# COLLECTIVE LEARNING OF CONCEPTS USING A ROBOT TEAM

Ana Cristina Palacios-García, Angélica Muñoz-Meléndez, Eduardo F. Morales
*National Institute for Astrophysics, Optics and Electronics*
*Luis Enrique Erro No. 1, Santa María Tonantzintla, Puebla, México*
*cpalacios@ccc.inaoep.mx, munoz@inaoep.mx, emorales@inaoep.mx*

Keywords: Robotics and automation; mobile robots and autonomous systems; vision, recognition and reconstruction; network robotics.

Abstract: Autonomous learning of objects using visual information is important to robotics as it can be used for local and global localization problems, and for service tasks such as searching for objects in unknown places. In a robot team, the learning process can be distributed among robots to reduce training time and produce more accurate models. This paper introduces a new learning framework where individual representations of objects are learned on-line by a robot team while traversing an environment without prior knowledge on the number or nature of the objects to learn. Individual concepts are shared among robots to improve their own concepts, combining information from other robots that saw the same object, and to acquire a new representation of an object not seen by the robot. Since the robots do not know in advance how many objects they will encounter, they need to decide whether they are seeing a new object or a known object. Objects are characterized by local and global features and a Bayesian approach is used to combine them, and to recognize objects. We empirically evaluated our approach with a real world robot team with very promising results.

## 1 INTRODUCTION

The design of robot teams is a very active research domain in the mobile robotics community. Robot teams have effectively emerged as an alternative paradigm for the design and control of robotic systems because of the team's capability to exploit redundancy in sensing and actuation.

The research on robot teams has focused on developing mechanisms that enable autonomous robots to perform collective tasks, such as strategies for coordination and communication (Asada et al., 1994; Matarić, 1997); exploration, mapping and deployment (Howard et al., 2006); sensing, survillance and monitoring (Parker, 2002); and decentralized decision making (Wessnitzer and Melhuish, 2003). In these works, a robot team can reduce time to complete a complex task that is allocated among its members.

Despite constant research on the design of robot teams, very little attention has been paid so far to the development of robot teams capable of learning from their interaction with their environment. In addition to their capability for accelerated learning, learning robot teams can be used to acquire a much richer and varied information compared to the information acquired by single learning robots.

Learning is a key issue to achieve autonomy for both, single robot and robot teams. Learning capabilities can provide robots flexibility and adaptation needed to cope with complex situations. In the context of robot teams, the most common machine learning approach has been reinforcement learning, where the idea is to learn optimal policies using a set of robots to improve the coordination of individual actions in order to reach common goals (Asada et al., 1994; Matarić, 1997; Parker, 2002; Fernández et al., 2005).

In this work we use visual information to learn, with a team of robots, descriptions of objects placed in a particular environment. Learning to recognize particular objects in an environment is important for robotics as it can be used for local and global localization tasks as well as for simple service tasks such as searching for objects in unknown places. Contrary to previous approaches, in our learning setting, the robots are not told the number or nature of the objects to be learned.

Vision is a primary source of perception in robotics and provides different features that can be used to classify objects. In general, using a particular set of features can be adequate for particular tasks but inadequate for other tasks. In this work, objects are characterized by two complementary features: (i) SIFT features (Lowe, 2004) and (ii) information about the silhouettes of objects. Other features could be used as well, but the main objective in this work is to show the different cases and possible confusions that can arise in the recognition of objects and merging of concepts, and how they can be addressed.

Numerous difficulties arise in robot teams when learning as well as sharing concepts that represent concrete objects. Some of these issues are discussed by Ye and Tostsos (1996) and include, how do robots represent their local views of the world, how is the local knowledge updated as a consequence of the robot's own action, how do robots represent the local views of other robots, and how do they organize the knowledge about themselves and about other robots such that new facts can be easily integrated into the representation. This article addresses the individual and collective representation of objects from visual information using a team of autonomous robots.

The rest of the paper is organized as follows. Section 2 reviews related work. Sections 3 y 4 describe, respectively, the stages of individual learning and collective learning of concepts. Section 5 describes our experimental results, and Section 6 provides conclusions and future research work.

## 2 RELATED WORK

Interesting experiments where physical mobile robots learn to recognize objects from visual information have been reported. First we review significant work developed for individual learning, and then we review learning approaches developed for robot teams.

Steels and Kaplan (2001) applied an instance-based method to train a robot for object recognition purposes. In this work objects are represented by color histograms. Once different representations have been learned from different views of the same object, the recognition is performed by classifying new views of objects using the KNN algorithm (Mitchell, 1997).

Ekvall et al. (2006) used different learning techniques to acquire automatically semantic and spatial information of the environment in a service robot scenario. In this work, a mobile robot autonomously navigates in a domestic environment, builds a map, localizes its position in the map, recognizes objects and locates them in the map. Background subtraction techniques are applied for foreground objects segmentation. Then objects are represented by SIFT points (Lowe, 2004) and an appearance-based method for detecting objects named Receptive Field Co-occurrence Histograms (Ekvall and Kragic, 2005). The authors developed a method for active object recognition which integrates both local and global information of objects.

