5 Proposed Experiments

For a vision to become reality, realistic intermediate steps are necessary. It would not help, if we build a vision system now but then had to wait until a human level soccer robot is available. So we propose a sequence of experiments that, without a humanoid robot, ultimately allows to verify that the proposed system is appropriate for human level soccer (Fig. 8).



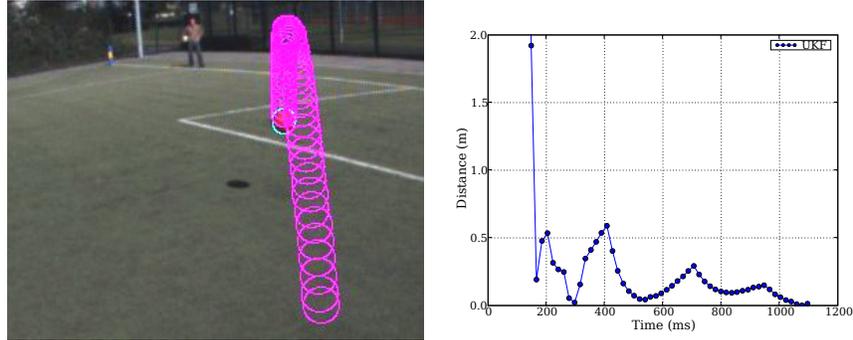
**Fig. 8.** Our proposed experiment: Mount a camera and an inertial sensor on the head of a human soccer player and use them to extract all the information, a humanoid soccer robot would need to take the human's place.

### 5.1 Helmet Camera with Inertial Sensor

The basic idea is to let a human soccer player wear a helmet with a camera and an inertial sensor and verify, that the information extracted by the vision system from the sensor data, would allow a humanoid robot to take the human's place.

As a first experiment, we propose to record data from a soccer match and run the vision system on that data offline. Since it is hard to obtain ground-truth data, we would use our expert's judgment to assess, whether the result would be enough for a humanoid robot to play soccer. It is very advantageous to work on recorded data allowing to reproduce results for debugging and analysis and to run the system even if it is still not real-time. Overall, it allows to first concentrate on functionality and robustness instead of computation time and integration.

We have already conducted a very first experiment [29, 28], where the ball and the field lines are manually extracted from the recorded images (available on request). The ball's trajectory is predicted by least-square estimation using the likelihood functions 1. and 2., as well as corresponding equations for how the inertial sensor observes the free motion of the camera (Fig.9). The results indicate that if the ball can be detected in the image with about one pixel precision, the prediction would be precise enough. We believe that these kinds of studies which deliberately discard essential aspects, such as integration, real-time computation, or autonomy are undervalued by the community who favors full system approaches. But even from a full system perspective, it is much more valuable to obtain an extensive result on a subsystem which then can guide the full system design than to do another increment on a full system.



**Fig. 9.** Predicting the trajectory of a flying ball from a moving camera-inertial system. As an initial study, the ball, the lines, and the goal corners have been manually extracted from the images. From this data, the trajectory of the ball is predicted (left). The right plot shows the error of the predicted touchdown point varying over time. It shows, even though the camera is moving, the prediction is roughly precise enough for interception.

## 5.2 Motion Capture Suit

Departing from the experiment above, one might ask whether more sensors are needed than just camera and inertial. Both human and humanoid robot can derive their motion from the joint angles. This provides the horizontal motion (odometry) and the height over ground. The horizontal motion facilitates localization and the height derived from vision is much less precise. Indeed, we experienced that the uncertain height is a major part of the error in Fig. 9.

An intriguing idea is to equip the human player with a *tracker-less* motion capture suit [34] measuring joint angles. Apart from providing the kinematic information discussed above, it also provides the trajectory of both feet. If the player hits the ball, one can compare the predicted ball trajectory with the real foot trajectory and evaluate the precision. This is important since in such an experiment, ground truth is not available.

## 5.3 Virtual Reality Display

The experiments above have the drawback that they are evaluated by an expert looking at the vision system's output. The most direct proof that this is all you need for playing soccer would be to give a human just that output via a head mounted display and see whether s/he can play.

