

Local Feature Histograms for Object Recognition from Range Images

B. Leibe¹, G. Hetzel², and P. Levi²

¹ Perceptual Computing and Computer Vision Group, ETH Zurich
Haldeneggsteig 4, CH-8092 Zurich, Switzerland

leibe@inf.ethz.ch

² University of Stuttgart, IPVR
Breitwiesenstr. 20-22, D-70565 Stuttgart, Germany
{hetzel,levi}@informatik.uni-stuttgart.de

Abstract. In this paper, we explore the use of local feature histograms for view-based recognition of free-form objects from range images. Our approach uses a set of local features that are easy to calculate and robust to partial occlusions. By combining them in a multidimensional histogram, we can obtain highly discriminative classifiers without having to solve a segmentation problem. The system achieves above 91% recognition accuracy on a database of almost 2000 full-sphere views of 30 free-form objects, with only minimal space requirements. In addition, since it only requires the calculation of very simple features, it is extremely fast and can achieve real-time recognition performance.

Key Words. 3D object recognition, range images, histograms

1 Introduction

Range images are a valuable supplement to other information channels for object recognition. They reveal direct, illumination independent information about an object's 3D surface shape and make the figure-ground segmentation considerably easier. Due to their inherent geometric nature, research has mainly focused on using range images for scenery reconstruction [1], or recognition of geometric objects [2, 3]. The recognition process usually consists of a series of preprocessing steps to extract edges, segment the object into consistent patches, and search for correspondences with a known object model [4–6]. Many methods have been designed to improve the quality of the intermediate steps [6, 7], and good results have been achieved for regular objects [2, 3]. However, free-form objects still pose significant problems. For them, edge extraction is very difficult, and segmentation results may vary largely from one view to the next. The computational effort necessary to compensate for these effects is prohibitive for real-time applications.

Appearance-based approaches have been very successful in dealing with this problem on color and greyscale images. In recent years, several approaches that

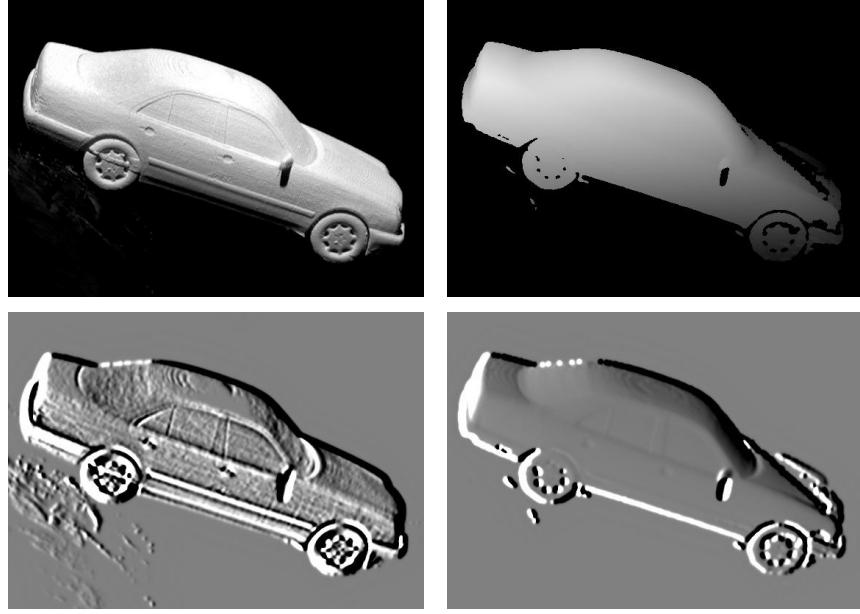


Fig. 1. A greyvalue and a range image of the same car (top), and the results when a gradient operator is applied to them (bottom). In the range image, the surface structure is almost completely lost.

work without segmentation have been proposed, using color histograms [8], eigen-pictures [9], local feature vectors [10, 11], gradient histograms [12], local curvatures [13], or curve segments [14]. Local histograms in particular have been shown to allow a powerful probabilistic framework that can achieve real-time performance, even under realistic viewing conditions [12].

