A Multimodal Attention Mechanism for Autonomous Mobile Robotics

Raúl Arrabales, Agapito Ledezma and Araceli Sanchis

***Abstract*— Whatever the mission of an autonomous mobile robot is, attention is a helpful cognitive capability when dealing with real world environments. In this paper we present a novel control architecture which enables an integrated and efficient filtering of multiple modality sensory information. The concept of *context* is introduced as the set of criteria that determines what sensory information is relevant to the current mission. The proposed attention mechanism uses these contexts as a mean to adaptively select the constrained cognitive focus of the robot within the vast multimodal sensory space available. This approach for artificial attention is tested in the domain of autonomous mapping.**

***Index Terms*— Physical agents, Attention, cognitive modeling, mobile robotics.**

I. Introduction

DESIGNING an autonomous control system for a mobile robot implies a decision on what inputs will be handled and what repertory of actions can be executed at any given time. The option of considering all the available sensory information as input for the core control of the robot is usually both unnecessary and extremely expensive in computational terms. Analogously, not all possible robot behaviors are appropriate at any given time. Instead of considering all physically plausible behaviors, the robot control system should take into account its current situation and assigned mission in order to build a shorter list of eligible behaviors. A simplistic definition of artificial attention can be drawn from the problem described above. Hence, let us say that an efficient artificial mechanism for attention would solve the problem of filtering relevant sensory information and selecting relevant behaviors.

According to the former definition, we need to specify what does *relevant* mean in terms of implementing an efficient attention mechanism. Relevant sensor data and relevant behaviors are those that could be both useful to accomplish the mission and adapted to the world in which the robot is situated. Attention has been typically applied to artificial vision systems taking the human visual attention mechanisms and its related eye movement control (foveation) as inspiration [6]. Visual attention has been extensively applied in robotics, e.g. [2]. However, much less effort has been put in pure multimodal attention mechanisms [7]. Usually attention mechanisms for robots focus in great degree in visual sensory information; nevertheless, some salient examples incorporate data from other sensors in the attention mechanism. For instance, laser range finders [9]. In this work we present a purely multimodal attention mechanism, in which vision could be eventually incorporated, but has not been used for preliminary testing. Instead, bumpers and sonar range finders have been applied.

In the next sections we discuss the implementation of an attention mechanism able to fulfill the requirement of selecting relevant sensorimotor information. Section II covers the definition of the attentional contexts that are used to form sets of sensory and motor data. Section III is dedicated to explain how the proposed mechanism allows the integration of different modality sensory information into the same context. The incorporation of the attention mechanism into a three-layer control architecture is described in section IV. Section V illustrates the application of the proposed technique to the domain of autonomous mapping. Finally, we conclude in section VI with a discussion of the benefits and possible areas of application of the attention mechanism in the field of cognitive robotics.

II. Definition of Attentional Contexts

Our proposed artificial attention mechanism is inspired in the concept of *context* as defined in the Global Workspace Theory (GWT) [5]. The GWT is a cognitive account for consciousness, and therefore it covers attention as a key characteristic of conscious beings. However, the GWT do not provide any algorithmic description of attention but just a metaphorical explanation. A theater spotlight simile is used to represent the focus of consciousness. This spotlight illuminates only a small part of the scene, which is considered the conscious content of the mind. The scene is actually built upon the subject's working memory. The movement of the spotlight, i.e. the selection of contents that will be used for volition and action, is directed by unconscious contextual systems. The aim of the work described in this paper is to design and test an implementation of such contextual systems, which are able to adaptively direct attention toward the interesting areas of the robot sensorimotor space.

From the point of view of perception, contexts are sets of percepts retrieved from the sensors. Percepts are considered the minimal information units obtained by the robot sensory

Raúl Arrabales is with University Carlos III of Madrid.
E-mail: raul.arrabales@uc3m.es

Agapito Ledezma is with University Carlos III of Madrid.
E-mail: ledezma@inf.uc3m.es

Araceli Sanchis is with University Carlos III of Madrid.
E-mail: masm@inf.uc3m.es

machinery [3]. Therefore, a sensory context is a complex percept composed of related single percepts. From the point of view of behavior, contexts are sets of actions available for execution. Hence, we can define behavioral contexts as compositions of related actions. In order to generate an efficient robot behavior, both sensory contexts and behavioral context have to be adaptively generated.

