GENERIC 3D OBJECT RECOGNITION FROM TIME-OF-FLIGHT IMAGES USING BOOSTED COMBINED SHAPE FEATURES

Doaa Hegazy, Joachim Denzler
Institute of Computer Science, Friedrich-Schiller-University in Jena , Ernst-Abbe-Platz 2, D-07743 Jena, Germany
{*hegazy , denzler*}*@informatik.uni-jena.de*

Keywords: Generic 3D object recognition, Object category database, Boosting, Range data.

Abstract: Very few research is done to deal with the problem of generic object recognition from range images. With the upcoming technique of Time-of-Flight cameras (TOF), for example the PMD-cameras, range images can be acquired in real-time and thus recorded range data can be used for generic object recognition. This paper presents a model for generic recognition of 3D objects from TOF images. The main challenge is the low resolution in space and the noise level of the data which makes careful feature selection and robust classifier necessary. Our approach describes the objects as a set of local shape specific features. These features are computed from interest regions detected and extracted using a suitable interest point detector. Learning is performed in a weakly supervised manner using RealAdaBoost algorithm. The main idea of our approach has previously been applied to 2D images, and, up to our knowledge, has never been applied to range images for the task of generic object recognition. As a second contribution, a new 3D object category database is introduced which provides 2D intensity as well as 3D range data about its members. Experimental evaluation of the performance of the proposed recognition model is carried out using the new database and promising results are obtained.

1 INTRODUCTION

Generic object recognition (i.e. object class recognition) has been an important topic of the computer vision research in recent years (e.g. (Fergus et al., 2003)). However, most of the successful approaches developed up to date have concentrated on the generic recognition of objects from 2D data, and very little attention has been paid to the use of 3D range data in this task.

Range images have the advantage of providing direct information about the shape of objects which makes them suitable for recognition of objects from their shape as well as 3D object recognition. Therefore, range data have been used mostly in *specific* 3D object recognition (e.g. (Hetzel et al., 2001)). The term *specific* object recognition means the recognition of a certain object, regarding only its own characteristics (e.g shape, color or texture) and at the same time the recognition model is not able to classify any new instance of the same visual class [1] of this object.

However, generic recognition of objects from their shape using range images is a difficult task. One reason for this is that surface shape representation is very important in a recognition procedure from range data but it is not clear which representation is more suitable for learning shapes of object classes. Moreover the currently available object category databases do not support the recognition of object categories using range images because they provide only 2D images of their object categories.

This paper has two main contributions. First, a novel 3D object category database is introduced. The database provides 2D/3D data about its object classes. The construction of the database is done using a 3D Time-of-flight PMD camera (Lange, 2000). Second, a

[1] Objects could be divided according to their real life visual appearance into visual classes or according to their function into functional classes. Generic object recognition concerns with recognizing object which belong to the same visual class.

recognition model for generic 3D objects from range images is presented. This model consists of three main steps. First, an affine interest point detector is applied to the intensity image to detect a set of interest regions. The detected interest regions are extracted together with their corresponding 3D depth data. Second, simple local surface shape features are computed from the extracted 3D regions. Finally, boosting, namely RealAdaBoost algorithm (Schapire and Singer, 1999), is used to learn these simple shape features for each class. The idea of the proposed model, which is combining and boosting interest point detector together with local descriptors for recognition, is normally used for the generic recognition tasks using 2D images and has never been used with range images.

The outline of the paper is as follows. The related work is summarized in section 2. Section 3 describes the new 3D object category database. The proposed generic 3D object recognition model is described and explained in section4. Experimental evaluations and results obtained are presented in section 5. Conclusions are finally drawn in section 6.

