A Joint Segmentation and Classification of Object Shapes with Feedback for 3D Point Clouds

Frauke Wübbold and Bernardo Wagner
Institute of Systems Engineering, Real Time Systems Group, Leibniz Universität Hannover, Appelstraße 9A, D-30167 Hannover, Germany

Keywords: Object Classification, Segmentation, Feedback, 3D Point Cloud, 3D Shape.

Abstract: Limited knowledge and limited deduction abilities are among the main restraints of autonomous robots for acting truly autonomously. This especially becomes obvious in the area of object recognition and classification, where many methods rely on knowledge teached manually in a prior setup step. Self-generating this knowledge from environment perception with a set of rules would significantly increase the robots autonomy as well as supersede manual training effort. In this paper, we propose a novel approach to rule-based classification for 3D point clouds by means of object shape, which additionally overcomes typical problems from a separate prior segmentation by integrating classification feedback into the segmentation process. Although it is still in its conceptual state, we explain in detail why we consider this approach to be very promising.

1 INTRODUCTION

Limited knowledge and the missing ability to detect and deduce like humans do are the main restraints for autonomous robots to act truely autonomously. In consequence, each autonomous robot has been built and trained for a set of specific tasks. This especially becomes clear in the area of object detection, where it is common practise to actively train an algorithm for recognising a set of specified objects. The set is chosen and trained manually and contains the objects which are considered most important for the tasks the robot is to fulfill. Traditionally, providing as well as training the set is a time-consuming task. Widely used training algorithms are Support Vector Machines (Chen and Ellis, 2011) (Knopp et al., 2010). Over the last years, increased efforts to improve classification autonomy by reducing manual preparation and supervision have resulted in training sets largely consisting of web data (Lai and Fox, 2009). They propose a classification which is able to transfer object information between different domains, e.g. with different backgrounds or pictures taken with different cameras, based on a small subset of object samples from both domains. (Endres et al., 2009) eliminated the training process by adapting unsupervised learning techniques from information retrieval and proposed a method which successfully and truly autonomously classifies objects into a predefined, i.e. manually predicted, number of classes. The major drawback of this approch it that initial knowledge of the number of classes is required, which is impossible to provide for a robot operating in an incompletely known environment. For a truly autonomous classification no prerequisite information should be necessary, because it is diffucult to specify already in advance the objects to be encountered or the number of classes to be seen. In an ever-changing environment, where changes could occur e.g. due to a human carrying an exemplar of a new object class into the robot's working area, it is impossible to provide these information. Instead we propose to use more general information like a set of rules much less dependant on the environment. Applying these rules to sensed environment data the approach will be able to generate object class definitions, to classify new objects and to permanently enhance existing class definitions by new classified objects.

Typically classification approaches require a previous segmentation of the sensor data into segments corresponding to real-world objects. As a result they are prone to incorporating segmentation errors without providing methods for their detection and correction. Thus, the success of a classification is directly linked to the quality of its preceding segmentation. We believe that a joint segmentation and classification scheme which feeds back classification results to the segmentation process would solve this problem. Ac-

Wübbold F. and Wagner B..
A Joint Segmentation and Classification of Object Shapes with Feedback for 3D Point Clouds.
DOI: 10.5220/0004121703880393
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2012), pages 388-393
ISBN: 978-989-8565-22-8

cording to our knowledge few work exist that fully or partially incorporate this idea in the combined area of object detection, recognition and classification, considering image-based approaches as well as those for 3D point clouds. Section 2 provides an overview of related approaches and section 3 gives a detailed description of our concept. Future work is outlined in section 4 and we offer our conclusions in section 5.

2 STATE OF THE ART

(Triebel et al., 2010) successfully integrate the generation of object hypotheses by means of segmentation into their repetetive object detection. Starting with the over-segmentation of a 3D point cloud into homogeneous surface parts, the authors take advantage of multiple occurrences of a group of segments for deducing the individual segments of which belong to a real-world object. Thus, object detection is robust to random and some perception-point specific segmentation faults. As the authors focus on detection of repetitive objects, single occurrences of a real-world object are neglected.

