Ground Truth Acquisition of Humanoid Soccer Robot Behaviour

Andrea Pennisi, Domenico D. Bloisi, Luca Iocchi, and Daniele Nardi

Dept. of Computer, Control, and Management Engineering,
Sapienza University of Rome, via Ariosto 25, 00185, Rome, Italy
{pennisi,bloisi,iocchi,nardi}@dis.uniroma1.it

Abstract. In this paper an open source software for monitoring humanoid soccer robot behaviours is presented. The software is part of an easy to set up system, conceived for registering ground truth data that can be used for evaluating and testing methods such as robot coordination and localization. The hardware architecture of the system is designed for using multiple low-cost visual sensors (four Kinects). The software includes a foreground computation module and a detection unit for both players and ball. A graphical user interface has been developed in order to facilitate the creation of a shared multi-camera plan view, in which the observations of players and ball are re-projected to obtain global positions. The effectiveness of the implemented system has been proven using a laser sensor to measure the exact position of the objects of interest in the field.

1 Introduction

A ground truth system is a necessary tool to evaluate and improve algorithms dealing with a series of challenging tasks in the RoboCup competitions. In particular, visual data collected from a global view of the environment can be used for validating innovative methods concerning the following aspects:

- Coordination;
- Localization;
- Game strategies;
- Feedback for adaptive methods;
- Quantitative measurements;
- Debugging.

Indeed, since humanoids robots have only a local and limited point of view of the environment in which they operate, tasks like coordination and localization are still open problems [6] and there is the need of creating and adopting publicly available benchmarks in order to validate and quantitatively evaluate different solutions.

In this paper, an open source software for monitoring humanoid soccer robot behaviours in the Standard Platform League (SPL) is described. The software

S. Behnke et al. (Eds.): RoboCup 2013, LNAI 8371, pp. 560–567, 2014.

is part of an easy and fast to set up system, designed to give a global point of view of the observed scene that is useful in order to recognize and evaluate the behaviour of the robots.

The system aims at providing:

1. A simple and fast calibration set up;
2. The foreground mask for each captured frame;
3. 3D information about each player in the field as well as the position of the ball in each frame;
4. A multi-camera data fusion scheme;
5. The set of the tracks representing the objects of interest.

The proposed approach is robust to the presence of people on the field (e.g., the referees), illumination changes, shadows, and modifications in the background geometry (e.g., the audience around the field).

The remainder of the paper is organized as follows. Related work is analysed in Section 2, while Section 3 describes both the hardware architecture and the software modules of the system. The software tools as well as the evaluation of the results coming from a first data set are described in Section 4. Conclusions are drawn in Section 5.

2 Related Work

Zickler *et al.* in [12] propose a shared vision system for RoboCup competition and use it to estimate a ground truth for NAO robots. Two RGB cameras are used to span the field and a set of makers on the top of the robots are used to identify them. Since markers are not allowed in the competition games, such an approach is limited to collecting ground truth information for testing purposes.

De Morais *et al.* in [4] describe a method for estimating the position of indoor soccer players by using multiple cameras in real indoor soccer games. The proposed set up consists of four cameras positioned around the soccer court. A simple object detector runs on each camera and projects the detections onto a world plan representing the soccer court. The main drawback of such a method is the need of a long calibration time for each camera.

Niemüller *et al.* in [9] present a data collection system for NAO robots. Reflective markers attached to NAO's body are used in combination to a 6D professional vision-based body motion tracking system. The position and the orientation of the observed robots are tracked by using 15 pulsed infra-red cameras placed on the field. Although the acquired data have a high accuracy, the system is difficult to set up and calibrate. Moreover, markers are not allowed in the competition games and infra-red cameras are quite expensive sensors.

