PRECISE BEARING ANGLE MEASUREMENT BASED ON OMNIDIRECTIONAL CONIC SENSOR AND DEFOCUSING

Davide Scaramuzza, Agostino Martinelli, Roland Siegwart

Swiss Federal Institute of Technology Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
[davide.scaramuzza, agostino.martinelli, roland.siegwart]@epfl.ch

ABSTRACT

Recent studies on multi-robot localization have shown that the uncertainty of robot location may be considerably reduced by optimally fusing odometry and the relative angles of sight (bearing) among the team members. However, the latter requires the capability for each robot of detecting the other members up to large distances and wide field of view. Furthermore, robustness and precision in estimating the relative angle of sight is of great importance.
In this paper we show how all of these requirements may be achieved by employing an omnidirectional sensor made up of a conic mirror and a simple webcam.
We use different colored lights to distinguish the robots and optical defocusing to identify the lights. We show that defocusing increases the detection range by several meters, compensating the decay of resolution related to the omnidirectional view, without losing robustness and precision. To allow a real time implementation of light tracking, we use a recent tree-based union find technique for color segmentation and region merging.
We also present a self-calibration technique based on an Extended Kalman Filter (EKF) to derive the intrinsic parameters of the robot-sensor system. The performance of the approach is shown through experimental results.

1. INTRODUCTION

Many robotic applications require that robots work in collaboration in order to perform certain tasks. Therefore, precise relative localization (position and orientation) of each member is needed for mobile robot autonomy. In the last decades, several solutions for cooperative localization have been proposed (see [1], [2], [3], [4], [5] and [6]).

In general the relative observations among robots may be categorized into three special cases: relative bearing (i.e. the direction of the observed member in the reference of the observer); relative distance; relative orientation (i.e. the orientation of the observed in the reference of the observer). When the three observations are simultaneously combined, they contain all the necessary information to estimate the configuration of the observed in the reference of the observer.

Recently, in [7] it has been shown that the accuracy of the localization is strongly improved by only using the relative bearing rather than the whole relative configuration. Conversely, in [1], [2], [3], [4], [5] and [6], all global relative information is exploited.
The generally adopted way to get the relative configuration of the observed in the reference of the observer is some combination of range sensors, like laser range finders, sonar (see [3]), and single or stereo cameras. Instead in [6] a fixed overhead camera with ground truth is exploited to estimate all robot configurations, but this is only possible for small environments. Other solutions use single frontal cameras but the robots need to move in such a way that they can see each other at all times.

As mentioned above, the configurations of all robots may be inferred by only fusing odometry and relative bearing. Therefore, instead of using expensive equipment to measure the distances, we only need a sensor to measure the bearing angle. In order to implement this strategy the robots are required to maintain visual contact among them. For this, a 360° omnidirectional camera seems to be the best solution with respect to a standard pinhole camera. Conversely, many problems occur because of field distortion and decay of visual resolution, since a larger region of space (6-9 times bigger) is projected in the same camera image plane. Because of this, standard object recognition techniques cannot be employed, nor can colored landmarks be used since their identification would be restricted to a smaller distance.

In the next section of this paper we show how some of these difficulties may be overcome employing an omnidirectional sensor composed of a conic mirror and a simple webcam. In section III we present the robot detection method, which uses different colored lights to distinguish the robots. Then we show how to take advantage of optical defocusing to identify the lights and increase the detection range by several meters. By doing so, we compensate the decay of resolution related to the

omnidirectional view without losing robustness and precision. In section IV we briefly describe a tree-based union find technique for color segmentation and region merging, which we used to allow a real time implementation of light tracking. In section V, we introduce an experimental self-calibration technique, based on EKF, to derive the intrinsic parameters of the robot-sensor system; that is, how to automatically derive position and orientation of the camera in the robot reference frame. Finally, in the last section, the performance of the approach is shown through experimental results.