In the work of Mitri et al. (2004), a scheme for fast color invariant ball detection in the RoboCup context is presented. To ensure the color-invariance of the input images, a preprocessing stage is first applied for detecting edges using the Sobel filter, and specific thresholds for color removal. Then, windows are extracted from images and predefined spatial features such as edges and lines are identified in these windows. These features serve as input to an AdaBoost learning procedure that constructs a cascade of classification and regression trees (CARTs). The system is capable of detecting different soccer balls in RoboCup and other environments. The resulting approach is reliable and fast enough to classify objects in real time.

Concerning the problem of collective learning of objects using robot teams there are, as far as we know, very few works. Montesano and Montano (2003) address the problem of mobile object recognition based on kinematic information. The basic idea is that if the same object is being tracked by two different robots, the trajectories and therefore the kinematic information observed by each robot must be compatible. Therefore, location and velocities of moving objects are the features used for object recognition instead of features such as color, texture, shape and size, more appropriate for static object recognition. Robots build maps containing the relative position of moving objects and their velocity at a given time. A Bayesian approach is then applied to relate the multiple views of an object acquired by the robots.

In the work of O'Beirne and Schukat (2004), objects are represented with Principal Components (PC) learned from a set of global features extracted from images of objects. An object is first segmented and its global features such as color, texture, and shape are then extracted. Successive images in a sequence are related to the same object by applying a Kalman filter. Finally, a 3D reconstructed model of an object is obtained from the multiple views acquired by robots. For that purpose, a Shape From Silhouette based technique (Cheung et al., 2003) is applied.

In contrast to previous works, in our method each member of the robot team learns on-line individual representations of objects without prior knowledge on the number or nature of the objects to learn. Individual concepts are represented as a combination of global and local features extracted autonomously by the robots from the training objects. A Bayesian approach is used to combine these features and used for classification. Individual concepts are shared among robots to improve their own concepts, combining information from other robots that saw the same object, and to acquire a new representation of an unnoticed object.

# 3 INDIVIDUAL LEARNING OF CONCEPTS

The individual concepts are learned on-line by a robot team while traversing an environment without prior knowledge on the number or nature of the objects to learn. The individual learning of concepts consists of tree parts: object detection, feature extraction and individual training.

Individual concepts of objects are represented by Principal Component (PC) over the information about the silhouettes of objects and Scale Invariant Features (SIFT). Learned concepts are shared among robots.

## 3.1 Object Detection

Robots move through an environment and learn descriptions of objects that they encountered during navigation. Objects are detected using background substraction. In this paper we assume a uniform and static background. We performed morphological operations (closing - errode) to achieve better segmentation. Once an object is detected, it is segmented and scaled to a fixed size, to make the global PC features robust to changes in scale and position.

## 3.2 Feature Extraction and Individual Training

The segmented objects are grouped autonomously by the robots in sets of images containing the same object. Robots assume that they are observing to the same object while it can be detected, and they finish to see it when they can not detect objects in the captured images. Only one object can be detected in an image at the same time. For each set of images, the robot obtains an individual concept that represents the object.

(a) (b) (c)

Figure 1: Examples of the silhouette (b) and average silhouette (c) of an object (a).

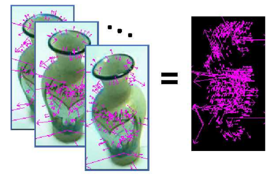

Figure 2: Examples of the SIFT features extracted from a set of images and the final set of SIFT features.

**Training using global features:** We applied Principal Component Analysis (PCA) over the average silhouettes that are automatically extracted from the set of images of a particular object. The average provides a more compact representation of objects and reduces segmentation errors. Figure 1 (a) shows an object used in the training phase, Figure 1 (b) shows its silhouette, and Figure 1 (c) illustrates the average silhouette obtained from a set of images that represent the object of Figure 1 (a). Once the robot has obtained an average silhouette, this is added by the robot to a set of known average silhouettes. After that, the robot uses PCA to reduce the dimensionality of all average silhouettes learned to get the PC features that represents them.

**Training using local features:** Each robot extracts local SIFT features of each image of the set of images, and groups them in a final set which contains all the different SIFT features that represent an object. In Figure 2 we show an example of the SIFT points obtained from a set of images of a *vase* and the final set of SIFT points obtained. The PC features and the SIFT features represent the individual concept of the observed object.

## 3.3 SHARING CONCEPTS

The concepts learned by robots are shared among them to achieve collective learning. This can be done off-line or on-line. In the case of collective off-line learning the robots share their individual concepts once they have learned all the training objects. On the other hand, in the collective on-line learning the robots share their individual concept as soon as a new object is learned.

## 4 COLLECTIVE LEARNING OF CONCEPTS

Collective learning of concepts enables robots to improve individual concepts combining information from other robots that saw the same object, and to acquire a new representation of an object not seen by the robot. Therefore, a robot can learn to recognize more objects of what it saw and can improve their own concepts with additional evidence from other robots.

A robot has to decide whether the concept shared by another robot is of a new object or of a previously learned concept. A robot can face three possibilities: coincident, complementary or confused information. The shared concepts are fused depending on the kind of information detected, as described below.

### 4.1 Pre-analysis of Individual Concepts

The concept learned by a robot is defined as follows:

$$C_k^i = \left\{Sil_k^i, SIFT_k^i\right\} \tag{1}$$

where $C_k^i$ is the concept $k$ learned by robot $i$, $Sil_k^i$ is the average silhouette, and $SIFT_k^i$ is the set of SIFT features that form the concept $k$.

