Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 41 (2012) 1307 – 1315

Engineering Procedia

www.elsevier.com/locate/procedia

International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012)

Color and Depth-Based Superpixels for Background and Object Segmentation

Islem Jebari[a,]*, David Filliat[a,b]

[a]*ENSTA ParisTech, 32 boulevard Victor 75015, Paris, France*
[b]*INRIA FLOWERS team, Bordeaux, France*

Abstract

We present an approach to multimodal semantic segmentation based on both color and depth information. Our goal is to build a semantic map containing high-level information, namely objects and background categories (carpet, parquet, walls …). This approach was developed for the Panoramic and Active Camera for Object Mapping (PACOM)[†] project in order to participate in a French exploration and mapping contest called CAROTTE. Our method is based on a structured output prediction strategy to detect the various elements of the environment, using both color and depth images from the Kinect camera. The image is first over-segmented into small homogeneous regions named "superpixels" to be classified and characterized using a bag of features representation. For each superpixel, texture and color descriptors are computed from the color image and 3D descriptors are computed from the associated depth image. A Markov Random Field (MRF) model then fuses texture, color, depth and neighboring information to associate a label to each superpixel extracted from the image. We present an evaluation of different segmentation algorithms for the semantic labeling task and the interest of integrating depth information in the superpixel computation task.

Keywords: image segmentation; Markov Random Field.

1. Introduction

The Panoramic and Active Camera for Object Mapping (PACOM) project addresses the understanding of how an autonomous embodied system can build and extract information from sensory and sensory-motor data and generates plans and actions to explore and navigate in typical indoor environmental settings. In particular, we seek to extract high-level semantic information that is easy to understand and interesting to the robot users such as surrounding objects and the environment structure. The project goal is to participate in the CAROTTE challenge that takes place in an arena of approximately 120m². Several kinds of objects are present, either isolated or gathered, in multiple specimens, which must be detected, located, and identified or characterized by the robot. The environment contains several rooms typically 10 or more, with variable grounds and various difficulties (fitted carpet, tiling, grid, sand, stones…).

* Corresponding author. Tel.: +331 45 52 70 50.
E-mail address: islem.jebari@ensta-paristech.fr

[†] The PACOM project is supported by DGA in the frame of the "CAROTTE" competition and funded by ANR under the subvention 2009 CORD 102. CAROTTE is organized by the French research funding agency (ANR) and the French armament procurement agency (DGA). Website: <http://www.defi-carotte.fr>

1877-7058 © 2012 Published by Elsevier Ltd.Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.07.315

We developed a multi-sensor system in support of the PACOM project (see Fig. 1) based on a pioneer 3 dx from Mobile Robots Inc [15]. The robot was fitted with a horizontal laser rangefinder used for the 2D localization, a ring of sonar sensors to avoid obstacles, a Pan-Tilt-Zoom camera to identify some critical obstacles like the gravel, and three on-board computers. We use a Kinect camera from Microsoft to construct 3D point clouds representing the environment.

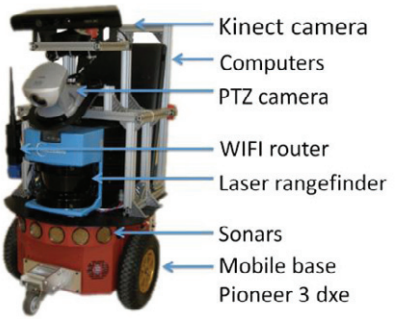


Fig. 1. The PACOM robot.

The semantic segmentation problem could be defined as follows: given an input image, assign a label to each one of the pixels; the labels are associated to high-level concepts that give a semantic interpretation to the scene; adjacent components with the same label constitute the "semantic segments" and are associated to the real world objects indicated by the label.

The problem can be approached from different perspectives. One alternative is to directly assign labels to the pixels and then find the connected components that constitute the semantic segments. Another alternative is to find first a segmentation of the image and then assign a label to each segment. In this paper we follow an intermediate approach, first we found an over-segmentation of the image, and then labels are assigned to each small region (called superpixel). Later, contiguous superpixels with the same label can be merged to form the final segments.