A. Context Criteria

We have designed the process of context formation as the application of predefined criteria in order to calculate the degree of relation between the potential elements of a given context. Basically, a context should be constructed in a way that it can become a meaningful representation of the reality, i.e. situatedness must be enforced by a proper definition of both sensory and behavioral contexts. The very basic factors that need to be considered in the correct representation of robot situation in the world are time and location. Nevertheless, other factors can be considered depending on the problem domain and internal state representation richness.

In the work described here only time and location have been considered as criteria for context formation. The time criterion refers to the exact moment at witch a stimulus is perceived. Therefore, it should be taken as an important criterion to relate one percept to another. Given that different sensors and their associated device drivers can take different time intervals to process the sensory information, a mechanism for time alignment is required. It has been demonstrated that such a time alignment mechanism is present in biological brains [10, 13]. Although visual and auditory stimuli are processed at different speeds, the time gap between different processed signals, whose physical originators were acquired at the same time, is automatically removed by the brain [14].

Location is another fundamental criterion for context formation as the representation of the position of objects in the world is a requirement for situatedness. Furthermore, the location of an object relative to the robot body (or any other reference frame) is required for generating adaptive behaviors. The relative location of any element in the sensory world is necessary for the integration of complex percepts; additionally, it allows the selection of a given direction of attention toward the most relevant places. The presence of space coding neurons and the use of reference frames (like somatotopic or head-centered) has been demonstrated in the mammal brain [8, 4].

Following the principles discussed above, we have used time and location as fundamental contextualization criteria for the formation of:

- *Sensory contexts* as composition of single percepts, and
- *Behavioral contexts* as composition of simple actions.

In order to generate these contexts, both single percepts (which are built from data packages obtained from sensors) and simple actions (which are defined as part of the robot control system) are required to incorporate estimated time and location parameters (see Fig. 1). In our proposed architecture there are two modules designed to calculate the time and location parameters: the *Timer* module maintains a precision clock (1 millisecond resolution) that represents robot's current execution age, and the *Propioception* module that maintains all the required information to calculate the exteroceptive sensors position. This information is necessary to estimate the relative location of an object or event detected by an exteroceptive sensor. The time and location parameters provided by the Timer and Propioception modules are used by the two preprocessor modules in charge of generating the single percepts and simple actions. The *Sensor Preprocessor* takes a given sensor reading as input, then calculates the relative position of the source of the reading and the instant when it took place using the information provided by the Timer and Propioception, and finally, it creates a single percept packing together the sensor reading with its contextualization information. The *Action Preprocessor* takes as input an action generated by the Self-Coordination module (this module and the way it works is described elsewhere [3]), and applies the same approach as in the Sensor Preprocessor in order to build the Simple Action representations.

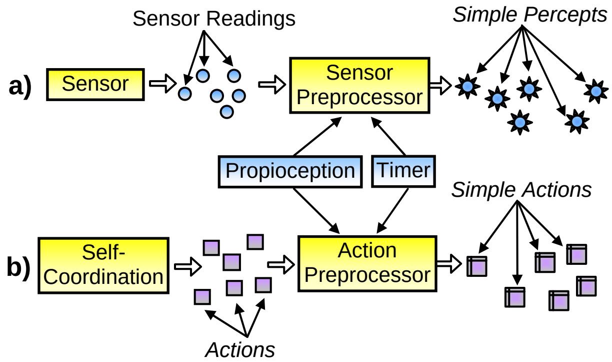

Figure 1. Creation of single percepts and simple actions.

More parameters should be added to single percepts and simple actions if other contextualization criteria are to be applied. In the work described in the present paper, the following parameters have been used:

- ***Timestamps:*** two different timestamps are recorded in single percepts. The first timestamp is set when the sensory data is collected from the sensor. Usually this timestamp is directly assigned by the sensor hardware and retrieved in the control system thought the sensor driver. The second timestamp is set when the percept is actually used in the control system. The time span between these two timestamps can be significant when a sensor is incessantly notifying readings and there is not enough onboard processing power to dispatch all the incoming data. Actually, the time span value can be used to discard too old sensory data which are not significant to the current robot state. Similarly, two timestamps are logged in the case of simple action. The first one is set when the simple action is created and enqueued in the control system. The second timestamp is set when the action enters the core execution cycle, i.e. when the action is actually dequeued and dispatched (begins physical execution). The time span between these two timestamps can be used to detect delays in the execution queue and eventually abort too old actions.

