Object Recognition in Swarm Systems: Preliminary Results

Paolo Stegagno, Caterina Massidda, and Heinrich H. Bülthoff

Abstract— **Object recognition is a fundamental topic for the development of robotic systems able to interact with the environment. Most existing methods are based on vision systems and assume a broad point of view over the objects, which are observed in their entirety. This assumption is sometimes difficult to fulfill in practice, and in particular in swarm systems, constituted by a multitude of small robots with limited sensing and computational capabilities. We have developed a method for object recognition with a heterogeneous swarm of low-informative spatially-distributed sensors employing a distributed version of the naive Bayes classifier. Simulation results show the effectiveness of this approach highlighting some nice properties of the developed algorithm.**

I. INTRODUCTION

The current miniaturization trend of electronic components, sensors and actuators will eventually bring to the effective employment of multi- and many-robot systems in everyday life. A multitude of small cheap and simple robots can be used to perform high level tasks as exploration, patrolling, cooperative transportation. The analysis and control of a swarm of this kind still pose a range of research challenges that must be addressed for robust real world applications. Among others, object recognition is fundamental in order to interact with the environment (e.g.: for the identification of the object of a common action). However it is mostly seen as a computer vision topic and several peculiarities of robotic swarms that can bring advantages and disadvantages in the recognition are ignored.

First, the typical dimension of the objects in the environment (cars, trees, buildings) is usually one or more orders of magnitude greater than the physical dimension of the platforms considered in swarm robotics. In addition, the exteroceptive sensor equipment of the robots is limited due to constraints on payload, power consumption and computational power. Hence, each robot is usually able to observe only some details (e.g.: one or two edges) or some specific characteristic (e.g.: color or material) of the sensed objects. On the other end, each small robot will have a different point of view, test a different characteristic, or even try to interact with the object to infer useful information.

Most of the state-of-the-art algorithms for object recognition (e.g.: [1], [2], [3]) involves highly-informative sensors as cameras or 3D range finders, assume a broad point of view over the observed objects (hence one single sensor is enough for the recognition), perform the computations in a centralized fashion and are computationally expensive and real-time-unfeasible on limited platforms. The usage of objects' local features in the recognition process has been thoroughly studied in literature [2], [3], [4], [5], but it has mostly been exploited by mean of a centralized entity and considering only visual features. Even more important, most methods selects informative local features of the observed objects and use their spatial relationship, thanks to their broad point of view, whereas each small agent of a swarm can only rely on its limited, non-selected point of view.

A generalization of the classical approach studies the problem of understanding the environment given the images collected from different points of view. State-of-the-art algorithms (e.g.: [6]) usually take advantage of common features from the different views. In [7] the authors selects an optimal number of images from different viewpoints. In [8] the authors reconstruct the cluttered parts of the environment and discern the subject from the background. Many authors [9] [10] [11] investigate the distribution of the computation among several camera stations, encoding the features used for the recognition before the transmission to a base station. In [12], the authors propose a recognition method to classify the object observed by a network of smart cameras. Again, each camera has a broad point of view over the object, and the methods benefit from overlapping in the measurements.

For the best of our knowledge, our work is the first to focus explicitly on object recognition in swarm systems. Our goal is the development and analysis of ad-hoc distributed algorithms to overcome and exploit the peculiarities of swarm systems. Hence, we aim at turning the intrinsic structure of the swarm as a multitude of local limited sources of information into an opportunity, exploiting the intrinsic similarities of objects of the same type, as for example similar colors, physical properties, patterns and textures and the same constituent materials. In fact, one of our objective is to develop a system able to recognize all the objects of certain type, and not only the single objects whose models are already known by the system. For this reason, in the simulations we will test the system with different objects (but of the same type) w.r.t. the ones used to build the models.

Finally, we want to overcome the idea that object recognition is strictly a computer vision topic, presenting a robotic system that only partially employs cameras and proposing a method for the fusion of information coming from several different types of sensors.

II. SYSTEM OVERVIEW

We consider a heterogeneous swarm system A of n agents $A = \{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n\}$ living in a generic environment and surrounding an object ω in such environment. Each $\mathcal{A}_i$ is equipped with an exteroceptive sensor and gathers a measurement z_i of ω. In general, different robots can be

P. Stegagno, C. Massidda and H. H. Bülthoff are with the Max Planck Institute for Biological Cybernetics, Department of Human Perception Cognition and Action, Spemannstraße 38, 72076 Tübingen, Germany {paolo.stegagno, caterina.massidda, hhb}@tuebingen.mpg.de.