In this paper we explore a multimodal technique based on two main steps: first, the over-segmentation of the scene in superpixels, and second, the assignment of labels to theses superpixels using a Markov Random Field (MRF) model. We compare different approaches for superpixels computation using either the color information, the depth information or both and we characterize them using bag of features representations. An energy function is then defined over the MRF using four main elements: the color image conditional probability, the depth conditional probability, the probability to assign a given labeling to two adjacent superpixels and the a priori label probability. We provide a detailed analysis of the importance of color and depth information at the two levels (superpixels and MRF) for the overall performances.

This paper is organized as follows: next section overviews the related work. Section 3 describes our multimodal semantic segmentation approach and Section 4 analyses the performance gain obtained when using color and depth information in our algorithm.

2. Related work

We are interested in the problem of semantic segmentation, i.e. assigning each pixel in an image to one of several pre-defined semantic categories. This is a supervised learning problem in contrast to low-level unsupervised segmentation which groups pixels into homogeneous regions based on features such as color or texture.

The existing works on semantic segmentation typically differ in the choice of elementary regions for which the labels are sought, the types of features which are used to characterize them, and means of integrating the spatial information. Instead of working directly at the pixel level, one strategy to address the semantic segmentation problem is to compute features that involves bigger entities than pixels (called superpixels), to assign labels to these entities according to a trained classification model, and finally, to group them into semantic objects. [7], [4] and [8] rely on small blob-based superpixels represented by descriptors such as color, texture [7], 2D frequency planes [2] or SIFT [4]. Once the superpixels are characterized, the descriptor set can be reduced by generating a specific non-redundant vocabulary of elementary features. In [4] and [6], the bag of features model was adapted based on hierarchical K-means; [1] built a randomized forest decision that uses simple pixel comparisons, performing an implicit hierarchical clustering into semantic textons.

It is then necessary to build a classifier for the labeling task using these descriptors and incorporating contextual information. Some approaches are based on a probabilistic framework such as a Random Field (RF), mainly the Markov Random Field (MRF) [16] and the Conditional Random Field (CRF) [17]. While the MRF is generative in nature, the CRF

models directly the conditional probability of labels given features thus simplifying the use of more complex features. There also exist approaches to enforce local consistency without RF models, including the forest of spanning tree method of [16] and the contextual empirical Bayes approach of [18]. [8] proposed a MRF model incorporating local data interaction in unsupervised parameter learning. This model includes computational efficiency by using superpixel structure and its ability to integrate local knowledge in the learning process. [5] trained a CRF on heterogeneous descriptors extracted at different scales and locations in the image. [7] proposed an SVM-MRF framework to model features and their spatial distributions (SVM is applied to represent conditioned feature vector distributions within each cluster, and MRF is used to model the spatial distributions of the semantic labels). [6] presented a MRF based multivariate segmentation algorithm called "multivariate iterative region growing using semantics" (MIRGS): the impact of interclass variation and computational cost are reduced using the MRF spatial context model incorporated with adaptive edge penalty and applied to regions.

[3] presented a learning-based unified image retrieval framework to represent images in local visual and semantic concept-based feature spaces. In this framework, a visual concept vocabulary (codebook) is automatically constructed by utilizing self-organizing map (SOM) and statistical models are built for local semantic concepts using SVM. The features are unified by a dynamically weighted linear combination of similarity matching scheme based on the relevance feedback information. [2] proposed a novel method which integrates principal component analysis (PCA) and SVM neural networks for analyzing the semantic content of natural images.

The recent wide availability of depth cameras [11] has spurred further progress in labeling 3D scenes. Several properties should be captured: local properties (visual appearance, shape, and geometry), visual context and geometric context (on top of, in front of, convexity …). [19] addresses the problem of segmenting 3D scan data into objects. The applied segmentation framework is based on a subclass of MRF which supports efficient graph-cut inference. [20] presents a contribution to the problem of 3D point cloud classification onboard a mobile vehicle using a CRF for scene interpretation and environment modeling. It is shown how efficient learning of a random field with higher-order cliques can be achieved using subgradient optimization. [21] uses a CRF model to discover and exploit contextual information, classifying planar patches extracted from the point cloud data.