- ***J-Index***: for the representation of the location parameter of both single percepts and simple actions we have decided to use the robot body center of mass as reference frame. The term *J-Index* refers to a structure able to represent or map the relative position of an object or event within a biological brain [1]. We have adapted and enhanced the original definition of the *J*-Index representation with the aim of representing both the relative position and relative dimensions of the object. Hence, our *J*-Indexes are implemented as a composition of several *n*-dimensional vectors. The main vector is called the *j* referent vector, and is used to calculate the relative position of the geometrical center of the percept's source or the geometrical target of an action. Depending on the nature of the sensor that is reporting the sensory data, more positional vectors can be calculated in order to estimate the size of the percept (examples for sonar range finder and bump panel arrays are described below).

The timestamp parameters are easily acquired using the robot's control system precision timer. However, the *J*-Index parameters require more elaboration, particularly in the case of movable sensors. In the case discussed here, we have used a Pioneer 3DX robot (see Fig. 2a) with fixed position sensors: a frontal sonar array (see Fig. 2c) and frontal and rear bump panels (see Fig. 2b). In the experiments that we have carried out so far, *J*-Indexes have been calculated for sonar readings and bump panels contact and release notifications. The *J*-Indexes are calculated as a function of the transducer (fixed) position and orientation (relative to the robot front).

Although the *J*-Index parameter can be primarily represented by a three-dimensional vector, for the problem of 2D mapping a two-dimensional *j* referent vector can be considered, where $(X,Z) = (0,0)$ represents the subjective reference frame of the robot (see Fig. 2b and 2c). Nevertheless, a *Y* coordinate (height) is usually calculated even though it is not used to generate the 2D floor plan representation.

The calculation of the *j* referent vector is different depending on the sensor. In the case of bump panels, as they are located at angles around the robot (see Fig. 2b), the *j* referent vector is calculated using equation (1). Where, *BR* is the bump panel radius, i.e. the distance from the center of mass of the robot to the bumper contact surface (see Fig. 2b). BA is the bump panel angle to the front of the robot (Pioneer 3 DX bump panels are located at angles -52º, -19º, 0º, 19º, and 52º). *BH* is the height at which the bumpers are mounted.

$$j = (X, Y, Z) = \begin{pmatrix} BR * Cos(BA), \\ BH, \\ BR * Sin(BA) \end{pmatrix} \quad (1)$$

Additionally, two more vectors are calculated to be associated to a bumper percept: the left-*j* referent and the right-*j* referent (see Fig. 3). These two vectors represent the dimensions of the percept (the width assigned to the collision).

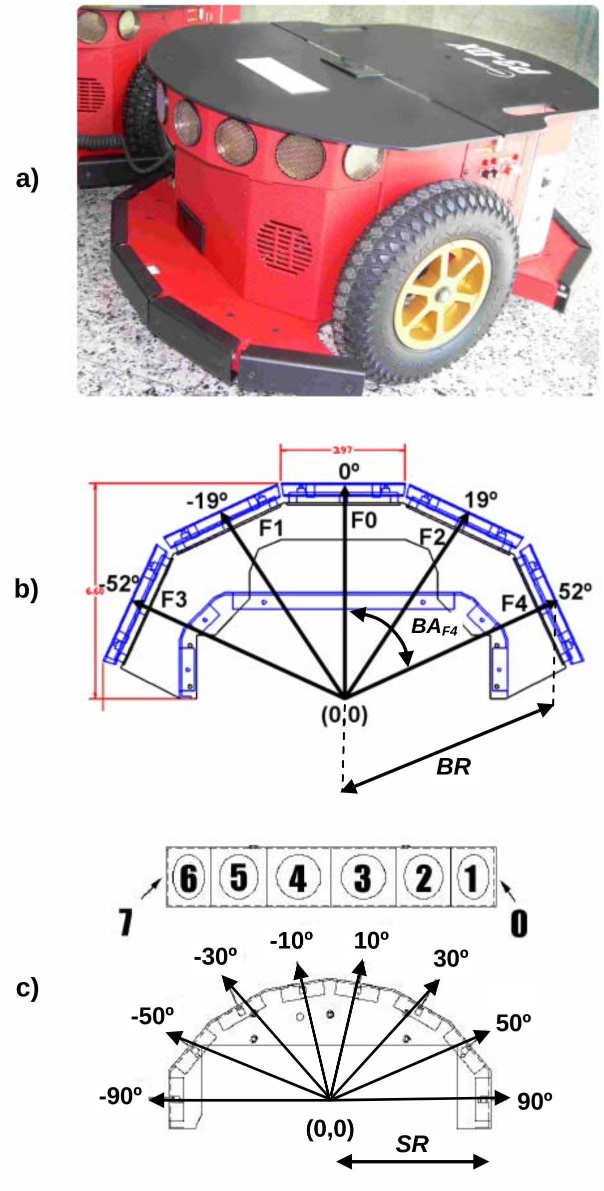

Figure 2. MobileRobots Pioneer 3 DX Robot, frontal bumper panel, and frontal sonar ring.

