Model-Based Relative Localization for Cooperative Robots Using Stereo Vision

A. Milella[1], F. Pont[2], R. Siegwart[3]
[1]Department of Mechanical Engineering, Politecnico of Bari, viale Japigia 182, 70126 Bari, Italy
[2,3] Autonomous Systems Lab, Swiss Federal Institute of Technology Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Abstract

In the last years, Multi-Robot Systems (MRS) have been receiving great attention, as they can be effectively employed in several fields. Generally, for a collaborative behaviour to be successful, a precise localization strategy is required. A number of collective positioning schemes are available in literature, which mainly differ depending on the sensors and on the cooperation strategies adopted. In this work, we propose a model-based relative localization method using stereo vision, which enables a complex agent, equipped with a stereo head, to simultaneously detect and localize several small robots, navigating in a coordinated manner for a common task. The paper describes the method in detail and presents experimental tests performed on a real multi-agent system, proving the method to be accurate and effective for multi-robot localization, and environment exploration and mapping.

1 Introduction

Multi-Robot Systems (MRS) constitute, nowadays, an important field of investigation within Robotics and Artificial Intelligence, as they can be effectively employed in several domains. Exploration of hostile environments, terrain mapping, as well as space, military and rescue operations are only a few examples of real world applications [5].

Generally, for a MRS to successfully accomplish an assigned task, a reliable localization method of each member is needed ([2], [3]).

When dealing with team-oriented behaviours, two positioning problems can be distinguished: absolute localization, in which each robot determines its pose with respect to some external global reference frame, and relative localization, in which each robot estimates the pose of every other robot in the team relative to itself [8].

Absolute localization, using either GPS or landmark methods, is a well-studied topic ([2], [4], [11]) and will be not addressed here.

This work focuses, instead, on relative localization, which is more important for many cooperative behaviours [14]. Specifically, a method for getting relative observations is proposed, employing a stereo vision device and a visual model of the robots.

The method fits very well in the context of a heterogeneous multi-robot system, where a team of several small vehicles navigates in a coordinated manner to achieve a common goal, supervised by a complex agent, equipped with a stereo head. Using the proposed localization strategy, the supervisor is able to simultaneously detect and localize all the small robots with respect to itself. This provides a reliable surveying system, without error growing phenomena, which can be usefully exploited for cooperative behaviours, like flocking and formation maintenance [5], and in all situations in which each member of the team of simple vehicles has to be precisely localized.

Real world applications include target localization in dangerous environments where the simple and cheap robots can be sacrificed while the advanced and expensive agent must remain in safe areas.

The problem of measuring the relative configuration between two robots for cooperation purposes has been addressed by some authors. Most of the proposed solutions employ range sensors ([7], [10], [12], [15]), video sensors ([12], [18], [19], [21]), or their combination [8]. The observed robot is usually equipped with some target pattern, such as retro-reflective fiducials or visually distinguishable landmarks, so as to be uniquely identified and localized by the observer.

For instance, in [10] each robot is provided with a laser range finder that automatically searches and traces corner cubes located on the top of the other robots, thus determining relative distances and azimuth angles. In [18], a monocular camera is mounted on the observer, while the observed brings a helix target pattern that

allows estimating relative position and orientation using visual line detection methods. An omnidirectional camera is proposed in [21] for precise bearing angle measurement. Coloured lights positioned on the robots are tracked employing colour segmentation and region merging algorithms. Finally, a system combining a single camera and a laser range finder can be found in [8]. Fiducials that are both retro-reflective and colour-coded are opportunely placed on the robots, in order to determine range, bearing, orientation, and identity of each of them.

The use of target patterns simplifies the segmentation of the scene and reduces the computational time. However, it often entails elaborated setup. Also, in the case of vision-based methods, standard colour segmentation algorithms do not provide sufficient robustness. Systems combining laser range finders and video sensors are more reliable, but they result in expensive and complex devices. Hence, more efficient and flexible relative localization methods need to be researched, yet.

In this work, we describe a relative localization strategy, which uses a combination of feature-based and appearance-based approaches for robust robot tracking, with 3D stereo information for robot pose estimation.