3. Multimodal semantic segmentation

We present the three semantic segmentation algorithms that we used for evaluating the interest of color and depth information for semantic segmentation. The first is only based on color images (color-based segmentation), the second is only based on depth images (depth-based segmentation) and the third integrates both color and depth information with several different approaches. These three approaches are based on the Markov Random Field framework.

3.1. Superpixel over-segmentation

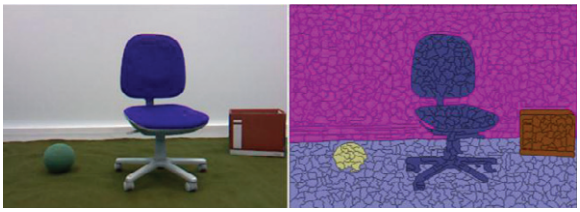
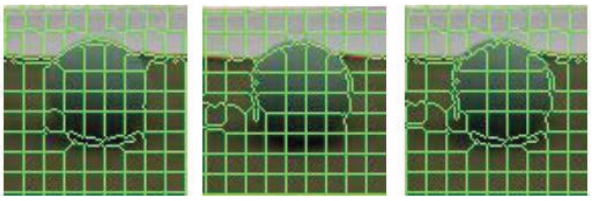


Fig. 2. Semantic segmentation of an image using superpixels. On the left: the original image. On the right: the over-segmented image in which each superpixel has been assigned a semantic label indicated by a color.

For our superpixel over-segmentation, we were inspired by the algorithm used in [4]. The algorithm uses watershed segmentation applied on the image Laplacian based on uniformly distributed seeds along a regular grid. The image Laplacian is obtained using either color, depth or by fusing the two information. The depth Laplacian is computed by applying a bilateral filtering on the depth image to reduce noise before computing the Laplacian. The color Laplacian is defined by converting the color image in grayscale before computing the Laplacian. The fusion Laplacian is defined by taking the maximum of the Laplacian of the color image and the depth one.

The fusion approach improves the contrast for objects having a color similar to the background that are usually badly segmented when using color information only.This method gives superpixels that offer a better delimitation for the objects (see Fig. 3), but presents also the disadvantage of producing more small superpixels. This occurs in particular if there are small shifts between the color image and the depth one, which can happen in cases of bad calibration or temporal shift between the two images if the robot moves during acquisition. As we will further see it, this new segmentation in superpixels produces however an overall positive effect.



Color Depth Fusion

Fig. 3. Example of improvement related to the integration of both color and depth information for the supepixels computation. Fusion better delimits the object, but provides superpixels sometimes more disturbed.

3.2. Color superpixel representation

For our experiments, we compute a feature vector per superpixel from the color image. First, we find the 'morphological center' that is defined as the maximum of a distance map computed in the superpixel with respect to the superpixel boundary. Second, we compute a SIFT descriptor for the superpixel's center. This generates 128 feature values, which are coupled with the average color of the superpixels, represented in the L*a*b color space. In total, this produces 131 features per superpixel.

3.3. Depth superpixel representation

We implemented and tested the descriptor proposed by [14] which is employed in the Xbox console from Microsoft to predict 3D positions of body joints from a single depth image. This descriptor computes, for each considered pixel, the difference in depth between two close pixels characterized by offsets $\theta = (u, v)$ which are normalized by the pixels depth to ensure that features are depth invariant. At a given point on the object, a fixed world space offset will result whether the pixel is close or far from the camera [14], [9]. A simple depth comparison features is employed, inspired by those in [10]. At a given pixel x, the features compute:

$$f_\theta(I, x) = d_I\left(x + \frac{u}{d_I(x)}\right) - d_I\left(x + \frac{v}{d_I(x)}\right) \quad (1)$$