In order to calculate the *j* referent vector corresponding to a given sonar reading, Equation (2) is used. Note that the calculation of *j* referent vectors is dependent on the type of sensor being considered.

$$j=(X,Y,Z)=\begin{pmatrix}(R+SR)*Cos(SA),\\ SH,\\ (R+SR)*Sin(SA)\end{pmatrix} \quad (2)$$

Where, R is the maximum range measured by the sonar transducer, SR is the distance from the center of mass of the robot to the sonar transducer, and SA is the angle at which the particular sonar transducer is located. Note that sonar transducers are located at angles -90º, -50º, -30º, -10º, 10º, 30º, 50º, and 90º to the front of the robot (see Fig. 2c). Therefore, each transducer is able to measure the free space available within a three-dimensional 15º wide cone (this cone aperture corresponds to the *SensComp* 600 transducer).

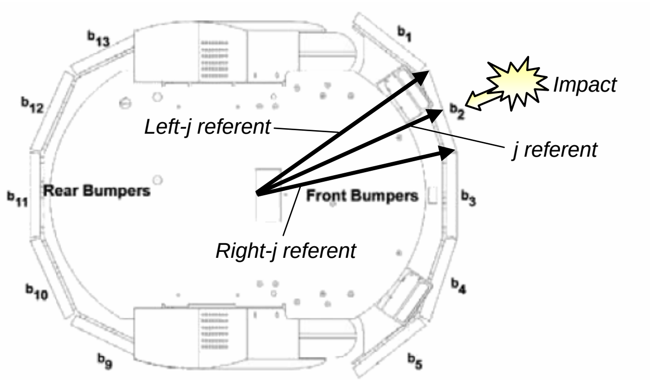

Figure 3. Vectors calculated to build the *J*-Index of a single bump panel contact percept.

Taking into account that the ultrasonic beams emitted by the sonar transducers take the form of a symmetric three-dimensional cone, at least one additional *j* referent vector has to be calculated in order to estimate the dimensions of the single transducer sonar percept, i.e. the open space perceived in front of that particular sonar transducer. The main *j* referent vector calculated using Equation (2) represents the cone bisector. Additionally, two more vectors: the left-j referent vector and right-j referent vector represent the lateral 2D boundaries of the percept (see Fig. 4).

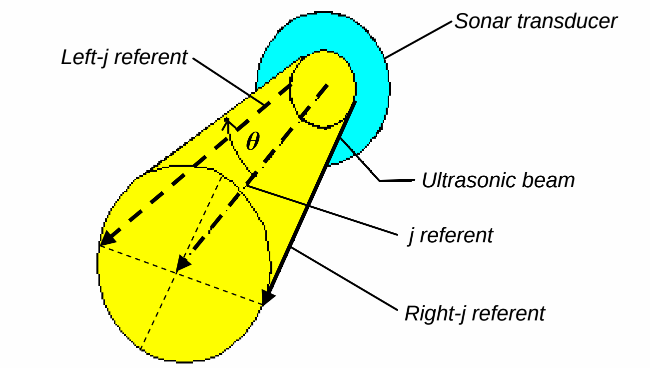

Figure 4. Vectors calculated to build the *J*-Index of a single sonar transducer percept.

The representations of *J*-Indexes for both sonar and bumpers have been designed as described above with the aim of implementing an attention algorithm. Although some of the calculated reference vectors are expendable, they are useful to pre-calculate the regions of the world affected by a given percept. Besides, this representation is also particularly useful for the subsequent task of 2D mapping.