2 RELATED WORK

Most of the recent researches and approaches in generic object recognition have focused on modeling the appearance and shape variability of objects with limited number of changes in viewing point (e.g. (Fergus et al., 2003; Leibe et al., 2004)). One main reason is that most of the current object category datasets contain images with small variations in viewing point (e.g. Caltech 4 and UIUC cars). A small number of research have investigated the problem of generic 3D object recognition. One of these approaches is presented by Savarese and Fei-Fei (Savarese and Fei-Fei, 2007). In their approach, a model of an object category is captured by linking together diagnostic parts of the objects from different viewing points. These parts are large and discriminative regions of the objects and consists of many local invariant features. To form a model of the object class, the parts are connected through their mutual homographic transformation. The resulting model is a summarization of both appearance and geometry information of the object class. In addition to that, (Savarese and Fei-Fei, 2007) introduced a new 3D object dataset. However, the approach presented in this paper is totally different from the approach of (Savarese and Fei-Fei, 2007). The main difference is that range images are used in our proposed approach, which is not the case in (Savarese and Fei-Fei, 2007) as they use 2D images. Furthermore, only surface shape features are used here to represent the instances of the object classes while no appearance information is used.

Another approach, which is closer to the work presented in this paper, is described in (Ruiz-correa et al., 2003). The approach developed to recognize objects belonging to a particular shape class in range images. In their approach, first, shape class components are learnt and extracted from range images. Then, the spatial relationships among the extracted components are encoded using a shape representation called symbolic surface signature. This results in forming a shape class model that consists of three-level hierarchy of classifiers where the first two levels of the hierarchy extract the component and the third one verifies their geometric relationships. The dataset used for the purpose of learning and classifying the model is range images of objects made of clay. The dataset is then enlarged by applying deformations to the original clay objects to offer intra-class variabilities.

Although our proposed approach agrees with the approach of (Ruiz-correa et al., 2003) in that surface shape descriptors are used to represent the object classes in real range images, there exist main important differences between the two approaches. First, a combination of three different simple local surface features is used in our approach as a representation of the instance of the different object categories. Second, learning is performed here using boosting which is different from the learning technique, namely Support Vector Machines (SVM), used in (Ruiz-correa et al., 2003). Moreover, a dataset of real range images and of real different object categories is used in our approach. The dataset contains large intra-class as well as inter-class variabilities, so it is not necessary to apply any deformation to enlarge it.

3 3D OBJECT CATEGORY DATASET

An object category database of 936 2D/3D images (2D grayscale as well as range data) of 26 objects (36 images per object) is built using a 3D Time-of-Flight PMD camera (Lange, 2000). The objects are instances of three main visual categories (classes): cars, motors and animals. A fourth class is constructed to be used as a background or a negative class during training and testing. This background class consists of objects which are visually different from the objects instances of the three main classes.

Due to the difficulty to record different outdoor

views of real objects using the PMD camera [2], human made objects (toys) are used to build the database. The instances of each object class are chosen with different sizes and appearances to achieve large intra-class variability as much as possible.

3.1 Dataset Acquisition

A 3D PMD camera was fixed to a rigid stand about 1.1 meters from its base. A motorized turntable was placed about 2 meters from the base of the stand. It is noticed by experiments that, by placing the turntable closer than 2 meters from the camera, the resultant images contain inaccurate distance measurements[3]. The camera was set in a way that the objects appear in the center of the image when placed at the center of the turntable. White background was provided by placing the turntable in front of a white wall. The normal lighting condition of the room was used.

Each object was placed in a stable configuration at approximately the center of the turn table. The turntable was then rotated through 360 degrees about the vertical axis and 36 2D/3D images were acquired per object; one at every 10 degrees of rotation. Figure 1 shows different database images of the three classes.

4 A GENERIC 3D OBJECT RECOGNITION MODEL

In this section, the main idea of the proposed generic 3D object recognition model is explained. Figure 2 provides a semantic view of the main components of the proposed model.

4.1 Preprocessing and Interest Regions Detection

Preprocessing: The range data of a TOF chip (in this paper PMD) has statistical noise. In order to filter this noise and smooth the range data, a preprocessing step by applying median filter is first performed. Furthermore, an initial histogram normalization is applied to the PMD grayscale images to enhance their low contrast and improve the interest points detection process.
Interest Regions: An implementation of the Hessian affine-invariant region detector developed by (Mikolajczyk and Schmid, 2002) is used to detect and extract interest regions from the 2D grayscale images.

[2]Settings required to use a PMD camera make it difficult to acquire outdoors views of real objects.

[3]For this reason, the size of the objects within the images is relatively small.