where $d_I(x)$ is the depth at pixel x in image I, and parameters $\theta = (u, v)$ describe offsets u and v. If an offset pixel lies on the background or outside the image bounds, the depth probe $d_I(x)$ is given a large positive constant value. To characterize a superpixel, we calculate a vector of 496 different values obtained with 496 θ configurations. These configurations are obtained by taking all the possible pairs of points among the pixels located 4 pixels around the superpixel's center in the eight principal directions (see Fig. 4 (c)). The design of these features was strongly motivated by their computational efficiency as no depth image preprocessing is needed.

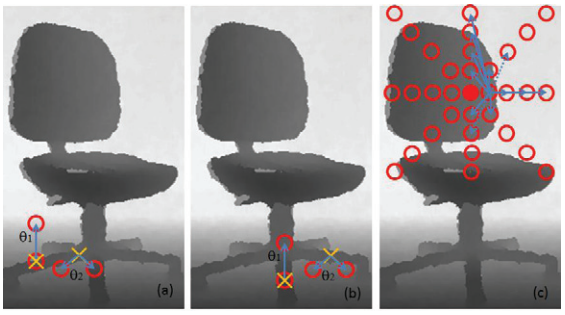


Fig. 4. Depth image features. The yellow crosses indicate the pixel x being classified. The red circles indicate the offset pixels as defined in Eq. 1. In (a), the two example features give a large depth difference response. In (b), the same two features at new image locations give a much smaller response. In (c), Illustration of a part of the neihboring considered around the superpixel's center for depth descriptors.

3.4. A Markov Random Field model for multimodal labeling

The color and depth-based semantic segmentation algorithm uses a MRF model to assign labels to superpixels. The process is divided in two main phases: training and testing. During training, the MRF is trained using a set of labeled, color

and depth images. During testing, the MRF is used to assign labels to new images. The overall process is illustrated in Fig. 5.

The training process proceeds as follows:

1) For each color and its corresponding depth image, a superpixel extraction algorithm is applied to find an over-segmentation using one of the three approaches described in section 3.1.

2) For each superpixel in each color image, a 131-feature vector is computed. For each superpixel in each depth image, a 496-feature vector is computed.

3) The set of all color feature vectors is used to build a color Bag-of-Features (BoF) codebook. The set of all depth feature vectors is used to build a depth Bag-of-Features (BoF) codebook. BoF codebooks are created by applying Learning Vector Quantization [22].

4) All the images of the training dataset are represented by the corresponding code-words for each superpixel. A superpixel neighborhood graph is computed (two superpixels are said to be neighbors if they share one or more boundary pixels.).

5) A MRF model is trained by computing the probability distributions that correspond to the model parameters. These distributions are:

