Towards Shape-Based Visual Object Categorization for Humanoid Robots

D. Gonzalez-Aguirre, J. Hoch, S. Röhl, T. Asfour, E. Bayro-Corrochano* and R. Dillmann
Karlsruhe Institute of Technology, Adenauerring 2, Karlsruhe-Germany.
{gonzalez, julian.hoch, roehl, asfour, dillmann}@ira.uka.de
**CINVESTAV-Unidad Guadalajara, Av. Científica 1145, Mexico*
edb@gdl.cinvestav.mx

***Abstract*— Humanoid robots should be able to grasp and handle objects in the environment, even if the objects are seen for the first time. A plausible solution to this problem is to categorize these objects into existing classes with associated actions and functional knowledge. So far, efforts on visual object categorization using humanoid robots have either been focused on appearance-based methods or have been restricted to object recognition without generalization capabilities.**

In this work, a shape model-based approach using stereo vision and machine learning for object categorization is introduced. The state-of-the-art features for shape matching and shape retrieval were evaluated and selectively transfered into the visual categorization. Visual sensing from different vantage points allows the reconstruction of 3D mesh models of the objects found in the scene by exploiting knowledge about the environment for model-based segmentation and registration. These reconstructed 3D mesh models were used for shape feature extraction for categorization and provide sufficient information for grasping and manipulation. Finally, the visual categorization was successfully performed with a variety of features and classifiers allowing proper categorization of unknown objects even when object appearance and shape substantially differ from the training set. Experimental evaluation with the humanoid robot ARMAR-IIIa is presented.

I. Introduction

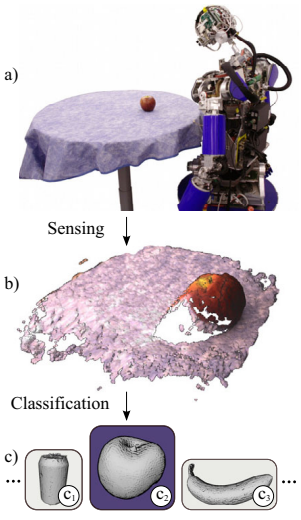


Fig. 1: Shape-based visual object categorization. a) The humanoid robot ARMAR-IIIa is exposed to an unknown object. b) Visually reconstructed scene. c) Object categories represented by shape-models from the training set.

In everyday scenarios, humanoid robots need to interact with a wide variety of objects in the environment in order to perform complex tasks. Reliably recognizing these objects is a problem which has been actively researched. As a consequence, there are now well established appearance-based methods for object recognition [1][2] where only previously learned objects have to be recognized.

However, an autonomous humanoid robot also needs to deal with objects that it has never encountered before. This generalization skill is the purpose of object categorization, where the encountered objects are assigned to previously defined categories. These categories are *"basic level classes"* which can be mostly characterized by their shape, see [3]. For example, object instances of the category *bottles* have different shapes, sizes, textures and colors. Still the robot should be able to correctly categorize a bottle even if it has not seen a particular exemplar before.

Furthermore, categories like *fruits* or *vegetables* change their appearance (color and texture) with progressing ripeness but maintain their overall shape. By using shape-based representations, it is possible to simultaneously deal with these circumstances and provide the essential grasping and manipulation information.

In this approach, a set of predefined object categories with several training samples was created. The training samples consist of labeled, visually reconstructed 3D mesh models for each category. Afterwards, a variety of classifiers was trained using shape features extracted from these 3D mesh models. Subsequently, in the on-line phase, the humanoid robot acquired 3D meshes by means of stereo vision in order to extract shape features. Finally, the trained classifiers were applied to categorize unknown objects, see Fig.1.

II. Related work

The focus of this work is to properly categorize small, rigid and graspable objects typically found in a human household environment while coping with the challenges of the limited visual sensing capabilities such as sensor's dynamic range, resolution, noise and self-occlusion.

Classic approaches for object recognition in robot vision solely use object appearance [4][5]. Besides, in order to be able to manipulate an object, the 6D pose has to be determined. While it is possible to obtain the object pose with appearance-based methods when using the depth information from stereo cameras [6], a more common approach is to match stored 3D models of the scene [7].

Appearance-based categorization approaches include the *Bag-of-Features* [8], which determines the distribution of local features in the feature space, and part-based approaches [9], which model objects as a collection of image parts or features.