While most of the existing schemes employ landmarks opportunely placed on the robots or some knowledge of the environment, our method does not require any particular setup of the robots, nor previous information about the surroundings. The only assumption is that the supervisor is always able to keep the visual contact with the small vehicles, based on some coordinated navigation strategy. The proposed scheme involves three main steps: 1) Automatic visual learning, 2) Multi-robot recognition, and 3) Multi-robot localization.

The first phase aims at constructing a model of the small robots. Generally, two principal approaches to visual object modelling can be distinguished ([13], [22 Trucco and Verri]): feature-based and appearance-based. Feature-based modelling uses features, such as geometrical primitives, object contours or regions of interest. The advantage of these methods is that they rely on compact object descriptors, they are relatively robust against occlusions and they provide some invariance to illumination and pose variations. Disadvantages include the fact that they require robust feature extraction. Conversely, appearance-based models represent an object through one or more images. Therefore, they allow the direct comparison of images and models, and complex objects can also be modelled in a quite simple manner. A drawback is that illumination and pose variations alter the images.

Here, an effective combination of both approaches is adopted. The model of the robots is, in fact, automatically generated extracting significant shape features and appearance characteristics from a database of images, which depict a sample vehicle at different orientations and distances from the sensor. The learning process is accomplished off-line, just one time. As it only requires the acquisition of a set of images of the object, this process is flexible and can be easily applied also in the case of heterogenous teams and complex shaped robots.

The model is used for real-time robot recognition. A combination of a minimum distance classifier [6 Gonzalez and Woods] based on shape information and of the PCA recognition method ([16], [17], [22 Trucco and Verri]) is implemented. Image processing is accomplished in HSV colour space, as that allows exploiting colour information more efficiently. Robustness is achieved by processing separately the left and right images of each acquired stereo pair, and using the epipolar constraint to find corresponding objects.

Once all the robots have been recognized, a dense disparity map [9] is computed for each of them, and finally the position of every robot relative to the supervisor is estimated.

Experimental results obtained with a real multi-robot system are presented, proving the effectiveness of the recognition algorithm, as well as the accuracy of the localization module. The potential capability of the method to reconstruct the trajectory of each robot and eventually provide a map of the surroundings is also shown. Finally, a practical application in which the robot team explores an indoor environment is described.

The paper is structured as follows. Section 2 details the various steps of the method. Section 3 reports experimental results. Section 4 shows the multi-robot application. Finally, section 5 contains the conclusions of the presented work.

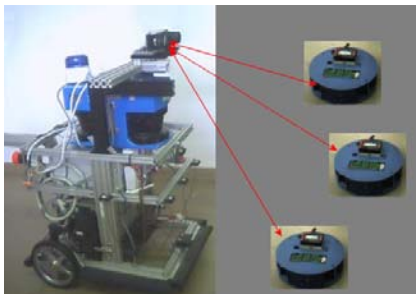


Fig 1. The Biba Robot (left) and the Smartease Robots (right)

2 Description of the method

In this section, a relative localization method for cooperative mobile robots is presented, which enables a complex agent, equipped with a stereo head, to supervise a team of small vehicles, navigating in a coordinated manner for exploration and mapping tasks.