In order to determine if a shared concept is previously known or not to a robot, it evaluates the probabilities that the PC features and SIFT features are previously known by the robot. The probability vectors of PC features calculated by robot $i$, $v_P^i$, indicate the probability that a concept shared by robot $j$, $C_k^j$, is similar to the concepts known by robot $i$, $C_1^i, \ldots, C_{numObjs}^i$, given the global features. $numObjs$ is the number of concepts of objects known by robot $i$. The process to obtain the probability vector PC is described as follows:

- A temporal training set of silhouettes is formed by adding the average silhouettes of concepts known by robot $i$ or actual robot, $Sil_1^i, \ldots, Sil_{numObjs}^i$, and the average silhouette of the shared concept $Sil_k^j$.

- The PCA is trained using the temporal set of average silhouettes. The projection of the average silhouettes know by robot $i$ is obtained as a matrix of projections, $matProys$. The projection of the average silhouette $Sil_k^j$ is obtained in a vector, $vectProys$.

- The Euclidean distance ($dE$) is calculated between each vector of the matrix $matProys$ and the vector $vectProys$ as shown in formula 2, i.e, we obtain the distance between all the projections already computed and the projection of the new silhouette.

$$dE_l^i = \sqrt{\sum_{r=1}^{nEigens} (matProys_{(l,r)} - vectProys_{(1,r)})^2} \tag{2}$$

where $nEigens$ is the number of eigenvectors used during the PCA training ($nEigens = numObjs^i - 1$), and $l$ is the index of the distance vector, where the maximum size of the vector $dE^i$ is $numObjs^i$.

- The distance value $dE^i$ is divided by a maximum distance value, $ThresholdMax$, determined experimentally to obtain a similarity metric also called the probability vector $v_P^i$ as shown in formula 3.

$$v_{P_l}^i = 1 - \frac{dE_l^i}{ThresholdMax} \tag{3}$$

If $dE_l^i$ is bigger than the $ThresholdMax$ value, then the probability will be fixed as shown in formula 4, which indicates that the projections of the object $j$ and the one of the object $i$ are completely different.

$$v_{P_l}^i = \frac{1}{numObjs} \tag{4}$$

The value of the SIFT similarity metric also called the probability vector SIFT at the position $v_{S_l}^i$, is obtained calculating the number of coincident SIFT, $n_{coin}$, between the individual SIFT concept $SIFT_l^i$ learned by robot $i$, and the individual SIFT concept $SIFT_k^j$ shared by robot $j$. If the number $n_{coin}$ is bigger than an average of coincidences determined experimentally, $AverageCoin$, then the probability will be fixed to $v_{S_l}^i = 1.0$, which means that both concepts contain the same local features SIFT. In other case, the probability will be calculated using equation 5.

$$v_{S_l}^i = \frac{n_{coin}}{AverageCoin} \tag{5}$$

The constant $AverageCoin$ represents the average of coincidences between two sets of SIFT points of the same object from different perspectives.

### 4.2 Analysis and Fusion of Individual Concepts

This section describes how to detect if the shared concept is coincident, complementary or confused, and how the individual concepts are fused to form collective concepts depending on the kind of detected concept.

#### 4.2.1 Coincident Concepts

A coincident concept is detected when two or more robots of the robot team learned individual concepts from similar views of the same object. A shared concept is classified as coincident if $v^i_{P_l} \geq \alpha$ and $v^i_{S_l} \geq \alpha$. That is, if both probabilities (PC and SIFT) of a previously learned concept are greater than a predefined threshold value ($\alpha$). If a shared concept is determined as coincident it is merged with the most similar known concept as follows:

**PCA fusion:** It is obtained by evaluating a new average silhouette from the average of the known $Sil^i_l$ and new $Sil^j_k$ silhouettes. After that, it is necessary to re-train the PCA substituting the concept $Sil^i_l$ with the new average silhouette which contains information of the concept learned by robot $j$.

**SIFT fusion:** It is obtained by adding the complementary SIFT points of concept $SIFT^j_k$ to the set of SIFT points of concept $SIFT^i_l$. Also, each pair of coincident SIFT points of both concepts is averaged in terms of position and their corresponding SIFT descriptors.

The main idea to fuse individual concepts is to improve their representation.

#### 4.2.2 Complementary Concepts

A concept $C^j_k$ contains complementary information if it differs with all known concepts by robot $i$, i.e., if both shape and local features are different to all known concepts by robot $i$, $C^i_1, \ldots, C^i_{numObjs}$. That is, if $v^i_P < \alpha$ and $v^i_S < \alpha$.

A complementary concept $C^j_k$ is fused with the collective concepts known by the robot $i$ as follows:

**PCA fusion:** The new average silhouette is added and the new PC concepts are obtained by re-training the PCA using the updated set of average silhouettes.

**SIFT fusion:** The new SIFT features are simple added to the current set of SIFT concepts known by the robot $i$.

#### 4.2.3 Confused Concepts

There are two types of confusion that can occur between concepts:

**Different shape and similar local features** (type 1): This type of confusion occurs when the new concept $C^j_k$ is complementary by shape, $Sil^j_k$, to all the concepts known by the robot $i$, $Sil^i_1, \ldots, Sil^i_{numObjs^i}$ but it is coincident by local SIFT features, $SIFT^j_k$, with at least one concept known by the robot $i$. That is, $v^i_{S_l} \geq \alpha$ and if $v^i_P < \alpha$.