When objects of different categories only differ in shape but not in texture (for example a wooden saltshaker and a wooden trivet), appearance-based methods reach their limits. With a model-based approach, the object's 3D shape can be incorporated into the categorization process.

In [10] and [11], point clouds were obtained from the objects using structured light projection and stereo camera on a mobile robot. The Fast Point Feature Histograms

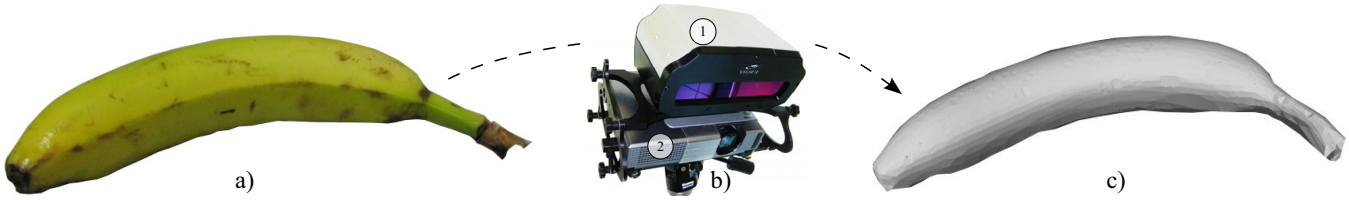


Fig. 2: The off-line 3D mesh models acquisition for training. a) Training sample object. b) The system used to digitize the training objects: (1)-Stereo camera, (2)-Projector. c) The digitized 3D polygon representation.

were calculated from the object's surface points and the objects were classified using Support Vector Machines and Conditional Random Fields. They achieved accurate results, 96.69% category identification, although they used the same objects for training and testing. In this approach only four categories with little inter-class variance were used. However, since most humanoid robots do not have light projectors, this sensing approach is not a viable option for on-line categorization systems for humanoid robots.

Another common approach is to use *spin image* features, which describe the global shape of the object from the perspective of local points on the object's surface. Spin images are shape descriptors which have been applied to surface matching [12], object recognition [13][14], 3D registration [15] and 3D object retrieval [16]. In [17], objects were modeled consisting of three parts which were categorized by spin images using the recognized part classes. The input data consisted of noise-less (simulated laser scanner) point clouds, achieving 96% accurate results when categorizing cars into eight categories. In [14], spin images were used in a 3D object detection system with the humanoid robot HRP-2 [18]. The scene was captured with stereo cameras and converted into a point cloud for the 3D mesh construction. Random scene points were selected and their corresponding spin images were calculated. These points were matched to previously calculated spin images of the model to be localized. This approach was designed to detect a single known object in the scene and does not deal with generalization such as the categorization of unknown objects.

Many features have been proposed for 3D model-based object recognition, for instance, *multidimensional table representations* [19], *spherical harmonic representations* [20], *shape distributions* [21], *coarse features* [22] and *conformal factors* [23]. Only few of those features degrade gracefully when dealing with occlusion and sensor noise expected in real applications. Because of their promising properties and superior noise degradation, i) spin images, ii) shape distributions and iii) coarse features were selected in this work to visually categorize objects by the humanoid robot through stereo vision.

In addition, from the different representations for 3D models (point clouds, voxel representations, octrees or collections of primitives such as boxes [24]) the 3D polyhedron models were selected due to their efficient construction and calculation of a wide variety of features. There are different algorithms to reconstruct polyhedral from point clouds, the most prominent are the *power crust* [25] and the *tight cocone* [26].

III. Methodology

The categorization is based on supervised learning to infer from known training objects to unknown observed objects. Classifiers were trained on a set of labeled training objects and applied to the objects that the robot encountered in the environment. The approach consists of the following phases:

- *Training*: Creates training data and trains classifiers.
- *Acquisition*: The robot gathers, segments and registers images of an unknown object to be categorized.
- *Reconstruction*: Obtains 3D meshes from the images.
- *Categorization*: Manages the shape feature extraction and the evaluation of the trained classifiers.

A. Training

The training of reliable classifiers requires a sufficiently large database of labeled training objects. Although there are public databases available with labeled object models, like the Princeton Shape Benchmark [27] or the KIT ObjectModels Web Database [28], these were not suitable.