This motivates us to explore how local feature histograms can be used for range images. As range images have different properties, different features are needed. Current range sensors have problems with transparent or reflecting materials and do not work equally well for all surface orientations. As a result, most real range images contain error regions resembling shadows and partial occlusions. The following section will analyze which features can be used under these circumstances and how they can be adapted for the use in histograms. We will then present experimental results to show that our chosen feature histograms allow fast and accurate recognition (Section 3). A discussion of our work and of future additions will conclude our work.

2 Feature Analysis

Most approaches on greyvalue images use Gaussian derivatives, alone or in combinations, for their recognition [10–12]. However, Figure 1 shows that this does

not work on range images. Instead, we can make use of the advantages of range images, namely that they provide direct information about the object's shape. We should therefore give preference to features that capture different aspects of this shape.

In the following, we will analyze three shape-specific local features: intensities, 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 Intensities

Pixel intensities are the simplest available feature. For greyvalue images, they are largely illumination dependent and thus not very useful for recognition. For range images, however, the intensity value corresponds directly to a distance to the object. The intensity distribution of an object can therefore provide valuable cues about its shape.

Intensity histograms are invariant against translations and image plane rotations. Since range images are often normalized to the range [0,255] in order to achieve scale invariance, they can be very sensitive to the perceived depth range, though. If there are large and abrupt changes in the depth range, e.g. due to occlusion effects, the whole histogram will be shifted and recognition might no longer be guaranteed. For this reason, intensity histograms can only be relied on for the recognition of surfaces with sufficient depth range. This condition can be easily detected, and if other features are available, they can take over in those cases.

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 having a bias for certain regions.

There are two main representations commonly used for this purpose. The first just discards the z-component of the normal vector and represents it as a pair (x, y) . This corresponds to a projection of the orientation hemisphere on the unit circle (see Figure 2a). The second possibility is a representation as a pair of angles (ϕ, θ) in sphere coordinates, as shown in Figure 2b. The angles can be calculated as follows:

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

This representation is harder to calculate and responds nonlinearly to image plane rotations. But, as can be seen from Figure 3, it has a larger usable his-

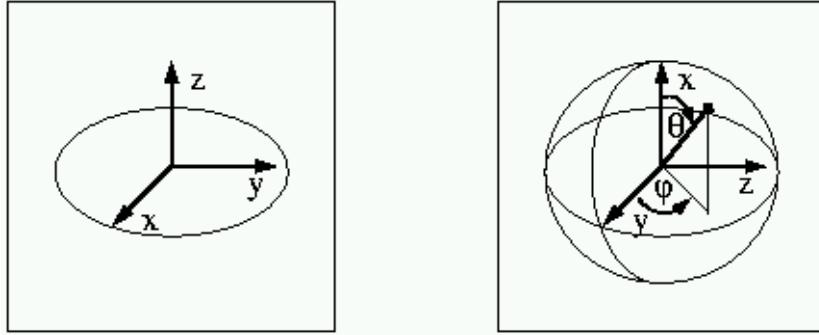


Fig. 2. Representation of normals in projective (left) and sphere coordinates (right)

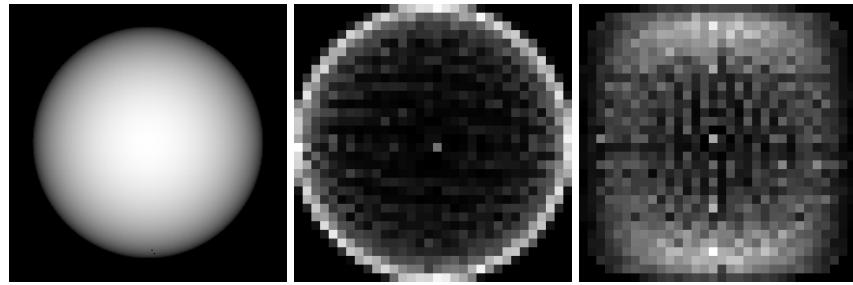


Fig. 3. In order to compare the projective and sphere coordinates, we took a range image of a sphere (left), and calculated the histograms for projection (middle) and sphere coordinates (right). The latter has a larger range and a more even distribution.

togram range and can provide a much more uniform distribution over the histogram cells. For this reason, we will only consider the (ϕ, θ) representation in our experiments.