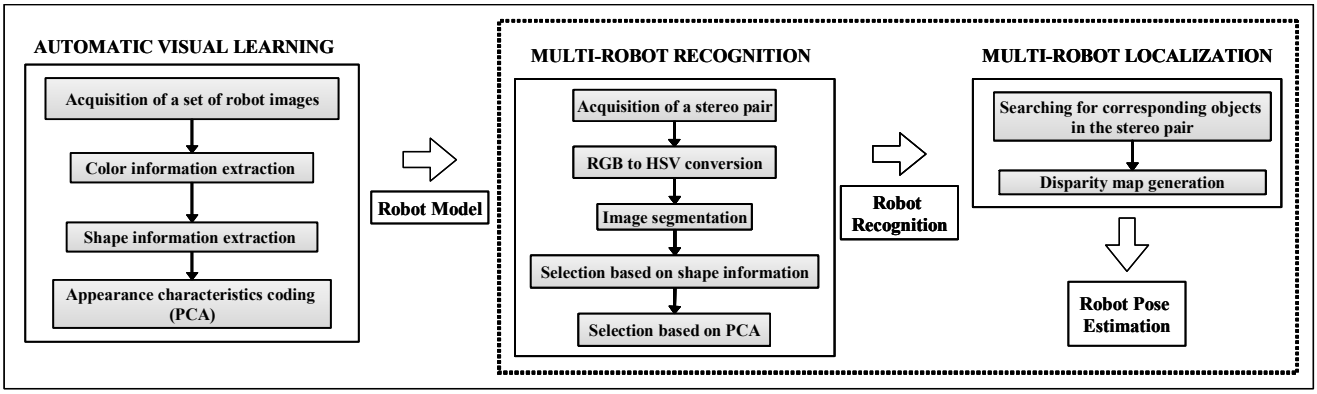


Fig 2. Block diagram of the method

The proposed method consists of three main phases: 1) Automatic visual learning, 2) Multi-robot recognition, and 3) Multi-robot localization. In the first phase, a visual model of the small vehicles is constructed in an automatic way, i.e. with limited human intervention, based on image processing techniques. The second phase aims at recognizing the small robots within the scene, using the visual model. Finally, in the third phase, the pose of each robot relative to the supervisor is estimated, exploiting 3D stereo information.

A detailed description of each step is given in the remainder of this section for a team of multiple robots, namely the Smartease robots, supervised by the so-called Biba robot equipped with a Videre Design stereo head.

The robotic platforms are shown in Fig 1. A flowchart of the whole localization method is instead illustrated in Fig 2.

2.1 Automatic visual learning

The first step of the method is an automatic visual learning process, which allows generating a model of the small robots, using a set of images depicting a sample vehicle at several configurations in the environment. In this phase, a fixed background is used, so that the image portion which contains the robot can be easily extracted by background subtraction. HSV colour coding is employed, for improving image segmentation. Assuming that the left and right cameras of the stereo device have a very similar field of view, only the left (or the right) images can be used.

The learning process is as follows. For each robot image, a histogram is, first of all, traced in the hue and saturation colour planes, to obtain mean values and standard deviations for colour representation.

Then, a vector of shape-descriptors is built for each binary image of the object that we denote as SV_i, for i=1, 2, …, P, where P is the number of images, i.e. of poses of the object. Descriptors have to be chosen depending on the specific form of the object to model. Here, due to the simple geometry of the vehicle only the percentage ratio of the object area to the total image area, the Heywood circularity factor, and the aspect ratio are taken into account [20 Russ].

Successively, images are rearranged to form a set of vectors normalized with respect to scale (see Fig 3) and brightness, and the Principal Component Analysis technique ([16], [17], [22 Trucco and Verri]) is employed to store the principal appearance characteristics of the object at each configuration.

This technique, also known as Karhunen–Loeve transform, uses the eigenvectors of the set of images as orthogonal basis for representing each image in a compressed manner. It will be shown later that the eigenspace compressed representation is a great advantage for object recognition purposes.

The key idea for compression is that, though a great number of eigenvectors are needed to represent images exactly, only a few are sufficient for capturing the gross appearance characteristics of the object. These are the K eigenvectors corresponding to the K largest eigenvalues, and constitute the so-called eigenspace. As K is usually much less than N, where N is the dimension of an image vector, the eigenspace provides for a compressed representation of the original image set.

For colour images, different solutions can be adopted [17]. Here, the colour bands in HSV colour space are used separately. Specifically, two eigenspaces are built, in the hue plane and in the saturation plane respectively, so that two sets of P eigenspace images are in the end available, one for each colour band.

Denoting with E_{hue} and E_{satur} the KxN matrices constituting the hue eigenspace and the saturation eigenspace, each robot image can be represented in a compressed manner by two K-dimensional vectors, h_i and s_i, defined as:

$$h_i = E_{hue}(H_i - A_{hue}) \quad (1)$$

$$s_i = E_{satur}(S_i - A_{satur}) \quad (2)$$

where H_i is the N-dimensional image vector in the hue plane, A_{hue} is the mean vector of all the hue image vectors, S_i is the N-dimensional image vector in the saturation plane, and A_{satur} denotes the mean vector of all the saturation image vectors. Fig 4(a) and 4(b) show the first three eigenvectors obtained in the hue and saturation planes, respectively.

