Combining Wrapper and Filter Approaches for Learning Concepts from Images provided by a Mobile Robot

Nicolas Bredeche NICOLAS.BREDECHE@LIP6.FR
LIMSI - University Paris 11, BP 133, 91403 Orsay Cedex, France

Lorenza Saitta SAITTA@MFN.UNIPMN.IT
Università del Piemonte Orientale, Dipartimento di Scienze e Tecnologie Avanzate, Corso Borsalino 54, 15100 Alessandria, Italy

Jean-Daniel Zucker JEAN-DANIEL.ZUCKER@LIP6.FR
LIP6 - University Pierre Et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex, France

Abstract

To efficiently identify properties from its environment is an essential ability of a mobile robot who needs to interact with humans. Successful approaches to provide robots with such ability are based on ad-hoc perceptual representation provided by AI designers. Instead, our goal is to endow autonomous mobile robots (in our experiments a PIONEER 2DX) with a perceptual system that can efficiently adapt itself to the context so as to enable the learning task required to anchor symbols. Our approach is in the line of meta-learning algorithms that iteratively change representations so as to discover one that is well fitted for the task. The architecture we propose may be seen as a combination of the two widely used approach in feature selection: the Wrapper-model and the Filter-model. Experiments using the PLIC system to identify the presence of Humans and Fire Extinguishers show the interest of such an approach, which dynamically abstracts a well fitted image description depending on the concept to learn.

1. Introduction: Anchoring symbols, detecting and identifying objects

Recent works in both Robotics and Artificial Intelligence have shown a growing interest in providing mobile robots with the ability to interact and communicate with humans. One of the main challenges in designing such robots is to give them the ability to perceive the world in a way that is useful or understandable to us. One approach is to give the robot the ability to identify physical entities and relate them to perceptual symbols that are used by humans (to refer to these same physical entities). To perform this task, the robot has to ground these symbols to its percepts (i.e., its sensor data). Recently, the term of Anchoring (Coradeschi & Saffiotti, July 2000) has emerged to describe the *building and maintenance of the connection between sensor data and the symbols used by a robot for abstract cognition.* As a matter of fact, anchoring is an important issue for any situated robot performing abstract reasoning based on physically grounded symbols. Amongst others, anchoring plays an important role to communicate or relate to either other robots (Steels, 1999) or humans (Thrun et al., 1999).

There are tasks, such as object manipulation or functional imitation, where anchoring requires explicitly recognizing objects and localizing them in the three-dimensional space. Fortunately, such an *object recognition* task is not necessarily required to achieve anchoring. In applications such as human/object tracking, face and object identification, or grounded robot-human communication, *object identification* is enough. Informally, to recognize an object often requires from the robot both identifying from its percepts what an object is, and using a model of the object to localize it. This task has been studied for a few decades now and is known to be difficult in unknown environments (Stone, 1993). On the contrary, identifying the sole presence of an object is simpler. Moreover, there exist many easy to use and reliable descriptions for

characterizing the presence of an object. To identify the presence of a fire in a room, one does not have necessarily to recognize it. Smelling smoke, hearing cracks, feeling heat, seeing dancing shapes on a wall are different ways of identifying the presence of a fire. For an autonomous robot, the ability to identify objects is a first step towards more complex tasks. *Object detection* (detecting a fire) may be built by regularly checking whether the object is identified. Identifying objects is therefore a simple form of anchoring symbols (such as fire) to its percepts.

In this paper, we are concerned with a practical task, where a PIONEER 2DX mobile robot has to rely on its limited vision sensors to anchor symbols such as human being, mobile robot or fire extinguisher that it encounters while navigating in our laboratory. Anchoring is then used to support human/robot or robot/robot communication. For instance, an interaction may be engaged if a human being is identified, or a rescue operation may be initialized if a non-responding PIONEER 2DX is identified. Identifying a fire extinguisher may allow the robot to respond to a query formulated by a human. To design an autonomous robot, living in a changing environment such as our laboratory, with the identification ability described above is a difficult task to program. As such it is a good candidate for a Machine Learning approach, which may be easily recasted as a classical *concept learning task*. To teach the robot to anchor symbols using Machine Learning has proven successful (Klingspor et al., 1996). To use machine learning techniques, the designer has to both define learning examples and a representation language based on the robot percepts to describe them.