The approach is of course fascinating and direct, but we have some concerns regarding safety. Anyway, this experiment becomes relevant only after we are convinced in principle, that the system is feasible. So this is something to worry about later.

## 6 Conclusion

In this position paper, we have outlined the road to a vision system for a human-robot soccer match. We claim that, since soccer is a structured environment, the basic techniques are available and the goal could be reached within a decade. The main challenge will be robustness, which we propose to address by optimizing a global likelihood function working on a history of raw images. We have outlined a sequence of experiments to evaluate such a vision system with data from a camera-inertial system mounted on the head of a human soccer player.

The reason for being confident such a system can be realized within a decade is the insight that it does not need general common-sense-reasoning AI. This is good news for the RoboCup 2050 challenge. But it suggests that, even when we meet that challenge, it does not imply we have realized the dream of a thinking machine, the whole challenge had started with.

That would not be the first time.

## References

1. Frese, U., Laue, T.: (A) vision for 2050: The road towards image understanding for a human-robot soccer match. In: Proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics. (2008)
2. Kitano, H., Asada, M.: RoboCup Humanoid Challenge: That's One Small Step for A Robot, One Giant Leap for Mankind. In: International Conference on Intelligent Robots and Systems, Victoria. (1998) 419–424
3. RoboCup Federation: RoboCup Official Site (2008) <http://www.robocup.org>.
4. Burkhard, H.D., Duhaut, D., Fujita, M., Lima, P., Murphy, R., Rojas, R.: The Road to RoboCup 2050. IEEE Robotics and Automation Magazine **9(2)** (2002) 31–38
5. Honda Worldwide Site: Honda World Wide — Asimo (2008) <http://world.honda.com/ASIMO/>.
6. Haddadin, S., Laue, T., Frese, U., Hirzinger, G.: Foul 2050: Thoughts on Physical Interaction in Human-Robot Soccer. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. (2007)
7. Haddadin, S., Laue, T., Frese, U., S.Wolf, Albu-Schäffer, A., Hirzinger, G.: Kick it with elasticity: Safety and performance in human-robot soccer. Robotics and Autonomous System - Special Issue on Humanoid Soccer Robots (2009) to appear.
8. Stone, P., Sutton, R.S., Kuhlmann, G.: Reinforcement Learning for RoboCup-Soccer Keepaway. Adaptive Behavior **13(3)** (2005) 165–188
9. Kuffner, J.J., Kagami, S., Nishiwaki, K., Inaba, M., Inoue, H.: Dynamically-stable Motion Planning for Humanoid Robots. Auton. Robots **12(1)** (2002) 105–118
10. Davies, E.R.: Machine Vision. Theory, Algorithms, Practicalities. Morgan Kaufmann (2004)
11. Guru, D., Shekar, B.: A simple and robust line detection algorithm based on small eigenvalue analysis. Pattern Recognition Letters **25(1)** (2004) 1–13
12. Price, K.: The annotated computer vision bibliography (2008) <http://www.visionbib.com/>.
13. Ramanan, D., Forsyth, D.: Finding and tracking people from the bottom up. In: IEEE Conference on Computer Vision and Pattern Recognition. (2003)

14. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2) (2004) 91 – 110
15. Beetz, M., v. Hoyningen-Huene, N., Bandouch, J., Kirchlechner, B., Gedikli, S., Maldonado, A.: Camerabased observation of football games for analyzing multiagent activities. In: *International Conference on Autonomous Agents*. (2006)
16. Beetz, M., Gedikli, S., Bandouch, J., Kirchlechner, B., v. Hoyningen-Huene, N., Perzylo, A.: Visually tracking football games based on tv broadcasts. *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India* (2007)
17. Laue, T., Röfer, T.: Particle filter-based state estimation in a competitive and uncertain environment. In: *Proceedings of the 6th International Workshop on Embedded Systems, VAMK, University of Applied Sciences; Vaasa, Finland* (2007)
18. Ullman, S.: Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex* **5**(1) (1995)
19. Binnig, G.: Cellenger automated high content analysis of biomedical imagery (2004)
20. Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., Belongie, S.: Objects in Context. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2007)
21. Kumar, S., Hebert, M.: A Hierarchical Field Framework for Unified Context-Based Classification. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2005)
22. Hoiem, D., Efros, A.A., Hebert, M.: Putting Objects in Perspective. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2006)
23. Ess, A., Leibe, B., Gool, L.V.: Depth and Appearance for Mobile Scene Analysis. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2007)
24. Leibe, B., Schindler, K., Cornelis, N., Gool, L.V.: Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(10) (2008) 1683–1698
25. Frese, U., Bäuml, B., Haidacher, S., Schreiber, G., Schaefer, I., Hähnlle, M., Hirzinger, G.: Off-the-shelf vision for a robotic ball catcher. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Maui*. (2001) 1623 – 1629
26. Röfer, T., et al.: GermanTeam RoboCup 2005 (2005) <http://www.germanteam.org/GT2005.pdf>.
27. Welch, G., Bishop, G.: An introduction to the kalman filter. Technical Report TR 95-041, University of North Carolina (1997)
28. Birbach, O.: Accuracy analysis of camera-inertial sensor based ball trajectory prediction. Master's thesis, Universität Bremen (2008)
29. Kurlbaum, J.: Verfolgung von ballflugbahnen mit einem frei beweglichen kamera-inertialsensor. Master's thesis, Universität Bremen (2007)
30. Wikipedia: Priming (psychology) — wikipedia, the free encyclopedia (2008) [Online; accessed 10-December-2008].
31. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P., Stein, L.: OWL web ontology language reference. Technical Report REC-owl-ref-20040210/, W3C (2004)
32. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C* **34**(3) (2004) 334–352

33. Bar-Shalom, Y., Li, X., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation. John Wiley & Sons, Inc. (2001)
34. Xsens Technologies B.V.: Moven, Inertial Motion Capture, Product Leaflet. (2007)

## Authors' Vitae



**Udo Frese** received his Ph.D. from University of Erlangen-Nürnberg in 2004. From 2004 until 2008 he has been a post-doc researcher in Bernd Krieg-Brückner's lab at University of Bremen and in the German Research Center for Artificial Intelligence (DFKI). He is now assistant professor for real-time computer vision at University of Bremen and further on affiliated with the DFKI. His research topics are real time computer vision, in particular simultaneous localization and mapping (SLAM) and computer vision for sport robotics.



**Tim Laue** received his diploma in computer science from the University of Bremen, Germany, in 2004. Until 2008 he has been a researcher in Bernd Krieg-Brückner's working group at the University of Bremen. Currently, he is with the German Research Center for Artificial Intelligence at the research group for Safe and Secure Cognitive Systems, Bremen, Germany. His research and development fields are simulation and localization for autonomous robots, i. e. for wheelchairs as well as for different soccer robots participating in RoboCup competitions.



**Oliver Birbach** received the diploma degree in computer science from University of Bremen in 2008. He is currently a junior researcher at the German Research Center for Artificial Intelligence in Bremen. His main research areas are multi-target tracking, context-based computer vision and computer vision in sports environments.



**Jörg Kurlbaum** studied computer science at the University of Bremen with personal preference in robotics. After he received the diploma in 2007, he started to work in the SFB/TR8 Spatial Cognition group on map building and representation for real-time operator applications. Jörg is affiliated with Udo Frese in his real-time vision group.



**Thomas Röfer** received his diploma in computer science and his Dr.ing. degree from Universität Bremen, Germany, in 1993 and 1998, respectively. He is in the Executive Committee of the RoboCup Federation, and he is member of the Transregional Collaborative Research Center SFB/TR 8 "Spatial Cognition" at Bremen. He is currently with the German Research Center for Artificial Intelligence at the research group for Safe and Secure Cognitive Systems, Bremen, Germany. His research interests include rehabilitation robotics, robot soccer, real-time computer vision, world modeling, and humanoid robots.