2. OMNIDIRECTIONAL CONIC SYSTEM

Omnidirectional imaging systems are typically arranged by positioning a standard camera below a convex reflective surface. The properties of the image and the magnitude of the field of vision will depend on the chosen mirror profile, which is usually spherical, parabolic [8], hyperbolic [9] or equiangular [10], [11]. These families of mirrors provide views with wide angles of elevation (e.g. 130° in the equiangular one), which are well-suited to robot navigation, environment exploration and representation, but they are not appropriate for object detection and recognition because the objects would appear too small to be correctly identified. Moreover, because of the precision involved in their manufacturing, they can be very expensive, especially if many are needed, as in multi-robot tasks.

Yagi et al. [12] were the first to propose the use of a conic projection image sensor (COPIS) for vision-guided navigation. It consists of a conic mirror and a camera, with its optical axis aligned with the cone axis as shown in Fig. 1.

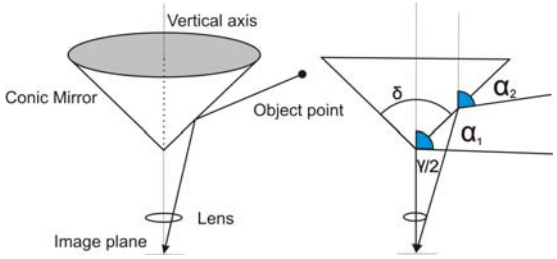


Figure 1: *Arrangement of the cone-camera system.*

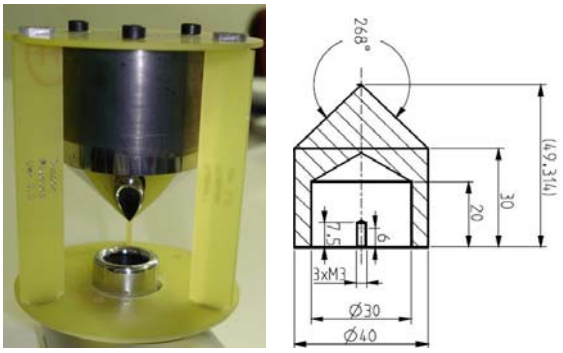
By this arrangement the visual field covers a range between two view angles α_1, α_2 which are determined from the visual angle γ of the camera and the vertex angle δ of the conic mirror as:

$$\alpha_1 = \delta$$
$$\alpha_2 = \delta - \frac{1}{2}\gamma \quad (1)$$

Thus the angle of sight provided by the sensor is $\alpha_1 - \alpha_2 = \gamma/2$. Note that, because of the linear mirror profile, the visual space is not distorted along vertical planes passing through the cone axis. Conversely, the field of vision is smaller than other catadioptric configurations, but this leads to a greater angular resolution.

Moreover, the advantages of using a conic mirror are its high quality, its inexpensiveness, and its easy manufacture. When the axes of the camera and the mirror are aligned, the real bearing angle can be directly computed from the image because there is an invariant relation with the distance from and the height of the object.

The parameters α_1, α_2 determine respectively the minimum and maximum distance from which objects can be seen. In order to have an angle of view independent of the visual angle of the camera (that generally ranges between $\gamma = 30°$ and $\gamma = 70°$ for different camera models) we set $\alpha_1 = \delta = 92°$, that is a cone with a 92° vertex angle. In this way, the sensor provides a field of vision extending from the horizon and upwards. In fact, $\alpha_1 = 90°$ would be sufficient to assure the desired field of vision, but by choosing $\alpha_1 = 92°$ we can prevent any manufacture errors or camera-mirror misalignments. The conic mirror that has been designed is shown in Fig. 2.



a) b)

Figure 2: *Our conic mirror placed above the webcam (a). Design parameters (b).*

This arrangement of course forces us to fix the landmark right above the mirror (see Fig. 4). As a result, its projection on the camera plane of another robot will appear at the edges of the image when the robots are close to each other and will approach the image center when they are far. In our arrangement we used a USB webcam with a horizontal field of vision of $\gamma = 40°$ and the minimum distance for which the object is detectable is 40 cm.