It is clear that a great part of the success of the learning task per se depends on the representation chosen (Saitta & Zucker, 2001a). Having an AI designer providing the robots with an adequate representation has a major drawback: it is a fixed, ad-hoc representation. Any change of setting (a museum instead of an AI lab) may require a new perceptual description. In order to overcome this drawback, our main objective is to endow an autonomous robot with the ability to dynamically abstract from its percepts different representations, well suited to learn different concepts. The intuitive idea is to have the robot explore the space of possible examples descriptions (with various colors, resolution, representation formalisms, etc.) so as to discover for each concept a well-fitted representation. The underlying intuition being that for anchoring the symbol human being a robot does not need the same visual resolution that might be necessary to identify a power-plug on a wall.

Section 2 presents a concrete setting in which this problem occurs and pinpoints why adapting ones representation may be useful to increase learning accuracy. In Section 3, related works about vision and anchoring in several research fields are quickly reviewed. Then, Section 4 explains our approach based on abstraction operators applied to visual information provided by the robot. Finally, a set of real world experiments describes the interest of such an approach and outlines the difference between two representations, each one fitted to a different concept (the presence of a human and the presence of a fire extinguisher).

2. Problem settings

The practical task we are concerned with is part of a wider project called Microbes (Picault & Drogoul, 2000), whose goal is to have a colony of eight robots co-habit with AI researchers. We aim at providing each PIONEER 2DX autonomous mobile robot with the ability to identify -but not recognize- objects or living beings encountered in its environment. Each robot navigates during the day and when resource are available it takes snapshots of its field of vision with its video camera. The snapshots are taken either randomly from time to time, or upon a specific human request [1]. At the end of each day, the robot may report to a supervisor and "ask" her/him what objects (whose symbols may or may not belong to a pre-defined lexicon) are to be identified on a subset of taken pictures[2]. It then performs a learning task in order to create or update the connection between sensory data and symbols which is referred to as the anchoring process. Figure 1 describes this process. The learning task associated to the anchoring is therefore characterized by a set of image descriptions and attached labels. It corresponds to a multi-class concept learning task.

A key aspect of the problem lies in the definition of the learning examples (the images) used by the robot during the anchoring process. In effect, a first step in any anchoring process is to identify (relevant) information out of raw sensory data in order to reduce the complexity of the learning task.

The PIONEER2DX mobile robot provides images thanks to its LCD video camera while navigating in the corridors. The images are 160×120 wide, with a 24 bits color information per pixel. Humans, robots,

[1] Thanks to *active learning* techniques, the robot may also take snapshots of scenes that appear to be interesting w.r.t. enhancing the detection accuracy of a known object (e.g. ambiguous images).

[2] Again *active learning* techniques may be used by the robot to select the most *informative* images.

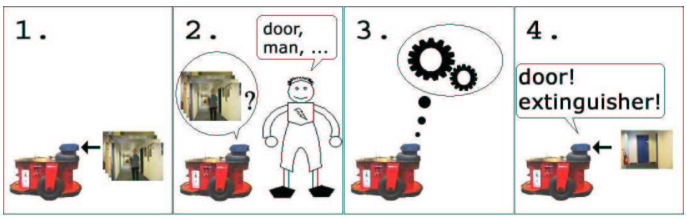

Figure 1. The four steps toward lexicon anchoring. As a first step, the robot takes a snapshot of its environment, and a supervisor labels it with the interesting content. The robot tries to associate the provided label(s) with its percept, and, after a number of such steps take place, it shall be able to autonomously label a new environment.

Figure 2. Two snapshots taken by the robot. Left: image labeled with "fire extinguisher" and "door". Right: image labeled with "human" and "door".

doors, extinguishers, ashtrays and other possible targets can be seen among the images as shown in Figure 2. All these possible targets, as they appear in the images, are of different shape, size, orientation and sometimes they are partially occluded. Finally, each image is labeled with the names of the occurring targets.

3. Changing the Representation of Images

3.1. Initial Perceptual Representation

We define the role of the robot's perceptual system as to extract *abstract percepts* out of *low-level percepts*, such as a set of pixels, from the video camera or sonar values. These abstract percepts provide a representation of the perceived world on which further computation will be based. They can be anything from sets of clustered colored regions to a matrix resulting from a Hough transform. The choice of a representation is motivated by finding a good trade-off that reduces the size of the search space and the expressiveness of the abstract percepts.

