3D Object Localization via Stereo Vision using an Omnidirectional and a Perspective Camera

Uwe-Philipp Käppeler, Markus Höferlin and Paul Levi

***Abstract*— This paper describes our stereo vision method which is combining an omnidirectional and a perspective camera. It was developed for our robot soccer team *1. RFC Stuttgart*, which attends RoboCup competitions every year. The common approach to stereovision leads to high deviations from the real positions when it is not possible to synchronize the cameras for tracking moving objects and when the orientations of the cameras of a soccer playing robot can be decalibrated during a game. Therefore we introduce an object localization for the RoboCup scenario combining the most accurate position information from each camera system. Our method for a reliable three-dimensional position estimation can be used to track a flying ball after being kicked by a robot.**

I. Introduction

In the last years the RoboCup soccer games became faster and faster [1], [2]. It showed that an omnidirectional drive and an omnidirectional vision system using a hyperbolic mirror allow a successful participation in the tournaments. Therefore almost every participating team adopted an accordant robot configuration. The ball is lying or rolling on the ground of the soccer field most of the time during a game but powerful kicker devices of the robots, which arised in the last years, make it possible to kick the ball in a way that it lifts-off the ground and flies over several meters.

The localization of a flying ball is a weak point of the omnidirectional camera system. A three-dimensional localization of the ball is difficult because the limited resolution of the camera is used to observe the whole environment of the robot at once and therefore the estimation of the distance to the ball based on its size in the image is inaccurate. As shown in the upper images of Figure 1, which were recorded by the omnidirectional camera of the robot, the size of the ball in the images differs only a few pixels even if the distance of the ball to the camera varies a lot. According to the images from the robots camera, the lower images show the position of the ball in both situations. On the left side, the ball is lying on the ground of the soccer field. On the right side the ball is put on a socket and is positioned on the line between its first position on the ground and the mirror of the omnidirectional camera. This results in equal pixel coordinates of the ball centers in the images recorded by the robot, as can be seen in the upper image. Assumed the ball is lying on the ground the robot estimates the same position of the ball in both situations.

U. Käppeler and Paul Levi are with the Institute of Parallel and Distributed Systems, University of Stuttgart, 70569 Stuttgart, Germany Uwe-Philipp.Kaeppeler@ipvs.uni-stuttgart.de

M. Höferlin is with the Visualization Research Center, University of Stuttgart, 70569 Stuttgart, Germany Markus.Hoeferlin@vis.uni-stuttgart.de

The schema in Figure 2 shows the erroneous position estimation when assuming the ball is lying on the ground. The distance on the ground d_g is a function of the real distance d_r which is measured parallel to the ground, the height of the camera h_c and the height of the ball h_b and given by

$$d_g = \frac{d_r \cdot h_c}{h_c - h_b}$$

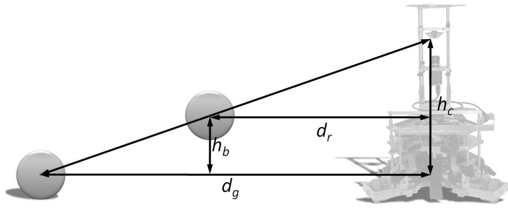


Fig. 2: Schema of the position estimation assuming the ball is lying on the ground

Figure 3 shows a scene of the RoboCup finals 2009 in Graz. The upper image shows the scene as recorded by our automatic cameraman with overlays showing the states of our robots [3]. The referee is positioning the ball on the center of the field but it is still above ground. The 2D Visualization of the field in the lower image shows our robots and the estimated positions of the ball which are marked by bright yellow circles as the positions are communicated between the robots via a wireless network. The overestimated distances of the balls to the robots are visible.