**Similar shape and different local features** (type 2): This type of information occurs when concept $C^j_k$ is coincident by shape, $Sil^j_k$, to at least one concept known by the robot $i$, but it is complementary using its local SIFT features, $SIFT^j_k$. That is, if $v^i_{P_l} \geq \alpha$ and $v^i_S < \alpha$.

In both types of confusion, type 1 or type 2, there can be two options:

a) **Different objects:** Both concepts correspond to different objects.

b) **Same object:** Both concepts correspond to the same object but they were learned by robots from different points of view.

In our current approach, both types of confusions are solved as if it was a new object (complementary). The reason is that robot $i$ cannot distinguish between the two options (different objects or same object) using only the individual and the shared concepts, because it needs more information to solve the confusion to decide if it is about the same object or if it is about a new one. To solve the disambiguation, as future work each robot should built autonomously a map and locate its position in the map. In addition, for each learned object, robots will locate them in the map. If an object is confuse, a robot can move to the position of the object marked in the map to see the object from different perspectives in order to solve the conflict.

## 5 EXPERIMENTS AND RESULTS

We performed several experiments to demonstrate the proposed algorithm. In section 5.1, we show the results of a general experiment that demonstrates the main features of the proposed approach. In section 5.2 we present the accurracy of the collective concepts versus the individual concepts.

In these experiments we used a robot team of two homogeneous Koala robots equipped with a video camera of $320 \times 240$ pixels. For more than two robots our method is applied strighforward. The only difference is that robots will perform the pre-analysis and the analysis and fusion of individual concepts for each shared concept by the other members of the team.

### 5.1 Concept Acquisition and Testing

The mobile robots learn on-line a representation for several objects while following a predetermined trajectory without prior knowledge on the number or nature of the objects to learn. The idea of using preplanned trajectories instead of making the robots wandering randomly, is that we can ensure that robots will see an object at a time with its video camera, because it is an important point for the correct performance of our method. The pre-planned paths do not imply that

(a) (b) (c) (d) (e) (f) (g)

Figure 3: Training objects. a) *vase*, b) *water bottle*, c) *can*, d) *dolphin*, e) *soda bottle*, f) *bottle* and g) *cone*.

robots are going to see an object from the same point of view, scale or orientation, and that the trajectories of robots are always the same in the experiments, as will be seen later.

Each robot shares its individual concept as soon as it is learned to improve the representation of this concept or to include a new concept in the other robot. Figure 3 shows the training objects used in this experiment. As can be seen in the figure, some objects have the same shape but different texture, some have the same texture but different shape, some others are not symmetric in their shape. The objective of this experiment is to show the performance of the system to detect coincident, complementary and confused information under a wide variety of conditions.

Robot 1 (R1) learned during individual training concepts for: *dolphin*, *can*, *water bottle* and *vase*. Robot 2 (R2) learned individual concepts for: *vase*, *soda bottle*, *bottle* and *cone*. Note that some objects are learned by both robots while others are only learned by one robot.

While learning a new concept, each robot has to decide whether to fuse the current concept with a previously known concept or include it as a new one. Tables 1 and 2 show the probability vectors of the PC features based on shape ($v_P^1$) and of the SIFT features ($v_S^1$) obtained by Robot 1. Tables 3 and 4 show the probability vectors of the PC ($v_P^2$) and SIFT ($v_S^2$) features obtained by Robot 2. In these tables the coincident information is represented in bold.

We used the defined criteria in Section 4.2 to recognize coincident, complement or confused concepts, with $\alpha = 0.65$ as threshold value, and the probability vectors of Tables 1, 2, 3 and 4.

Tables 5 and 6 show the results of the analysis performed by each robot. As can be seen from these tables, each robot encountered the three types of possible information and fuse its concepts accordingly.

For instance, Table 1 shows how are the probabilities of objects of R1 affected using only PCA over shapes of objects, as both robots encounter and learn concepts while traversing the environment. In the first row, R1 learns about the concept *dolphin* and acquires

Table 1: Probability vectors PC ($v_P^1$) obtained by R1.

| | New (collective concepts R1) | | | | | | |
|---|---|---|---|---|---|---|---|
| Objects | *Dolphin* | *Vase* | *Can* | *Soda bottle* | *Water bottle* | *Bottle* | *Cone* |
| *Dolphin*$_{R1}$ | - | - | - | - | - | - | - |
| *Vase*$_{R2}$ | 0.19 | - | - | - | - | - | - |
| *Can*$_{R1}$ | 0.31 | 0.26 | - | - | - | - | - |
| *Soda bottle*$_{R2}$ | 0.36 | 0.28 | 0.58 | - | - | - | - |
| *Water bottle*$_{R1}$ | 0.43 | 0.28 | 0.53 | **0.73** | - | - | - |
| *Bottle*$_{R2}$ | 0.31 | 0.17 | 0,56 | 0.61 | 0.58 | - | - |
| *Vase*$_{R1}$ | 0.25 | **0.69** | 0.42 | 0.43 | 0.41 | 0.32 | - |
| *Cone*$_{R2}$ | 0.31 | 0.01 | 0.28 | 0.28 | 0.33 | 0.43 | - |