As mentioned in the previous section, the problem we consider is that of automatically finding a representation of a set of labeled images that is well adapted to the learning of concepts. Let us underline that our goal

Concept to learn	number of positive examples	number of negative examples	Size of file
Fire extinguisher	175	175	20,1 Mb
Human being	60	60	9,7 Mb

Table 1. The two learning sets of images associated to the concept "human" and "fire extinguisher"

is not to achieve the best performance on the particular learning task mentioned in the previous section. To obtain the best performance would require that experts in the field build an ad-hoc representation for each concept to learn. On the contrary, we are interested in having a robot find by itself the good representation, so that, if the context changes or the concept to learn is different, it has the ability to discover by himself the good level of representation. We therefore consider the representation provided by the sensors as an *initial* representation.

From the robot's point of view, each pixel from the camera is converted into a *low-level percept*. In the initial image representation, where each pixel is described by its position (x,y), its *hue* (the tint of a color as measured by the wavelength of light), its *saturation* (term used to characterize color purity or brilliance) and its *value* (the relative lightness and darkness of a color, which is also refereed to as "tone"). The initial description of an image is therefore a set of 19200 (160 x120 pixels) 5-uple (x,y,h,s,v). Each image is labeled by symbols following the process described in Section 2. The positive examples of a given concept (e.g., "presence of a fire extinguisher") to learn correspond to all images labeled positively for this concept. The negative examples are the images were the target concept does not appear. The number of examples for two of the concepts we considered are given on Table 1.

The initial representation of images, consisting of hundreds of thousands of pixels, is clearly a too low-level representation to be used by Machine Learning algorithms.

We have chosen the multiple-instance representation to represent information from the images. Within the *multiple instance* setting, objects are represented by *bags of feature vectors*. Feature vectors are also called *instances* and as in the traditional setting features may be numeric as well as symbolic. The size of a bag b is noted $\sigma(b)$ and may change from one object

to another. Its instances are noted $b_1 \ldots b_{\sigma(b)}$. The multiple-instance representation is an in-between representation, more expressive than feature-vector but for which efficient algorithms do exist, compared to algorithms used with a relational description.

3.2. Dimensions of abstraction

In the perspective of automatically exploring the set of possible representations of an image, we propose to identify particular operators and to experiment with them. There are countless operators that could be applied to an image hoping for more accurate learning. Operators changing the *contrast*, the *resolution*, the *definition* [3] are all possible candidates.

To improve the learning of concepts, we are interested in transformation that are abstractions in the sense that they decrease the quantity of information contained in the image(Saitta & Zucker, 2001b). The two main dimensions for abstraction that we shall study are:

- the resolution of the image, i.e., its granularity.
- The structure of the image, i.e., the smallest individually accessible portion of the image to consider, be it a pixel or a complex region.

For each of these dimensions, we have defined an abstraction operator: respectively, *associate* and *aggregate*. The associate operator consists in replacing a set of pixels with a unique (mega)pixel that has for its (h,s,v) values the average of the pixels that were associated. This operator is a built-in operator for the robot as it corresponds to a particular *sub-sampling*. The aggregate operator consists in grouping a set of pixels to form a region or a pattern. This operation is also referred to as "term construction" in the literature(Giordana & Saitta, 1990). The region does not replace the pixels it is composed of, and therefore the resolution or granularity of the image is not changed. What changes is the structure of the image. The aggregate operator may be either data-driven (this is the case for region growing algorithms) or model based.

[3]The contrast measures the rate of change of brightness in an image; high contrast suggests content consisting of dark blacks and bright whites; medium contrast implies a good spread from black to white; and low contrast implies a small spread of values from black to white. The resolution is a measure of the proportion of the smallest individually accessible portion of a video image to the overall size of the image. The higher the resolution, the finer the detail that can be discerned. The definition corresponds to the clarity of detail in an image and is dependent upon resolution and contrast (Drury, 1990)