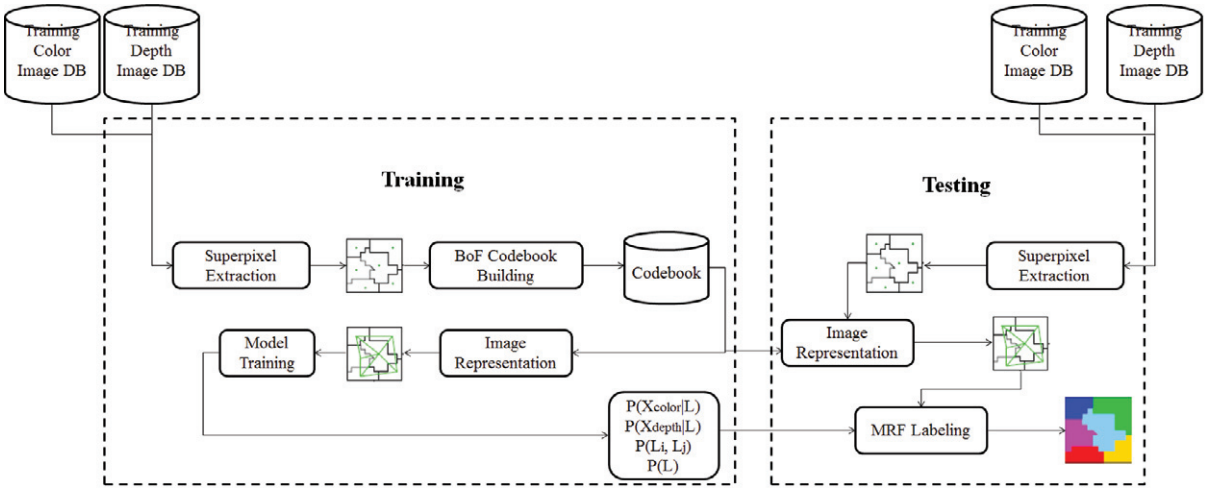


Fig. 5. Multimodal semantic segmentation process

- $P(x_i^{color} \mid l_i)$: the color conditional probability.
- $P(x_i^{depth} \mid l_i)$: the depth conditional probability.
- $P(l_i \mid l_j)$: the neighboring label join probability.
- $P(l_i)$: the a priori label probability.

The test process is as follows: for a particular image, the superpixel extraction and image representation processes are applied as described in the training process; then, the MRF energy optimization algorithm (described in the following subsection) is applied, using the parameters learned during the training phase, to find the superpixel labels.

A MRF is a graph, (V, E_d) where each graph node $l_i \in V$ corresponds to a random variable. The MRF satisfies the following property: $P(l_i \mid V \setminus l_i) = P(l_i \mid N_i), \forall i \in V$ where N_i represents the set of neighbors of l_i. This is called the locality property and basically stands that the random variable l_i is conditionally independent of the rest of variables given its neighbors. Usually the MRF's variables take values in a discrete set of labels $\Lambda = \{\lambda_1, ..., \lambda_m\}$. It is common also to associate each variable l_i with a variable x_i. In this case, the variable l_i is called a latent variable, that means that it cannot be measured directly, and it has to be inferred by the values of the observed variables x_i. In this particular framework, the problem is: given a set of observations to infer the most probable assignations for the latent variables, which can be state as:

$$\max_L P(L \mid X) = \max_L P(l_1,...,l_n \mid x_1,...,x_n) \quad (2)$$

This problem is in principle, a hard problem to solve since the space of possible assignments for L grows exponentially with the size n of the graph. However, there are efficient algorithms that exploit the particular structure of MRF to find optimal or close to optimal solutions to this problem. In general, all the algorithms exploit the so called factorization property of the joint probability: $p(V) = \frac{1}{Z}\prod_C \Psi_C(V_C)$ where Z is a normalization constant, C runs over the maximum cliques of the graph, and ψ_C is function over the variables of the corresponding clique called a potential function. Usually, the potential functions take the form: $\psi_C(C) = e^{-E_C(V_C)}$ where E is an energy function. In this case the joint probability can be expressed as: $p(V) = \frac{1}{Z}e^{-E(V)} = \frac{1}{Z}e^{-\sum_C E_C(V_C)}$. As a result, the MRF probability distribution is determined by specifying the energy function, and minimizing it is equivalent to maximizing the joint probability.

The problem of semantic segmentation is modeled using a MRF model as follows:

1) The vertexes of the graph V correspond to the set of superpixels extracted from one image; the edges E are determined by the superpixel adjacency relationship.
2) The labels of the superpixels are modeled by the l_i latent variables. The observed variables x_i are broken in two variables x_i^{color} and x_i^{depth} that correspond respectively to the color and depth superpixel's information.
3) The MRF energy function is defined as follows:

$$E(L) = \alpha E_{color}(L) + \beta E_{edge}(L) + \gamma E_{prior}(L) + \delta E_{depth}(L) \quad (3)$$

where:

