Semantic Perception using Spatial Potential Fields

Malgorzata Goldhoorn[1] and **Ronny Hartanto**[2]

Abstract. Dealing with knowledge from the human natural environment is one of the most important capabilities that robotic systems should be equipped with to act autonomously. Moreover, a robot cannot perform a given task properly without knowing the meaning of the contained objects. In addition, semantic perception is very challenging and robots must be able to deal with uncertainties and ambiguities that may occur partially caused by noisy sensors. In this paper the concept of a novel approach for semantic object recognition for indoor scenes is presented. In our method, spatial potential fields are defined to model probabilistic spatial relations between objects. These relations can then be used to recognise specific objects and find their most probable position in a given environment.

1 INTRODUCTION

Intelligent robots able to perform everyday tasks in human living environments are certainly going to play an important role in the society in the future. Especially indoor service robots that could support humans in their everyday live will be of great use. One of the main challenges is that such robots need to act and understand their environment, which, however is very dynamic and rather designed for humans. A system's ability to detect objects in its surrounding environment and assign them meaningful descriptions is mandatory for an agent to support a human or interact with him. In this paper a concept of a novel idea for semantic object recognition and indoor scene labelling based on spatial potential fields and probability is presented. We introduce a new method, in which the spatial potential fields (*SPF*) and maximum field intensity (*MFI*) are used for the object recognition process. Furthermore, the definition of a new type of probabilistic spatial relations between objects is introduced.

2 RELATED WORK

In recent years, a number of studies have been done in the area of semantic perception for robotic systems [10, 5, 7, 9]. While the increasing availability of low-cost RGB-D sensors has likely contributed to this trend, the main reason may be found in the fact that robots need these capabilities to perform complex tasks in human living environments [6, 8]. In one recent study [1] contextually guided semantic labelling and search are presented. In this method a graphical model with geometrical features and contextual relations between objects is used. To obtain a better view of the objects in the scene, active object recognition is performed. Aydemir et al. [3, 2] describe an approach for active visual search. In this process topological relations between objects are used to create a potential search action and find a given object. Similar to Anand et al. [1] the next best view algorithm is applied to deal with occlusion. Other recent work which covers the challenges of object labelling in indoor environments using RGB-D data has been presented in [11]. The authors developed and evaluated a method for scene labelling that combines RGB-D features and contextual models by using MRFs (Markov Random Fields) and segmentation.

3 SPATIAL POTENTIAL FIELDS

Some tasks of a domestic service robot focus on supporting humans in their everyday life. Most of those involve object manipulation tasks, such as retrieving a certain object such as a bottle of milk or a cereal box. Another scenario would be preparing a breakfast or lunch table. Nevertheless, such scenarios require that a robot is able to detect and recognise objects in the given task. To achieve this, *a priori* knowledge about the objects and their relations in a given environment is required. Thus, the robot is able to utilise this information for deriving a search strategy for a given object. As most target objects are relatively small in comparison to the environment in which the robot is performing the task, semantic and spatial knowledge about the object could reduce the area to be searched by the robot. An approach for modelling the spatial information of the objects is designed by using a new type of potential fields: *spatial potential fields*. Potential fields have so far been used in the area of robots navigation [4] for obstacle avoidance. In this area, the potential fields are calculated from the detected objects or the given map of the environment. The magnitude of a field depends on the distance of the respective object to the robot. The calculated potential fields are used for avoiding obstacles on the robot pathways.