3. DETECTION

3.1. Light Detection

The decay of visual resolution and the distortion due to the non-planar mirror profile make standard object recognition techniques unsuitable for omnidirectional cameras. Similarly, colored blobs placed above the robot, such as the ones used in the RoboCup domain, are also inappropriate. Fig. 3 shows two pictures taken respectively

by an equiangular (a) and a conic (b) sensor after unwrapping. The colored blobs (indicated by the arrows) are only 1 m away from the camera. Because of the major angular resolution of the conic mirror, in (b) the blobs are more detectable. However, because of different lighting conditions and shadows, the color intensity diffused by the blobs may change and, as a result, robustness is compromised. Moreover, the greatest distance for which the object is still visible is 2-3 m because the size of the object image decays inversely with the distance.

a)

b)

Figure 3: *Unwrapped images taken by means of an equiangular (a) and a conic mirror (b). The colored blobs used are indicated by the arrows. They are located only 1 m away from the camera.*

For our purposes, the robots should maintain visual contact with one another for several meters and with the minimum number of false positives. Thus, the above strategy is not practical.

Instead of using colored blobs we exploited colored lights placed on the robot mirror. Each of these consists of a set of 24 high luminosity and monochromatic LED emitters (2 V, 20 mA), having 20° of directivity. They are positioned along the perimeter of a circle in order to have an omnidirectional beam (Fig. 4). Different colored lights are used to distinguish more robots.

Figure 4: *The circular set of LEDs forming the light source, positioned on the mirror.*

3.2. CCD Saturation

The biggest difficulty that we found using colored light detection was CCD saturation. Saturation is a normal part of CCD operation. Each pixel of the CCD can only contain several hundred thousand electrons. Once this capacity is exceeded, these excess electrons begin to pour over the potential barriers that define the pixel and into adjacent pixels [13]. Consequently, the values associated to the saturated pixels reach the maximum (e.g. 255 for standard 24 bit RGB cameras) and appear white. As a result, color information is lost. Moreover, these saturated white pixels are surrounded by a glow or a halo, due to other optical phenomena like the internal reflection of the camera lens and the aperture of the iris.

This was exactly the problem we encountered in our system, because the bright light coming from the LED beam points directly towards the conic mirror, thus saturating the camera (Fig. 5).

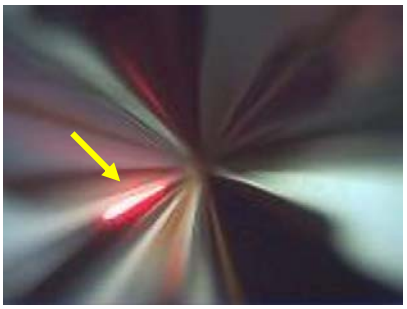


Figure 5: *The light source used in our experiments (pointed to by the arrow) as seen through the conic mirror. It appears white because of saturation. The source is located 5 meters away from the sensor.*

3.3. Optical Defocusing

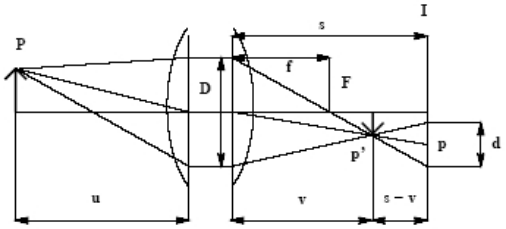


Figure 6: *The camera geometry and camera parameters.*

Because of saturation, standard color detection techniques cannot be applied. In order to have the original light color correctly reproduced into the CCD, the light should be very far from the camera. Thus, to avoid the saturation problem, we experimented the optical defocusing, that is, we only modify the distance between CCD and lens, by acting manually on the camera objective. The intuitive reason for doing so is that defocusing spreads the light coming into the CCD into a larger region with less intensity. Theoretically, this may be explained in terms of spatial convolution. In fact, in optics, the blur of an image due to defocusing can be described by a point spread function [14]. Let *P* be a point in the scene and let *p'* be its focused image (Fig. 6). If *P* is not in focus then the image of *P* becomes a circular image *p* called the *blur circle*. The point spread function represents the precise structure of the blur circle. Due to diffraction, variations of light wavelength and other optical

phenomena, this is usually described using a two-dimensional Gaussian function, which can be thought as the impulse response of the lens system (i.e. the blurring filter):

$$h = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}. \quad (2)$$

Where σ is the spread parameter, which depends on the focal length *f*, the lens diameter *D*, the real distance *s* of the image from the lens, and the distance *u* from the real object [14]:

$$\sigma \propto Ds \cdot \left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right). \quad (3)$$