A low-cost and portable system using a single Kinect has been proposed by Khandelwal and Stone in [5]. The system is designed to find the location of robots and ball in the SPL environment. No special identifiers on the robots are required, although 22 known landmarks must be placed on the field to carry out the calibration. The correspondences between the point cloud acquired from the

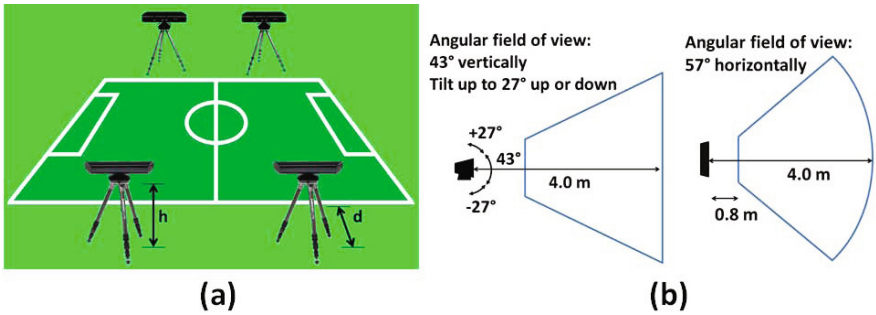

Fig. 1. Hardware architecture of the proposed system. a) Each Kinect is mounted on a tripod at an height h from the ground (in our setup $h = 1.05\ m$) and at a distance d from the long side line (in our setup $d = 0.8\ m$). b) Vertical and Horizontal field of view for the Kinect.

Kinect and the position of the landmarks are obtained through a graphical user interface (GUI). The user clicks a pixel in the image and the system performs the association. Only the points above 0.30 m of height are considered and a Euclidean method for cloud clusterization is used. However, the system cannot provide the orientation of the robots and it allows to monitor only a part of the field.

Compared to related work, the main improvements of our approach are: 1) A network of low-cost sensors to cover the entire field is used; 2) No markers are needed on the field or on the robots; 3) An accurate foreground mask is extracted for each frame; 4) The orientation of each robot is estimated from its 3D point cloud. The details of our method are described in the remainder of the paper.

3 System Description

In this section the hardware architecture of the proposed system is described, followed by the details of each software module involved in the ground truth registration.

3.1 Hardware Architecture

The hardware architecture of the system is shown in Fig. 1a. Four Kinects [7] are placed along the long side lines of the field, two for each side. The choice of this setting derives from the physical limitations of the sensors [8], that are reported in Fig. 1b in terms of the vertical and the horizontal angular field of view. Given the above limitations and the dimensions of the field to be monitored ($9 \times 6\ \ m^2$ for the SPL [10]), we find that a reasonable set up is the following. Each Kinect is mounted at about 1 m height, with the tilt angle set to the maximum achievable down value (-27°), at a distance of 0.8 m from the side line. This is an easy

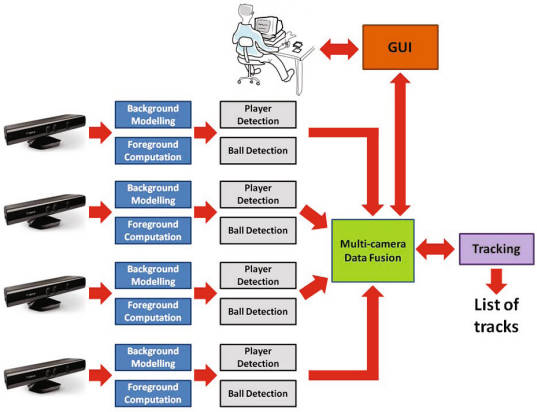

Fig. 2. Software modules

and reproducible set up, specially with respect to other approaches that require to place the sensor(s) on the top of the field and/or at a considerable height (e.g., [9]). Furthermore, the given distances are only suggested and can vary from sensor to sensor, because the system can receive those measures as input (see Section 4). In order to acquire the images, we connected the two Kinects on the same side of the field to a commercial notebook (CPU Intel Core i7, 2.4 GHz, 8 GB RAM).