In our novel idea, we define and use *spatial potential fields* (*SPFs*) for calculating the spatial information of the objects. This information is used further on in the process of semantic object recognition. *SPFs* can be seen as a new method to describe contextual topological relationships between pairs of objects in a probabilistic manner. In our concept we distinguish between two types of objects in the environment. The first type consists of socalled *reference* objects R_O such as wall, floor, ceiling, door or a furniture. These objects are used as context information in the search for other items. The second type contains *target* objects T_O. This type involves entities which are usually smaller than reference objects, like food articles or office supplies. The spatial potential fields are determined between reference objects and target objects using spatial relations. Such relations could be for instance *below*, *near*, or *above*. These relations can be seen as abstractions of the object configuration in space, such as their distances, directions or topological relationships. Therefore, potential fields are spread out over the objects and describe the degree of the intensity for a given spatial relation. *SPFs* are defined as

[1] Malgorzata Goldhoorn, Department of Computer Science, University of Bremen, Robotics Group, 28359 Bremen, Germany, email: malgorzata.goldhoorn@informatik.uni-bremen.de

[2] Ronny Hartanto, DFKI Bremen (German Research Center For Artificial Intelligence), Robotics Innovation Center, 28359 Bremen, Germany, email: ronny.hartanto@dfki.de

ellipsoidal forms, where two factors are determined from the reference object dimension, e.g. width and height of the reference object. The third factor of the ellipsoidal field is variable and determined by the distance from the reference object. These values are used for determining the potential field value/factor. For example, the field indicates how close an object is to another objects. Figure 1 shows an example of spatial potential fields for the relations *near* and *above*.

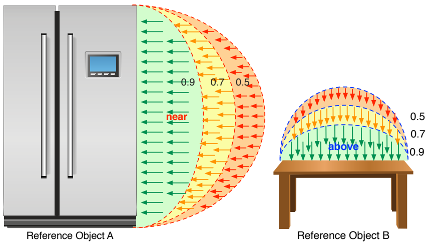

Figure 1. Two dimensional illustration of the spatial potential fields concepts for *near* (A) and *above* (B) relations

The spatial potential field of an object with a relation to a reference object $SPF_{relation}$ is formally defined as:

$$SPF_{relation} = \frac{\sum_{i=1}^{n} max(PF_{x_i})}{n} \tag{1}$$

where: PF_{x_i} is the possible potential field for the given point x_i with the given relation and n is the number of all points of the object being tested for the given relation. The possible potential field (PF) is defined as a value between 0 and 1, cf. Figure 1. As a result, the calculated $SFP_{relation}$ is a valid value for further processing in the probabilistic reasoning method.

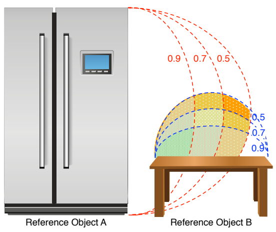

Figure 2. Two dimensional illustration of the maximum field intensity concept of relation *near* of reference object A (refrigerator) and relation *above* of reference object B (table)

The potential fields have another important role in our approach, namely, they are used to calculate the most probable position of a given target object T_O. This can be done by means of *Maximum Field Intensity*. It is fairly likely that an object can have more than one relation to a number of reference objects. The *MFI* specifies where the fields of given spatial relations of objects combine to produce the highest intensity. An example of *MFI* resulting from two relations is presented in Figure 2.

One major advantage of our potential fields approach is the possibility to combine a number of *SPF* relations into an *MFI* value. Instead of calculating the *MFI* directly using a combination of the relations of the object with other reference objects, the *MFI* is calculated as a superposition of those individual *SPFs* of each relation. This advantage is also algorithmically efficient, as *SPF* calculations are done once for each object's relation. Any number combinatorial *MFI* can be calculated formally as follows:

$$MFI = \frac{\sum_{j=1}^{m} SPF_j}{m} \tag{2}$$

where SPF_j is the spatial potential field for the relation j and m is the number of *SPF* relations being combined.

4 PROBABILISTIC SPATIAL RELATIONS

In order to distinguish and detect objects, prior knowledge about typical objects in the environment is needed. This knowledge must be expressive enough and contain sufficient information to enable the definition of various classes of entities. Furthermore, using such knowledge, it must be possible to deal with uncertainties and ambiguities which may occur. Through probabilities many different cases of object appearance and relations can be described. However, more relevant is that reasoning can still be done even if part of information is missing, which in turn is not possible by using first order or description logic exclusively.