$$E_{depth}(L) = -\sum_{l_i \in V} \log P(x_i^{depth} \mid l_i)$$

$$E_{color}(L) = -\sum_{l_i \in V} \log P(x_i^{color} \mid l_i)$$

$$E_{edge}(L) = -\sum_{(i,j) \in E_d} \log P(l_i \mid l_j)$$

$$E_{prior}(L) = -\sum_{i \in V} \log P(l_i)$$

This definition is motivated by an expression of the conditional probability of the labeling given by:

$$P(L \mid X) = \frac{P(X \mid L)P(L)}{P(X)} = \frac{P(X^{depth} \mid L)P(X^{color} \mid L)P(L)}{P(X)} \simeq P(X^{depth} \mid L)P(X^{color} \mid L)P(L) \quad (4)$$

The last expression is motivated by the fact that the evidence probability $P(X)$ is the same for all the different labels. Since we are interested in the maximum a posteriori estimation, it is enough to take into account only the numerator.

The computational problem is to find the labels that maximize the posterior probability (2). Recently, different efficient algorithms have been proposed to solve this problem including: graph cuts, loopy belief propagation and tree-re-weighted message passing [12]. In our implementation we used a general algorithm to solve the max-sum problem in graphs based on linear programming [13].

4. Experimental evaluation

In this section, we present an experimental evaluation of our multimodal semantic segmentation algorithm and compare it with color-based segmentation and depth-based segmentation algorithms for backgrounds and objects classification.

The proposed system was evaluated on a specifically created database: a collection of 137 labeled images that associate each pixel with one of 7 semantic classes. The semantic classes are the following: (1) carpet floor, (2) white wall, (3) file box, (4) chair, (5) ball, (6) lino floor, (7) wooden wall. (1), (2), (6) and (7) represent the backgrounds, including floors and walls. (3), (4) and (5) represent various objects. Our goal with this preliminary evaluation was to evaluate the performance gained by using a multimodal approach before applying our system to a larger database.

We run experiments with these 7 classes, but also with 5 classes, the 3 objects being gathered in only one class "Object" with the goal of separating objects from background. The training step is done on 127 images chosen randomly, and tests are done on the 10 remaining images. We run the three algorithms with respectively 7 classes and 5 classes and evaluate the

interest to have gathered the various objects in only one class in order to have more significant results. In order to limit the effects of the training images choice, this procedure is carried out 10 times and the average performances are reported.

For classification, we report the per-class average accuracy, i.e. the diagonal average of the confusion matrix between the ground truth label part and the most likely inferred part label.

We also investigate the effect of several MRF parameters on the classification accuracy by using a grid search on these parameters and reporting the values obtained with the best parameter set.

For the color-based segmentation experiments, we used the following energy function:

$$E(L) = \alpha E_{color}(L) + \beta E_{edge}(L) + \gamma E_{prior}(L)$$

the parameter γ was kept equal to 0.2 and (α, β) were varied (from 0 to 1.0 with step of 0.2).

For the depth-based segmentation, we used:

$$E(L) = \alpha E_{depth}(L) + \beta E_{edge}(L) + \gamma E_{prior}(L)$$

the parameter γ was kept equal to 0.2 and (α, β) were varied (from 0 to 1.0 with step of 0.2).

For the multimodal segmentation experiments, we used:

$$E(L) = \alpha E_{color}(L) + \beta E_{edge}(L) + \gamma E_{prior}(L) + \delta E_{depth}(L)$$

with $\gamma = 0.2$, $\delta = 1 - \alpha$ and (α, β) were varied (from 0 to 1.0 with step of 0.2).