3.2 Software Modules

The software architecture of the system is shown in Fig. 2. For each captured frame, the foreground mask is extracted and the positions of the ball and players are estimated (Fig. 3). A data fusion module re-projects all the detected positions on a shared plan view of the field (Fig. 4), while a tracking module generates the list of tracks in output.

To compute an accurate foreground mask both RGB and depth information are used [3]. An RGB image (Fig. 3a) and a 16 bit depth map are stored for each captured frame; they are labelled with a timestamp and a sequence number (for synchronization purposes). A statistical approach, called IMBS [1], is used to create the background model that is updated every 30 seconds for dealing with illumination changes.

The positions of the players are obtained by using the foreground mask (Fig. 3b) in combination with the depth map (Fig. 3c). A set of blobs (*observations*) is extracted by means of a height filter that considers as valid observations only the blobs under 0.70 m (in this way, humans on the field are filtered out). Moreover, the pixels that belong to shadows are also suppressed since they are recognized as ground points on the basis of their height values.

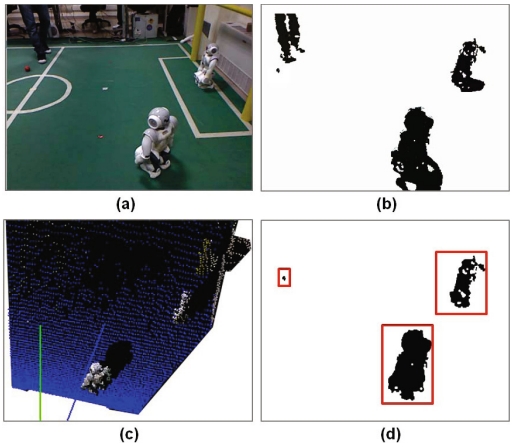

Fig. 3. Player and ball position estimation. a) Original frame. b) Foreground extraction. c) 3D scene reconstruction. d) Player and ball detection.

An example of player detection is shown in Fig. 3d. In order to estimate the robot position to be re-projected on the plan view, the lower point of the blob is considered and it is transformed in world coordinates.

A surface normal approach [11] is used to detect the orientation of the robot. Given the set of 3D points of each blob, the problem of determining the normal to a point on the surface is approximated by estimating the normal of a plane tangent to the surface, thus resulting in a least-square plane fitting estimation problem. Therefore, surface normal estimation is reduced to an analysis of the eigenvectors and eigenvalues of a covariance matrix created from the nearest neighbours of the query point, where the sign of the normal is assigned by knowing the view point of the scene.

A color based approach is used for detecting the ball. The RGB image is converted into the HSV color space and the ball is detected searching for a red coloured blob. Then, the same transformation applied to the players' positions is used.

Multi-camera data fusion is carried out thanks to four homography matrices that are used to generate the plan view of the field (Fig. 4). The homography matrices are calculated through a GUI (that is described in the next section) and are used to generate a warp image for each Kinect. Then, all the four warp images are stitched together to reconstruct the whole field using the Stitcher class provided by OpenCV.

For tracking both the players and the ball a multi-hypothesis approach based on a set of Kalman Filters has been chosen [2]. Data association is used to determine relationships between observations and tracks, but multiple hypotheses are maintained when observations may be associated to more than one track.

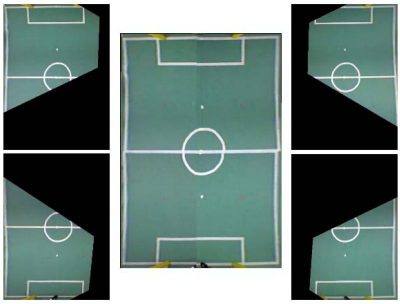

Fig. 4. Plan view. Multiple camera data fusion is carried out merging the four warp images generated using the GUI.

4 System Outcomes

We created the web site `http://labrococo.dis.uniroma1.it/gnao` containing the C++ source code for all the software modules and for the GUI, whose features are described in the following. In addition, we made available for downloading also a first data set used to obtain the measurements discussed in this section.

