3D Object Recognition from Range Images using Local Feature Histograms

Günter Hetzel[1] Bastian Leibe[2] Paul Levi[1] Bernt Schiele[2]

[1] IPVR
University of Stuttgart
D-70565 Stuttgart, Germany
{hetzel, levi}@informatik.uni-stuttgart.de

[2] Perceptual Computing and Computer Vision Group
ETH Zurich
CH-8092 Zurich, Switzerland
{leibe, schiele}@inf.ethz.ch

Abstract

This paper explores a view-based approach to recognize free-form objects in range images. We are using a set of local features that are easy to calculate and robust to partial occlusions. By combining those features in a multidimensional histogram, we can obtain highly discriminant classifiers without the need for segmentation. Recognition is performed using either histogram matching or a probabilistic recognition algorithm. We compare the performance of both methods in the presence of occlusions and test the system on a database of almost 2000 full-sphere views of 30 free-form objects. The system achieves a recognition accuracy above 93% on ideal images, and of 89% with 20% occlusion.

1. Introduction

Recognition of free-form objects from range data is a challenging problem. Segmentation is ill-defined for arbitrarily curved surfaces, and the computational effort necessary to compensate for this is prohibitive for real-time applications. Even when only one object is present in the image, most real range images contain erroneous regions resembling shadows and self-occlusions (Figure 1). These artifacts are due to inherent limitations of current triangulating scanning techniques and cannot be avoided. Practical object recognition systems that work on range images must therefore be robust to occlusions.

Many classic approaches to object recognition use methods that are either global, like eigenpictures [11] or eigenshapes [1], or that rely on an initial segmentation of the object [6, 20, 3]. Those methods obtain good results on clean images, but their reliance on global properties makes them vulnerable to occlusions. A notable exception are Johnson's and Hebert's "spin images" [7], object-centered local histograms of surface locations, which have been shown to yield good results with cluttered or occluded objects. As this method is based on finding correspondences between image and model regions, it is rather time intensive, though. [2] gives a good overview about current global and local approaches on range images.

Figure 1. *(left) Ideal range image of a rubber duck, (right) real scan with self-occlusion.*

On color and greyvalue images, segmentation-free approaches have been very sucessful in dealing with occlusions. In recent years, several approaches based on local features have been proposed, using color histograms [17], local feature vectors at key points [14, 16], local gradient histograms [15], surface shape histograms [19], curvatures [5, 12], local appearance [4], or curve segments [13]. Local feature histograms in particular have been shown to provide a powerful probabilistic framework that can handle occlusions very well [15].

This motivates to explore how local feature histograms from range images can be used for efficient object recognition. We are particularly interested in the behavior with missing sensor values, which we simulate to varying degrees in order to assess the quality of the recognition method.

Of course, as range images have different properties, different features are needed. The following section discusses appropriate local range data features and how they can be adapted for the use in histograms. Section 3 introduces two recognition methods for histograms, and experimental re-

0-7695-1272-0/01 $10.00 © 2001 IEEE

sults presented in section 4 show that the resulting feature histograms allow fast and accurate recognition under varying degrees of occlusion. A discussion of the results and of future additions concludes the paper.

2. Feature Analysis

Most approaches on greyvalue images use various Gaussian derivatives for recognition [14, 16, 15]. Range images, on the other hand, can provide much more detailed information about the object's shape. We should therefore give preference to features that capture different aspects of this shape.

In the following, we analyze three shape-specific local features: pixel depth, surface normals, and curvatures. Our goal is to find features that are easy to calculate, robust to viewpoint changes, and that contain discriminant information. We will show that the three features mentioned above fulfill these criteria. In addition, we will demonstrate how they can be represented in histograms.

2.1. Pixel Depth

Pixel intensities are the simplest available feature. For greyvalue images, they depend on illumination and are thus not very useful for recognition. For range images, however, the intensity value corresponds directly to the distance to the object. The distribution of these distances can provide valuable cues about the object's shape.