A combined approach to image segmentation and classification based on pattern recognition and using shape models is introduced by (Lecumberry et al., 2010). By fitting these models to images the combined framework identifies the best matches between an image region and a shape model, thus simultaneously conducting segmentation and classification. As the shape models are obtained in an initial training step, this approach requires a priori object knowledge and classification is restricted to the trained classes only.

(Farmer and Jain, 2004) propose an approach towards closed-loop segmentation and classification for images. After an initial background removal step the authors suggest to apply multiple segmentations each time using a different feature. For each segmentation groups of segments are classified according to the most probable class with regard to the previously trained classes. By evaluating the best-matching classified groups regarding their subjacent segmentation the most feasible segmentation feature is identified. The authors do not state clearly how the knowledge about which feature performs best is integrated into the overall process, i.e. how exactly the loop is closed. Additionally, this approach is limited to classification into the pre-trained classes.

For segmenting overlapping chromosomes in images (Schwartzkopf et al., 2005) introduce a maximum-likelihood test. Based on an erroneous segmentation, the authors achieve a reliable localisation and classification of the chromosomes due to prior knowlege about the total number of chromosomes expected in human cells as well as their composition.

(Guyet et al., 2007) suggest a collaborative human-machine learning approach based on a multi-agent feedback system for interpreting medical images. A pattern recognition segments the image into the currently known classes which subsequently are updated in order to adapt to the latest perceptions. Feedback from classification to segmentation is provided by adding a new pattern for each new class. The overall process is supervised by a human and thus incongruous for autonomous robots.

3 CONCEPT

This paper presents a novel concept of joint segmentation and classification based on a set of rules which is visualised in figure 1. Its benefits will be i) improved robustness to segmentation errors due to feedback from the classification and ii) significantly increased robot autonomy owing to autonomous class generation. Designed for autonomous service robots, our approach enables the robot's main task, e.g. a transportation assignment, neither obstructing nor delaying it. Therefore, the classification has to cope with the well-known results of driving past an object, i.e. partial views, and hence to use local features for both segmentation and classification.

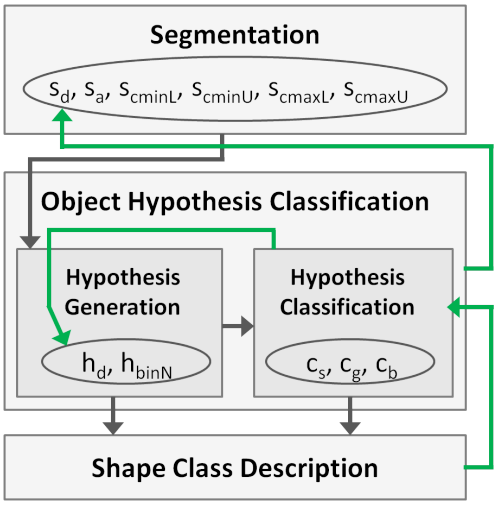

Figure 1: Schematic of the joint segmentation and classification with feedback visualised in Green. All variables are defined in the related subsections of section 3.

In human object classification, 3D object shape is an important feature providing good classification results even when the other features like colour, texture or smell are missing, and/or the object shape is

merely partially visible. Taking advantage of this observation, our classification relies on the 3D object shape. Note that we do not exactly intend to classify into real-world object classes as they are known to humans, but into classes of similar object shapes. A detailed sampling of surface shapes can easily be obtained by commertial 3D sensors like the XBox Kinect or the Velodyne HDL-64E, both providing high-resolution 3D point clouds.

The main components of our approach are i) the autonomous generation of shape class descriptions and ii) the feedback part for improving the overall result. The first component will be described in subsections 3.1 to 3.3, covering the segmentation, the generation and classification of object hypothesis and the generation of shape class descriptions. Subsection 3.4 details the feedback from classification to segmentation as well as to object hypothesis generation.