Fig. 3: Erroneous position estimation of the ball above ground during the RoboCup finals 2009 in Graz

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 7-12

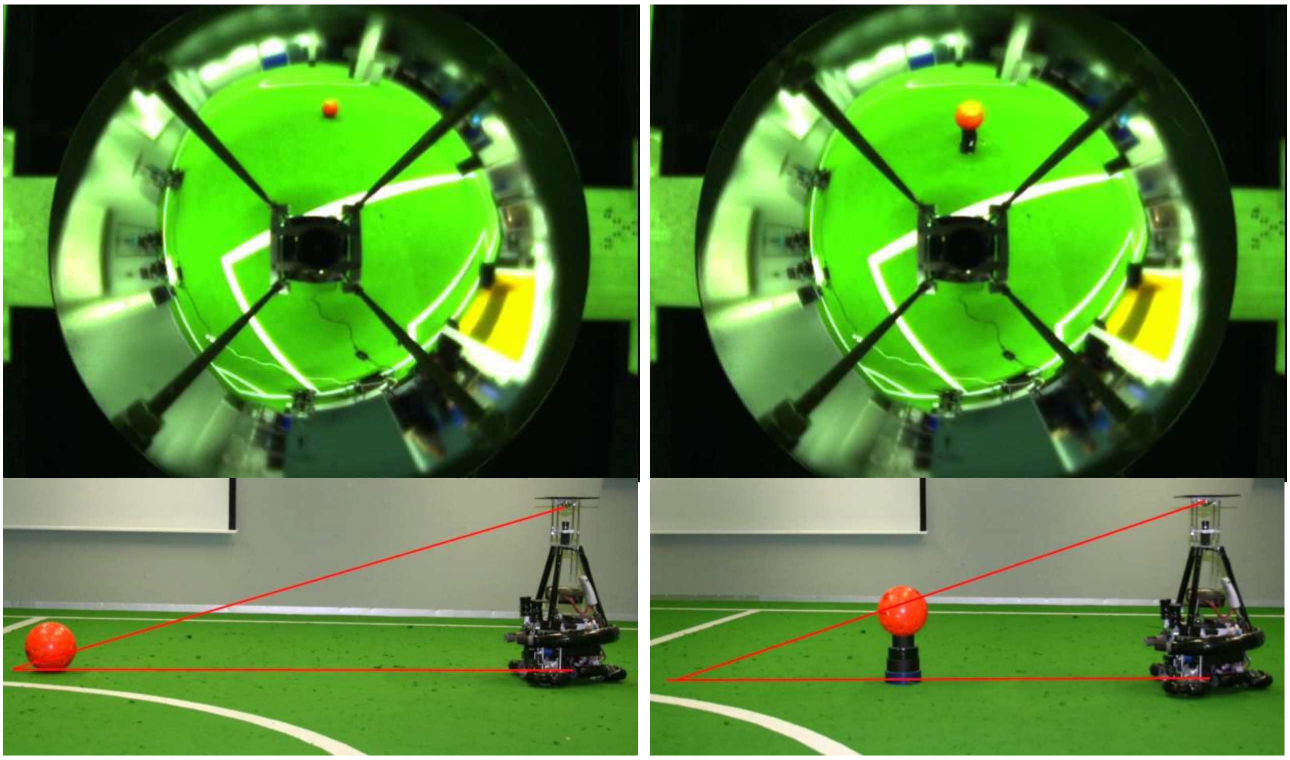


Fig. 1: Comparison of images recorded by the omnidirectional camera system: Different positions of the ball (lower images) result in the same position in the image of the omnidirectional camera system (upper images). On the left side the ball is lying on the ground, on the right side it is put on a socket.

Of course, most robocup teams take advantage of this circumstance and especially most of the goal kicks are performed shooting the ball above the height of the robots. It is difficult for a goalkeeper robot to stop a flying ball heading towards the goal when its position is estimated faultily and its high velocity makes it appear in only a few images taken by the camera. Therefore, the localization of the ball and the vision system had to be extended. To reliably detect and localize flying balls we mounted a second camera on our robots and developed a stereovision system combining object coordinates obtained by an omnidirectional [4] and a perspective camera [5]. During a RoboCup soccer game the orientation of the cameras can be decalibrated for example when the camera mount is hit by a kicked ball. Therefore our 3D localization is based on the most reliable information provided by the omnidirectional and perspective camera.