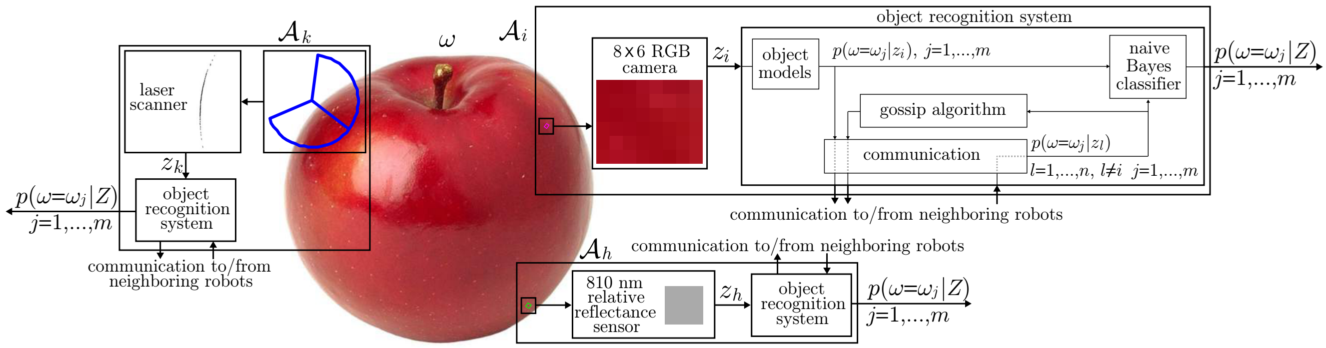

Fig. 1: An outline of the object recognition system. Three robots $\mathcal{A}_i$, $\mathcal{A}_h$, $\mathcal{A}_k$ are equipped respectively with a 8×6 RGB camera, a 810 nm relative reflectance sensor and a laser scanner. The details of the object recognition system are depicted for robot $\mathcal{A}_i$.

equipped with different sensors. Possible types of sensors include cameras, laser range finders, sonar and IR arrays, material detectors, temperature and stiffness sensors among others. In this work, as a mean of example and for simulation purpose we have considered three types of sensors.

The first is a 8×6–pixel RGB camera, used as a color sensor. As a full size image from a normal-resolution camera contains too many data to be efficiently processed in case of limited computational power, still visual data can provide useful information for the recognition, for example the color of a small portion of the observed object. The second sensor is a laser range finder, and is used to measure a small portion of the profile of the object. Finally, we also consider a 1-pixel near-infrared (NIR) relative reflectance sensor, which is able to measure the amount of electromagnetic radiation at a given wavelength (810 nm in our case) reflected by a small portion of the surface of the object, with respect to the total electromagnetic radiation at that wavelength impacting on that surface. In general, we will denote with z_i the measurement collected by the robot $\mathcal{A}_i$, independently from the type of $\mathcal{A}_i$'s sensor.

We assume the robots to be able to move in the environment as a swarm maintaining network connectivity. We assume also that in the environment there is one and only one object ω out of a set of m possible objects $\Omega = \{\omega_1, \ldots, \omega_m\}$. Then, whenever the robots encounter ω on their path, their task is to identify which type of object ω is.

As outlined in Fig. 1, each $\mathcal{A}_i$ collects a measurement z_i of the object, and uses it to feed the object recognition system, whose details in Fig. 1 are specified only for $\mathcal{A}_i$. Using previously known models of the objects in Ω, each $\mathcal{A}_i$ computes an estimate of the probabilities $p(z_i|\omega = \omega_j), j = 1, \ldots, m$ that the observed object ω is of type ω_j given its own measurement z_i. This information is used inside a naive Bayes classifier, along with the probabilities $p(z_l|\omega = \omega_j), j = 1, \ldots, m, l = 1, \ldots, n, l \neq i$ obtained through communication. Since the communication graph is not complete, an appropriate gossip algorithm is enforced to spread the knowledge of all $p(z_l|\omega = \omega_j)$.

III. Simulations

We have tested the developed system on a database of 12 types of objects (leaf, banana, sunflower, apple, starfish, butterfly, grape, hammer, pineapple, strawberry, wrench, scissors). For each ω_j, we have used four images to build the models, hence the entire database is built using 48 images. In a typical simulation, the image of an object is in the scene, and multiple robots are randomly deployed over it (image and reflectance measurements) or in its proximity (laser scanner). The simulated collected measurements are used to feed the system.

We have conducted extensive simulations on a set of additional 60 images, 5 for each object, varying the total number of robots and the type of sensors equipped. We have tested the system with 6, 15 and 30 robots, and for each of those simulations, we have tested the cases of all robots equipped with cameras, laser scanners or reflectance sensors, half of the robots equipped with cameras and half with laser scanners, and finally one third of the robots equipped with each type of sensor, for a total number of 15 different configurations. For each of them we have performed 20 simulations for each image in the testing set (then 100 simulations for each ω_j, and 1200 simulations to test each configuration of the swarm).