Table 2: Probability vectors SIFT ($v_S^1$) obtained by R1.

| | New (collective concepts R1) | | | | | | |
|---|---|---|---|---|---|---|---|
| Objects | *Dolphin* | *Vase* | *Can* | *Soda bottle* | *Water bottle* | *Bottle* | *Cone* |
| *Dolphin*$_{R1}$ | - | - | - | - | - | - | - |
| *Vase*$_{R2}$ | 0.09 | - | - | - | - | - | - |
| *Can*$_{R1}$ | 0.12 | 0.12 | - | - | - | - | - |
| *Soda Bottle*$_{R2}$ | 0.28 | 0.11 | 0.40 | - | - | - | - |
| *Water bottle*$_{R1}$ | 0.15 | 0.59 | 0.20 | 0.20 | - | - | - |
| *Bottle*$_{R2}$ | 0.08 | 0.15 | **0.65** | 0.04 | 0.12 | - | - |
| *Vase*$_{R1}$ | 0.16 | **1.00** | 0.23 | 0.10 | 0.08 | 0.09 | - |
| *Cone*$_{R2}$ | 0.05 | 0.28 | 0.43 | 0.10 | 0.14 | 0.09 | - |

it. In the second row, R2 then learns about *vase* and shares this concept to R1. The probability, according to the PCA features to be a *dolphin* is 0.19 (second row). R1 learns the object *can*, which has a probability of 0.31 to be a *dolphin* and a probability of 0.26 to be a *vase*, which was learned by R2 and shared to R1 (third row). In the fifth row, R1 learns about a *water bottle* but it confuses with the *soda bottle* learned and shared before by R2. As can be seen from Figure 3, both objects have the same shape and consequently the PCA features are not able to discriminate between these two objects. This is not the case for the SIFT features, which prevent R1 to consider it as the same object (as explained below). In the seventh row, R1 learns about *vase* which was already learned

Table 3: Probability vectors PC ($v_P^2$) obtained by R2.

| | New (collective concepts R2) | | | | | | |
|---|---|---|---|---|---|---|---|
| Objects | *Vase* | *Dolphin* | *Soda bottle* | *Can* | *Bottle* | *Water bottle* | *Cone* |
| $Vase_{R2}$ | - | - | - | - | - | - | - |
| $Dolphin_{R1}$ | 0.19 | - | - | - | - | - | - |
| $Soda\ bottle_{R2}$ | 0.28 | 0.36 | - | - | - | - | - |
| $Can_{R1}$ | 0.26 | 0.31 | 0.58 | - | - | - | - |
| $Bottle_{R2}$ | 0.17 | 0.31 | 0.61 | 0.56 | - | - | - |
| $Water\ bottle_{R1}$ | 0.29 | 0.44 | **0.73** | 0.54 | 0.58 | - | - |
| $Cone_{R2}$ | 0.01 | 0.31 | 0.28 | 0.28 | 0.43 | 0.33 | - |
| $Vase_{R1}$ | **0.69** | 0.25 | 0.43 | 0.42 | 0.32 | 0.41 | 0.01 |

Table 4: Probability vectors SIFT ($v_S^2$) obtained by R2.

| | New (collective concepts R2) | | | | | | |
|---|---|---|---|---|---|---|---|
| Objects | *Vase* | *Dolphin* | *Soda bottle* | *Can* | *Bottle* | *Water bottle* | *Cone* |
| $Vase_{R2}$ | - | - | - | - | - | - | - |
| $Dolphin_{R1}$ | 0.18 | - | - | - | - | - | - |
| $Soda\ bottle_{R2}$ | 0.11 | 0.28 | - | - | - | - | - |
| $Can_{R1}$ | 0.12 | 0.12 | 0.04 | - | - | - | - |
| $Bottle_{R2}$ | 0.15 | 0.08 | 0,04 | 0.64 | - | - | - |
| $Water\ bottle_{R1}$ | 0.59 | 0.15 | 0.20 | 0.20 | 0.12 | - | - |
| $Cone_{R2}$ | 0.09 | 0.05 | 0.10 | 0.43 | 0.09 | 0.14 | - |
| $Vase_{R1}$ | **1.00** | 0.16 | 0.10 | 0.23 | 0.09 | 0.08 | 0.12 |

and shared by R2, and in this case both concepts are merged.

To test the performance of the individual concepts and the collective concepts acquired by each robot, the concepts were used in an object recognition task. Each robot followed a predefined trajectory to recognize objects in the environment. The objects were detected by the robot team in the following order: *cone*, *water bottle*, *vase*, *bottle*, *soda bottle* and *dolphin*. Once an object is detected, the robot ($i$) evaluates its class using the PC ($v_P^i$) and SIFT ($v_S^i$) probability vectors and combines both probabilities using a Bayesian approach:

Table 5: Detected information by R1 for each own and shared individual concepts.