II. Related Work

In this paper, we present a method for three-dimensional object localization based stereo vision that combines object coordinates obtained from different camera systems. 3D object localizatino based on only one omnidirectional camera is described by Taiana *et al.* [6], Jamzad *et al.* [7] or Olufs *et al.* [8]. Our method for combining several cameras differs from algorithms that compute complete depth images based on epipolar geometry as described and compared for example by Badino [9], Badino *et al.* [10] or Richard Hartley and Andrew Zisserman [11]. A method for stereo vision using two cameras on pan-tilt units is described by Gehrig *et al.* [12], where inconsistent fields of view are combined. Methods for stereo vision based on omnidirectional camera systems and epipolar geometry are described by Zhu [13] and Cagnoni *et al.* [14]. Since it is difficult to mount two omnidirectional camera systems on our robot due to constraints defined in the RoboCup rules, our approach uses an omnidirectional camera [4] and a classical, perspective camera [5]. The combination of these cameras is used for a three-dimensional object localization.

III. Stereo vision combining most accurate measurements

A. Camera systems and ball detection

Two different camera systems are mounted on our robot. The first camera system is omnidirectional. The camera itself is directed upwards towards a hyperbolic mirror which is mounted 12 cm above the camera. The height of a robot in RoboCup is limited to 80 cm which therefore is the height of our mirror above ground. The second camera is mounted parallel to the ground and heading to the front. Its height above ground is 50 cm. The field of view of the perspective camera is covered by the field of view of the omnidirectional camera except for the upper part of the image showing obstacles above the height of the mirror. The image processing algorithms for the ball detection based on color segmentation and pattern recognition in the images of the perspective camera and the omnidirectional camera are described by Burla [15]. The ball is a symmetric object. Therefore, it is possible to determine its center point and base point from any direction if the ball is not partially occluded by any other obstacle. The base point of the ball can be determined in each image and it can be used as a reference point for comparing the positions.

B. Stereo vision

A base of stereo vision algorithms is the problem of correspondence as described by Badino [9]. A three-dimensional localization of an object is based on determining corresponding reference points of that object in both images. Afterwards the coordinates of the object can be determined by epipolar geometry [11]. Straight lines in three-dimensional coordinates can be computed based on the image coordinates of the reference points. Each straight line is defined by one point according to the position of the camera or the camera's mirror and the direction given by the position of the reference point in the image. Afterwards the three-dimensional coordinates of the object can be determined by the intersection of both straight lines. Based on inaccuracies it is possible that the straight lines are skew and do not intersect. The corresponding coordinates of the object have to be determined as the point which is the closest to both straight lines. Especially the tracking of moving objects results in skew lines, when it is not possible to synchronize the cameras. Timestamps of the images and filters for estimating the positions in the images for identical points in time by information aging can be used to reduce the deviations [16]. The filter for information aging can be applied to the position of the reference points in the image for the object localization. The schema for a triangulation and straight line intersection for the combination of a perspective camera and an omnidirectional camera is depicted in Figure 4.

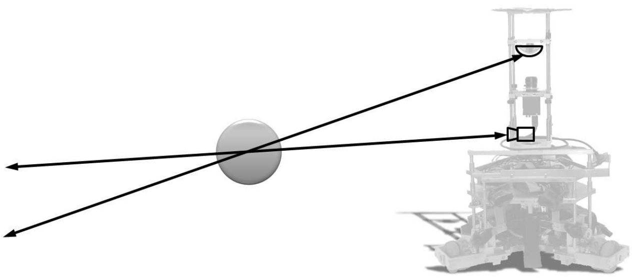


Fig. 4: Schema of the three-dimensional position estimation based on triangulation

Imprecise calculations of ball positions in image coordinates lead to additional inaccuracies for determining the straight lines between the object and the cameras. The limited resolution of the omnidirectional camera influences the accuracy of the three-dimensional object localization. The direction to a reference point, which is described by the pan angle ϕ in Figure 5, can be determined with a resolution of less than one degree and the determination of this direction is not influenced by the height of the object above ground or a small decalibration of the midpoint of the mirror in the image. But the tilt angle of the line from the mirror to the object has to be determined by the distance of the reference point to the center of the image. This distance has a resolution of only 160 pixels in our image. Small decalibrations concerning the midpoint of the mirror in the image lead to deviations in this angle.