B. Actions Context Composition

As Single Percepts and Simple Actions include the contextualization parameters (timestamps and *J*-Indexes) it is straightforward to calculate similarity distances between them. Therefore, contexts can be defined based on the dimensions of relative time and relative location. Each sensory context is used to build a representation structure called *complex percept* (see Fig. 5a). Complex percepts enclose the required information to represent the meaning of the associated sensory context as required by the subsystems of the autonomous control system. As behavioral contexts are formed they may trigger the generation of the corresponding complex behaviors, which are representations that enclose sequences of actions specified by the behavioral context (see Fig. 5b). In the present work, the behavioral context formation has been oversimplified because the only available actuator is the Pioneer 3DX differential drive. Two basic operations have been defined for the control of the differential drive:

- ***RotateInPlace***: this operation takes an angle in degrees as input parameter (positive values mean counterclockwise rotation) and triggers the robot rotation in position until it completes the consigned angle.
- ***MoveStraight***: this operation takes a speed in meters per second as input parameter (positive values mean move forward) and triggers the robot movement towards the current heading (or backwards for negative speed values).

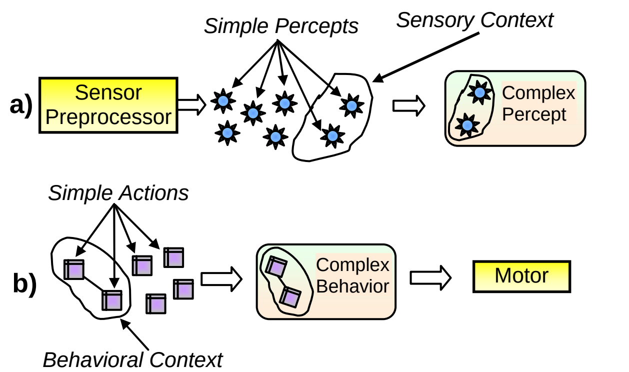

Figure 5. Formation of complex percepts and complex behaviors.

Attending to the relative direction specified by the attention mechanism (a composition of *J*-Indexes representations), an angle parameter is calculated for the

RotateInPlace operation in order to set the robot heading towards the object that "called the robot's attention". Also, a speed parameter is calculated as a function of the distance to the object. This means that the typical minimum behavioral context is formed by a sequence of simple actions like a *RotateInPlace* operation followed by a *MoveStraight* operation.

III. Multimodal Integration

Combining multiple monomodal sensory data sources is a typical problem in mobile robotics, also known as multisensory integration or sensor data fusion [11]. Actually, in the present work we are also approaching the problem of fusing propioceptive and exteroceptive sensor data. Neuroscientists refer to the binding problem [12], as the analogous problem of how to form a unified perception out of the activity of specialized sets of neurons dealing with particular aspects of perception. From the perspective of autonomous robot control we argue that the binding problem can be functionally resolved by applying the proposed contextualization mechanism.

A. Monomodal Context Formation

Taking the bump panel percepts as example, we can illustrate how a sensory context gives place to a monomodal complex percept. Using the aforementioned criteria, time and location, if the bumper driver of our robot reports contact events in bump panels *b2*, *b3*, and *b4* simultaneously (see Fig. 6), a context is automatically created if these three independent notifications have close enough timestamps. Therefore, the three single percepts are associated by a temporal context. Additionally, as *b2*, *b3*, and *b4* are located side by side, the corresponding contact percepts *J*-Indexes will indicate proximity, thus forming an additional spatial context that again associates these three single percepts. The newly created complex percept, which is a composition of three single percepts, also holds a representation of a *J*-Index. This complex percept *J*-Index is calculated as a function of the reference vectors of the former single percepts (note that Fig. 6 depicts with solid lines the *J*-Index referent vectors of the formed complex percept, and dashed lines represent the referent vector of the old single percepts).

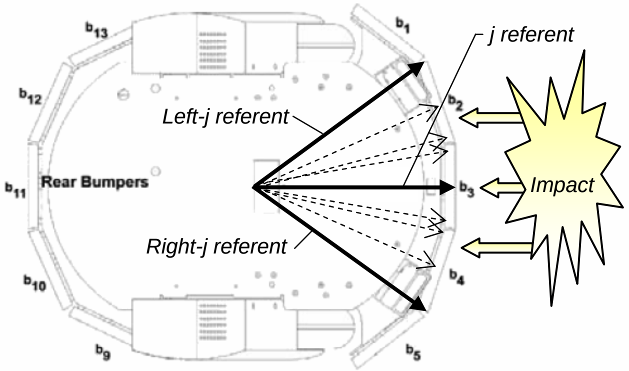

Figure 6. Vectors calculated to build the *J*-Index of a complex bumper contact percept.

The way the *J*-Index of a complex percept is calculated depends on the nature (shape, dimensions, etc.) of the single percepts that take part in the context that gave place to it. The composition of *J*-Indexes is trivial when all the single percepts belong to the same modality (as illustrated in Fig. 6). However, the composition can be complex when several different modalities are involved.