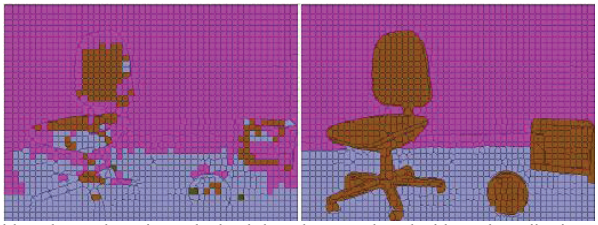
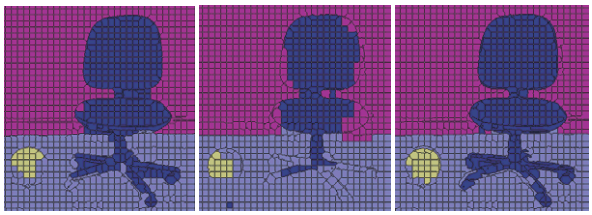


Fig. 6. Example of results obtained with 5 classes, by using only depth-based approach and without the neiboring term of the MRF (on the left) and with the optimal parameters using the multimodal approach (on the right).



(a) color-based approach (b) depth-based approach (c) multiomodal approach

Fig. 7. Example of results obtained with 7 classes. (a) corresponds to the color-based approach with parametrs $(\alpha, \beta, \gamma) = (0.4, 0.4, 0.2)$. (b) corresponds to the depth-based approach with parameters $(\alpha, \beta, \gamma) = (0.6, 0.4, 0.2)$. (c) corresponds to the multimodal approach (taking into account both, color and depth images for superpixel over-segmentation) with parameters $(\alpha, \beta, \gamma, \delta) = (0.4, 0.4, 0.2, 0.6)$

Fig. 6 and Fig. 7 illustrate the results typically obtained. With the best parameters, the recognition is of very good quality. As shown in table 1, we obtained around 80% of recognition rate. The recognition is overall better for the background classes than for the objects taken either in isolation or as a single class. The superpixels computation integrating both depth and color information, associated with a MRF using also depth and color information gives the best results. The regrouping of the various objects in only one "object" class improves the total rate of recognition appreciably, but doesn't distinguish between the objects anymore and decreases a little bit the floors and walls recognition rates. This approach is however interesting because it improves the average recognition accuracy of the "object" class, i.e. that it is possible to better distinguish the objects from the background, knowing that the segmented objects can be identified thereafter using reliable methods, such as those used in [15].

The gain obtained by the use of depth and color for superpixels extraction is positive, but quite small in practice. This is linked to the fact that these superpixels are only bringing improvements for the borders of objects that are difficult to perceive in the color image, which only happens for a few objects in our database.

The gain obtained by using color and depth in the MRF is more sensitive and improves the overall performances by 1%.

Table 1. Per-class and average recognition accuracy (%) of the various semantic classes (5 and 7 classes respectively) for various algorithms. The last vertical colomn corresponds to the global average accuracy (%).

algorithm	carpet floor	white wall	object	lino floor	wood wall	global average accuracy
Color-based segmentation (color superpixels)	80.32	82.48	76.82	70.05	70.87	79.03
	81.38	82.77	73.52	70.94	68.91	78.73
Depth-based segmentation (depth superpixels)	76.91	81.79	58.52	4.36	4.49	63.88
	77.29	81.83	43.62	3.71	4.22	61.19
Multimodal segmentation (color superpixels)	80.70	82.84	77.57	77.83	77.38	80.06
	81.44	82.91	75.41	79.88	77.40	80.64
Multimodal segmentation (depth superpixels)	80.31	82.72	77.38	76.59	74.35	80.15
	81.04	82.82	75.22	78.46	76.15	80.26
Multimodal segmentation (color and depth superpixels)	80.70	82.87	**77.63**	77.86	77.79	80.69
	81.63	**83.14**	73.64	**79.69**	**78.27**	**80.49**

5. Conclusion & perspectives

We proposed a semantic segmentation algorithm based on color and depth information and evaluated the influence of various parameters of this algorithm. Applied in an indoor environment, the use of depth and color information for superpixels segmentation and semantic labeling effectively improves the accuracy of the segmentation of the environment into backgrounds and objects classes, compared with algorithm using only color or only depth.

For the next months, we plan to evaluate other depth descriptors, make more representative evaluations with more complex databases acquired in real conditions during the CAROTTE competition and compare our results with another algorithm using only 3D information for object/background segmentation [15].