For the perspective camera, we have different inaccuracies. A decalibration of the heading of the camera leads to errors in determining the angle to an object while the determination of the distance to an object based on its size in the image is not affected. Since the field of view of the perspective camera leads to a bigger size of the object in the image, the determination of the distance based on this size is more accurate when using the image of this camera instead of the image of the omnidirectional camera with a bigger field of view.

Using both camera systems it is possible to calibrate the pan and tilt of the perspective camera while the ball is near the robot, not moving and lying on the ground. In this case, the angle to the object derived from the image of the omnidirectional camera can be used for calibrating the heading of the perspective camera and thereby for improving the accuracy of the localization based on stereo vision and triangulation.

C. Estimating direction, distance, and height

Combining the advantages of both camera systems in determining different parts of the coordinates of an object we suggest a new approach for stereo vision based on an omnidirectional and a perspective camera as depicted in Figure 5 where the point of origin is the center of the robot and the center of the mirror. The image of the perspective camera is used to determine the distance d to the object by its size in the image. The angle ϕ to the object is determined by its position in the image of the omnidirectional camera.

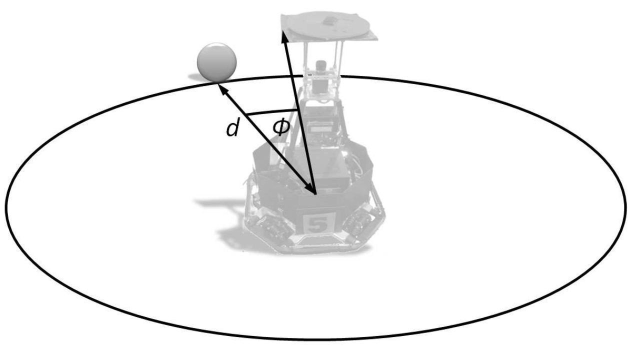


Fig. 5: Schema of the position estimation based on angle and distance

This combination already provides a two-dimensional object localization, which is more reliable than a localization based on only one of the camera systems. In addition the height for the three-dimensional object localization can be obtained accordingly to the schema in Figure 2 by combining the distance derived from the image of the perspective camera and used as distance d_r with the path of the straight line defined from the omnidirectional image by the position of the mirror and the direction towards the object.

IV. Evaluation

For the evaluation of the described methods for stereo vision, we created two different test scenarios. The first test is the localization of a static ball on different positions with distance to the robot ranging from one meter to three meters. In this test, 14 measurements have been taken, six of them have been made while the ball was laying on the ground, which is the assumption of the localization based on only the omnidirectional camera. Mean absolute and mean relative errors of the localization of the ball obtained

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 7-12

from all methods are listed in Table I and Table II. The relative error is defined in relation to the real value, which was measured manually. For each localization method, the x and z coordinates describe the position of the ball relative to the robots position. Column h lists the estimated height of the ball above ground. The distance of the ball and the angle to the ball in relation to the orientation of the robot are given in column d and ϕ. Both camera systems have been calibrated independently before the localization tests. The last column of the table with relative errors is the average of all errors of a single method. This average can be used as a quality ratio to compare the accuracy of all methods where smaller values refer to higher accuracies of the localization method.

coordinate	x [m]	z [m]	h [m]	d [m]	ϕ [°]
perspective	0.36	0.16	0.15	0.07	-
omnidirectional	0.26	0.69	-	0.75	0.02
triangulation	0.18	0.64	0.10	0.68	0.26
combination	0.012	0.054	0.035	0.053	0.30

TABLE I: Absolute errors of all methods for the localization of the ball

coordinate	x %	z %	h %	d %	ϕ %	avg %
perspective	64.8	6.0	99.8	2.4	-	43.3
omnidirectional	45.7	27.4	-	27.3	0,8	25.3
triangulation	32.7	23.9	64.7	24.0	1,6	29.4
combination	1.3	1.9	16.5	0.2	1,9	4.4