Histograms of pixel distances are invariant against translations and image plane rotations. Normalization of the range values to the interval [0,255] makes them invariant to scale but very sensitive to the perceived depth range. For this reason, distance histograms should only be relied on for surfaces with sufficient depth range. Distance histograms are also problematic in situations where the depth range can be influenced by other objects or background clutter and should not be used in applications where these situations can occur.

2.2. Surface Normals

Surface normals can be easily calculated from first derivatives of the image. After the usual normalization, only two components of the resulting vector are relevant. We therefore have to search for a two-dimensional representation that is spread over as much as possible of the available histogram range without exhibiting a bias for certain regions.

Our previous research has shown that a representation as a pair of angles (ϕ, θ) in sphere coordinates fulfills both of these requirements [10]. The angles can be calculated as follows:

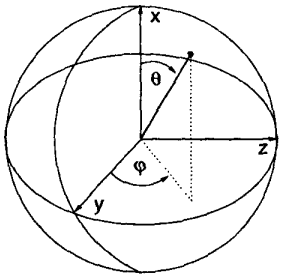

Figure 2. *Representation of normals in sphere coordinates.*

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

2.3. Curvature

Surface curvature can be calculated either directly from first and second derivatives, or indirectly as the rate of change of normal orientations in a certain local context region. The usual pair of Gaussian curvature K and mean curvature H only provides a very poor representation, since the values are strongly correlated [8, 12]. Instead, we use them in the form of the "shape index", introduced by Koenderink and modified by Dorai and Jain [8, 5, 12]:

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

with $k_{min}(p)$ and $k_{max}(p)$ denoting the principal curvatures around the point p. The shape index S_I has the range $[0, 1]$, and every distinct surface shape corresponds to a unique value of S_I (except for planar surfaces, which will be mapped to the value 0.5, together with saddle shapes). The shape index is invariant to translations, but due to the limited resolution, noise is introduced in the presence of image plane rotations and scale changes.

3. Histogram comparison methods

Each of the features described in the previous section captures a specific aspect of a small region on the object's surface. By combining them in a multidimensional histogram, we can directly model the probability distribution of different feature combinations and thus of certain shape patches.

Given the distributions for all objects in the database, the recognition problem reduces to the task of finding the distribution that best explains the measurements taken from the test object. This section presents two methods for this

purpose: histogram matching, and a maximum a posteriori probability estimation. As our tests in the following section show, both methods produce comparable results on ideal test images, but they differ in their capability to deal with partial occlusions.

3.1. Histogram Matching

The main motivation of histogram matching for object recognition is its low computational cost. Since similar shape patches are always assigned to the same histogram cells, there is no need to solve a correspondence problem. Instead, we can evaluate the contents of corresponding cells by calculating a comparison measure.

The formal statistical method for assessing the dissimilarity between two probability distributions is the χ^2-test [15]. Starting from the null hypothesis that two data sets are drawn from the same distribution, the goal is to disprove the hypothesis. Since we typically do not assume exact knowledge of the model distribution, we employ a modified version which compares two observed histograms Q and V. The first step is to calculate the χ^2-divergence:

$$\chi^2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \tag{3}$$

Based on this result, the χ^2-test, as described in the statistics literature, requires the evaluation of a significance estimate. In the context of object recognition, this is hardly ever done, since viewpoint changes and visibility constraints usually lead to very low significance values. For this reason, we compare histograms only based on the χ^2-divergence.

In addition, we want to mention two other comparison measures that are often used in the literature. The intersection measurement, introduced by Swain and Ballard for the comparison of color histograms [17], provides a very fast and easy way to quantify the common parts of two histograms. The intersection of two histograms V and Q is defined as:

$$\cap(Q,V) = \sum_i min(q_i, v_i) \tag{4}$$

Another popular measure, often used in information theory, is the Kullback-Leibler divergence [9]. In its symmetric version, it is defined as:

$$KL(Q\|V) = \sum_i (q_i - v_i) \ln \frac{q_i}{v_i} \tag{5}$$

Since this measure requires the calculation of a logarithm for every comparison of histogram cells, it is much slower than the other two variants, though.