2.3 Curvature

Surface curvatures 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 [15, 13]. Instead, we will use them in the form of the "shape index", introduced by Koenderink and modified by Dorai and Jain [15, 16, 13]:

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

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, it will vary with image plane rotations and scale changes. In addition, the reliance on second-order derivatives makes it very sensitive to noise, in particular on regular surfaces with only little local shape. Thus it works much better for free-form than for geometric objects.

3 Test Results

In order to evaluate the quality of the proposed features, we have conducted a series of experiments with different feature combinations and histogram resolutions. Since our goal was to find out how well the features were suited to the recognition task, we have used, at this stage, only a simple recognition strategy of histogram matching using the well-known χ^2 divergence [12] between two histograms Q and V :

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

Our test database consists of 30 free-form objects¹ (Figure 4). 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 very accurate polygonal representations of 3D objects can be obtained from relatively few (10-15) scans [1]. By rendering these models into a depth buffer, we can get range images from arbitrary viewpoints. All test images in this work are ideal scans, without shadows and occlusions. A future version of our work will examine the influences of realistic occlusions and compare the results to those obtained from real images.

The training set contains 1980 images, 66 from each of the 30 objects, distributed evenly over the whole viewing sphere with angles of $23 - 26^\circ$ 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. All histograms are normalized to a uniform sum in order to compensate for differently sized objects.

In a first test on a subset of 20 objects, we compared the performance of intensities, normals, and shape index alone (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 image intensities are not nearly as good. However, this changes when we combine them with normals in a second experiment on the full database of 30 objects (Table 2). This combination is able to achieve over 91% recognition² with a very small histogram size (only 128 cells). Taking

¹ The complete database with over 10000 range images is available at <http://range.informatik.uni-stuttgart.de>

² 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.

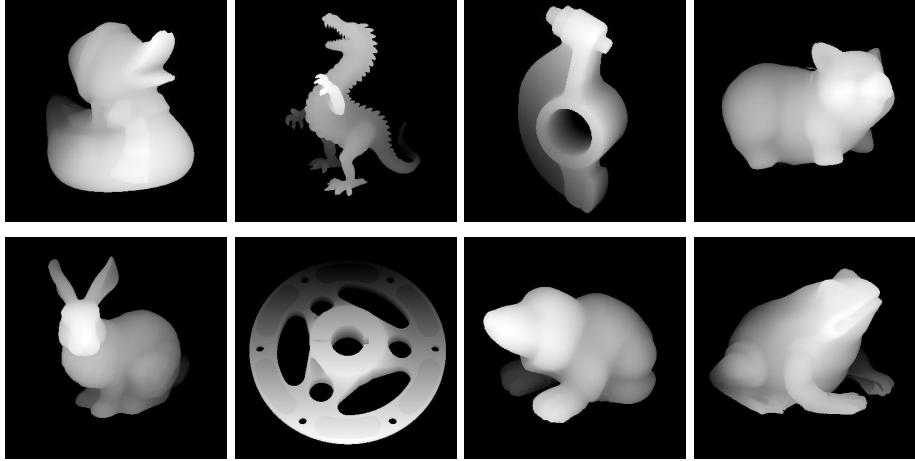


Fig. 4. Some examples of objects from the test database

features	histogram size	identification (1-3)	pose estimation (1-3)
i	32	43.80%	58.59%
n	8-8	80.60%	89.56%
s	64	82.55%	91.22%
$s + i$	16-16	80.05%	89.24%

Table 1. Recognition results of intensities (i), normals (n), and shape index(s) with first and best 3 matches (20 objects). Only the best histogram resolutions are shown.

into account the relatively large spacing of the viewpoints, this is a very good result. Compared to this, the combination of all three features brought only a minor increase in performance to 93%.