| Individual concepts | Related objects | $v_P^1$ | $v_S^1$ | Kind of info. |
|---|---|---|---|---|
| $Dolphin_{R1}$ | - | - | - | Complementary |
| $Vase_{R2}$ | All $l$ ($l = 1$ to $numObj^1$) | $v^1_{P(2,l)} < 0.65$ | $v^1_{S(2,l)} < 0.65$ | Complementary |
| $Can_{R1}$ | All $l$ | $v^1_{P(3,l)} < 0.65$ | $v^1_{S(3,l)} < 0.65$ | Complementary |
| $Soda\ bottle_{R2}$ | All $l$ | $v^1_{P(4,l)} < 0.65$ | $v^1_{S(4,l)} < 0.65$ | Complementary |
| $Water\ bottle_{R1}$ | *Soda bottle* | $v^1_{P(5,4)} = 0.73$ | $v^1_{S(5,i)} < 0.65$ | Confuse type 2 |
| $Bottle_{R2}$ | *Can* | $v^1_{P(6,l)} < 0.65$ | $v^1_{S(6,3)} = 0.65$ | Confuse type 1 |
| $Vase_{R1}$ | *Vase* | $v^1_{P(7,2)} = 0.69$ | $v^1_{S(7,2)} = 1.00$ | Coincident |
| $Cone_{R2}$ | All $l$ | $v^1_{P(8,l)} < 0.65$ | $v^1_{S(8,l)} < 0.65$ | Complementary |

$$P^i_{B_l} = \frac{v^i_{P_l} \times v^i_{S_l} \times P_u}{\left(v^i_{P_l} \times v^i_{S_l} \times P_u\right) + \left((1 - v^i_{P_l}) \times (1 - v^i_{S_l}) \times (1 - P_u)\right)} \qquad (6)$$

where $P_u$ is a uniform probability distribution ($P_u = \frac{1}{numObjs^i}$), $v_P^i = p(PC\ \ projection \mid Class = i)$, $v_S^i = p(SIFT\ \ matching \mid Class = i)$, $P_B^i$ is the Bayesian probability vector ($p(Class = i \mid PCprojection, SIFTmatching)$, and $l$ is the index of the Bayesian probability vector, where the maximum size of the probability vector is $numObjs^i$.

Figures 4, 5 and 6 show the average probabilities obtained during the object recognition task for each set of images of the same class, using the individual and collective learned concepts. The dotted bars indicate the classification errors. A classification error is produced when a robot classifies an unknown object with a probability $\geq 0.6$. The unknown objects for robots 1 and 2 are those which were not learned during their individual training.

The classification errors of Robot 1 in Figure 4 occur when the objects *cone*, *bottle* and *soda bottle* are classified as *dolphin*, *water bottle* and *water bottle*, respectively. The classification errors of Robot 2 occur when the objects *can*, *water bottle* and *dolphin* are classified as *bottle*, *soda bottle* and *soda bottle*,

Table 6: Detected information by R2 for each own and shared individual concepts.

| Individual concepts | Related objects | $v_P^2$ | $v_S^2$ | Kind of info. |
|---|---|---|---|---|
| *Vase*$_{R2}$ | - | - | - | Complementary |
| *Dolphin*$_{R1}$ | All $k$ | $v^2_{P_{(2,k)}} < 0.65$ | $v^2_{S_{(2,k)}} < 0.65$ | Complementary |
| *Soda bottle*$_{R2}$ | All $k$ | $v^2_{P_{(3,k)}} < 0.65$ | $v^2_{S_{(3,k)}} < 0.65$ | Complementary |
| *Can*$_{R1}$ | All $k$ | $v^2_{P_{(4,k)}} < 0.65$ | $v^2_{S_{(4,k)}} < 0.65$ | Complementary |
| *Bottle*$_{R2}$ | All $k$ | $v^2_{P_{(5,k)}} < 0.65$ | $v^2_{S_{(5,k)}} < 0.65$ | Confuse type 1 |
| *Water bottle*$_{R1}$ | *Soda Bottle* | $v^2_{P_{(6,3)}} = 0.73$ | $v^2_{S_{(6,k)}} < 0.65$ | Confuse type 2 |
| *Cone*$_{R2}$ | All $k$ | $v^2_{P_{(7,k)}} < 0.65$ | $v^2_{S_{(7,k)}} < 0.65$ | Complementary |
| *Vase*$_{R1}$ | *Vase* | $v^2_{P_{(8,1)}} = 0.69$ | $v^2_{S_{(8,1)}} = 1.00$ | Coincident |

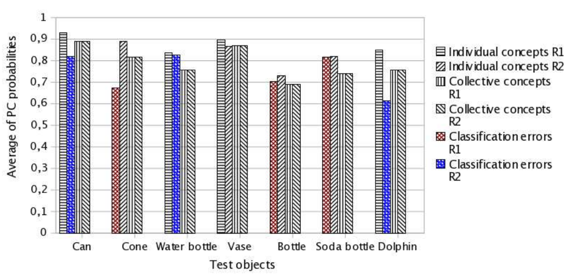


Figure 4: Average PC classification probabilities for the object recognition task using the individual and collective PC concepts.

respectively. For the *vase* there is no classification error because both robots learn individual concepts of it.

Although the probability bars presented in the previous figures show a higher probability for individual concepts than for collective concepts, in reality the collective concepts are more robust as they represent the probabilities considering a larger number of objects. This will be discussed in Section 5.2.