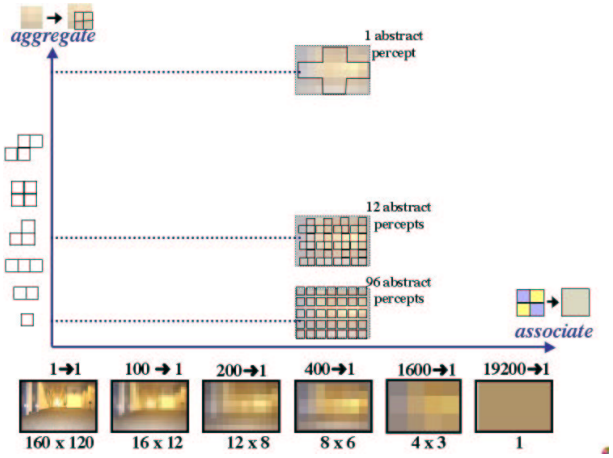

Figure 3. The space of image representation obtained by applying the associate operator (changing the *resolution*) and the aggregate operator (changing the *structure*). Three examples of representations obtained after having applied the aggregate operators and the associate operator.

For already mentioned reasons of efficiency required by the use of a robot we have considered an aggregate operator that is applied to contiguous pixels forming a particular shape. Figure 3 depicts the space of representation changes associated to these two operators.

We shall refer to the initial concept as *low-level percepts*[4]. The ones obtained after applying the abstract operators will be referred to as *abstract percepts*, since they will be used as percepts for further processing. For clarity, abstract percepts obtained by applying the aggregate operator will be referred to as *s-percepts* and ones obtained by applying the associate operator will be referred to as *r-percepts*[5] .

4. Automatically changing the representation for learning

In the previous section two abstraction operators to change the representation of images were presented. The parameter of the associate operators we have considered is the number of pixels that are associated to form a (mega)pixel. The parameter of the aggregate operator is the pattern or region structure.

With respect to the learning task described in Section 2, a key issue is to analyze the impact of representa-

[4]We will depart from the traditional use of the term "ground" for the initial representation(Sacerdoti, 1974) as for a robot the notion of "grounding" corresponds to another notion

[5]s-percept, as in *structural percept* and r-percepts as in *resolution percept*

tion changes on learning. The main question is related to the choice of one operator and its parameters. In Machine Learning, the abundant literature on feature selection shows that approaches fall in two broad categories: the `wrapper` and the `filter` approach. Intuitively, the `wrapper` approach uses the performance of the learning algorithm as a heuristic to guide the abstraction. The `filter` approach uses a a priori knowledge to select appropriate abstractions. In Machine Learning the combination of this two approaches has not yet been explored. In the following, we present how these two approaches can be combined. As it is an approach that attempt to learn from the learning process itself it is also refered to as a *meta-learning* approach.

Since resolution changes the information contained in an image, we have used an information-based filter approach to choose an a-priori good resolution to start with for learning. The filter approach is therefore used to explore the horizontal dimension of the space of Figure 3. To explore different possible patterns to apply with the aggregate operator, we have used a `wrapper` approach. The PLIC system is the result of the combination of these two approaches.

4.1. A Filter-Based exploration of the abstract spaces

Let us suppose that a snapshot with N pixel is taken, and that we are looking for the localization of an object with r pixels. Let us suppose, for now, that each pixel is either activated (has an intensity value 1) or not activated (has an intensity value 0). The level with N pixels is the more detailed one, whereas the maximum confusion is reached when the activity value is averaged over all the N pixel, obtaining thus a single, large percept. Let τ be the homogeneous value of this largest percept. Let, moreover, σ be the standard deviation of the intensities over the N pixels.

If groups of k pixels are associated into a single percept, its average intensity will be computed from the intensities of the component pixels, and the object we are looking for will appear more or less blurred, depending on how many activated pixels are included in the k ones. Let us assume that the object can still be discovered if its average intensity is greater than $\alpha \times \tau$, where α is a number greater than one, which depends on the sensor sensibility. For instance, if we can distinguish objects whose average intensity differs by the global average τ by p standard deviations, we will have: $\alpha = 1 + (p\sigma/\tau)$.

Let us now consider a generic percept containing an association of k pixels. Depending on the position of this percept w.r.t. the object, its average intensity will change. Among all the positions, there will be at least one that reaches a maximum of intensity, and will allow the object to be optimally individuated. If $k \nmid r$, the optimal new percept is included in the object picture, and its average intensity will be 1. If $k > r$, the optimal percept will include r pixels with intensity 1 and (k r) pixels whose average intensity can be set equal to τ. Then, the percept average intensity will be:

$$I(k) = [r + (kr)\tau]/k \quad (1)$$

In order to locate the object, it must be: $I(k) \geq \alpha\tau$

From the above condition we obtain a maximum number of pixels for the new percept:

$$k \leq r(1-\tau))/(\tau(\alpha - 1)) \quad (2)$$

It is clear the intensity of the new percept tends to τ when k tends to N.