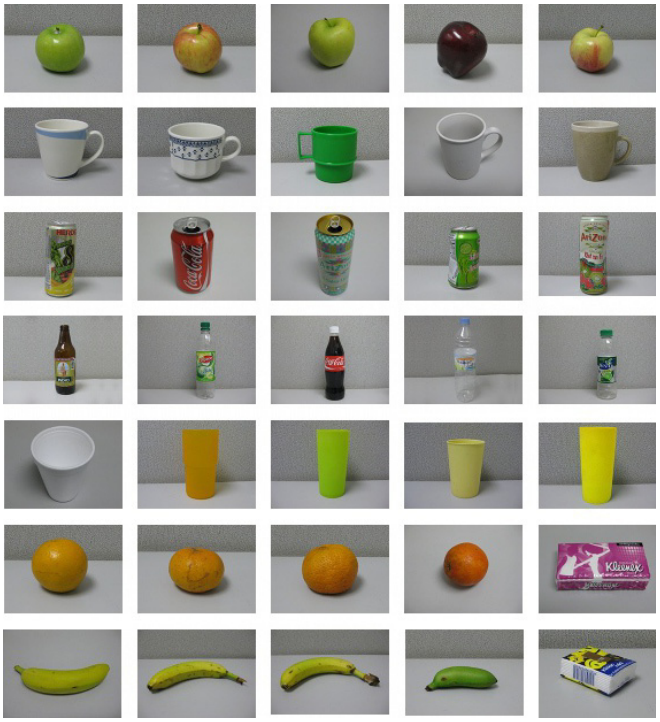


Fig. 3: The training set. Objects for each category with different shapes and sizes were selected to capture the natural intra-class variance.

Often, as it is the case in the Princeton Shape Benchmark, models represent artificial objects or are simplified versions of real objects. Therefore, they do not contain the real object characteristics nor the proper dimensions, which can carry important information for the categorization process.

Other databases contain the real object characteristics and proper dimensions, but unfortunately they lack the necessary variety of shapes needed for classifier training, for instance, the KIT object models Web Database is more oriented towards appearance-based approaches and disposes of a large variety of box and cylinder models, but for other shapes there are only single models available.

Due to these limitations, a new training database was created, see Fig.2. This model database comprises 35 objects belonging to 8 categories with sufficiently different shapes: *apples*, *mugs*, *beverage cans*, *oranges*, *bottles*, *bananas*, *beakers* and *tissue packages*. Notice that some categories with similar shapes were chosen (like *apples* and *oranges*), to determine if small differences between the categories are discriminative in presence of larger intra-class variance.

The 35 training objects were scanned using a 3D camera system StarCam™ . It projects structured light on the object and captures the resulting patterns with a stereo camera, see Fig.3. It densely samples the surface of an object from different angles to create a 3D reconstruction.

For each object, a 3D point cloud with approximately 5000 points was obtained and a watertight surface representation was created using the power crust algorithm, see Fig.2-c. The resulting meshes consist of approximately 10,000 to 15,000 convex polygons, which is small enough for fast feature extraction in less than one second. This database of real world objects was used to extract discriminant features in order to train different classifiers, see Sec.III-D.

B. Visual Acquisition

In the on-line evaluation, the humanoid robot ARMAR-IIIa (see Fig.1) obtained the 3D object reconstruction. Since from a vantage point only a part of the object is visible, the object was circumnavigated by the robot and several stereo views of the object were captured, see Fig.4. These views were used to create a 3D surface model of the object.

Because the stereo reconstruction is sensitive to lighting effects such as over-exposure, under-exposure and gloss, additional image preprocessing was performed before the reconstruction. For each view, several images with different exposures were captured and fused into a tone mapped HDR image [29]. This robust visual sensing method improves the image quality and preserves local contrast, see Fig.5.

Finally, the object to be categorized was segmented in the registered input images using a model-based environmental segmentation algorithm, which exploits the CAD model of the table and the relative pose of the humanoid robot during the acquisition, see Fig.6.

C. Reconstruction

For each set of images captured from one specific position, a point cloud of the scene was calculated using stereo reconstruction. For correspondence analysis, an extension of the Hybrid Recursive Matching method was used [30].