So, if $g(x,y)$ is the observed image of an object on the screen and $f(x,y)$ is the corresponding focused image, then g is given by the convolution

$$g(x,y) = h(x,y) * f(x,y). \quad (4)$$

In the presence of saturation, $f(x,y)$ tends to a Dirac function. The more we increase the defocusing (and therefore σ) the more the saturation degrades. At a certain level of defocusing and at a given distance from the light source, it completely vanishes (see Fig. 7).

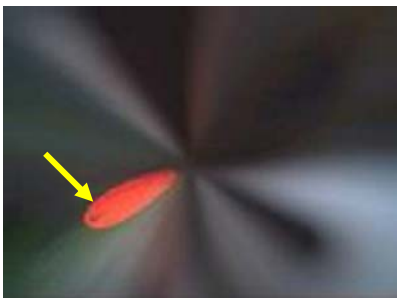


Figure 7: *The same image as in Fig. 5 but after defocusing. Now the light source (pointed to by the arrow) appears in its original color. It also looks like an ellipse. The source is located 5 meters away from the sensor.*

Fig. 7 is the same as Fig. 5 but after defocusing. Observe that the image of the light source has been spread into an ellipse; this may be easily explained by observing that the ellipse is the convolution of a rectangle (the previous saturated image) and the Gaussian impulse response. Note that now saturation is annihilated. The light source clearly appears in its original color (red) and its edges are also well defined.

This technique of avoiding saturation via defocusing has proved to correctly and robustly identify the light source within a range between 40 cm and 5 meters. In fact, because of the spreading, the bright point appears larger and thus may be identified up to several meters. For instance, the picture depicted in Fig. 7 was obtained with the light source located 5 meters away from the vision sensor.

4. FAST COLOR SEGMENTATION AND MERGING

The most important steps in our color vision task are *segmentation* and *merging*. The first one classifies each pixel in an image into one of a discrete number of color classes. The other one merges pixels belonging to the same color class into connected regions. Since these operations are crucial in real-time mobile robot applications, we implemented the tree-based union find strategy described in [15]. In this approach, very fast color segmentation is performed by using look-up-tables. Then, the first stage is to compute a run length encoded version of the segmented image. This is done by scanning horizontal rows and searching for a run of adjacent pixels of the same color class (see Fig. 8).

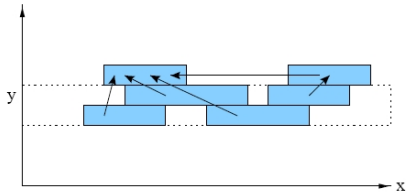


Figure 8: *The process of region merging by addressing the run's parent elements. Each rectangle represents a run that is a set of adjacent pixels of the same color class.*

Subsequently, a second pass is needed to merge the classified runs. This is performed by assigning to each run an identifier indicating the run's parent element, that is, the upper leftmost member of the region. Initially, each run labels itself as its parent. Then the merging procedure scans all runs and merges those which are of the same color class and which overlap under four-connectedness. At the same time, it updates the run's parent element identifier by pointing it towards the upper leftmost member of the same color region. The process is illustrated in Fig. 8. In our equipment (Pentium IV, 1.7 GHz), this algorithm takes about 7 ms to segment and merge up to 32 different color regions for images of size 640x480 pixels.

5. SELF CALIBRATION BASED ON EKF

As we said at the beginning, the bearing angle measurements in the local robot frame can be directly inferred from the angle of each light source in the defocused image. The camera reference frame is chosen with the *x* and *y* axes parallel to the horizontal and vertical borders of the image. The axes origin is fixed at the intersection of the radial edges in the omnidirectional image because we assume that the axes of the camera and the cone are perfectly aligned. We verified this hypothesis to be satisfied.