From the analysis in section 2, we know that intensities and shape index are best suited for different kinds of images. By taking only the best results from the two combinations “normals + intensities”, and “normals + shape index”, we can get a recognition rate of up to 94.9%. This indicates that these two feature combinations can form a good supplement and compensate for one another’s individual weaknesses.

An interesting result is that the best feature combinations need only very small histograms. Using the combination of normals and intensities, we can get a recognition rate of 91% 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, for example, it takes only 0.1 CPU seconds to

features	histogram size	identification (1-3)	pose estimation (1-3)
$n + s$	8-8-16	87.75%	92.90%
$n + i$	4-4-8	91.60%	96.01%
$n + s + i$	4-4-8-8	93.16%	96.96%

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

match a test image with the 1980 histograms in the database on a Sun Blade 1000 (600MHz).

4 Probabilistic Recognition

Simple histogram matching is still a very crude 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 (resulting from viewpoint changes). This estimate is necessary when we want to combine different feature channels. A probabilistic approach, as described in Schiele's work [12] can provide much better results.

Instead of calculating an abstract distance measure, this approach directly estimates the posterior probability of an object hypothesis o_n , given a particular set of independent measurement vectors m_1, \dots, m_k . Using the Bayesian theorem, and assuming that all objects are equally likely, we obtain:

$$p(o_n | \bigwedge m_k) = \frac{\prod_k p(m_k | o_n)}{\sum_i \prod_k p(m_k | o_i)} \quad (4)$$

where $p(m_k | o_n)$ designates the likelihood of measurement vector m_k given the object o_n . This probability can be estimated directly from the histogram saved for o_n . Schiele's results indicate that only a relatively small number of measurement vectors (20-33%) is necessary to reliably detect and identify objects [12]. We applied the probabilistic recognition to the feature combination nsi . First results show that this can further improve the recognition performance.

5 Future Work and Conclusion

In this paper, we have shown the usefulness of local feature histograms for view-based object recognition from range images. Our approach achieves recognition rates above 91% on a database of almost 2000 full-sphere views of 30 objects while using only very small histograms. The system is very fast and achieves real-time performance. In the future, we plan to extend it with a probabilistic framework to compensate for occlusions and evaluate it on real range images.

References

1. G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Computer Graphics Proceedings, Annual Conference Series, 1994, ACM SIGGRAPH*, pages 311–318, 1994.
2. J.H. Yi and D.M. Chelberg. Model-based 3d object recognition using bayesian indexing. *Computer Vision and Image Understanding (CVIU)*, 69(1):87–105, 1998.
3. Y.K. Ham and R.-H. Park. 3d object recognition in range images using hidden markov models and neural networks. *Pattern Recognition*, 32(5):792–742, 1999.
4. R. Hoffman and A.K. Jain. Segmentation and classification of range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5), 1987.
5. Y. Lamdan, J.T. Schwarz, and H.J. Wolfson. Object recognition by affine invariant matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'88)*, pages 335–344, 1988.
6. X. Jiang and H. Bunke. *Dreidimensionales Computersehen, Gewinnung und Analyse von Tiefenbildern*. Springer-Verlag, 1997.
7. M.D. Wheeler and K. Ikeushi. Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3), 1995.
8. M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
9. H. Murase and S.K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
10. R.P. N. Rao and D.H. Ballard. Object indexing using an iconic sparse distributed memory. In *Fifth International Conference on Computer Vision (ICCV'95)*, pages 24–31, 1995.
11. C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'96)*, 1996.
12. B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–52, 2000.
13. C. Nastar. The image shape spectrum for image retrieval. Technical report, INRIA Rocquencourt, 1997.
14. R.C. Nelson and A. Selinger. A cubist approach to object recognition. In *Fifth International Conference on Computer Vision (ICCV'95)*, pages 614–621, 1998.
15. J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–565, 1992.
16. C. Dorai and A.K. Jain. Cosmos - a representation scheme for free-form surfaces. In *Fifth International Conference on Computer Vision (ICCV'95)*, pages 1024–1029, 1995.