4.2 Local Features Computation

Range images have the advantage of providing direct information about the shape of objects. Therefore, it is wise to make use of this advantage and give preference to features that capture different aspects of this shape. For this reason, shape-specific local feature histograms are used in our model. These features presented and used in (Hetzel et al., 2001) for the task of free-form specific 3D object recognition. The features are namely: pixel depth, surface normals and curvature. The main advantages of these features are that they are easy to calculate, robust to viewpoint changes and contain discriminative information (Hetzel et al., 2001).

4.2.1 Pixel Depth

The distance to the object provided by the PMD camera is the simplest available feature. Computing a histogram of pixel distances provides a simple feature which is invariant against translations and image plane rotations and at the same time gives valuable cues about the shape of the object. In this paper, a histogram of 64 bins of pixel distances is calculated and used.

4.2.2 Surface Normals

A representation of surface normals as a pair of two angles (ϕ, θ) in sphere coordinates is presented in (Hetzel et al., 2001). This representation is shown to spread over as possible of the available histogram range without having a bias for certain regions (Hetzel et al., 2001). The angles can be calculated as follows:

$$\phi = \arctan(\frac{n_z}{n_y}), \theta = \arctan \frac{\sqrt{(n_y^2 + n_z^2)}}{n_x} \quad (1)$$

A two dimensional histogram of size 8 x 8 bins of the of two angles is computed and used.

4.2.3 Curvature

The shape index representation depends on the surface curvature (Hetzel et al., 2001). Its representation is given as follows:

$$S_I = \frac{1}{2} - \frac{1}{\pi} * \arctan \frac{k_{max}(p) + k_{min}(p)}{k_{max}(p) - k_{min}(p)} \quad (2)$$

where $k_{max}(p)$ and $k_{min}(p)$ denoting the principle curvatures around the point p. The shape index S_I has

Figure 1: Example images of the database for the three used visual classes.

Figure 2: The proposed generic 3D object recognition model.

the range of $[0,1]$, and every distinct surface shape corresponds to a unique value S_I (except for planar surfaces, which is mapped to the value 0.5, together with saddle shapes) (Hetzel et al., 2001). A histogram of shape index of 64 bins is used.

4.3 Learning Model

The learning model is based on the AdaBoost with confidence-rated prediction algorithm (Schapire and Singer, 1999) (RealAdaBoost). RealAdaBoost takes a training set $I = \{I_1, ..., I_N\}$ and their associated labels $l = \{l_1, ..., l_N\}$, where N is the number of training images and $l_i = +1$ if the object in the training image I_i belongs to the class category and $l_i = -1$ otherwise.

Since more than one feature type is used, each training image I is represented by a set of features $\{F_{i,j}(t_{i,j}, v_{i,j}), j = 1...n_i\}$ where n_i is the number of features in image I_i, $t_{i,j}$ indicates the type of the feature (d for pixel depth, c for surface normals and s for shape index) and $\mathbf{v}_{i,j}$ is the feature vector. RealAdaBoost algorithm puts weights on the training images and requires construction of a weak hypothesis h_k which, relative to the weights, has discriminative power. The algorithm is run for a certain number of iterations T. In each iteration k, one weak hypothesis is selected and the weights of the training images are updated. A linear combination of the weak hypotheses together with their weights is used as a strong hypothesis to classify new images.

5 EXPERIMENTS AND RESULTS

Two sets of experiments are performed to validate the proposed recognition approach. These two experiments allow to investigate the categorization ability of the approach as well as its performance with respect to clutter and occlusion. The first set of experiments considers scenes with single class member while the second one considers scenes with multiple objects containing background clutter and occlusion. Training the model is performed only once using images containing a single class member. Due to the lack of established research in generic 3D object recognition, it is difficult to obtain a standard dataset to compare the results with. Therefore, all experiments are performed using our 3D object category dataset. A total number of 200 images is used for training the model: 100 training images of a randomly selected instances of each object class in addition to 100 training images of the background class. RealAdaBoost algorithm is run for $T = 150$ iterations [4]. The model's performance is evaluated using the Receiver-Operating-Characteristic curve (ROC). Moreover, The ROC-equal-error rate is computed for each curve. This error rate gives a nice trade-off value between the true positives and false positives and is defined as the point on the ROC curve where the true positive rate = 1-false positive rate.