In order to measure the angle, we actually compute the barycentre of each color blob in the omnidirectional image. In fact, recall that due to defocusing, the color blob appears as an ellipse, thus, the angle to the barycentre is the same as the departure direction of the light. However, we also have to take into account the position and orientation of the sensor in the robot reference frame (see

Fig. 9). In fact, an error of 1 cm in estimating the position of the camera with respect to the robot's frame would produce a bearing error of 0.2° if the distance between the robots is 3 m.

Furthermore, regarding the observed robot, we have to consider the position of the light source in its own reference frame, which is in general not exactly at the origin. However, we fixed the light source on the sensor mirror and therefore only three parameters (ϕ_O, ρ_O, θ_O in Fig. 9) need to be estimated. To this end, we introduce a method based on an EKF. In particular, we describe a simple procedure to estimate the parameters characterizing the transformation between two reference frames, respectively attached to the robot and to the camera. Since we assume that the mirror axis is aligned with the camera's optical axis, this transformation is between two 2D references and therefore is fully characterized by the three parameters ϕ_O, ρ_O and θ_O. The two reference frames are shown in Fig. 9. To evaluate the three parameters, we introduce an EKF, which simultaneously estimates the state $(d, \theta, \phi_O, \rho_O, \theta_O)^T$, where d and θ are respectively the distance between the robot center and the light source and the angle between the robot orientation and the line connecting the light source with the robot center (see Fig. 9).

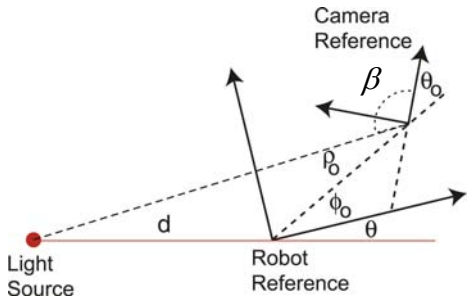


Figure 9: *The two reference frames respectively attached to the robot and to the camera. The five parameters estimated through an EKF and the bearing observations β are also indicated.*

The filter uses the encoder readings as inputs and the bearing angles of the light source in the camera reference as outputs. We chose to characterize the system with a state containing the previous five parameters since it can be proved that, with the previous inputs and outputs, this state is observable [16]. On the contrary, the state containing the whole robot configuration together with the three parameters ϕ_o, ρ_o, θ_o is not observable [17]. The dynamics of our system are described through the following equations:

$$\begin{cases} d_{i+1} = d_i + \delta\rho_i \cos(\theta_i) \\ \theta_{i+1} = \theta_i + \delta\theta_i - \dfrac{\delta\rho_i}{d}\sin(\theta_i) \\ \phi_{o_{i+1}} = \phi_{o_i} \\ \rho_{o_{i+1}} = \rho_{o_i} \\ \theta_{o_{i+1}} = \theta_{o_i} \end{cases} \quad (5)$$

where $\delta\rho_i$ and $\delta\theta_i$ are respectively the robot translation and rotation. Since our robot is equipped with a differential drive system, their expressions are:

$$\delta\rho = \frac{\delta\rho_R + \delta\rho_L}{2} \qquad \delta\theta = \frac{\delta\rho_R - \delta\rho_L}{b} \quad (6)$$

where $\delta\rho_R$ and $\delta\rho_L$ are the right and left wheel translations, directly returned by the encoders, and b is the distance between the wheels.

The bearing angle of the light source in the camera reference frame corresponding to the available observations has the following analytical expression:

$$\beta = \text{atan2}\left(-\rho_o \sin(\theta+\phi_o), -d - \rho_o \cos(\theta+\phi_o)\right) - \theta - \phi_o - \theta_o \quad (7)$$