B. Multimodal Context Formation

Focusing on the mentioned fundamental criteria for contextualization (time and location), all percepts, independently of their modality, can be compared with each other, thus allowing a simple mechanism to create perceptual contexts. The key point is that selected criteria for context formation must be common to all available sensory modalities. The contexts formed following this method can have significant meanings. For instance, *"all objects within the reach of the robot"* (context formed applying the criterion of location and estimating that the relative location is below a given threshold, like the robotic arm reach distance in this case), or *"all events that took place between five and ten minutes ago"* (context formed applying the criterion of time and estimating that the relative timestamp of the events fall within the given interval). Similarly, more criteria could be used in order to build more specific contexts. Note that the formation of these sorts of contexts can involve different modality sensors, like laser range finders, sonar range finders, camera sensors, etc.

A common application of multimodal sensory information fusion is the disambiguation or refutation of contradictory sensor data. In the case under study in this paper, contradictory information happen to be processed when the sonar transducers fail to detect a sharp solid corner (the ultrasonic beams are diverted, and do not come back to the transducer, failing to provide a realistic range measurement). In such a scenario, the last resorts are the bumpers. When the robot base is too close to the sharp corner, bumpers will contact the obstacle and notify single percepts, which in turn will become complex percepts. However, during the process of complex percepts formation, potential contradictory information has to be handled. The time criteria for context formation will associate the roughly simultaneous readings from both sonar and bumpers. But, in the case of a bad sonar reading the single percepts available are not consistent. Therefore, a policy has to be established in order to build a significant complex percept out of conflicting single percepts. A single but effective approach is to apply a level of confidence to each sensor modality depending on the situation. In the case described here, we have just assigned more confidence to bumper contact notifications than sonar measurements.

The attention mechanism proposed in this work is designed to be highly dynamic and configurable. Following the principles described above, as many contexts can be created as criteria have been defined, and one or more contexts can be used to form a complex percept. The concrete definition of criteria and context is to be selected based on the specific problem domain. Nevertheless, the necessity of integrating this attention mechanism into a control architecture is always present, independently of the problem domain.

IV. Layered Attention

Typically, autonomous robot control architectures are structured in layers. Each layer usually represents a different level of control, from lower reactive levels to higher deliberative levels. In the case of the proposed attention mechanism, a three level control architecture called CERA (*Conscious and Emotional Reasoning Architecture*) has been considered (see Figure 7) [3].

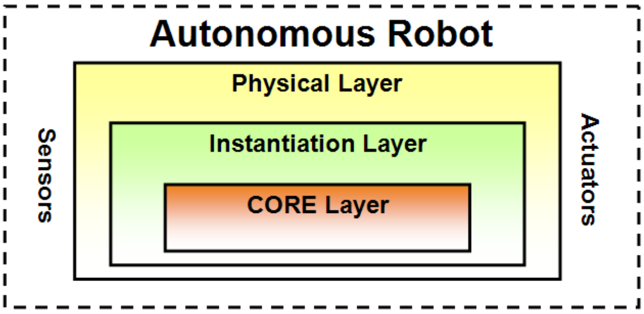

Figure 7. CERA Control Architecture Layers.

CERA Physical Layer provides the required functionality in order to interface with the robot hardware. In other words, it provides access to physical (or simulated) sensors and actuators. Additionally, as the CERA architecture has been designed to host the proposed attention mechanism, the physical layer is also in charge of calculating the *J*-Indexes of percepts and actions. From the point of view of the attention mechanism, the CERA Physical Layer is the domain of Single Percepts and Simple Actions. As the Physical Layer is specific to a given hardware it has to be changed or adapted if the underlying physical robot is replaced by a different model. The composition of percepts and actions forming Complex Percepts and Complex Actions takes place in the CERA Instantiation Layer. This is the place where mission-specific contexts are to be applied, and therefore mission-specific complex percepts and behaviors are generated. As the Instantiation Layer is designed specifically for a given problem domain it can be replaced by a different problem instantiation without changing the existing Physical and Core layers. Finally, the CERA Core Layer is where a machine consciousness model is implemented based on several modules that represent higher cognitive functions. One of these functions related to consciousness is attention. The attention module implemented in the Core Layer is designed to activate the most appropriate contexts at any given time.

At the level of the CERA Core Layer learning mechanisms could be applied in order to improve the attention selection technique. Moreover, the attention mechanism is to be integrated with other Core Layer modules, like memory and self-coordination managers in order to use the required related information for the activation of appropriate contexts in the instantiation layer.