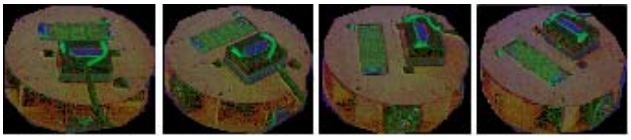


Fig 3. Four HSV model images of the Smartease robot normalized with respect to scale

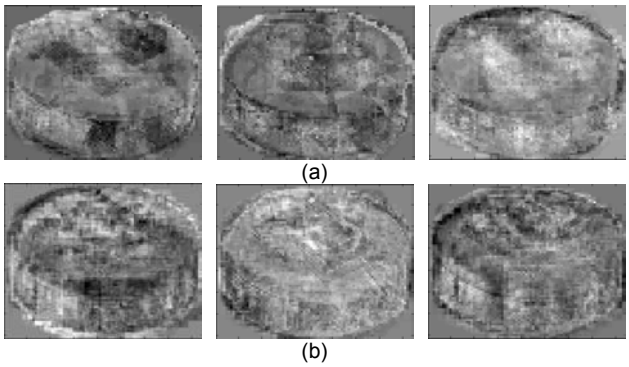


Fig 4. (a) First three eigenspace vectors in the hue plane, and (b) in the saturation plane

2.2 Multi-robot recognition

The information inferred in the learning phase can be effectively employed for real-time robot recognition. The following process is applied independently to the left and right images of each stereo pair.

1. *RGB to HSV colour space conversion* (Fig 5(a)-5(b)), in order to enhance the thresholding operation described below.
2. Independent *thresholding* in the hue and saturation planes, based on colour information acquired in the learning phase.
3. Application of *Boolean and morphological operations* for binary shape reconstruction, followed by a *labelling algorithm*.
4. *Selection of the labelled objects using shape information* (Fig 6(a)): a first selection is accomplished based on the computation of a shape distance factor (SDF), defined as the Euclidean distance between the shape-descriptor vector of the l-th labelled object, SV_l, for l=1, 2, …, L, where L is the number of labelled objects, and each shape-descriptor vector, SV_i, stored in the database. The condition for the object to be selected as a possible candidate is that:

$$\min_i \left(\| SV_l - SV_i \| \right) \le T_{SDF} \qquad (3)$$

where T_{SDF} is a user-defined threshold value.

5. *Selection based on the PCA recognition method* (Fig 6(b)): objects selected according to the feature-based model are further analysed using the PCA recognition method. The key idea is that the Euclidean distance in eigenspace is equivalent to image correlation, with the advantage that the eigenspace vectors have a much smaller dimension than image vectors. Specifically, maximizing correlation corresponds to minimizing distance [22 Trucco and Verri]. Based on this criterion, the image of each candidate object is firstly separated into its colour bands. Let us denote with H_l the hue image vector of the l-th object and with S_l its saturation image vector, both normalized with respect to scale and brightness. These vectors are projected onto the hue and saturation KxN eigenspaces, thus obtaining the K-dimensional vectors h_l and s_l, according to (1) and (2). For an object to be finally considered as an instance of the robot, the following conditions have to be both verified:

$$\min_i \left(\| h_l - h_i \| \right) \le T_{hue} \qquad (4)$$

$$\min_i \left(\| s_l - s_i \| \right) \le T_{satur} \qquad (5)$$

where T_{hue} and T_{satur} are user-defined thresholds.

Fig 5. (a) Left image of a stereo pair in RGB; (b) the same image in HSV

Fig 6. (a) Objects selected using colour and shape information; (b) object recognition using PCA

2.3 Multi-robot localization

After one or more instances of the robot have been identified in the left and right images of the stereo pair, the epipolar constraint is exploited to find corresponding

objects. A disparity map is also computed for each of them, using the SRI Stereo Engine algorithm [9]. It consists of an area correlation-based matching process, followed by a post-filtering operation which uses a combination of a confidence filter and left/right check. That allows obtaining a 3D point cloud, which can be exploited for localizing the object with respect to the supervisor.

In general, the 3D pose of the robot relative to a reference frame attached to the camera, i.e. to the supervisor, can be computed employing a registration method to align the 3D point cloud with a geometric model of the robot [1].

In this case, the problem can be simplified by making some assumptions, which are reasonable for many indoor multi-robot applications: first, it is supposed that all the robots move on a plane and that the orientation of this plane with respect to a reference frame attached to the supervisor is known; then, it is assumed that the small robots have a simple geometric shape, that can be approximated by a circular shape in the plane of motion.