TABLE II: Relative errors of all methods for the localization of the ball

The upper two rows describe the localization methods using only one camera system. As described the omnidirectional system itself does not allow a reliable estimation of the height of the ball. Therefore, our algorithm assumes that the ball is lying on the ground. The mean relative errors conform to our assumptions of different accuracies of the camera systems. The best measurements from the perspective camera are concerning the distance of the ball with a relative error of only 2.4% whereas the relative error of the height is almost 100%. This error arises from small inaccuracies in the estimation of the distance to the ball, which affects the interpretation of the height of the ball in the image. The different relative errors for the x and z coordinates show that the perspective cameras pan angle obviously was shifted which results in big errors of the localization of objects in long distances. The omnidirectional camera system computes an angle to the object which has to be localized and the relative error of this angle of only 0.8% shows that this value is the most accurate one derived from the omnidirectional camera system. But, together with the poor resolution for calculating the distance and the assumption that the ball is always lying on the ground, all other coordinates have relative errors above 27%.

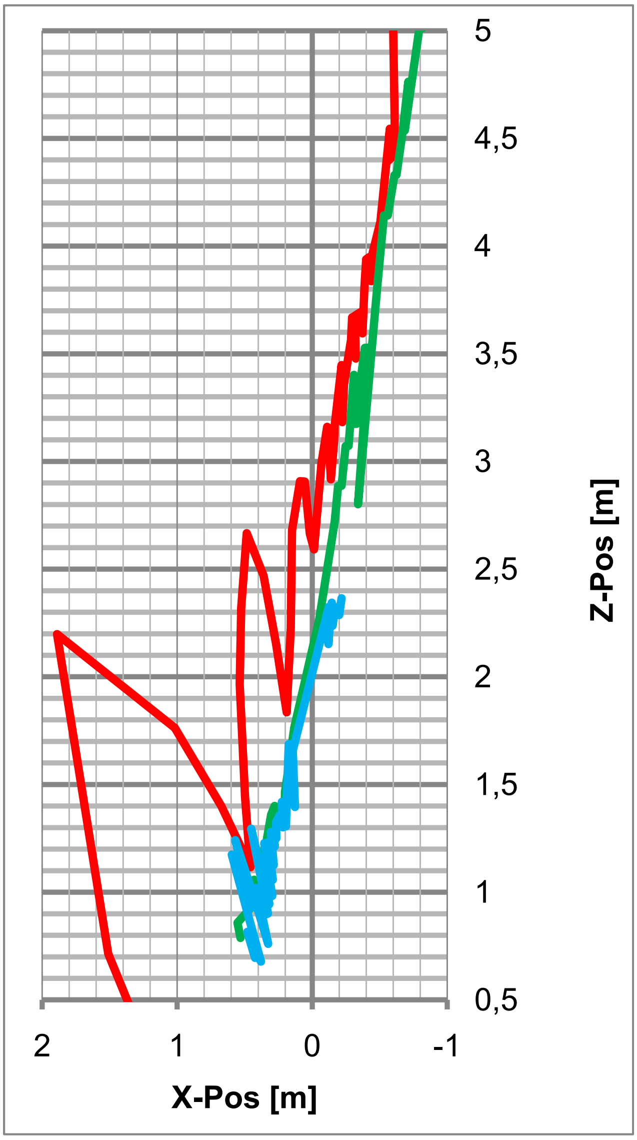
The lower two rows of the tables show the errors of the three-dimensional localization methods. The first one is the stereo vision by triangulation and intersection of the calculated straight lines from each image. The second one is the method combining the most reliable single coordinates from each camera system. The errors of the x and z value of the triangulation method already show an improvement of accuracy compared to the localization only based on the omnidirectional camera since the false assumption of the ball on the ground is not presumed by the stereo vision method. But relative errors greater than 23% of all coordinates besides the angle to the object evidence a poor accuracy of this localization. Especially the method combining only the most reliable information derived from each camera system leads to relative errors of less than 2% of all values which describe the 2D position of the ball and except for the estimated height of the ball. The absolute and relative errors of this method are smaller then the errors of the corresponding values of all other methods. Especially the relative error of only 0.2% of the d value describing the distance to the ball is advantageous in robocup where the different roles of the team are applied to the robots depending on their distance to the ball.