The results of the simulations, summarized in the confusion matrix (Table I) and in the overall percentage of correct recognitions (Table II) show the effectiveness of the approach. Remarkably, the performances in general increase with the number of robots, and whenever different types of measurements are used together the results are better than the results obtained with single sensors.

IV. Conclusions

In this work we have presented preliminary results of object recognition simulations using robotic swarms. For future works, one of the first task is to perform 3D simulations and experiments with larger data-sets. Moreover, we are interested in extending the number and types of sensors, evaluating also the impact and feasibility of temperature and stiffness sensors among others.

From a theoretical point of view, we plan to analyze the characteristics of the algorithm to optimize the number of sensors of each type and address the situation of multiple objects in the environment applying clusterization of the team members based on the object that they are observing.

	Le	**Ba**	**Su**	**Ap**	**St**	**Bu**	**Gr**	**Ha**	**Pi**	**Sr**	**Wr**	**Sc**
Le	**77**	0	0	0	0	0	1	0	0	0	0	0
Ba	20	**100**	0	0	1	0	0	0	0	0	0	0
Su	1	0	**100**	0	0	0	0	0	0	0	0	0
Ap	0	0	0	**75**	0	0	0	0	0	4	0	0
St	0	0	0	0	**99**	0	0	0	0	0	0	0
Bu	0	0	0	0	0	**100**	0	0	0	0	0	0
Gr	0	0	0	0	0	0	**90**	0	0	0	0	0
Ha	0	0	0	0	0	0	1	**97**	0	0	0	0
Pi	0	0	0	0	0	0	0	0	**100**	0	0	0
Sr	2	0	0	25	0	0	8	0	0	**96**	0	0
Wr	0	0	0	0	0	0	0	3	0	0	**99**	0
Sc	0	0	0	0	0	0	0	0	0	0	1	**100**

TABLE I: Confusion matrix for simulations with 30 robots, 10 equipped with cameras, 10 with laser scanners and 10 with reflectance sensors. Abbreviations: Le - Leaf; Ba - Banana; Su - Sunflower; Ap - Apple; St - Starfish, Bu - Butterfly; Gr - Grape; Ha - Hammer; Pi - Pineapple; Sr - Straweberry; Wr - Wrench; Sc - Scissors.

	C	**L**	**R**	**C+L**	**C+L+R**
6 robots	67.5	48.5	80.2	70.3	86.33
15 robots	69.7	56.9	86.3	78	93.6
30 robots	69.1	64.1	90.3	78	94.4

TABLE II: Percentage of correct associations for all 15 configurations of the swarm. C is camera, L is laser scanner, R is reflectance, C+L is camera and laser scanner, C+L+R is camera, laser scanner and reflectance sensor.

REFERENCES

[1] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7:1, pp. 11–32, 1991.

[2] D. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 1150–1157.

[3] C. Arth and H. Bischof, "Real-time object recognition using local features on a dsp-based embedded system," *Journal of Real-Time Image Processing*, vol. 3:4, pp. 233–253, 2008.

[4] B. Schiele and J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *International Journal of Computer Vision*, vol. 36, pp. 31–50, 2000.

[5] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3d object recognition from range images using local feature histograms," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2, 2001, pp. II–394–II–399 vol.2.

[6] A. Thomas, V. Ferrar, B. Leibe, T. Tuytelaars, B. Schiel, and L. Van Gool, "Towards multi-view object class detection," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006, pp. 1589–1596.

[7] S. Abbasi and F. Mokhtarian, "Automatic view selection in multi-view object recognition," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 1, 2000, pp. 13–16 vol.1.

[8] R. Diaz, S. Hallman, and C. C. Fowlkes, "Detecting dynamic objects with multi-view background subtraction," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[9] A. Yang, S. Maji, C. Christoudias, T. Darrell, J. Malik, and S. Sastry, "Multiple-view object recognition in band-limited distributed camera networks," in *Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on*, Aug 2009, pp. 1–8.

[10] D. M. Chen, S. S. Tsai, V. Chandrasekhar, G. Takacs, J. Singh, and B. Girod, "Tree histogram coding for mobile image matching," in *Proceedings of the 2009 Data Compression Conference*, ser. DCC '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 143–152.

[11] C. Christoudias, R. Urtasun, and T. Darrell, "Unsupervised feature selection via distributed coding for multi-view object recognition," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[12] N. Naikal, A. Yang, and S. Sastry, "Towards an efficient distributed object recognition system in wireless smart camera networks," in *Information Fusion (FUSION), 2010 13th Conference on*, July 2010, pp. 1–8.