Based on these hypotheses, the projection of the 3D points on the plane of motion is considered, and the position of a small robot relative to the supervisor is estimated as the position of the centre of the minimum enclosing circle that can be computed for these points.

Fig 7 gives a schematic representation of the positioning system. In this figure, (c, X_c, Y_c, Z_c) denotes the camera reference frame, (o, X_r, Y_r, Z_r) indicates the reference frame attached to the small robot, α is the known orientation of the camera relative to the plane of motion of the vehicles, and p represents the position of the small robot relative to the camera that has to be estimated.

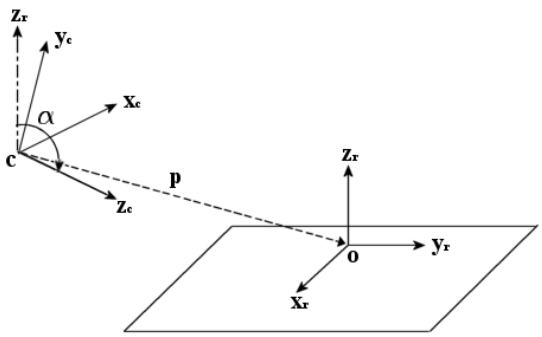
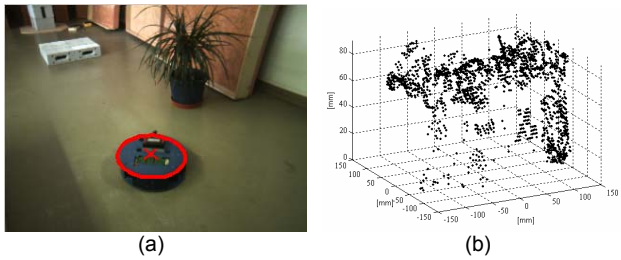


Fig 7. Model of reference frames

In Fig 8(a), the minimum enclosing circle and its centre are represented, reprojected onto the image plane, for a detected robot, showing the method to work properly. Fig 8(b) reports instead the 3D reconstruction of the robot.

Finally, the application of the method for the case of a multi-robot system is illustrated in Fig 9, where three Smartease robots are simultaneously localized, in the left and right images of a stereo pair. A name (R0, R1, R2, …) is automatically assigned to each robot, showing that corresponding objects have been fairly recognized. The centres of the circles computed for each robot are also shown, reprojected onto the left frame.



(a) (b)

Fig 8. (a) Minimum enclosing circle reprojected onto the image; (b) 3-D robot reconstruction

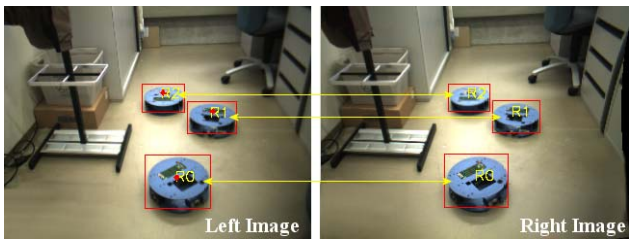


Fig 9. Multi-robot recognition and localization

3 Experimental results

In order to evaluate the effectiveness of the proposed method, several tests were carried out. First of all, a model of the Smartease robots was stored using 30 images, according to the process described in Section 2.1.

Then, a set of experiments was performed to verify the robustness of the recognition module against variations of the robot's pose and of the lighting conditions. Specifically, these tests aimed at evaluating the capability of the method to properly detect the robot while the latter is moving, assuming different configurations with respect to the stereo device, under various lighting conditions.

To this purpose, a sample vehicle was guided to reach several positions within different indoor environments. Various orientations of the camera relative to the plane of motion were also considered. The vehicle was properly recognized in 97% of cases. Failures occurred due to sharp shadows or in presence of large occlusions.

Fig 10(a) shows some images of the vehicle, detected at various distances and orientations with respect to the camera. Fig 10(b) shows that the recognition module continues to work correctly even for lighting reduction as much as 80% (L=0.2) of the optimal value (L=1).