The tests compare all measurements of the position of a static ball where the ground truth of the position is known. But in a RoboCup game the ball is moving and the accuracy of the localization of the ball, which is combining different camera systems, is reduced when it is not possible to synchronize the cameras. To compare both stereo vision methods we did not interpolate the measurements to cancel offsets by usage of the timestamps of the measurements. For testing the algorithms, we kicked a ball that lifts of and bounces on the ground several times. The ground truth of the position of the ball during the flight is unknown. Since the accuracies of the methods are compared already by quality ratios measuring the static ball we compare the stereo vision methods for the flying and bouncing ball by their three-dimensional trajectories. The trajectories of the estimated three-dimensional positions of the ball are shown in Figure 6 and Figure 7. The red trajectory in Figure 6 a) derived from the omnidirectional camera shows the effect of the assumption that the ball is lying on the ground. Its true height increases the distance to the robot which can be seen in the strong periodic deviations of the positions to the left compared to the other trajectories. The blue trajectory in Figure 6 b), which is derived from the stereo vision based on triangulation, shows that the error of the estimated height increases for higher distances of the ball (on the right side of the trajectory). This leads to the corresponding high relative error listed in Table II of 64.7%. In Figure 6 c) the blue trajectory derived from stereo vision makes the bouncing of the ball identifiable. The fact that we are not able to synchronise our cameras on the robot leads to the alternating deviations of the trajectory from a straight line which would match the real behavior of the flying ball. The alternating deviations are visible in Figure 6 a) and b). These alternating deviations are reduced when using the stereo vision based on combining the most reliable coordinates from the different camera systems. The blue trajectories in Figure 7 show smaller alternating deviations in comparison to the blue trajectories in Figure 6.

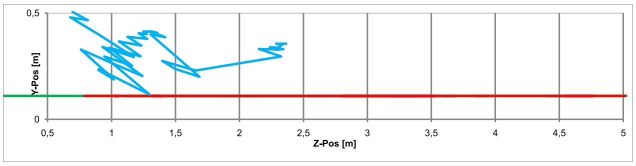
V. Conclusion

In this paper, we presented two different methods for stereo vision which combine an omnidirectional camera system and a perspective camera. The methods are used for three-dimensional object localization. The first method is based on epipolar geometry by intersecting straight lines which are defined by the positions of the cameras and the

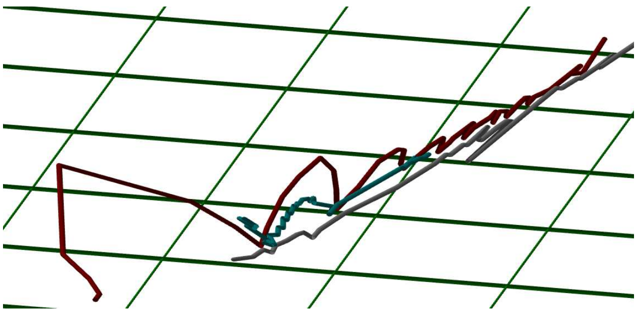
Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 7-12



a) x and z coordinates (position on the gound)