3.1 Segmentation

Many real-world objects can easily be imagined as groups of homogeneous shape parts, thus we adapt the idea of part-based shape description from e.g. (Marton et al., 2011). Instead of restricting shape parts to shape primitives (i.e. planes, cylinders and spheres), we intend to enable more flexible shapes by using surface orientation and principal curvatures as segmentation features as well as part descriptors. Orientation and principal curvatures, which express a 3D curvature at a surface point by its minimum and maximum 2D curvature, have repeatedly been proven powerful local surface descriptors. We consider the surfaces at two points to be homogeneous if they observe three threshold-based rules:

- The differences in principal curvatures are restricted by lower and upper bounds for the minimum as well as the maximum curvature: s_{cminL}, s_{cminU}, s_{cmaxL} and s_{cmaxU}, while curvature directions have to be similar.
- The angle between both surface normal vectors is to deceed s_a.
- The Euclidean distance between the points is less than s_d.

The second rule we found helpful in first experiments for surface partitioning at sharp corners in an overall curved surface due to our fast but rough estimation of the principal curvatures which is subsequently defined. The last rule serves as an initial guess for division into real-world objects. Figure 2 shows the homogeneous shape parts of a box-shaped object.

Taking advantage of our point clouds being ordered in a 2D matrix according to the horizontal and vertical sensor beam emitting angles, a search for neighbouring scan points is obsolete. Instead, point neighbourhoods can directly be extracted from the matrix. An approximation of the local surface normal vector at a distinct scan point typically is obtained by applying the Principal Component Analysis to a small neighbourhood, discarding all points which exceed the Euclidean distance threshold s_d. For principal curvature calculation, we approximate the local 2D curvature with $c = \frac{\Delta\alpha}{\Delta s}$, where $\Delta\alpha$ denotes the angle between the surface normals of two neighbour points and Δs the Euclidean distance of the points. Instead of calculating 2D curvatures for all planes containing the surface normal, we restrict these calculations to four pre-defined planes: the vertical, the horizontal and both diagonal planes in between. Thus, we achieve approximations for the principal curvature at a distinct point by evaluating the curvatures to each point of its direct neighbourhood while neglecting neighbours with a distance higher than s_d.

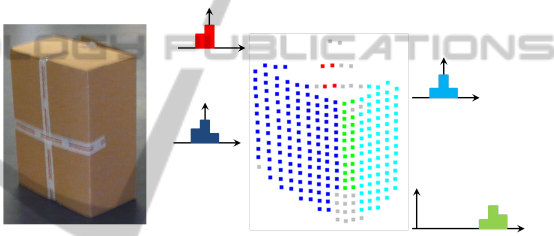

Figure 2: Homogeneous shape parts of a simple box-shaped object with reduced maximum curvature histogram for each surface part. Colours identify points and histograms of the different shape parts, while for grey points no curvature could be calculated because of too few neighbours.

Depending on point density and operating environment, very small and very large segments are excluded from further processing, as they either contain too few surface information or they most likely derive from huge homogeneous surfaces like the ground or walls of buildings. Note that this heuristic does not exclude uneven ground, which we do not consider at the present state of our research although it is most likely to occur in outdoor environments.

3.2 Object Hypothesis Classification

For an initial guess at the group of homogeneous surface parts belonging to the same real-world object, we cluster segments if the minimum point-wise Euclidean distance between the surface parts deceeds the threshold h_d. This simple relation serves as initial rule for object hypothesis generation and will be refined based on classification feedback. The resulting groups of surface parts are called object hypotheses. Aiming at autonomous generation of object shape

classes, we suggest clustering similarly shaped object hypotheses and define each cluster a shape class. A cluster will be generalised to an object shape descriptor representing the shape class as outlined in subsection 3.3. Subsequently generated object hypotheses will be clustered with other hypotheses, a shape class or both. Clustering is based on a similarity measure which firstly has to enable classifying partial object views thus encorporating local components. Secondly, it has to account for similar but not identical shapes where even a small number of surface parts might not match at all as visualised in figure 3, details given in subsection 3.3. We intend to achieve both goals by incorporating curvature information of each surface part as well as the relative positions between the parts of an object hypothesis. In the following an object hypothesis description feasible for fast shape comparison and a rule-based heuristic for shape similarity detection are introduced.