5.1 Experiment 1: Categorization Performance

In this set of experiments, the categorization ability of the recognition model is investigated. A test set of 100 images is used: 50 images of a novel instances of each object class and 50 images of the background class. Figure 3 displays the ROC curves for each object class while the ROC-equal-error rates are presented in table 1. The model achieves a high categorization performance on the three used object class. Although the used range images do not contain complex scenes, some difficulties are imposed on the recognition task due to the small size of objects in the images. Detailed variations between different object classes are not clear which makes categorization a hard task even for humans (see figure 1).

[4] We conclude this number by experiments where T is varied from 10 to 300. After T = 150, the test error remains constant.

Table 1: ROC-eqq-err rates of the categorization performance of the used three object classes.

Object class	ROC-equal-error
Cars	0.02
Motors	0.02
Animals	0.00

Figure 3: The ROC curves of the three classes on the categorization task.

Figure 4: Example of the images recorded for the task of categorization in complex scenes.

Figure 5: The ROC curves of the three classes on the categorization with the presence of clutter and occlusion task.

Table 2: ROC-eqq-err rates of recognition using complex scenes for the used three object classes.

Object class	ROC-equal-error
Cars	0.18
Motors	0.20
Animals	0.20

5.2 Experiment 2: Categorization in Complex Scenes

A new set of test images for each object class is recorded for this set of experiments (see figure 4) . These new test images contain occlusion and clutter by placing instances of each object class (different from the instances used in training) together with instances of new previously unused object classes. A total of 36 range images from different view points are then recorded for each object class. The ROC curves are shown in figure 5 and the ROC-equal-error rates are displayed in table 2.

Obviously, the performance in these experiments degrades than the previous experiments due to the presence of occlusion and clutter. Beside that, the low resolution of the intensity images of the PMD camera affects the detection performance of the point detector which influences in turn the categorization performance. Another important aspect concerning the recognition model is the computational time needed for the training the testing processes. The average training time of the model is approximately 26 minutes for each object class while the test time for a whole test set is approximately one minute for each class.

6 CONCLUSIONS

This paper has presented two contributions. First, a database for generic 3D object recognition has been presented. It has the advantage of providing range data as well as intensity information recored by a Time-of-Flight device of its object classes. The database will be made available for the public comparison of different approaches [5]. Second, a model for generic 3D object recognition from range images has been proposed. The main idea of the model is simple and has never been applied to range images before. The proposed model describes the objects as a set of simple local surface shape features computed from interest regions detected by a region detector. Learning is done using RealAdaBoost algorithm. Experiments have been performed using the new presented database and promising results have been obtained.

However, many improvements could be applied to the model in order to obtain better performance in the future. One of these improvements is the use of a point detector which is applied directly to range images (3D point detector). Another important issue is improving the quality of the intensity images delivered by the PMD camera by combining it with a high resolution 2D camera.

Finally, the extension of the 3D object category database by adding more object categories and providing high resolution intensity and color data about them, in addition to the 3D data, is an important step for the future work.

REFERENCES

Fergus, R., Perona, P., and Zisserman, A. (2003). Object Class Recognition by Unsupervised Scale-Invariant Learning. In *IEEE Computer Society Conference on computer vision and Pattern Recognition CVPR3*, volume 2, pages 264–271.

Hetzel, G., Leibe, B., Levi, P., and Schiele, B. (2001). 3d object recognition from range images using local feature histograms. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, volume 2, pages 394–399.

Lange, R. (2000). *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*. PhD thesis, University of Siegen.

Leibe, B., Leonardis, A., and Schiele, B. (2004). Combined object categorization and segmentation with an implicit shape model. In *In ECCV workshop on statistical learning in computer vision*, pages 17–32.

[5] Address of the database web page, removed for blind review.

Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In *7th European Conference on Computer Vision ECCV02*, pages 128–142.

Ruiz-correa, S., Shapiro, L. G., and Meil, M. (2003). A new paradigm for recognizing 3-d object shapes from range data. In *Proceedings of the IEEE Computer Society International Conference on Computer Vision 2003, Vol.2*, pages 1126–1133.

Savarese, S. and Fei-Fei, L. (2007). 3d generic object categorization, localization and pose estimation. In *ICCV07*, pages 1–8.

Schapire, R. E. and Singer, Y. (1999). Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37:297–336.