We show in Table 7 the precision of the object recognition task using the individual and collective concepts. The precision is presented in two ways, one considering the total number of objects and the other one taking into a count only the number of ob-

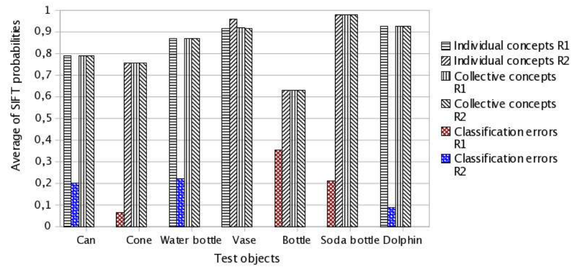


Figure 5: Average SIFT classification probabilities for the object recognition task using the individual and collective SIFT concepts. Any robot makes classification errors.

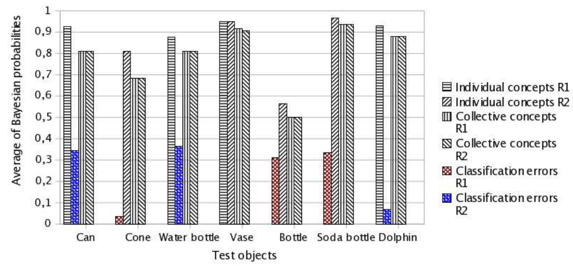


Figure 6: Bayesian classification probabilities which uses the Bayesian fusion of the PCA and SIFT classification probabilities.

jects used during the individual training (reported in parentheses). As can be seen the collective concepts produce a significantly better precision.

Table 7: Precision in the object recognition task using the individual and collective concepts acquired by each robot.

| | R1 | R2 | R1-R2 | R2-R1 |
|---|---|---|---|---|
| PCA | 55.69 % (100.00 %) | 49.98 % (94.82 %) | 86.15 % | 86.16 % |
| SIFT | 48.32 % (86.23 %) | 42.89 % (79.11 %) | 87.84 % | 87.84 % |
| Bayes | 52.59 % (94.20 %) | 51.68 % (81.54 %) | 80.73 % | 80.73 % |

## 5.2 Accuracy of the individual and collective concepts

In this section we compare the results of the individual concepts with that of collective concepts. In each experiment, a different set of objects was used, and both robots learned the same set of objects. Therefore, all the shared concepts were coincident, that is, robots learned both individually and collectively the same number of concepts. At the end of each experiment the robots learned four concepts that were proved by a test sequence. In Figure 7 we present the accuracy obtainded by the robots using the PC features of the individual and collective concepts in an object recognition task. Figure 7 shows the averages in accuracy of the number of images well classified under six experiments (*correct*), the average percentages of the number of non detected or non classified images (*no detected*), and the average percentage of false positives for each concept (*false* +). Figure 8 and 9 show, respectively, the accuracy obtained by the robots when using the SIFT vectors and the Bayesian approach.

As it can be observed in Figures 7, 8 and 9, the accuracy that indicate the quantity of well classified images (*correct*) using the collective concepts for the object recognition task, is in general better than the accuracy using the individual concepts. For PC, SIFT and Bayes there is an improvement in the accuracy up to 2.56 %, 13.79 % and **20.62 %**, respectively. This demonstrates that the collective concepts have better coverage than the individual concepts because they contain information acquired from different points of view, which allows a better recognition of test objects. Also, the percentages of the number of non detected images of collective concepts are smaller than the ones of the individual concepts. For PC, SIFT and Bayes there is a reduction in the percentage of non detected images of up to 2.56%, 13.79% and 20.62%, respectively. In Table 8 we present the average percentages of false positives for both, the individual and the collective concepts acquired by the robots. We conclude that the collective concepts have better quality than the individual concepts.

In general for both, the individual and the collective concepts, we observed an improvement in the accuracy when using the Bayesian approach. In Table 9 we present the average percentages of accuracy using the individual and collective concepts.

The average profit in the percentages of classification using the Bayesian approach using the collective concepts with regard to the individual concepts is of **14.63 %**.

Table 8: Average percentages of false positives.

| | PCA | SIFT | Bayes |
|---|---|---|---|
| Individual | 14.42 % | 0.64 % | 0.64 % |
| Collective | 13.14 % | 0.00 % | 0.00 % |

Table 9: Average percentages of accuracy.

| | PCA | SIFT | Bayes |
|---|---|---|---|
| Individual | 84.94 % | 67.88 % | 80.18 % |
| Collective | 87.18 % | 81.12 % | 94.81 % |

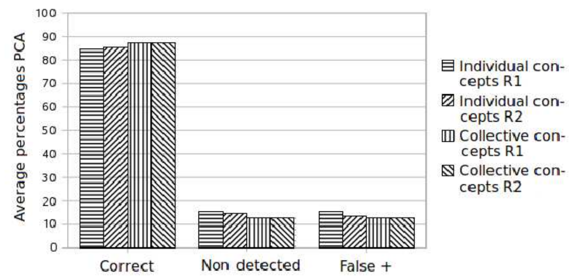


Figure 7: Accuracy in coverage using the part PC of concepts.

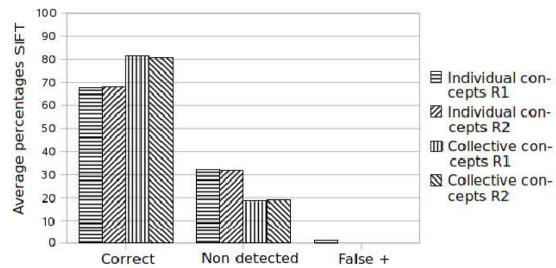


Figure 8: Accuracy in coverage using the part SIFT of concepts.