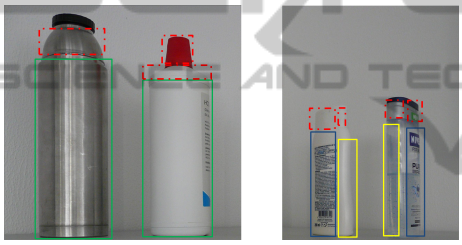

Figure 3: Objects belonging to two shape classes: mostly clyindrical (left), more free-form (right). Similarly curved shape parts are marked with solid lines in matching colours, dashed red lines indicate not-matching curvature indicators while neglecting small shape parts.

The object hypothesis description bases on a neighbourhood graph where the nodes contain information on the segments and the edges connect adjacent segments satisfying the distance threshold h_d. Surface part information consist of i) the mean positon, ii) an approximation of its size, and iii) a curvature indicator for fast detection of similar surface curvature. The mean position is relevant for generating the shape class description as well as for the feedback (subsections 3.3 and 3.4, respectively). The size enables a scale-variant classification and is obtained by an oriented bounding box.

For a compact description of the local shape part curvature we suggest principal curvature histograms, which will serve as curvature indicator and offer a simple method for curvature similarity measure by comparing the number of elements in analog histogram bins. As the point density on a surface - among others - depends on the distance to the sensor and the angle of incident, we normalise the bins by the total number of points in the segment. The resulting histograms can efficiently be represented by i) discarding the entries of a (normalised) bin if its value deceeds a mimimum threshold h_{binN}, and ii) storing only the continuous set of remaining filled bins, starting at the lowest and stopping at the highest non-empty bin. h_{binN} ensures to discard noisy values and to focus on the most prominent curvatures. A high threshold additionally facilitates comparison, as only a small number of bins has to be considered. For simplicity, we are neglecting the curvature orientations at the current state of research, although they might improve the overall result. For each homogeneous shape part of figure 2, the reduced representation of the maximum curvature is visualised.

Based on this object hypothesis description, a pair of hypotheses or an hypothesis and a shape class are considered to be similar if they comply with the following threshold-based rules: i) the deviation in size deceeds c_s, and ii) more than c_g percent of the overlapping neighbourhood graphs are similar. The latter rule states that more than c_g percent of the overlapping graphs contain nodes of similar curvature with matching adjacent nodes. Similar curvature is denoted by no more than c_b different bins in each corresponding pair of principal curvatures.

Object hypotheses which contain very few shape information, e.g. because they consist of a single shape part, are improved by adding information from subsequent scans of the (assumed) same real-world object. Restricting our approach to a static environment at the current state of our research, identification of hypotheses belonging to the same real-world object relies on the positions of their shape parts. Hypothesis enhancement involves detection and merging of similar parts, the first of which is covered by the hypothesis similarity detection; the latter is part of generating of shape class descriptions and detailed in the following subsection.

3.3 Shape Class Description

Combining a cluster of object hypotheses into a consistent shape class description significantly reduces the amount of data which has to be stored in order to retain meaningful shape information. Additionally, the fusion of multiple hypotheses enables a speed-up in classification due to a reduced number of shape comparisons. Presuming measurement noise to suffice a Gaussian distribution, the combined description is less susceptible to measurement noise whilst incorporating all distinct shapes of the object hypotheses assigned to this class and thus providing a compact representation of its homogenous parts. Furthermore, a shape class descriptor has to contain information on how relevant each of its surface parts is for the shape

class. Shape parts, which are quite similar in all object hypotheses assigned to a shape class, can be considered to contain relevant information of the overall shape and thus to be very descriptive for this class. In contrast, shape parts with a large variety of different shapes among the hypotheses of this class do not contain much information about the overall shape. Figure 3 shows four differently shaped objects belonging to a total of two object shape classes: primarily cylindrical bottles in the left and more free-form flasks in the right. Shape parts with matching curvature indicators are emphasised with solid lines of the same colour. Most pairs of shape parts can be considered as descriptive for their shape class except for the pair marked in Yellow whose curvatures in vertical direction differ significantly. Although keeping in mind the wide variety of object shapes and the resulting difficulty of precise boundaries of shape classes, our intention is to prove our approach in principle. Thus, we currently focus on a limited number of well-distinguishable shape classes like those shown in figures 2 and 3 permitting inner-class shape variations.