V. Paying Attention to Mapping

Autonomous mapping of unknown office-like environments has been selected as preliminary problem domain for the testing of the proposed attention mechanism. It provides a valid real world scenario where the sensors and actuators described above can be used alone to achieve the mission goal: to obtain an approximate floor plan of the surroundings. The simulated scenario that we have initially configured for the testing provides perfect odometry (there is no noise induced in the robot wheel encoders); therefore, we have focused on the problem of mapping, neglecting the localization estimation problem for the time being. Figure 8 shows a screen capture of the simulated environment we have used for initial testing.

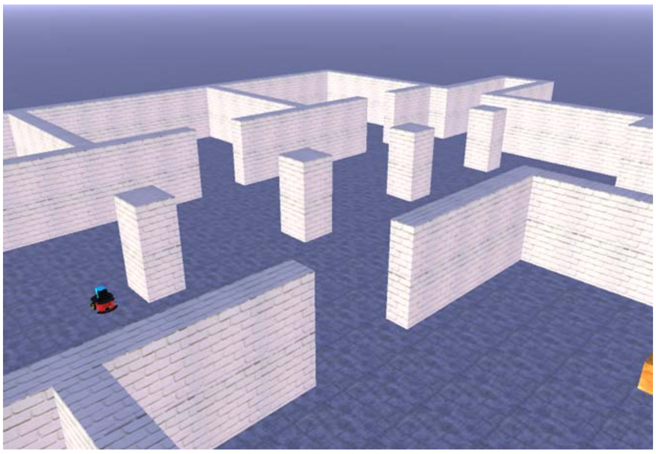

Figure 8. Simulated indoor environment.

One of the objectives of the proposed attention mechanism is to offer an effective policy for selecting the next action (or complex behavior) as part of the robot's main control loop. In the case of unknown environment exploration, spatial contexts are defined in order to estimate the best heading that the robot should take. A specific CERA Instantiation Layer has been coded with the aim of representing the particular complex percepts that are required for the mapping mission. Concretely, sonar single percepts that represent an obstacle are combined, using their *J*-indexes as combination criteria, into complex percepts that symbolize walls. Figure 9 shows an illustration of part of the internal state maintained by the Instantiation Layer. Wall complex percepts are represented with black solid lines, and white areas represent free space.

Based on the internal map representation, which is continuously updated with new sensed percepts, the attention module calculates a set of possible regions of interest, i.e. areas where the robot should pay attention to. Following the same design principle as explained above for the single and complex percepts, a *j* referent vector is calculated for each region of interest. Taking into account that minimizing exploration time is a requirement of the autonomous mapping mission, attention should be focused on those areas which have not been previously visited. Therefore, region of interest *j* referents are periodically calculated as a function of available unexplored space around the robot.

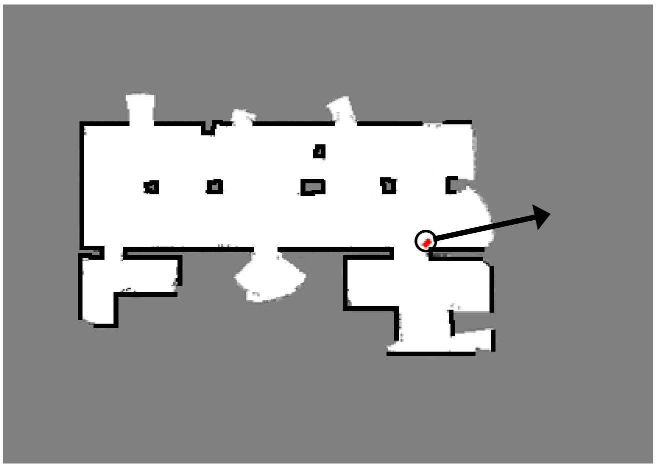

Figure 9. Map automatically generated by the robot.

The attention module performs a 360º scan, applying 45º steps, over the map generated so far. Note that these are not real sonar scans physically performed by the robot, but just a logical scan performed over the internal robot map representation. This internal representation of a map is built using the sonar range data as the robot wander around the environment. The example illustrated in Figure 9 shows that a 22.5º inclination *j* referent has been selected. This vector indicates that a region of interest exists in that direction.