4.1 Graphical User Interface (GUI)

The GUI supports the calibration of the four Kinects through a two step procedure (Fig. 5). In the first step, the GUI allows to load up to four different Kinect views (Fig. 5a).

Pressing the button "Execute" under each Kinect view it is possible to start the homography based calibration (Fig. 5b). A homography is a relation between points that belong to two different spaces S and S', in which each point of S coincides to only one point of S'. Since the problem of 2D homography consists

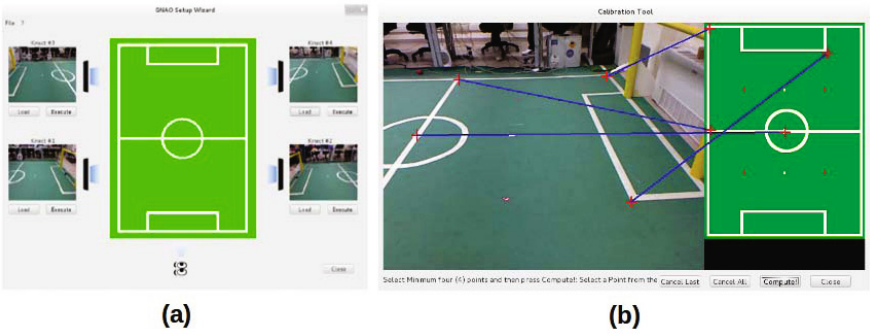

Fig. 5. Graphical User Interface. a) Camera assignment. b) Point selection.

in finding a transformation that converts the points from an image plane to the points of another plane, the GUI allows the user to select a point in the image coming from a Kinect and its corresponding point in a virtual representation of the field (right side of Fig. 5b). The user is required to select at least four correspondences in order to compute the homography, although selecting more points can produce a more accurate transformation. When the user ends the selection of the points, a text file that contains the homography related to the selected Kinect is saved by pressing the button "Compute". The point selection procedure must be repeated for each Kinect and it is also possible to save the tilt angle and the height of each sensor.

4.2 Results on a First Data Set

We registered a first data set of ground truth data capturing three NAOs by using four Kinects on a 4×3 m^2 training field. We made three runs (each one about 20 minutes long) obtaining a computational speed of about 15 frames per second acquiring 640×480 images. In Fig. 6 two frames from the collected data set show the output of the robot orientation extraction routine.

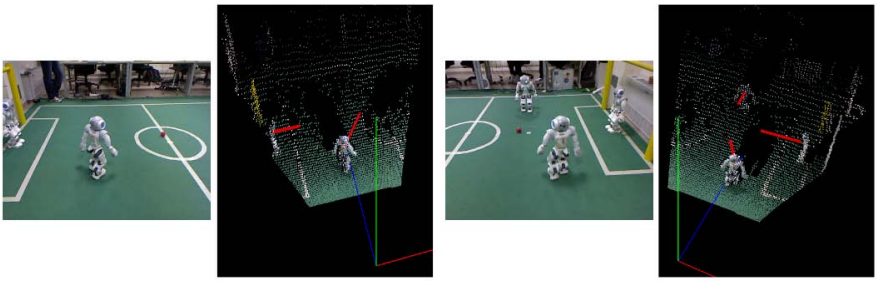

Fig. 6. Extraction of the robot orientation

Table 1. Quantitative evaluation for the player estimation method

Experiment	Ground Truth	Estimated Position	Error
	3.398 m	3.489 m	0.09
1	4.078 m	4.1268 m	0.049
	2.932 m	2.681 m	0.25
	3.349 m	3.357 m	0.008
2	4.156 m	4.125 m	0.03
	2.761 m	2.578 m	0.182
	3.566 m	3.344 m	0.221
3	4.002 m	4.08 m	0.078
	2.794 m	2.578 m	0.215
	2.784 m	2.658 m	0.125
4	2.76 m	2.562 m	0.197
	2.914 m	2.635 m	0.278
	3.627 m	3.781 m	0.154
5	2.85 m	3.008 m	0.158
	3.133 m	3.609 m	0.476
	2.56 m	2.54 m	0.018
6	3.135 m	3.015 m	0.119
	4.101 m	3.952 m	0.148