The above reasoning can be extended to values associated to the pixels different from a binary intensity. For example, in the application considered in this paper, one of the averaged value can be the color hue. In this case, we notice that the red color of the fire estinguisher, for instance, can be detected by the system at a resolution (8×6), which does not allow the color to be detected by the human eye.

4.2. A Wrapper-based exploration of the abstract spaces

Once a resolution has been selected by the filter-based approach described in the previous sub-section, the wrapper-based component of the PLIC system explores different image structure iteratively. PLIC is a reformulation tool that acts as a wrapper according to given rules in order to find the best granularity and structure for describing the images. In these experiments, PLIC acts as follow : an initial structure is chosen, and the image is reformulated in a multiple-instance representation using this structure; then, the concepts are learnt using this representation. Based on the results on cross-validation [6] of the learning algorithm, a new structure is devised by PLIC. As for now, the search for a good structure is done in an exhaustive manner from the simplest one (i.e., one pixel at the chosen resolution) and exploring all the connected shapes of k pixels before increasing k. The following figure is a synthesis of this wrapper approach. The multiple instances rule learner RIPPERMI(Chevaleyre & Zucker, ECML2001)

[6] a widely used data-oriented evaluation of the learning generalization error that consists in dividing the learning set into a learning set and a training set

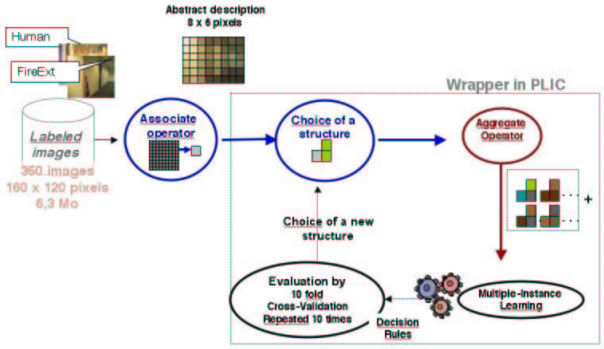

Figure 4. The PLIC system Wrapper component.

was used on the descriptions obtained from these images with a ten-fold cross validation. Moreover, each experiment is repeated 10 times in order to get a good approximation of the results. RIPPERMI returns a set of rules that covers the positive examples. PLIC interacts with RIPPERMI in order to evaluate and create descriptions.

5. Experiments

5.1. Experimental Setup

To evaluate the interest of abstracting visual percepts from a Machine Learning point of view, a number of different experiments have been carried out. The experiments presented are based on the images acquired by a PIONEER2DX mobile robot. The attributes used for image description are : *hue, saturation, value* for each pixel, and *hue, saturation, value* and corresponding *standard deviations* for each r-percept. The targets (be it a human or a fire extinguisher) as they appear in the images are different in shape, size, orientation and are sometimes partially occluded. Labeling with the names of the occurring targets was done by a supervisor (as explained in Section 2), and a noisy set of labels was produced as well (wrong labels were given on purpose). Two sets of experiments are presented, the first one illustrates the impact of the operator aggregate used by the wrapper-based component of PLIC and the second the impact of the associate operator used by its filter-based component. For each of these experiments the results for the concept "human" and "fire extinguisher" are given.

5.2. Evaluating automatic changes of granularity

The results obtained by the system PLIC are presented in Tables 3 and 2 below. Figures 6 and 5 respectively plot the evolution of the learning accuracy depending on the resolution chosen for the extinguisher class and the human class. These experiments on granularity illustrates the impact of changing the resolution but do not use the filter-based exploration described in section 4.1. The idea is to explore all possible granularity and choose the best according to the experiments.