In our approach, we use probability for the description of the spatial relations between objects. In this way different possible relations of an object can be represented. For instance, it can be defined that a given target object T_A is located on reference object R_B or on other reference object R_C with a given probability. That means the target object can be found not only in one place but in many different areas, which is almost always the case in human living environments. These probability values may changes according to the given environment. The relations have the general form:

$$\beta : \varphi \times \eta \times \eta \rightarrow [0; 1] \tag{3}$$

With $\eta = \{\eta_1, ..., \eta_n\}$ and $\eta_n \in [T, R]$ as an object type and $\varphi = \{\varphi_1, ..., \varphi_n\}$ as a set of relations. Based on formula (3) the following relations can be defined:

$$\beta(above, T_A, R_B) = 0.2 \tag{4}$$

$$\beta(near, R_C, R_B) = 0.6 \tag{5}$$

These relations can also be considered as a spatial heuristic which can be used in the search for a given target object. By means of this heuristic the search space can be reduced, by searching for the object at the position with the highest probability first. Furthermore, the heuristic yields all probable locations for the object's occurrence.

5 CONCLUSION AND FUTURE WORK

In this paper a novel idea of using the spatial potential fields for semantic object recognition has been presented. We have described a concept of our spatial potential fields and how they can be applied to create a hypothesis for object occurrences and derive additional knowledge for object recognition. In addition, we have provided a definition for probabilistic spatial relations between objects. The next step of our work will be the further implementation of our novel idea, analysing and validating of our concept. We would like to extend our algorithm to new spatial relations and other domains apart from the kitchen environment used in the example presented here and perform tests in real human living environments with a mobile robot.

ACKNOWLEDGEMENTS

This work has been supported by the Graduate School SyDe, funded by the German Excellence Initiative within the University of Bremen's institutional strategy.

REFERENCES

[1] A. Anand, H. S. Koppula, T. Joachims, and A. Saxena, 'Contextually guided semantic labeling and search for three-dimensional point clouds', *The International Journal of Robotics Research*, (2012).

[2] A. Aydemir, A. Pronobis, M. Gobelbecker, and P. Jensfelt, 'Active visual object search in unknown environments using uncertain semantics', *Robotics, IEEE Transactions on*, **29**(4), 986–1002, (Aug 2013).

[3] A. Aydemir, K. Sjoo, J. Folkesson, A. Pronobis, and P. Jensfelt, 'Search in the real world: Active visual object search based on spatial relations', in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 2818–2824, (May 2011).

[4] J. Borenstein and Y. Koren, 'The vector field histogram-fast obstacle avoidance for mobile robots', *Robotics and Automation, IEEE Transactions on*, **7**(3), 278–288, (Jun 1991).

[5] M. Eich, M. Dabrowska, and F. Kirchner, 'Semantic labeling: Classification of 3d entities based on spatial feature descriptors', in *IEEE International Conference on Robotics and Automation, ICRA*, (2010).

[6] C. Galindo, J. Fernández-Madrigal, J. González, and A. Saffiotti, 'Robot task planning using semantic maps', *Robot. Auton. Syst.*, **56**(11), 955–966, (November 2008).

[7] R. Hartanto, *A Hybrid Deliberative Layer for Robotic Agents: Fusing DL Reasoning with HTN Planning in Autonomous Robots*, volume 6798 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, 2011.

[8] D. Pangercic, M. Tenorth, D. Jain, and M. Beetz, 'Combining perception and knowledge processing for everyday manipulation', in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 1065–1071, (2010).

[9] A. Quattoni and A. Torralba, 'Recognizing indoor scenes', *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2009).

[10] R. B. Rusu, Z.C. Marton, N. Blodow, A. Holzbach, and M. Beetz, 'Model-based and learned semantic object labeling in 3d point cloud maps of kitchen environments', in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pp. 3601–3608, (2009).

[11] R. Xiaofeng, B. Liefeng, and D. Fox, 'Rgb-(d) scene labeling: Features and algorithms', in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2759–2766, (June 2012).