In an initial series of tests, we experimented with all three measures, but did not find a significant difference in recognition performance. In the following, we report only the results of the χ^2-divergence because of its statistical relevance.

3.2. Probabilistic Recognition

Simple histogram matching is still a very coarse recognition method. Its main two deficiencies are that it cannot deal with partial occlusions too well, and that the usual χ^2 significance estimate fails when we compare slightly shifted histograms (for example resulting from viewpoint changes). These estimates are necessary if we want to combine different feature channels. A probabilistic approach, as described in Schiele's work [15] can provide much better results.

The main idea is that we no longer calculate an abstract distance measure between two histograms, but instead we directly estimate the posterior probability of an object hypothesis, given a particular set of independent measurement vectors $m_1, ..., m_k$. Using the Bayesian theorem, and assuming that all objects are equally likely, we obtain:

$$p(o_n | \bigwedge m_k) = \frac{\prod_k p(m_k \mid o_n)}{\sum_i \prod_k p(m_k \mid o_i)} \tag{6}$$

where $p(m_k|o_n)$ designates the likelihood of measurement vector m_k given the object o_n. This probability can be obtained from the histogram saved for o_n.

As this technique directly calculates the posterior probability of an object given the data, it not only allows to determine the best recognition result. It also provides a reliable confidence estimate specifying how much this result can be trusted. This is especially important in applications where false positives must be avoided, or where the recognition result shall be combined with other channels (like color information) for a more reliable recognition.

4. Experimental Results

In order to assess the quality of the proposed features and comparison methods, we have conducted a series of experiments with different feature combinations and histogram resolutions.

Our test database consists of 30 free-form objects (Figure 3). Because of the huge effort necessary to obtain full-sphere range images of real objects, we have decided to create the images synthetically. One of the advantages of range imagery is that accurate polygonal representations of 3D objects can be obtained from relatively few (10-15) scans [18]. By rendering these models into a depth buffer, we can get range images from arbitrary viewpoints.

The use of synthetic data also allows to simulate the influences of varying degrees of occlusion and missing data on recognition performance. In order to simulate self-occlusion, we block a certain part (20%-80%) of the image and only collect feature vectors from the remaining regions.