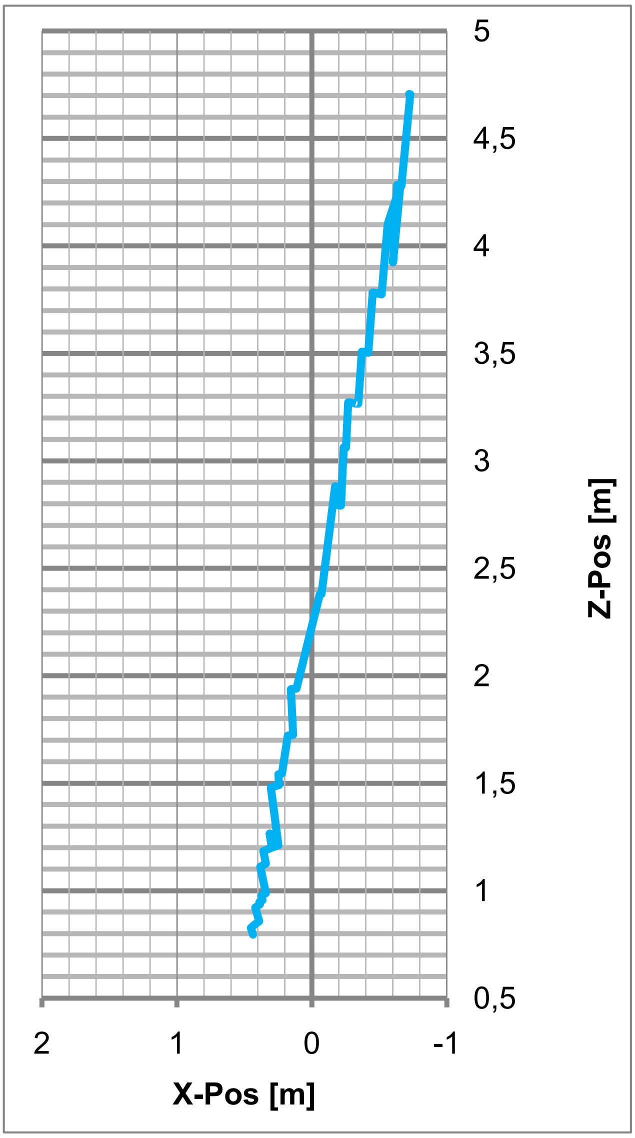


b) y and z coordinates (height of the object)

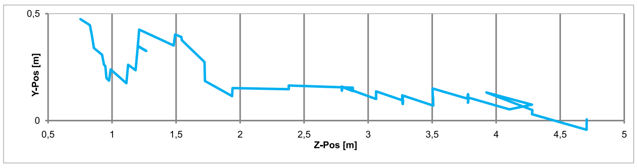


c) perspective plot showing the trajectory of the ball

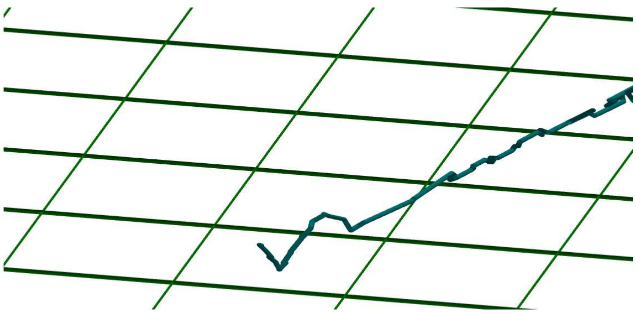
Fig. 6: Trajectory of the ball obtained by the stereo vision method based on triangulation compared to trajectories estimated using only one camera system (red: omnidirectional camera, green: perspective camera, blue: stereo vision)



a) x and z coordinates (position on the gound)



b) y and z coordinates (height of the object)



c) perspective plot showing the trajectory of the ball

Fig. 7: Trajectory of the ball obtained by the stereo vision method based on combination of the most reliable coordinates from different camera systems.

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 7-12

directions towards the corresponding reference points. The second method for stereo vision is based on the combination of the most reliable values from both camera systems. Experiments identified the most reliable values from the localizations based on each single camera system. The most accurate distance to an object can be measured by the perspective camera. The direction towards an object can be derived reliably by the omnidirectional camera. The orientation of the cameras can be decalibrated, when the robot is hit by a kicked ball. Therefore we combine the information from both cameras which is influenced less by the decalibration. An experiment with a static ball and known ground truth positions was carried out to compare all the methods for localization using single cameras or stereo vision by means of a quality ratio based on average deviations. In a second experiment with a flying ball, the methods are compared based on plotted trajectories of the estimated positions of the ball. The experiments showed that the stereo vision method that is combining only the most accurate values from each camera system is the most reliable. All coordinates obtained from this method except for the height of the ball showed a mean relative deviation of less than 2% from the true values in experiments with distances ranging from one meter to three meters. The estimation of the height in these experiments showed a mean deviation of 16.5%. The introduced stereo vision method used on our soccer robot not only allows a three-dimensional localization of the ball, but also improves the two-dimensional position estimation when the ball appears in the field of view of the additional perspective camera.

Acknowledgment

This work was partially funded by the DFG (Deutsche Forschungsgemeinschaft) as part of the Priority Program ”Scalable Visual Analytics” (SPP 1335) and the Center of Excellence ”Spatial World Models for Mobile Context-Aware Applications” (SFB 627).