The EKF estimates the state $(d, \theta, \phi_O, \rho_O, \theta_O)$ by fusing the information coming from the encoder data and the bearing angle observations. In order to implement the standard equations of this filter we need to compute the two Jacobians F_x and F_u of the dynamics in (5) respectively with respect to the state and with respect to the encoder readings $(\delta\rho_R, \delta\rho_L)$. Finally, we need to compute the Jacobian (H) of the observation function in (7) with respect to the state [18]. By a direct computation we obtain:

$$F_x = \begin{bmatrix} 1 & -\delta\rho\sin\theta & 0 & 0 & 0 \\ \dfrac{\delta\rho}{d^2}\sin\theta & 1-\dfrac{\delta\rho}{d}\cos\theta & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$F_u = \begin{bmatrix} \dfrac{\cos\theta}{2} & \dfrac{\cos\theta}{2} \\ \dfrac{1}{b} - \dfrac{\sin\theta}{2d} & -\dfrac{1}{b} - \dfrac{\sin\theta}{2d} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (9)$$

$$H = \left[\frac{-\rho_o \sin(\theta+\phi_o)}{d^2 + 2\rho_o d\cos(\theta+\phi_o) + \rho_o^2}, \frac{-d\rho_o\cos(\theta+\phi_o) - d^2}{d^2 + 2\rho_o d\cos(\theta+\phi_o) + \rho_o^2}, \right. \quad (10)$$
$$\left. \frac{-d\rho_o\cos(\theta+\phi_o) - d^2}{d^2 + 2\rho_o d\cos(\theta+\phi_o) + \rho_o^2}, \frac{d\sin(\theta+\phi_o)}{d^2 + 2\rho_o d\cos(\theta+\phi_o) + \rho_o^2}, -1 \right]$$

6. EXPERIMENTAL RESULTS

Before applying the calibration procedure described above, we needed to estimate the precision of our sensor in measuring angles from the light source. In order to do so, we moved the light along a circle centered on the sensor and compared the real displacement with the displacement

inferred from the bearing given by the sensor. We obtained very good precision with uncertainty less than 0.25°.

In the calibration phase, we set our sensor on the robot with the following settings: $\phi \approx -\pi$, $\rho \approx 0.07$ m, and $\vartheta \approx -\pi/2$, all of which we measured manually. Then, we let the robot move randomly along a winding trajectory of about 4.5 m. Finally, the encoder data and bearing measurements were fused according to the calibration procedure above. The values of the three parameters estimated during motion are plotted versus distance in Fig. 10. As you can see, after 3 meters of navigation they start to converge to a stable value.

Finally, the estimated parameters were $\phi_0 = -3.11$ rad, $\rho_0 = 0.074$ m and $\vartheta_0 = -1.58$ rad. Note that they are consistent with the values measured manually. Also, observe in Fig. 10 that the plot of ρ_0 starts at 0 since we took the robot origin as the initial value in the EKF. Nevertheless, its estimation converges to the expected value.

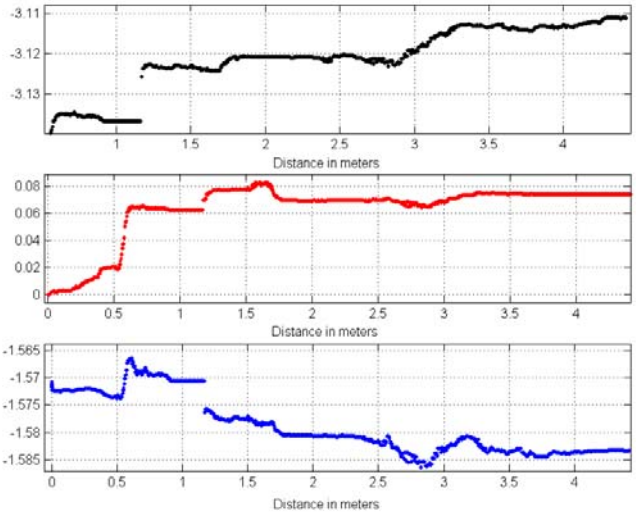


Figure 10: *The estimated ϕ_O, ρ_O, θ_O during motion, plotted versus distance.*

7. CONCLUSIONS

In this paper we introduced a new approach for estimating the bearing angle of a robot in the local reference frame of another robot. The proposed method uses an omnidirectional sensor made up of a conic mirror and a simple webcam. The design of this sensor is also described in the paper.