For combining shape descriptions we suggest to merge the points of the hypotheses with the 3dimensional Normal Distribution Transform (NDT) (Huhle et al., 2008) and to down-sample the resulting point cloud to the initial point density. For this new point cloud, homogeneous surface parts and a new shape parts graph have to be created. The shape class descriptor consists of the combined point cloud, the graph and, for each node in the graph, the number of object hypotheses involved. Thus, the description contains the combined shape information with reduced noise owing to matching and down-sampling. A new hypothesis can easily be integrated by repeating NDT merging while weighting the points of a class description part with its number of involved hypotheses and the points of the new hypothesis with 1. This ensures a true mean shape of the previously combined and the single new hypothesis.

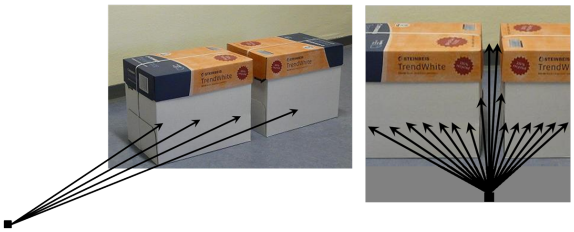

Figure 4: Relationships between incident angle of the sensor beam and point density at the surface. More dense surface sampling offers more reliable surface information.

The shape descriptiveness of a class part is reciprocal to its variety among a group of corresponding shape parts from different hypotheses and can be measured by the curvature variance, more specificly by the maximum deviation of the hypotheses shape parts to their corresponding part in the shape class descriptor. Applying the curvature indicator from subsection 3.2 for example would result in the maximum number of different bins for the minimum and the maximum curvature, respectively. This measure of shape part descriptiveness could be fed back into the hypothesis classification scheme to achieve a faster and more reliable classification, but is not integrated into our classification scheme at the present state of our research.

3.4 Classification Feedback

Perceptions from perspectives normal to the object surface in general contain the most reliable information. The more the perspective deviates from the surface normal, i.e. the more shallow the incident angle of e.g. the laser beam becomes, the lesser is the intensity of the reflected beam, which could lead to less reliable distance measurements. But more important, for more shallow angles the point density on the surface decreases significantly, as visualised in figure 4. In consequence perceptions from shallow incident angles contain considerable less information on the surface shape. Additionally, from a disadvantageous perspective, several object configurations might be percieved as a single object because the non-connected parts are not detected. On the one hand, sparse sampling at a shallow angle of incident could cause the erroneous seperation of a single object into several segments and thus under-segmentation. On the other hand, it could result in the erronous merging of multiple objects to a single segment, i.e. to under-segmentation, due to invisibility of non-connected parts. Both cases of segmentation faults would cause significant errors in the classification results and, as a consequence, in the shape class descriptors, if they were not detected and eliminated. For detection, we suggest to use the shape information of several perceptions of the (assumed) same real-world object from different perspectives redundant in most parts. As we restrict the environment to be static, perceptions of the same real-world object are those which cover the same global position. We suggest to evaluate differences in shape parts in combination with the reliability of the shape parts. In consistency with the above considerations, the steepness of the incident angle results in an approximation of the reliability of a surface part, which can be used for selecting the most reliable of several contradictory perceptions of the (assumed) same real-world object. If a less reliable shape contrasts a reliable one, the latter is chosen for further processing. Additionally,

the segmentation distance threshold s_d is adapted to prevent future false segmentations. As the reliability of a perception depends on its perspective, the new distance threshold s_{fd} is defined as a function of the perspective, strictly speaking, of the angle between surface normal and the connecting line of the position of the surface part and the sensor. Initialising s_{fd} with s_d for all angles, the value assigned to a certain angle has to be reduced in case of under-segmentation and to be increased in case of over-segmentation. Analogously, a threshold function h_{fd} has to be applied for object hypothesis generation instead of h_d.