Another set of experiments was realized to test the accuracy of the localization module. The vehicle was

placed at several known locations, spread over a rectangular area of 0.6 x 1 m. At each location i, for i=1, 2, …, n, where n is the number of positions (n=50), the percentage relative error E_i between the actual position p^i_a and the estimated position p^i_e was computed as:

$$E_i = \frac{\left\| p^i_e - p^i_a \right\|}{\left\| p^i_a \right\|} \times 100 \qquad (6)$$

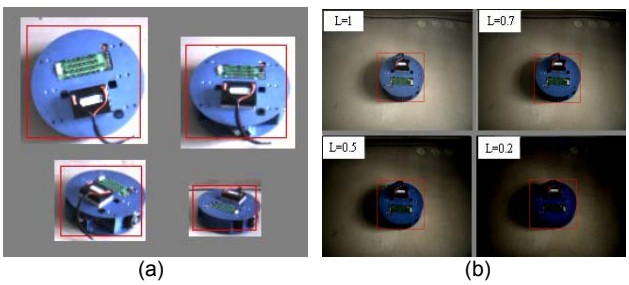
Fig 11 reports a graph of the estimated errors: an average error of 4.4% with a standard deviation of 3.4% was obtained.

Finally, the potential capability of the method to reconstruct the trajectory of each robot and eventually build a map of the environment was investigated, guiding one small robot to follow various paths.

Here, results for a rectangular path are reported. Fig 12 shows, the trajectory of the vehicle as estimated by the vision system, compared with that derived from the encoders mounted on onboard the small vehicle, during one run. A mean percentage relative error between encoder measures and vision-based measures at the j-th run, for j=1, 2, …, m, where m is the number of repeated runs (m=5), was defined as:

$$E_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(\frac{\left\| e_{j,i} - v_{j,i} \right\|}{\left\| e_{j,i} \right\|} \times 100 \right) \qquad (7)$$

where n_j is the number of samples, $e_{j,i}$ is the i-th encoder-derived position, and $v_{j,i}$ is the i-th vision-derived position. Table 1 reports the average errors at each run along with their statistical spread, showing that the average errors were always within 7%.



(a) (b)

Fig 10. (a) Robustness of the recognition algorithm to variations of robot's pose, and (b) of lighting conditions

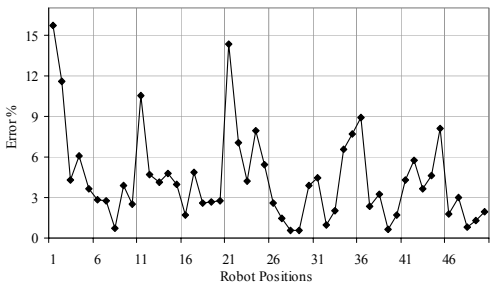


Fig 11. Percentage relative position errors

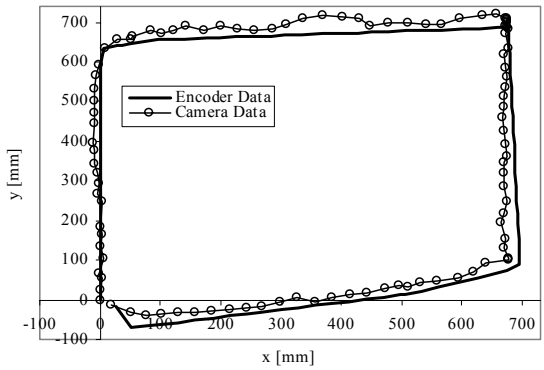


Fig 12. Robot trajectory: comparison between encoder data and camera data

Run	Average Error (E_j %)	Standard Deviation (%)
1	6.6	5.2
2	5.7	5.6
3	6.1	5.9
4	4.5	6.0
5	4.9	5.0

Table 1. Relative errors between encoder and camera data for a rectangular path over 5 runs

4 Coordinated multi-agent exploration

The method was successfully integrated in the context of a practical application, in which a team of Smartease robots supervised by a Biba robot explores an indoor environment made up of rooms and corridors. The Biba robot is able to localize itself precisely in the environment. Smartease robots are then localized relatively to the Biba robot, using the method described in this paper.