## 6 CONCLUSIONS AND FUTURE WORK

In this paper we have introduced a new on-line learning framework for a team of robots. Some of the main features of the proposed scheme are:

- The robots do not know in advance how many objects they will encountered. This pose several problems as the robots need to decide if a new seen object or shared concept, is of a previously learned concept or not.
- The representation of objects are learned on-line while the robots are traversing a particular environment. This is relevant for constructing autonomous robots.
- Three possible cases in which to merge concepts and how to merge them were identified.

The detection of coincident concepts avoids producing multiple concepts for the same object. The detection of complementary concepts allows to detect and learned unknown objects not seen by a particular robot. The detection of confused concepts allows to fuse information: 1) when the object have different shape and similar SIFT features, and 2) when the objects have similar shape and different SIFT features.

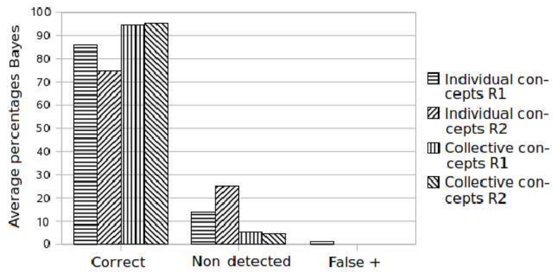


Figure 9: Accuracy in coverage using the Bayesian approach.

These cases are particularly difficult to deal with because the objects may be genuinely different or may be the same but seen from different points of view by the robots.

In general, the object recognition using the collective concepts had a better performance than using the individual concepts in terms of accuracy. This occurs because the collective concepts consider information from multiple points of view producing more general concepts.

As future work we propose to integrate schemes to object segmentation for dynamic environments. For instance, using an object segmentation based on distance as in Méndez-Polanco et al., 2009. Use a different set of features and identify possible conflicts between more that two kind of features. We also plan to incorporate planning of trajectories to autonomously allocate the environment among robots. We also plan to add strategies to solve some confusions in shared concepts by taking different views from these objects. Finally, we plan to incorporate our algorithm for robot localization and search of objects, and to test our work for robot teams with three or more robots.

## ACKNOWLEDGEMENTS

The first author was supported by the Mexican National Council for Science and Technology, CONACYT, under the grant number 212422.

# REFERENCES

Asada, M., Uchibe, E., Noda, S., Tawaratsumida, S., and Hosoda, K. (1994). Coordination of multiple behaviors acquired by vision-based reinforcement learning. *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 917–924.

Cheung, G. K. M., Baker, S., and Kanade, T. (2003). Visual hull alignment and refinement across time: A 3-D reconstruction algorithm combining Shape-From-Silhouette with stereo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, 2:375–382.

Ekvall, S., Jensfelt, P., and Kragic, D. (2006). Integrating active mobile robot object recognition and SLAM in natural environments. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pages 5792–5797.

Ekvall, S. and Kragic, D. (2005). Receptive field cooccurrence histograms for object detection. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '05)*, pages 84–89.

Fernández, F., Borrajo, D., and Parker, L. E. (2005). A reinforcement learning algorithm in cooperative multi-robot domains. *Journal of Intelligent and Robotic Systems*, pages 161–174.

Howard, A., Parker, L. E., and Sukhatme, G. S. (2006). Experiments with a large heterogeneous mobile robot team: Exploration, mapping, deployment and detection. *International Journal of Robotics Research*, 25:431–447.

Lowe, D. G. (2004). Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Matarić, M. J. (1997). Reinforcement learning in the multi-robot domain. *Autonomous Robots*, 4(1):73–83.

Méndez-Polanco, J. A., Muñoz-Meléndez, A., and Morales, E. F. (2009). People detection by a mobile robot using stereo vision in dynamic indoor environments. *Proceedings of the 8th Mexican International Conference on Artificial Intelligence (MICAI '09)*, 5845:349–359.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Enginering/Math.

Mitri, S., Pervolz, K., Surmann, H., and Nuchter, A. (2004). Fast color independent ball detection for mobile robots. *Proceedings of the IEEE International Conference Mechatronics and Robotics (MechRob '04)*, pages 900–905.

Montesano, L. and Montano, L. (2003). Identification of moving objects by a team of robots based on kinematic information. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '03)*, 1:284–290.

O'Beirne, D. and Schukat, M. (2004). Exploration and object recognition using cooperative robots. *Proceedings of the International Conference on Imaging Science, Systems and Technology (CISST '04)*, pages 592–598.

Parker, L. E. (2002). Distributed algorithms for multi-robot observation of multiple moving targets. *Autonomous Robots*, 3(12):231–255.

Steels, L. and Kaplan, F. (2001). AIBO's first words. The social learning of languaje and meaning. *Evolution of Comunication*, 4(1):3–32.

Wessnitzer, J. and Melhuish, C. (2003). Collective decision-making and behaviour transitions in distributed ad hoc wireless networks of mobile robots: Target-hunting. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J. T., and Ziegler, J., editors, *ECAL*, volume 2801 of *Lecture Notes in Computer Science*, pages 893–902. Springer.

Ye, Y. and Tsotsos, J. K. (1997). On the collaborative object search team: a formulation. In *Meets Machine Learning, Lecture Notes in Artificial Intelligence*, pages 94–116. Springer-Verlag.