4 FUTURE WORK

Currently, we are working on the object hypothesis generation, which partially includes the proposed segmentation. Thus, the main parts of the concept, class generation and feedback, have still to be set up, verified and evaluated. For the sake of simplicity, we will restrict our system to a static environment, which would not result in loss of generality. Additionally, we will start feedback verification with simply shaped objects like boxes and self-generation of object classes with a set of well distinguishable shapes with slight variations within each class, like those illustrated in figures 2 and 3. The final part of our future work will be the evaluation of the overall system. For further improvement of the classification scheme, it will include the feedback of shape part descriptiveness, derived while generating the shape class descriptions, into the hypothesis classification step.

5 CONCLUSIONS

In this paper, we introduced our idea for a joint segmentation and classification with feedback. We are confident that our approach can significantly contribute to a more robust as well as a more autonomous object classification thus overcoming traditional classification methods which either rely on an initial training set or some other specific information. As this kind of information has a priori to be provided by a human, the robot cannot act truly autonomously. Additionally, this kind of information is difficult to obtain even for humans, given the incompletely-known and ever-changing environment, in which the robot typically operates. In contrast, our approach autonomously generates object classes. Consequently, we expect a considerable improvement of autonomy due to our self-generating object classes fom environment perceptions based only on a simple set of rules. Furthermore, the joint segmentation and classification feeds back classification results into both the segmentation and the object hypothesis generation, and thus is able to prevent many cases of over- and under-segmentation, which typically occur due to incorrect assumptions and thresholds in both segmentation and hypothesis generation.

REFERENCES

Chen, Z. and Ellis, T. (2011). Multi-shape descriptor vehicle classification for urban traffic. In *Proceedings of Digital Image Computing: Techniques and Applications*. IEEE.

Endres, F., Plagemann, C., Stachniss, C., and Burgard, W. (2009). Unsupervised discovery of object classes from range data using latent dirichlet allocation. In *Proceedings of Robotics: Science and Systems*. The MIT Press.

Farmer, M. and Jain, A. (2004). A wrapper-based approach to image segmentation and classification. In *Transactions on Image Processing Volume 14 Issue 12*. IEEE.

Guyet, T., Garbay, C., and Dojat, M. (2007). A human-machine cooperative approach for time series data interpretation. In *Proceedings on Artificial Intelligence in Medicine*. IEEE.

Huhle, B., Magnusson, M., Strasser, W., and Lilienthal, A. (2008). Registration of colored 3d point clouds with a kernel-based extension to the normal distributions transform. In *Proceedings on Robotics and Automation*. IEEE.

Knopp, J., Prasad, M., and Van Gool, L. (2010). Orientation invariant 3d object classification using hough transform based methods. In *Proceedings of the ACM workshop on 3D object retrieval*. ACM.

Lai, K. and Fox, D. (2009). 3D laser scan classification using web data and domain adaptation. In *Proceedings of Robotics: Science and Systems*. The MIT Press.

Lecumberry, F., Pardo, A., and Sapiro, G. (2010). Simultaneous object classification and segmentation with high-order multiple shape models. In *Transactions on Image Processing Volume 19 Issue 3*. IEEE.

Marton, Z.-C., Pangercic, D., Blodow, N., and Beetz, M. (2011). Combined 2d-3d categorization and classification for multimodal perception systems. In *International Journal of Robotics Research*. SAGE Publications.

Schwartzkopf, W., Bovik, A., and Evans, B. (2005). Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images. In *Transactions on Medical Imaging Volume 24 Issue 12*. IEEE.

Triebel, R., Shin, J., and Siegwart, R. (2010). Segmentation and unsupervised part-based discovery of repetitive objects. In *Proceedings Robotics: Science and Systems*. The MIT Press.