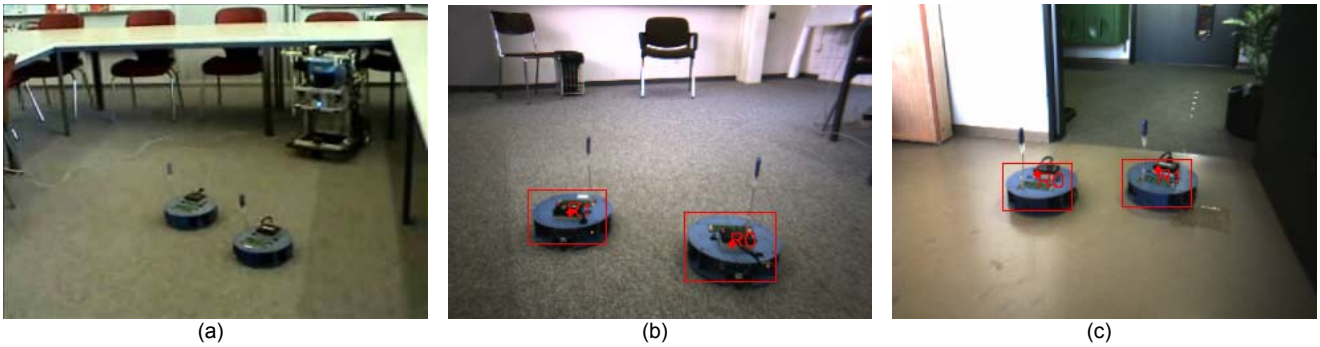
Smarteases are cylindrical robots with differential drive embedding a PC104 board running RTAI Linux on a Pentium 166MHz. Remote control was performed through a 802.11b ad-hoc wireless network between the robots and a control laptop.

The Biba robot embeds a PowerPC 750 clocked at 400MHz running RTAI Linux, a Pentium III running Windows and two SICK laser range finders mounted back to back. The navigation and localization software runs on the PowerPC, while the image processing method presented in this paper is executed on the embedded PC. The Biba robot relies on odometry data, an a-priori known map of the environment, and SICK laser scanner data to localize itself in the environment.

Fig 13(a) is a picture of the robot team; Fig 13(b) and 13(c) show instead the Smartease robots, localized while exploring two different rooms. In this experience, the small vehicles were remote-controlled, while a navigation strategy allowed the supervisor to modify its trajectory

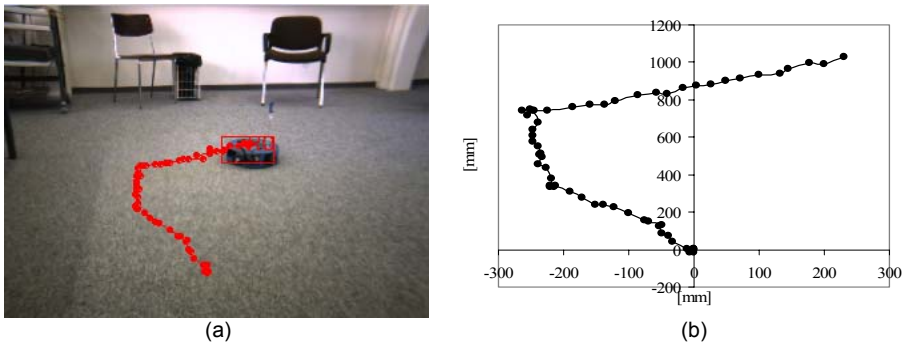
based on relative position information so as to keep the visual contact.

Finally, Fig 14 illustrates the application of the method for robot trajectory reconstruction and environment mapping. Specifically, Fig 14(a) reports the vehicle detected in one frame, while moving surveyed by the Biba robot maintained in a stationary condition. The estimated positions are also represented, reprojected onto the image plane in Fig 14(a), and plotted in real world in Fig 14(b).



(a) (b) (c)

Fig 13. (a) The multi-robot team; (b, c) Smartease robots exploring two different environments



(a) (b)

Fig 14. (a) Robot trajectory reprojected onto the image plane, and (b) in real world

5 Conclusions

A method of relative localization for multi-robot systems was presented, which was demonstrated successfully with a team of small vehicles navigating in a coordinated manner, supervised by a complex agent.

The method allows the supervisor to simultaneously detect and localize all the small robots, providing a reliable surveying system. It employs an effective combination of feature-based and appearance-based visual modelling and tracking algorithms for robot recognition, and 3D stereo information for multi-robot localization.