With the aim of an application for multi-robot localization, we used colored light sources to distinguish the robots, and optical defocusing to identify the lights. We showed that defocusing allows us to avoid CCD saturation and to increase the detection range by several meters, compensating the decay of resolution related to the omnidirectional view, without losing robustness and precision. To allow a real time implementation of light tracking, we used a fast technique for color segmentation and region merging.

We also presented a self-calibration technique, based on an extended Kalman filter, to derive the intrinsic parameters which describe position and orientation of the camera in the robot system. The performance of the approach was shown through experimental results.

8. REFERENCES

[1] D. Fox, W. Burgard, H. Kruppa, S. Thrun, Autonomous Robots 8, 325-344, 2000.

[2] R. Kurazume, S. Nagata, and S. Hirose, "Cooperative positioning with multiple robots," in *Proc. 1994 IEEE Int. Conf. Robotics and Automation*, vol. 2, Los Alamitos, CA, pp. 1250–1257, May 8–13, 1994.

[3] R. Grabowski, L. E. Navarro-Serment, C. J. J. Paredis, and P. K. Khosla,"Heterogeneous teams of modular robots for mapping and exploration," *Auton. Robots*, vol. 8, no. 3, pp. 293–308, June 2000.

[4] I. M. Rekleitis, G. Dudek, and E. E. Milios, "Multi-robot exploration of an unknown environment, efficiently reducing the odometry error," in *Proc. 15th Int. Joint Conf. Artificial Intelligence (IJCAI-97)*, vol. 2, M. Pollack, Ed., Nagoya, Japan, pp. 1340–1345, Aug. 23–29, 1997.

[5] I. M.Rekleitis, G.Dudek, and E. E.Milios, "On multiagent exploration," in *Proc. Vision Interface* Vancouver, BC, Canada, pp. 455-461, 1998.

[6] S.I. Roumeliotis and G.A. Bekey, Distributed Multirobot Localization, *IEEE Transaction On Robotics And Automation*, Vol 18, No.5, October 2002.

[7] Martinelli, A., Pont, F. and Siegwart, R., Multi-Robot Localization Using Relative Observations, in *Proc. 2005 IEEE Int. Conf. Robotics and Automation*, (ICRA05), 2005.

[8] S. K. Nayar, "Catadioptric Omnidirectional Camera,", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.

[9] T. Svoboda, T. Pajdla, V. Hlavac, Central Panoramic Cameras: Design and Geometry, In proceedings of the 3rd Computer Vision Winter Workshop, Gozd-Martuljek, Slovenia, February 1998.

[10] J. S. Chahl, M. V. Srinivasan, "Reflective surfaces for panoramic imaging," Applied Optics, vol. 36, no. 31, pp. 8275–8285, 1997.

[11] M. Ollis, H. Herman, S. Singh, Analysis and Design of Panoramic Stereo Vision Using Equi-Angular Pixel Cameras, tech. report CMU-RI-TR-99-04, Robotics Institute, Carnegie Mellon University, January, 1999.

[12] Y. Yagi, Y. Nishizawa, M. Yachida, Map Based Navigation of the Mobile Robot Using Omnidirectional Image Sensor COPIS, *Proceedings IEEE the International Conference on Robotics and Automation* (R&A-92),pp.47-52,Vol.1,No.5,1992.

[13] J.A. Conchello, Image Collection and Correction for Computational Optical Sectioning Microscopy, Anatomy and Neurobiology Department, Washington University, National Institutes of General Medical Sciences, National Institutes of Health, 2000.

[14] A. Horii, ``Depth from defocusing", technical report, ISRN KTH/NA/P--92/16--SE, June 1992.

[15] J. Bruce, T. Balch, M. Veloso, Fast and Inexpensive Color Image Segmentation for Interactive Robots. In Proceedings of IROS-2000, Japan, October 2000.

[16] A. Martinelli, D. Scaramuzza, R. Siegwart, Camera Extrinsic Parameters Estimation during Robot Motion, in preparation.

[17] A. Martinelli, R. Siegwart, Observability analysis for Robot Localization, in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Edmonton, Canada.

[18] Y.Bar-Shalom, T.E.Fortmann, "Tracking and data association, mathematics in science and engineering", Vol 179, AcademicPress, New York, 1988.