References

[1] J. Shotton, M. Johnson, R. Cipolla. Semantic Texton Forests for Image Categorization and Segmentation. In Proc. IEEE CVPR 2008.
[2] Chuan Y. Chang, Hung J. Wang, and Chi F. Li. Semantic analysis of real-world images using support vector machine. Expert Syst. Appl., 36(7):10560–10569, 2009.
[3] Md, Prabir Bhattacharya, and Bipin C. Desai. A unified image retrieval framework on local visual and semantic concept-based feature spaces. J. Vis. Comun. Image Represent., 20(7):450–462, June 2009.
[4] B. Micusik and J. Košecka. Semantic segmentation of street scenes by superpixel co-occurrence and 3D geometry. In IEEE Workshop on Video-Oriented Object and Event Classification (VOEC), held jointly with International Conf. on Computer Vision (ICCV), Japan, 2009. IEEE.
[5] Giuseppe Passino, Ioannis Patras, and Ebroul Izquierdo. Context awareness in graph-based image semantic segmentation via visual word distributions. Image Analysis for Multimedia Interactive Services, International Workshop on, 0:33–36, 2009.
[6] A. K. Qin and David A. Clausi. Multivariate Image Segmentation Using Semantic Region Growing With Adaptive Edge Penalty. IEEE Transactions on Image Processing, 19(8):2157–2170, August 2010.
[7] L. Wang and B. S. Manjunath. A semantic representation for image retrieval. In Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), pages II–523–6. IEEE, 2003.
[8] XiaoFeng Wang and Xiao P. Zhang. A new localized superpixel Markov random field for image segmentation. In ICME'09: Proceedings of the 2009 IEEE international conference on Multimedia and Expo, pages 642–645, Piscataway, NJ, USA, 2009. IEEE Press.
[9] Real-time Human Pose Recognition in Parts from Single Depth Images: Supplementary Material.
[10] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In Proc. CVPR, pages 2:775–781, 2005. 4
[11] Microsoft Corp. Redmond WA. Kinect for Xbox 360. 1, 2

[12] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A Comparative Study of Energy Minimization Methods for Markov Random Fields . In Ale¨s Leonardis, Horst Bischof, and Axel Pinz, editors, Computer Vision ECCV 2006, volume 3952 of Lecture Notes in Computer Science, pages 16–29, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
[13] TomasWerner. A Linear Programming Approach to Max-Sum Problem: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(7), 2007.
[14] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake Real-Time Human Pose Recognition in Parts from Single Depth Images, Proceedings of Computer Vision and Pattern Recognition, June 2011.
[15] Filliat, D., Battesti, E., Bazeille, S., Duceux, G., Gepperth, A., Harrath, L., Jebari, I., Pereira, R., Tapus, A., Meyer, C., Ieng, S., Benosman, R., Cizeron, E., Mamanna, J.-C., & Pothier, B. (2012) RGBD object recognition and visual texture classification for indoor semantic mapping. Proceedings of the 4th International Conference on Technologies for Practical Robot Applications (TePRA).
[16] Verbeek, J., & Triggs, B. (2007a). Region classification with Markov field aspects models. In CVPR, 2007.
[17] Verbeek, J., & Triggs, B. (2007b). Scene segementation with crfs learned from partially labeled images. In NIPS, 2007.
[18] Lazebnik, S. (2009). An empirical Bayes approach to contextual region classification. In CVPR, 2009.
[19] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, A. Ng. Discriminative Learning of Markov Random Fields for Segmentation of 3D Range Data. Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[20] D. Munoz, N. Vandapel, M. Hebert. Onboard Contextual Classification of 3-D Point Clouds with Learned High-order Markov Random Fields, ICRA 2009.
[21] Xiong and Huber. Using Context to Create Semantic 3D Models of Indoor Environments, BMVC 2010.
[22] T. Kohonen. Learning vector quantization. In: M.A. Arbib, editor, The Handbook of Brain Theory and Neural Networks., pages 537–540. MIT Press, Cambridge, MA, 1995.