Experimental tests, performed on a real multi-agent system, proved the method to be accurate and effective for multi-robot localization and environment exploration and mapping tasks.

Acknowledgment

The work presented in this paper has been supported in part by the IST project RECSYS (IST-2001-32515).

References

1. P.J. Besl, N.D. McKay: A method for registration of 3-D shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (2), February, 1992.

2. J. Borenstein, H.R. Everett, L. Feng: Where am I? Sensors and methods for mobile robots positioning, Technical Report UM-MEAM-94-21, University of Michigan, December 1994.

3. G.N. DeSouza, A.C. Kak: Vision for mobile robot navigation: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, February 2002, vol. 24, n. 2, pp. 237-267.

4. T. D'Orazio, F. P. Lovergine, M. Ianigro, E. Stella, A. Distante: Mobile robot position determination using visual landmarks, IEEE Transactions on Industrial Electronics, 41 (6), 1994, pp. 654-662.

5. A. Farinelli, L. Iocchi, D. Nardi: Multi-robot systems: a classification focused on coordination, IEEE Transactions on System Man and Cybernetics, part B. 34 (5), October 2004, pp. 2015-2028.

6. R.C. Gonzalez, R.E. Woods: Digital Image Processing, 2nd Edition, Prentice Hall, 2002.

7. R. Grabowski, L.E. Navarro-Serment, C.J.J. Paredis, P.K. Khosla: Heterogeneous teams of modular robots for mapping and exploration, Autonomous Robots, June 2000, vol. 8, n. 3, pp. 293-308.

8. A. Howard, M.J Matarić, G.S. Sukhatme: Cooperative relative localization for mobile robot teams: an ego-centric approach, Multi-Robot Systems: From Swarms to Intelligent Automata, Alan C. Schultz and Lunee E. Parker and Frank E. Schneider (eds), Kluwer, 2003, pp. 65-76.

9. K. Konolige: Small vision systems: hardware and implementation, Eighth International Symposium on Robotics Research, Japan, October 1997.

10. R. Kurazume, S. Hirose: An Experimental Study of a Cooperative Positioning System, Autonomous Robots, 2000, 8(1), pp. 43-52.

11. J.J. Leonard, H.F. Durrant-White: Mobile Robot Localization by tracking geometric beacons, IEEE Transactions on Robotics and Automation, June 1991, 7 (3), pp. 376-382.

12. R. Madhavan, K. Fregene, L.E. Parker: Distributed Cooperative Outdoor Multirobot Localization and Mapping, Autonomous Robots, 2004, 17(1), pp. 23-39.

13. E. Marchand, P. Bouthemy, F. Chaumette, V. Moreau: Robust visual tracking by coupling 2D motion and 3D pose estimation, Proc. International Conference on Image Processing, 1999.

14. A. Martinelli, F. Pont, R. Siegwart: Multi-Robot localisation using relative observations, Proc. IEEE Int. Conf. Robotics and Automation, 2005.

15. M. Moors, F. Schneider, D. Wildermuth: Relative position estimation in a group of robots, Proceedings of the 6th International Conference on Climbing and Walking Robots (CLAWAR), 2003, pp. 983-990.

16. H. Murase, S.K. Nayar: Visual learning and recognition of 3-D objects from appearance, International Journal of Computer Vision, 1995, vol. 14, pp. 5-24.

17. S.K. Nayar, S.A. Nene, H. Murase: Real-time 100 object recognition system, Proceedings of ARPA Image Understanding Workshop, San Francisco, February 1996.

18. I.M. Rekleitis, G. Dudek, E. Milios: On multiagent exploration, Proc. Vision Interface, Vancouver, BC, Canada, 1998, pp. 455-461.

19. S. Roumeliotis, G.A. Bekey: Distributed multi-robot localization, IEEE Trans. On Robotics and Automation, 2002, 18 (5).

20. J.C. Russ: The image processing handbook, 2nd Edition, CRC Press, 1994.

21. D. Scaramuzza, A. Martinelli, R. Siegwart: Precise bearing angle measurement based on omnidirectional conic sensor and defocusing, Proceedings of European Conference on Mobile Robots, 2005.

22. E. Trucco, A. Verri: Introductory techniques for 3D computer vision, Prentice Hall, 1998.