Resolution	Accuracy in %	std.dev	time (s)
$1x1^7$	64,02	1,29	0,03
4x3	59,11	1,09	0,13
8x6	62,47	1,27	0,79
16x12	67,54	1,23	4,03
32x24	66,45	1,45	19,12
64x48	65,48	1,28	87,33

Table 2. PLIC results on learning the "human" concept

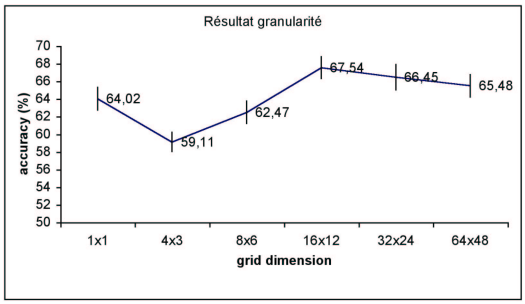

Figure 5. Experiments with the "human" concept

The evolution of the learning accuracy for the extinguisher class vs. the resolution shows that more complex resolutions are much more appropriate that the simpler ones (i.e., the histogram representation). As a matter of fact, the accuracy is enhanced by 12.51%, from 62.51% to 75.02%. However, this improvement is not linear over the set of possible resolutions, since the accuracy is better for the original 1×1 resolution than for the 4×3 resolution. This can be explained by the

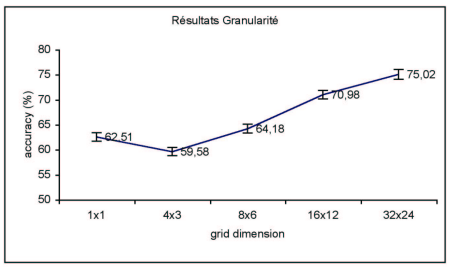

Figure 6. Experiments with the "fire extinguisher" concepts

Resolution	Accuracy %	std.dev	time (s)
1x1	62,51	0,85	0.39
4x3	59,58	0,82	1.62
8x6	64,18	0,89	9.22
16x12	70,98	0,86	37.68
32x24	75,02	0,96	128.31

Table 3. PLIC results on learning the "fire extinguisher" concept

fact that the histogram representation is somewhat fitted to capture some relevant information for image description (Stricker & Swain, 1994). As a consequence, using a more complex representation may cause the accuracy to decrease since the enhanced search space makes it more complicated to learn while it may not help to better discriminate the target concept.

Resolution enhancement proves to be profitable as soon as a 8×6 resolution and the accuracy improves in a nearly linear way from this point. Now, if we bring our attention to the evolution of the learning accuracy for the human class, we can see that the learning curve is about the same, with a smaller amplitude, except that at some point the accuracy reaches a maximum (67.54% with the 16×12 resolution), and then starts to decrease. This should tend to prove that the 16×12 resolution is the most fitted resolution for this concept, in this environment, given what examples have been observed. More complex resolutions cannot achieve a better representation for this concept, or at least the tradeoff between expressiveness and complexity is not relevant anymore. As a matter of fact, we can extrapolate to say that if the resolution is enhanced, accuracy should tend to reach 50% at some point, that is random prediction.

However the general shape of the accuracy curves for both concepts seem to share the same properties about the trade off between expressiveness and complexity, experiments shows that the best fitted granularity depends on the concept to learn.

5.3. Experiments on automatic changes of Structure

PLIC's wrapper tool was used to generate up to nine different s-percept's structural configurations that are shown on figure 7. Each structural configuration is then applied from every single r-percept to generate a learning sets based on a 8×6 resolution of each of the images of the image database, which is not the best resolution for either concepts but provides a common basis for evaluating learning accuracies. According to

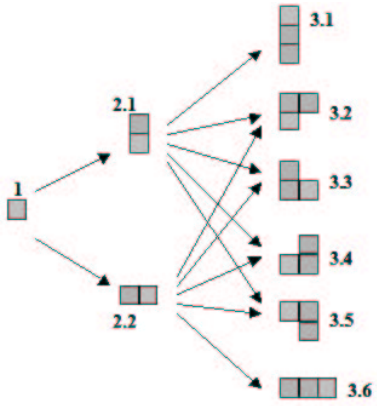

Figure 7. The nine structural configurations generated by the Wrapper

the heuristic used by PLIC, a given structural configuration is evaluated only if the corresponding learning accuracy is better than at the previous step, *and* if at least one attribute of each of the embedded r-percept appears in the decision rules produced by the learner (eg. *in figure 7, the configuration '3.1' will be evaluated only if configuration '2.1' matches these conditions*). Tables 5 and 4 respectively show the results for the human class and the fire extinguisher class.