This algorithm produces a dense disparity image by recursively estimating the disparity for each pixel in the left image from four candidates in the right image. These candidates are

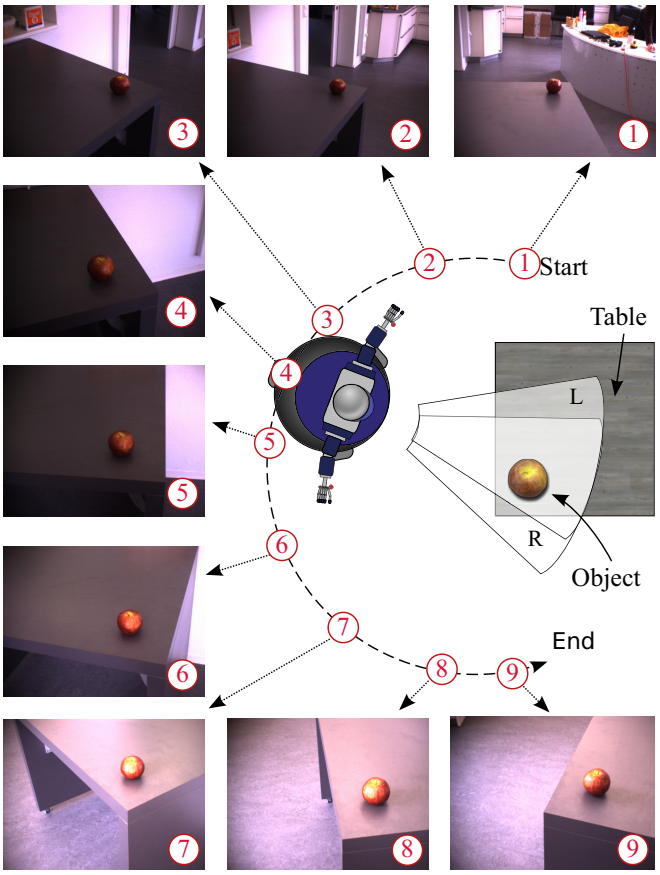


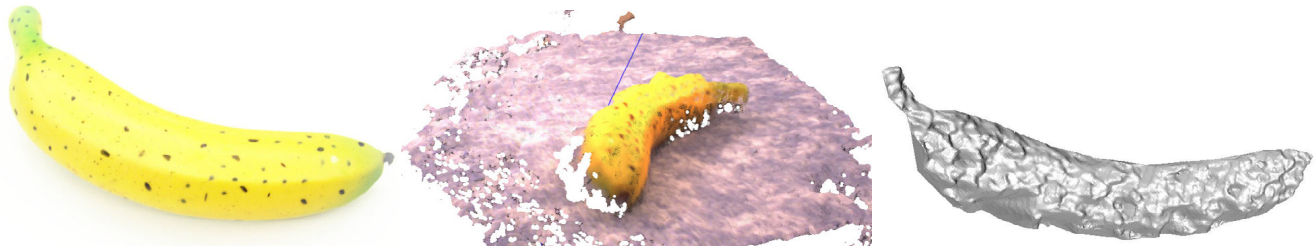
Fig. 4: The acquisition path around the test object defined by specifying several vantage points. The humanoid robot ARMAR-IIIa calculated an interpolated the path between the given positions and traversed capturing stereo images from the object.

determined in a two-stage process. i) Three candidates are found from the spatial neighborhood in the *block recursion*. This ensures a smooth disparity distribution especially in low textured regions. ii) The fourth candidate is calculated in the *pixel recursion* using optical flow techniques. The pixel recursion introduces disparity values in regions of discontinuity. For each candidate, the similarity is calculated using block matching. The candidate with the most similar neighborhood is set as the correspondence.

The outlier detection was done by a consistency check, where disparities of the left and right images were compared and rejected if their difference exceeds a threshold. Finally, the 3D coordinates of the image points were calculated by using the detected correspondences.

The resulting point clouds of the individual views were fused using environmental visual cues, namely known labeled table edges and region growing segmentation. Subsequently, the point cloud was preprocessed in order to remove outliers by calculating the object's centroid and the mean distance $\bar{d}$ between points belonging to the object and their centroid. Points for which the calculated distance d was greater than the mean distance ($d > \theta \cdot \bar{d}$ with θ set to 2.5) were removed from the point cloud.

Finally, the resulting point cloud was converted into a surface mesh using the power crust algorithm, see Fig.7. The tight cocone algorithm [26] was also evaluated, however, power crust produced models with fewer noise artifacts.



a) b) c)

Fig. 7: 3D Reconstruction of a banana by the humanoid robot stereo vision. a) Original object. b) Calculated point cloud. c) Reconstructed surface mesh.