There is a key difference between the *j* referent vectors that we defined for percepts and actions, and the *j* referent vector used for region of interests. While the former are relative to the robot center of mass and current heading, the latter are absolute in terms of map orientation. As our robot is not equipped with a compass, north is arbitrarily set up to the initial robot heading. Every time a sonar scan in used to update the map, a reference frame conversion is applied from the relative reference system of the robot to the absolute reference system used in the map representation. Conversely, when the *j* referent vector of a region of interest is selected by attention, it is converted to the robot's relative reference frame before building the corresponding complex behavior: rotate in place until the heading matches the region of interest direction, and then move straight.

VI. CONCLUSION

A novel attention mechanism for autonomous robots has been proposed and preliminary testing has been done in the domain of unknown environment mapping. The integration of the attention cognitive function into a layered control architecture has been demonstrated. Additionally, the problem of multimodal sensory information fusion has been addressed in the proposed approach using a generic context formation mechanism. The preliminary results obtained with the simulator show that this account is applicable to classical mobile robotics problems like autonomous mapping. Nevertheless, moving to a real world environment will imply dealing with the problem of imperfect odometry [15]. In such a scenario our proposed attention mechanism has to be integrated into an SLAM (Simultaneous Localization and Mapping) system. This work is currently underway, and the same described mechanism is expected to be used. As a final map cannot be used as internal representation of the real world environment, a typical probabilistic map should be used instead.

The system described in this paper is work in progress. More complex attentional contexts (and therefore more contextual criteria) have to be defined in order to face other problem domains. Perception is well covered for sonar range finder and bumpers. Nevertheless, the proposed technique can be applied to other sensors and actuators. Additional development would be required in the CERA Physical Layer in order to add other sensors like laser range finders or vision sensors. The definition of behavioral contexts and complex behaviors should be also enhanced to cope with more complex actuators and to generate more efficient behaviors.

ACKNOWLEDGMENT

This research has been supported by the Spanish Ministry of Education and Science under project TRA2007-67374-C02-02.

REFERENCES

[1] Aleksander, I. and Dunmall, B. Axioms and Tests for the Presence of Minimal Consciousness in Agents. *Journal of Consciousness Studies, 10,* 4-5 (2003).

[2] Arrabales Moreno, R. and Sanchis de Miguel, A. Applying machine consciousness models in autonomous situated agents. *Pattern Recognition Letters, 29,* 8 (2008), 1033-1038.

[3] Arrabales, R., Ledezma, A. and Sanchis, A. Modeling Consciousness for Autonomous Robot Exploration. In Proceedings of the International Work-Conference on the Interplay between Natural and Artificial Computation (*IWINAC)*. 2007.

[4] Avillac, M., Deneve, S., Olivier, E., Pouget, A. and Duhamel, J. Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience., 8,* 7 (2005), 941-949.

[5] Baars, B. J. *A Cognitive Theory of Consciousness.* Cambridge University Press, New York, 1993.

[6] Burt, P. J. Attention mechanisms for vision in a dynamic world. *9th International Conference on Pattern Recognition,* (1988), 977-987. Vol. 2.

[7] Crespo Mariño, J. L. *Una Arquitectura de Atención Distribuida para Agentes con Sensorización Multimodal.* Tesis Doctoral. Universidade da Coruña, 2007.

[8] Fogassi, L., Gallese, V., Pellegrino, G., Fadiga, L., Gentilucci, M., Luppino, G., Matelli, M., Pedotti, A. and Rizzolatti, G. Space coding by premotor cortex. *Experimental Brain Research, 89,* 3 (1992), 686-690.

[9] Frintrop, S., Nüchter, A. and Surmann, H. Visual Attention for Object Recognition in Spatial 3D Data. In Proceedings of the *2nd International Workshop on Attention and Performance in Computational Vision.* (Prague, Czech Republic). Springer, 168-182.

[10] Giard, M. H. and Peronnet, F. Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *J. Cognitive Neuroscience., 11,* 5 (1999), 473-490.

[11] Luo, R. C. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man and Cybernetics, 19,* 5 (1989), 901-931.

[12] Revonsuo, A. and Newman, J. Binding and Consciousness. *Consciousness and Cognition, 8,* 2 (1999), 123-127.

[13] Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S. and Woldorff, M. G. Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, *45*, 3 (2007), 561-571.
[14] Spence, C. and Squire, S. Multisensory Integration: Maintaining the Perception of Synchrony. *Current Biology*, 13 (2003), R519-R521.
[15] Thrun, S. Probabilistic Algorithms in Robotics. *AI Magazine, 21*, 4 (2000), 93-109.