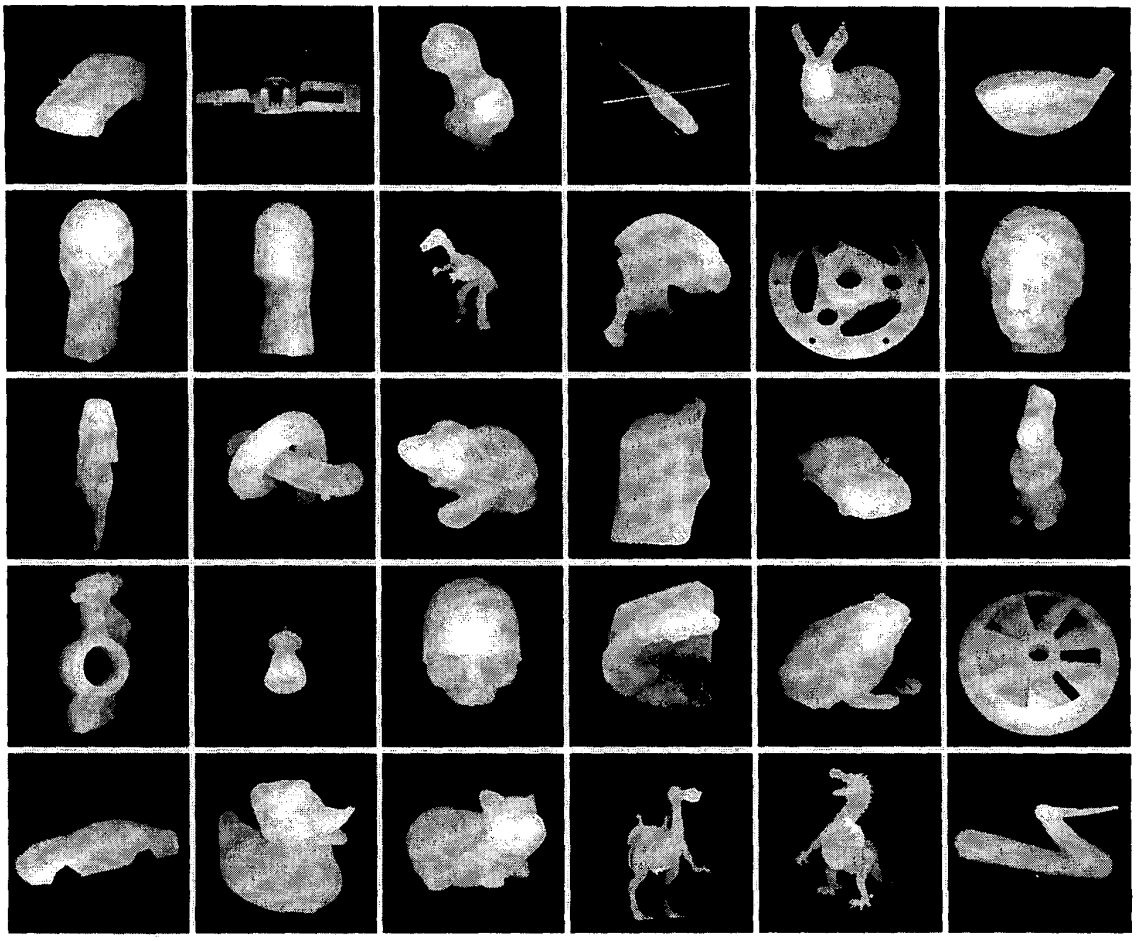

Figure 3. *The 30 objects from the test database (the complete database with 258 full-sphere views of each object is available at http://range.informatik.uni-stuttgart.de).*

Note that this method is not a sufficient simulation of real occlusion effects with other objects, which would add their own characteristics to the histogram representation. However, it is very similar to the type of measurement errors that occur in range images (as detailed in the introduction).

The training set contains 1980 images, 66 from each of the 30 objects, distributed evenly over the whole viewing sphere, with angles of $23 - 26°$ between viewpoints. The system is then tested on the 192 views per object lying halfway between the training views, for a total of 5760 images in the test set. This corresponds to a test under viewpoint shifts of $11.5 - 13°$. All histograms are normalized to a uniform sum in order to compensate for differently sized objects.

4.1. Tests on Ideal Range Images

In a first test on a subset of 20 objects, we compared the performance of pixel depth, normals, and shape index alone using the χ^2 divergence (Table 1). The high discrimination capabilities of these features can be observed from the result that both normals and shape index are sufficient to correctly recognize about 80% of the objects. With only 43% recognition, the pixel depths are not nearly as good.

However, this changes when they are combined with normals in a second experiment on the full database of 30 objects (Table 2). This combination is able to achieve over 89% recognition with a very small histogram size (only 128 cells). Taking into account the relatively large spacing of the viewpoints, this is a very good result. A further increase

f.	h. size	recog.	(1-3)	pose est.	(1-3)
d	32	43.80%	58.59%	21.43%	36.67%
n	8-8	80.60%	89.56%	27.60%	51.28%
s	64	82.55%	91.22%	39.97%	66.85%
s+d	16-16	80.05%	89.24%	19.67%	40.44%

Table 1. *Recognition and pose estimation results of pixel depths (d), normals (n), and shape index (s) with first and best 3 matches (20 objects). Only the best histogram resolutions are shown.*

f.	h. size	recog.	(1-3)	pose e.	(1-3)
n+s	8-8-16	89.24%	94.25%	74.57%	87.97%
n+d	4-4-8	89.20%	95.24%	76.18%	89.18%
nsd	4-4-8-8	93.18%	97.26%	80.94%	92.15%

Table 2. *Recognition results of higher-dimensional combinations of all three features with best histogram sizes (30 objects).*

in performance to over 93% recognition can be obtained by combining all three features. In these tests, our prime interest was in recognition performance. The results indicate, however, that a quite reliable pose estimate can be obtained as a nice by-product. The pose estimation scores are not accurate, though, since there are many unaccounted symmetries in our test database.