Results from the experiments show that in both case, the highest accuracy is achieved by one of the most complex structural configurations (64.81% for the extinguisher class, 71.56% for the human class). The benefits of structural reformulation is more relevant for the human class with a 3.5% gain than for the extinguisher class, where it is nearly useless. This can be explained by the intrinsic properties of both concepts: detecting a human may require expressing relations between parts (e.g. *head on body*) while an extinguisher is often viewed as a uniform rectangle red shape in the environment.

In these experiments, this structural configuration is quite simple because of the low number of abstract percepts considered. However, human occurrence detection was achieved by the robot more than 7 times out of 10 thanks to less than 48 s-percepts.

6. Related works

The problem setting described here have much in common with that of content-based image retrieval. As a matter of fact, results from popular approaches to image retrieval such as global color histogram comparison, which are known to achieve good classification results(Stricker & Swain, 1994), or region-based similarity(Wang, 2001), which provides an efficient way to

Structure Id	Accuracy %	std.dev	time (s)
1	64,18	0,89	10,43
2.1	62,04	0,715	7,475
2.2	63,795	0,83	10,185
3.1	64,335	0,695	9,02
3.2	64,405	0,755	8,655
3.3	63,21	0,715	9,465
3.4	63,46	0,75	9,475
3.5	64,81	0,77	9,615
3.6	64,175	0,79	12,08

Table 4. PLIC results on learning the "fire extinguisher" concept and changing structure

Structure Id	Accuracy %	std.dev	time (s)
1	68,06	1,35	1,44
2.1	70,645	1,375	1,785
2.2	65,665	1,355	2,27
3.1	66,575	1,24	2,095
3.2	66,48	1,38	2,36
3.3	67.01	1,335	2,37
3.4	68.945	1,24	2,235
3.5	71,56	1,17	2,235
3.6	64,87	1,36	2,58

Table 5. PLIC results on learning the "human" concept and changing structure

compare images using spatial properties between segmented regions, are relevant for this problem.

Image retrieval is concerned with classifying an image based on its content, without precisely identifying the location of the target concept in the image. Approaches to image retrieval include model-based classification, images description using Fourier transforms, wavelets(Wang, 2001), etc. A popular approach is based on matching sets of connected color regions between images (Hsieh & Fan, 2000). The goal is to find instances of a given spatial configuration between regions extracted from the images (i.e., using a region growing algorithm). The main drawbacks of this approach are the imprecision of region growing techniques, and the cost of the matching phase between undirected planar graphs, representing sets of connected regions. However, as mentioned previously good classification results also also achieved by simply comparing the global color histogram of each image (Stricker & Swain, 1994). In mobile robots, image classification into categories is used for creating landmarks or for navigating. In this case, image retrieval using global color histograms is particularly well fitted because it classifies quickly the whole image.

Our problem differs with that of image retrieval because we are concerned with checking if there is a specific property hidden in the image. As a matter of fact, the environment of the robot provides very similar images where global variations are not bounded to the target concept while image retrieval is about finding globally similar images among a set of very different images. Moreover, we intend to create a set of rules which is known to be much faster to apply than any similarity measure, leading to nearly costless image classification which can easily be implemented in a real-time operating mobile robot.

7. Conclusion

In this paper we have addressed the problem of using automatic abstraction of visual percepts by an autonomous mobile robot to improve its ability to learn *anchors*(Coradeschi & Saffiotti, July 2000). This work finds its application in a real-world environment within the MICROBES multi-robots project (Picault & Drogoul, 2000), where anchors provides a basis for communication between the PIONEER 2DX robot and its human interlocutor. In the approach we proposed the robot starts with the initial low-level representation of the images it perceives with its LCD video camera, and iteratively changes their representation so as to improve the learning accuracy. Between the low-level pixel representation and a global histogram representation there is an immense space of possible representations. To explore part of this abstract space of representation we have identified two operators. A first one changes the resolution and loose information by averaging the colour of squares of pixels. A second one that groups pixels without changing the resolution.