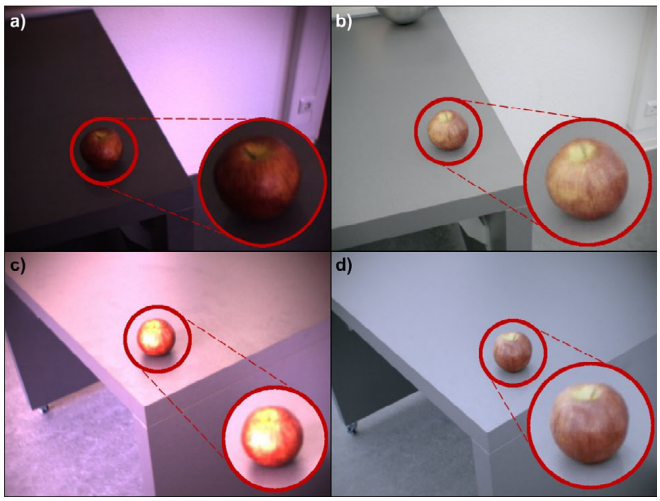


Fig. 5: HDR image acquisition. Captured images exhibit areas of a) under- and c) over-exposure. Tone mapped HDR images also show details in b) very dark or d) very bright areas.

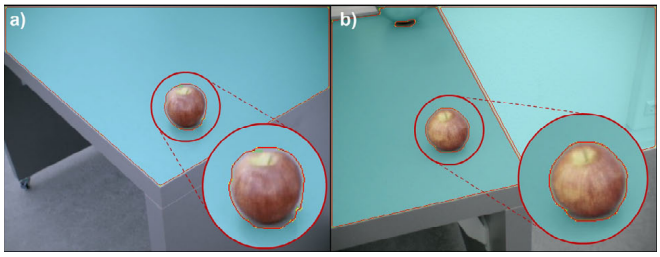


Fig. 6: Image segmentation. a) Adaptive region growing segmentation results from Fig.5-b. b) Results from Fig.5-d.

D. Categorization

Finding a significant set of features is crucial for object categorization. Adequate features should be high discriminative and robust to noise and other sources of variation. They also need to be efficient, pose invariant and capable of partial matching [31].

1) Spin Images: The composition and performance of spin images (see [12]) is influenced by three parameters:

- *Image size* s : Determines the size and resolution of the spin image, namely the number of bins s^2.
- *Bin size* b : Sets the distance between the different bins and determines the support distance $d = s \cdot b$. Increasing d results in a more global behavior of the spin images.
- *Support angle* α: Determines if the object's rear side is also considered for the calculation. Increasing the support angle leads to a more global behavior.

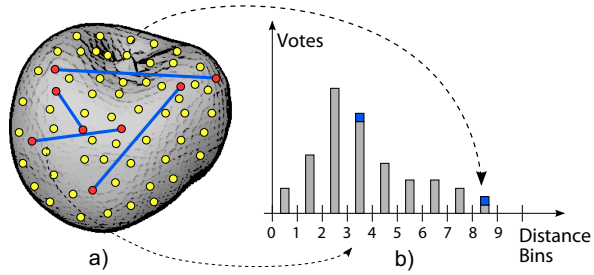


Fig. 8: The distance histogram calculation concept [21]. a) Object surface mesh with sampled surface points (yellow). Distance calculation for randomly selected point pairs (red). b) Distance histogram.

Spin images were used as features for object recognition by training standard classifiers like support vector machines and artificial neural networks. In the categorization phase, scene spin images were classified and the category that receives the most votes was selected. Notice that PCA dimensionality reduction of the spin images was applied prior to the training.

2) Shape Distributions: These representations [21] are histograms of a *shape function* that covers geometric properties of an object. Possible shape functions are for example the distance between two random points on the object's surface or the angle between three random surface points, see Fig.8. The shape function is evaluated for many random surface points and the resulting histogram can be used for matching using dissimilarity metrics or standard classifiers.

In the presented work, the $D2$ measure was implemented, which is the distance between two random surface points also called distance distribution. This measure yields the best categorization results according to [21]. Additionally, in order to obtain a smooth histogram, kernel density estimation was applied with the Epanechnikov kernel [32] because of its performance and theoretical properties, see Fig.9.