From the analysis in section 2, we know that intensities and shape index are best suited for different kinds of images. By trying the two combinations "normals + depths", and "normals + shape index" in parallel and assuming a perfect decision strategy to pick the best answer from the two, we can get a recognition rate of up to 94.9%. Finding such a decision strategy is a problem in itself, but the result indicates that these two feature combinations can form a good supplement and compensate for one another's individual weaknesses.

In general, our results are comparable to the ones reported by Campbell and Flynn for an eigenshape-based system [1]. They obtained similar recognition accuracies on a smaller database containing 20 free-form objects, but with a larger viewpoint spacing of about 32° between views. However, as we will show in the next section, our system is also very robust to missing data and still gives good results when only a small portion of the object is visible.

4.2. Tests with Occlusions

As already mentioned in the introduction, robustness to occlusion is a vital characteristic for any recognition method which shall be applied to real range images. In order to measure the influence of occlusion and compare its impact on the recognition performance of the different

visible	χ^2	χ^2 (1-3)	prob.	prob. (1-3)
100%	93.58%	98.23%	92.36%	97.53%
80%	86.77%	95.03%	89.13%	95.90%
60%	61.77%	81.07%	78.99%	90.69%
40%	27.81%	45.52%	59.55%	79.17%
20%	11.46%	23.85%	31.88%	52.01%

Table 3. *Recognition performance relative to the visible object portion.*

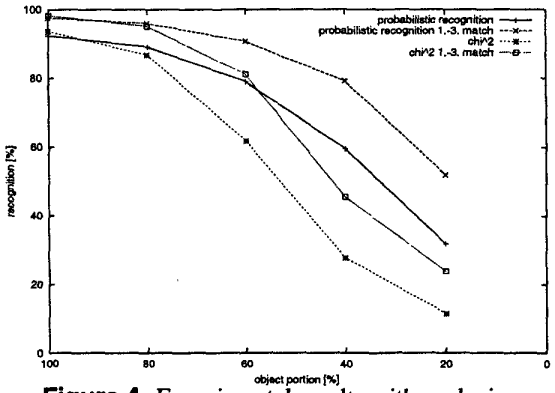

Figure 4. *Experimental results with occlusion.*

methods, we designed an additional experiment.

We used the complete training set (30 objects, 1980 images), but only a reduced test set of 15 randomly chosen objects (2880 test images). For each test image, we varied the visible object portion from 100% down to 20% and recorded the recognition results using the χ^2-divergence and the probabilistic recognition algorithm. All tests were done with a combination n-s-d of all three features and a histogram resolution of 4-4-4-4.

Figure 4 and table 3 summarize the recognition results for different visible object portions. With no occlusion the χ^2-test is still the the best recognition method. But this changes rapidly as the object is successively occluded. If less than 80% of the object is visible, the probabilistic recognition clearly obtains better results than the χ^2-divergence.

Using only 40% of the object surface, almost 80% of the test images are still recognized within the first 3 matches. Even if the system can only observe about 20% of the object surface, it still manages to recognize more than half of the test examples in the first 3 tries. This confirms that the probabilistic recognition is capable of reliable recognition in the presence of occlusion.

5. Discussion and Conclusions

An interesting result of the experiments is that good results can be obtained with quite small histograms. Using the

combination of normals and intensities, we can get a recognition rate of 89% with only 128 histogram cells. Thus, a whole object with its 66 training views can be represented by only $128 * 66 = 8448$ real values – significantly less space than is needed for the thumbnail image to visualize the object!