To guide the exploration of the space of possible abstractions, we have combined in the PLIC system two approaches, one based on a priori consideratyions, and one using the learning results. From a Machine Learning point of view, this architecture corresponds to the combination of the two widely used approaches in feature selection: the Wrapper-model (Kohavi & John, 1998) and the Filter-model (Kohavi & Sommerfield, 1995). The set of experiments that have been conducted show that both operators do impact on the learning accuracy. It is interesting to notice that the best resolution and structure (sort of coordinates in the abstract space) found by the system depends of the concept. Since less low-level information is required to detect the presence of a human than a fire extinguisher, it is not surprising that the optimum resolution is different. It is also clear that as the num-

ber of examples increase, different reformulation might perform better. Creating high-level abstract percepts does not only improve accuracy, it makes object detection faster for the robot. This is true as long as the abstraction process does not itself takes too much time. This is a known trade-off in the field of abstraction (Giunchiglia, 1996). As a matter of fact, abstracting regions by using region growing algorithm (Rehrmann & Priese, 1998) was a candidate abstract operators but its computation is too costly for online detection.

This study shows that for learning anchors, an approach that periodically searches for the most accurate representation, given the examples at hand, is a promising direction. Moreover, it appears that for each anchor that needs to be learnt, different abstractions might be more appropriate. These findings, although preliminary, raises several questions with respect to the robot architecture. The search for a better representation should be triggered by a decrease of performance of the acquisition of new examples ? How to compare the application of operators that change the resolution and operators that change the structure ? A central question for any lifelong learning system, integrating abstraction abilities, is to decide whether to continue to *exploit* its current representation or *explore* new representations at the risk of loosing resources if no better ones is found.

References

Chevaleyre, Y., & Zucker, J.-D. (ECML2001). A framework for learning rules from multiple instance data. *Proc. European Conference on Machine Learning (ECML2001).*

Coradeschi, S., & Saffiotti, A. (July 2000). Anchoring symbols to sensor data: preliminary report. *Proceedings of AAAI-2000.* Austin, Texas.

Drury, S. (1990). *A guide to remote sensing.* Oxford: The Kluwer International Series on Information Retrieval, 11.

Giordana, A., & Saitta, L. (1990). Abstraction: a general framework for learning. *Working notes of the AAAI Workshop on Automated Generation of Approximations and Abstraction* (pp. 245–256). Boston, MA.

Giunchiglia, F. (1996). Using abstrips abstractions where do we stand ? *Artificial Intelligence Review, 13*, 201–213.

Hsieh, I., & Fan, K. (2000). Color image retrieval using shape and spatial properties. *ICPR00, Vol.I: pp 1023-1026.*

Klingspor, V., Morik, K., & Rieger, A. D. (1996). Learning concepts from sensor data of a mobile robot. *Machine Learning, 23*, 305–332.

Kohavi, R., & John, G. (1998). The wrapper approach. *Feature Selection for Knowledge Discovery and Data Mining, H. Liu and H. Motoda (eds.), Kluwer Academic Publishers, pp33-50..*

Kohavi, R., & Sommerfield, D. (1995). Feature subset selection using the wrapper method: Overfitting and dynamic search space *International Conference on Knowledge Discovery and Data Mining.*

Picault, S., & Drogoul, A. (2000). The microbes project, an experimental approach towards open collective robotics. *Proc. of the 5th International Symposium on Distributed Autonomous Robotic Systems.* Springer-Verlag Tokyo Inc.

Rehrmann, V., & Priese, L. (1998). Fast and robust segmentation of natural color scenes. *Asian Conference on Computer Vision.* Hongkong, China.

Sacerdoti, E. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence, 5*, 115–135.

Saitta, L., & Zucker, J.-D. (2001a). A model of abstraction in visual perception. *Applied Artificial Intelligence. 15(8): 761-776.*

Saitta, L., & Zucker, J.-D. (2001b). A model of abstraction in visual perception. *Applied Artificial Intelligence, 15*, 761–776.

Steels, L. (1999). The talking heads experiment. volume 1. words and meanings. *Antwerpen.*

Stone, J. (1993). Computer vision: What is the object? *Prospects for AI, Proc. Artificial Intelligence and Simulation of Behaviour. Birmingham, England..* IOS Press, Amsterdam. pages 199–208.

Stricker, M., & Swain, M. (1994). The capacity and the sensitivity of color histogram indexing. *Technical Report 94-05, Communications Technology Lab, ETH-Zentrum.*

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Hhnel, D., Rosenberg, C., Roy, N., Schulte, J., & Schulz, D. (1999). Minerva: A second generation mobile tour-guide robot. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).*

Wang, J. Z. (2001). *Integrated region-based image retrieval.* Oxford: The Kluwer International Series on Information Retrieval, 11.