3) Coarse Features: These features are created by calculating geometric properties of a 3D model. In [22], the approach was used for a shape matching engine with measures such as volume, surface area, volume to surface area ratio, bounding box aspect ratio and derived values including the surface and volume of the object's convex hull.

In this work, 17 measurements were composed into a feature vector. They included the dimensions of the bounding box, the bounding sphere, object area and volume, area and volume of the convex hull, as well as several deduced features like convexity and compactness, see Tab.I.

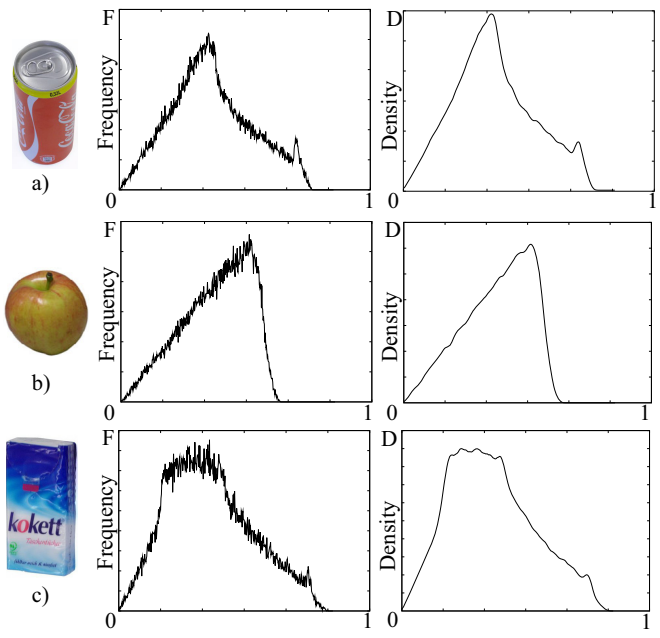


Fig. 9: The distance histogram and the estimated probabilistic distribution function using kernel density estimation. a) Cylindrical object. b) Spherical object. c) Box object.

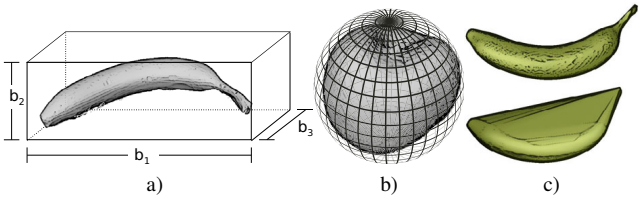


Fig. 10: Concepts used for coarse features calculation. a) Bounding Box. b) Bounding Sphere. c) Convex Hull.

Directly using the calculated feature vectors (from spin images, coarse features or shape distributions) as input for classifiers often leads to inferior results. This happens because if certain dimension in the feature vector contains very large values compared with the other dimensions, then this dimension tends to be over representative in the feature space. Consequentially in order to prevent this situations, it was necessary to rescale or normalize the feature vectors.

In this work the following classifiers were selected:

- Soft margin support vector machines with linear kernels and RBF kernels.
- Multilayer perceptrons with one hidden layer.
- K-nearest neighbor classifiers with different values for k and different distance metrics.

In order to estimate the optimal parameters for the different classifiers, a grid search [33] was done through a performance cross validation on the training database.

IV. Experimental Evaluation

A. Noiseless Data

The system was evaluated using only the noise free objects digitized with the 3D scanner (see Fig.2) in order to evaluate the pure discriminative power of the classifiers.

The classifiers were applied to the training set. Each experiment was performed by applying a cross validation:

Volume and Surface Area	Volume V, Area A
Bounding Box	Sides $b_1, b_2, b_3 (b_1 \geq b_2 \geq b_3)$ Ratios $b_1/b_3, b_2/b_3$
Cuboid Ratio	$V/(b_1 \cdot b_2 \cdot b_3)$
Bounding Sphere	Radius r_{bs}, Volume V_{bs}
Sphere Ratio	V/V_{bs}
Convex Hull	Volume V_{ch}, Area A_{ch}
Convexity	A/A_{ch}
Compactness	V/V_{ch}
Hull Packing	$1 - V/V_{ch}$
Hull Compactness	A^3_{ch}/V^2_{ch}

TABLE I: Elements in the coarse features vector, see details in Fig.10.