To quantitatively evaluate the accuracy of the proposed set up, we measured the exact position of the players in the field using a laser range finder. Then, we compared the results obtained by using the laser with the measurements provided by the system. The results, obtained monitoring three NAOs placed in six different relative positions, are reported in Tab. 1 (units in meters).

5 Conclusions

In this paper, an open source software for creating a ground truth acquisition system for evaluating humanoid soccer robot behaviours is described. The software is divided in modules and it is designed to work in a low-cost multi-camera architecture. We used four Kinect sensors that can be calibrated using a specifically designed GUI (also made available). Both the hardware architecture and all the software modules needed to achieve the goal have been discussed in detail. The main features of the proposed approach are as follows: 1) Fast and easy multi-camera set up; 2) No markers needed; 3) Real-time foreground mask computation; 4) Orientation of each robot estimated from its 3D point cloud. The proposed approach can be easily adapted to be suitable for other RoboCup leagues.

References

1. Bloisi, D., Iocchi, L.: Independent Multimodal Background Subtraction. In: CompIMAGE, pp. 39–44 (2012)
2. Bloisi, D.D., Iocchi, L.: ARGOS - A Video Surveillance System for Boat Traffic Monitoring in Venice. IJPRAI 23(7), 1477–1502 (2009)
3. Bloisi, D.D., Iocchi, L., Monekosso, D.N., Remagnino, P.: A Novel Segmentation Method for Crowded Scenes. In: VISAPP, pp. 484–489 (2009)
4. De Morais, F.E., Goldenstein, S., Roch, A.: Automatic Localization of Indoor Soccer Players from Multiple Cameras. In: VISAPP, 205–212 (2012)
5. Khandelwal, P., Stone, P.: A Low Cost Ground Truth Detection System Using the Kinect. In: Röfer, T., Mayer, N.M., Savage, J., Saranlı, U. (eds.) RoboCup 2011. LNCS, vol. 7416, pp. 515–527. Springer, Heidelberg (2012)
6. Li, X., Lu, H., Xiong, D., Zhang, H., Zheng, Z.: A Survey on Visual Perception for RoboCup MSL Soccer Robots. Int. J. Adv. Robotic Sy. 10, 1–10 (2013)
7. Microsoft Corp. Redmond WA: Kinect for Xbox 360
8. MSDN Library: Skeletal Tracking, `http://msdn.microsoft.com/en-us/library/hh973074.aspx#ID4ENB`
9. Niemüller, T., Ferrein, A., Eckel, G., Pirro, D., Podbregar, P., Kellner, T., Rath, C., Steinbauer, G.: Providing Ground-truth Data for the Nao Robot Platform. In: Ruiz-del-Solar, J., Chown, E., Plöger, P.G. (eds.) RoboCup 2010. LNCS (LNAI), vol. 6556, pp. 133–144. Springer, Heidelberg (2010)
10. RoboCup Technical Committee: RoboCup Standard Platform League (Nao) Rule Book, `http://www.tzi.de/spl/pub/Website/Downloads/Rules2013.pdf`
11. Rusu, R.B.: Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. Articial Intelligence (KI - Kuenstliche Intelligenz) (2010)
12. Zickler, S., Laue, T., Birbach, O., Wongphati, M., Veloso, M.: SSL-Vision: The Shared Vision System for the RoboCup Small Size League. In: Baltes, J., Lagoudakis, M.G., Naruse, T., Ghidary, S.S. (eds.) RoboCup 2009. LNCS (LNAI), vol. 5949, pp. 425–436. Springer, Heidelberg (2010)