References

[1] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara, “Robocup: A challenge problem for ai and robotics,” *Lecture Notes in Computer Science*, vol. 1395, pp. 1–19, 1998.

[2] O. Zweigle, U.-P. Käppeler, H. Rajaie, K. Häußermann, A. Tamke, A. Koch, B. Eckstein, F. Aichele, and P. Levi, “1. RFC Stuttgart Team Description 2009,” in *RoboCup 2009 International Symposium*, June 2009, pp. 1–8.

[3] A. Koch, A. Berthelot, B. Eckstein, O. Zweigle, K. Häussermann, U.-P. Käppeler, A. Tamke, H. Rajaie, and P. Levi, “Advanced data logging in RoboCup,” in *Proceedings of the AMS 2009, Autonome Mobile Systeme, Karlsruhe, Germany, December 3-4, 2009*, ser. Informatik Aktuell, Dillmann, Beyerer, Stiller, and Zöllner, Eds. Karlsruhe, Germany: Springer Verlag, December 2009.

[4] T. Buchheim, G. Kindermann, R. Lafrenz, H. Rajaie, M. Schanz, F. Schreiber, O. Zweigle, and P. Levi, “Team Description Paper 2004 CoPS Stuttgart,” in *RoboCup-Symposium 2004*, 2004.

[5] O. Zweigle, U.-P. Käppeler, H. Rajaie, K. Häußermann, R. Lafrenz, A. Tamke, F. Schreiber, M. Höferlin, M. Schanz, and P. Levi, “CoPS Stuttgart Team Description 2008,” in *Proceedings of the RoboCup International Symposium; Suzhou, China, July 14-20, 2008*, 2008.

[6] M. Taiana, J. Santos, J. Gaspar, J. N. abd A. Bernardino, and P. Lima, “Color 3d model-based tracking with arbitrary projection models,” in *Workshop on Omnidirectional Robot Vision (SIMPAR)*, 2008.

[7] M. Jamzad, A. Hadjkhodabakhshi, and V. Mirrokni, “Object Detection and Localization Using Omnidirectional Vision in the RoboCup Environment,” *Scientia Iranica*, vol. 14, no. 6, pp. 599–611, 2007.

[8] S. Olufs, F. Adolf, R. Hartanto, and P. Plöger, “Towards probabilistic shape vision in robocup: A practical approach,” *RoboCup Symposium*, pp. 171–182, 2006.

[9] H. Badino, “Comparitive Research of Stereo Vision Methods,” Diplomarbeit, DaimlerChrysler AG, 2003.

[10] H. Badino, U. Franke, C. Rabe, and S. Gehrig, “Stereo-vision based detection of moving objects under strong camera motion,” in *International Conference on Computer Vision Theory and Applications*, 2008, pp. 253–260.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*. Cambridge University Press, 2003.

[12] S. K. Gehrig, J. Klappstein, and U. Franke, “Active Stereo for Intersection Assistance,” in *Vision, modeling, and visualization 2004: proceedings, November 16-18, 2004, Standford, USA*. IOS Press, 2004, p. 29.

[13] Z. Zhu, “Omnidirectional stereo vision,” *Proc. of ICAR01 IEEE Conference*, pp. 22–25, 2001.

[14] S. Cagnoni, M. Mordonini, L. Mussi, and G. Adorni, “Spherical rectification based stereo vision applied to hybrid dual camera vision systems,” in *Workshop on Omnidirectional Robot Vision (SIMPAR)*, 2008.

[15] A. Burla, “3D Ball Detection in RoboCup,” Master Thesis, Universität Stuttgart, Fakultät Informatik, Elektrotechnik und Informationstechnik, Germany, 2007.

[16] O. Zweigle, U.-P. Käppeler, T. Rühr, K. Häußermann, R. Lafrenz, F. Schreiber, A. Tamke, H. Rajaie, A. Burla, M. Schanz, and P. Levi, “CoPS Stuttgart Team Description 2007,” in *Proceedings of the RoboCup International Symposium; Atlanta, Georgia, July 9-10, 2007*, Atlanta, 2007.