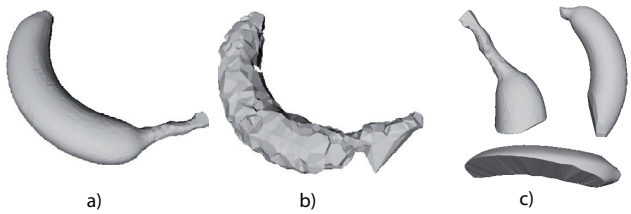


Fig. 11: Artificial model deterioration. a) Original model. b) Noisy model. c) Partially occluded object.

First, the database was randomly split into a training subset and a testing subset, with a training to testing ratio of 3:1. Afterwards, classifiers were trained on the training subset and evaluated on the testing subset. This procedure was repeated 1000 times, each time with a different testing and training subsets. Finally, the overall performance was used to calculate the statistical mean and standard deviation of the classifying accuracies, see results in Tab.II.

Classifier	CF	σ	$D2$	σ	SI	σ
SVM	85%	11%	68%	12%	83%	13%
MLP	86%	12%	60%	14%	76%	13%
kNN	83%	11%	65%	13%	83%	12%

TABLE II: Mean and standard deviation of the categorization accuracy results using the data set digitized with 3D scanner for Coarse Features (CF), Distance Distribution ($D2$) and Spin Images (SI).

The coarse features and the spin images attained good results. Most problems occurred due to the confusion between similar categories, like apples and oranges or mugs and beakers, see Tab.III.

	Banana	Can	Apple	Orange	Beaker	Tissues	Mug	Bottle
Banana	70%	–	–	–	–	–	–	10%
Can	–	30%	–	–	30%	50%	–	–
Apple	–	–	100%	70%	–	–	–	–
Orange	–	–	–	30%	–	–	–	–
Beaker	–	70%	–	–	50%	50%	10%	–
Tissues	–	–	–	–	–	–	–	–
Mug	–	–	–	–	20%	–	90%	–
Bottle	30%	–	–	–	–	–	–	90%

TABLE III: Averaged confusion matrix of 10 SVM classifiers using distance distributions features. Columns represent true labels, rows represent estimated categories by classifiers.

B. Noisy Data

In order to measure the influence of noise on the chosen features, a test was performed where a large amount of noise was added to the objects in the noiseless set. From the original point sets, 500 points were randomly chosen as the basis for the new objects. These points were then superimposed with 5 to 10 mm of evenly distributed noise. Even more, 3% of the points were chosen as outliers and

their positions were randomly shifted by 5 to 10 cm, see Fig.11-b. Subsequently, classifiers were trained on the (noise free) training set, and evaluated on the testing set with the added noise, see Tab.IV.

The deterioration in accuracy can be attributed to two causes. For one, the features for noisy objects and objects without noise are different. Although a common assumption is that it is generally preferable to use a training set without noise, in this case the same experiment performed with the training data also being noisy achieved higher accuracy (the difference was 5% to 10%). The introduced noise decreases the inter-class variance at the decision borders. The classifiers where unable to differentiate for example between the *orange* and *apple* categories and the *beverage can* and *beaker* categories became nearly indistinguishable.

Classifier	Coarse Features	Distance Distributions	Spin Images
SVM	61%	61%	63%
MLP	57%	59%	53%
kNN	59%	56%	57%

TABLE IV: Accuracy results for classifiers trained on noise free training set and tested on data set with artificial noise.

C. Partially Occluded Data

Experiments were also performed on artificially occluded objects by using only parts of the model, see Fig.11-c. Each object was used to create three occluded variants by cutting off a random part of the object. This was done by selecting a random plane intersecting the object and preserving only the points on one side of the plane. It was ensured that the resulting object had a length of 30% to 70% along the chosen plane's normal compared to the original object, see the categorization results in Tab.V.

Classifier	Coarse Features	Distance Distributions	Spin Images
SVM	53%	42%	75%
MLP	60%	68%	61%
kNN	55%	56%	70%

TABLE V: Accuracy results for classifiers trained on data set without occlusion and tested on occluded data set.

D. Real Data

In order to evaluate the categorization approach with the humanoid robot, classifiers were trained on the training set and applied to the models that were reconstructed by the humanoid robot. The testing set comprised an apple, two bananas, a beaker, a beverage can, a bottle, a tissue package and an orange. Each object was captured from several views. For each view a 3D point cloud was calculated. All point clouds were fused and used to create a 3D polygon model, see Fig.12. The trained classifiers were then applied to these 8 models, see results in Tab.VI.