With the small histogram sizes shown in the table, the system is also very fast. Using 256-cell histograms and the χ^2-divergence, for example, it takes only 0.1 CPU seconds to match a test image with the 1980 histograms in the database on a Sun Blade 1000 (600MHz). The intersection measurement is even faster and needs only 0.026 seconds. The probabilistic recognition method is not as fast as χ^2 or intersection, but with a recognition time of about 1 second to compare one image with the whole training set, it is still fast enough for most applications. In addition, its runtime mostly depends on the number of measurement vectors, not on the size of the histogram. Since only very simple features are calculated, the total runtime of the recognition system is suitable for real-time applications with all three recognition methods.

The test results show that the system can handle self-occlusion well. Even with 20% occlusion, the correct solution was within the first three results in over 95% of the test cases. From this, we expect that our system will be usable for recognition from real range images. In future work, we will concentrate on this aspect and test the system on real data.

We also estimate that we can further improve the recognition performance by employing an additional verification stage using a simple hypothesis checker.

Finally, we want to point out that the probabilistic recognition provides a principled way for obtaining reliable confidence estimates that allow the combination with other recognition channels, like color or greyvalue information. Many modern 3D scanners already offer both depth and photometric information, so a combination of these two imaging modalities would be a natural extension. The system presented here supports such a combination, and it will be rewarding to explore how it can be used to get a more reliable recognition.

Acknowledgements

Günter Hetzel's research was supported by the German Research Foundation under research grant DFG-SFB514. Bastian Leibe's research is part of the CogVis project, funded in part by the Comission of the European Union under contract IST-2000-29375, and the Swiss Federal Office for Education and Science (BBW 00.0617).

We are grateful to Greg Turk of Georgia Tech, the Georgia Tech Rapid Prototyping and Manufacturing Institute (RPMI), the Avalon public 3D archive (avalon.viewpoint.com), Cyberware (www.cyberware.com), and 3D Cafe (www.3dcafe.com) for making their models available online.

References

[1] R. Campbell and P. Flynn. Eigenshapes for 3d object recognition in range data. In *CVPR'99*, pages 505–510, 1999.

[2] R. Campbell and P. Flynn. A survey of free-form object representation and recognition techniques. *CVIU*, 81(2):166–210, 2001.

[3] O. Camps, C.-Y. Huang, and T. Kanungo. Hierarchical organization of appearance-based parts and relations for object recognition. In *CVPR'98*, pages 685–691, 1998.

[4] V. C. de Verdiere and J. Crowley. Visual recognition using local appearence. In *ECCV'98*, 1998.

[5] C. Dorai and A. Jain. Cosmos - a representation scheme for free-form surfaces. In *ICCV'95*, pages 1024–1029, 1995.

[6] R. Hoffman and A. Jain. Segmentation and classification of range images. *Trans. PAMI*, 9(5), 1987.

[7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *Trans. PAMI*, 21(5):433–449, May 1999.

[8] J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–565, 1992.

[9] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[10] B. Leibe, G. Hetzel, and P. Levi. Local feature histograms for object recognition from range images. Technical Report 6/2001, University of Stuttgart, http://www.informatik.uni-stuttgart.de/cgi-bin/NCSTRL_view.pl?id=TR-2001-06, August 2001.

[11] H. Murase and S. Nayar. Visual learning and recognition of 3d objects from appearance. *IJCV*, 14:5–24, 1995.

[12] C. Nastar. The image shape spectrum for image retrieval. Technical report, INRIA Rocquencourt, 1997.

[13] R. Nelson and A. Selinger. A cubist approach to object recognition. In *ICCV'95*, pages 614–621, 1998.

[14] R. N. Rao and D. Ballard. Object indexing using an iconic sparse distributed memory. In *ICCV'95*, pages 24–31, 1995.

[15] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *IJCV*, 36(1):31–52, 2000.

[16] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. In *CVPR'96*, 1996.

[17] M. J. Swain and D. H. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.

[18] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceedings of SIGGRAPH '94*, pages 311–318, 1994.

[19] P. Worthington and E. Hancock. Object recognition using shape-from-shading. *Trans. PAMI*, 23(5):535–542, May 2001.

[20] J. Yi and D. Chelberg. Model-based 3d object recognition using bayesian indexing. *CVIU*, 69(1):87–105, 1998.