Surprisingly, the best results were achieved by the coarse features, while spin images and distance distributions delivered less reliable results.

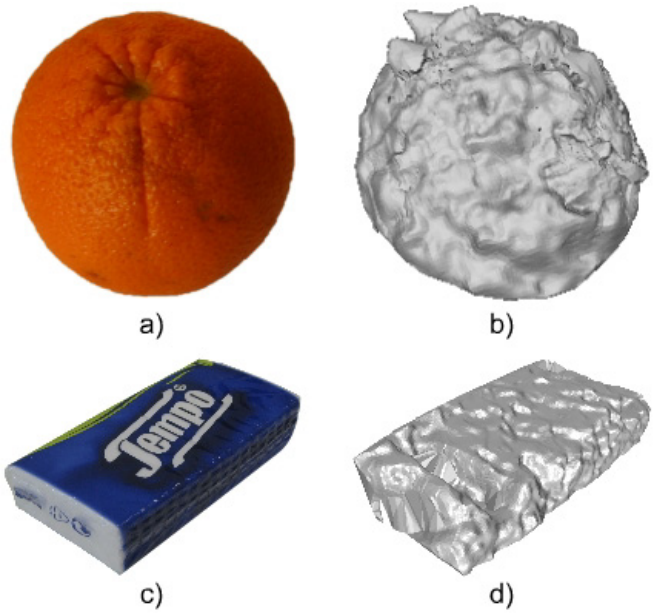


Fig. 12: Visual model reconstruction. a) Test object orange. b) Reconstructed model orange. c) Test object tissue package. d) Reconstructed model tissue package.

Classifier	Coarse Features	Distance Distributions	Spin Images
SVM	100%	35%	50%
MLP	88%	35%	61%
kNN	70%	34%	51%

TABLE VI: Accuracy results for classifiers trained on (scanned) training set and evaluated on models reconstructed by the humanoid robot. Notice that these results (especially the 100% accuracy of the SVM using Coarse Features) reflects the outcome of the experimets with the 8 mentioned objects.

E. Parameters

Proper parameters for the different feature vectors were empirically chosen. The histogram size for the distance distribution was set to 512, with the largest histogram bin corresponding to 3σ, where σ represents the mean distance.

The kernel density estimation was performed with an Epanechnikov kernel and the bandwidth was set to 0.1σ. For the histogram generation, 100,000 surface points were sampled and 100,000 point pair distances were calculated.

The spin image size was set to 10, with bin size 6 mm (resulting in support distance 6 cm). Support angle was set to 120°. The spin image stack for training was constructed by taking the 16 k-means cluster centers calculated from 1000 random spin images. Categorization was performed on 100 random spin images using majority voting.

For the coarse features, the best results were achieved by using the *bounding box dimensions*, *bounding sphere radius*, *volume*, *convex hull area* and *convex hull volume* as feature vector components. The estimated classifier parameters calculated by the grid search are presented in Tab.VII.

Classifier	Coarse Features	Distance Distributions	Spin Images
SVM	$_{\text{RBF}}\, C = 1, \gamma = 1$	$_{\text{RBF}}\, C = 1, \gamma = 1$	$_{\text{Linear}}\, C = 1$
MLP	64 Neurons	32 Neurons	20 Neurons
kNN	k=1	k=1	k=3

TABLE VII: Optimal empirical estimated parameters for different classifiers with selected features.

V. Conclusion

Shape-based object categorization is a challenging problem, especially with the limited visual sensing capabilities of a humanoid robot. Up to now, no pure vision-based systems were capable to generalize objects using their 3D shape. The difficulties of this task include varying lighting conditions, unfavorable object surfaces, context and self-occlusions, which results in incomplete or noisy reconstructions.

By exploiting the environmental knowledge while fusing several HDR stereo views from different vantage points and applying robust reconstruction techniques, the humanoid robot is able to acquire sufficiently detailed 3D models of small objects in real application conditions such as difficult surfaces and unfavorable lighting.

The careful selection and proper transfer of the coarse features method from 3D shape retrieval to the object categorization task enabled the categorization of unknown objects by generalizing from known digitized samples. This promising results corroborate that model-based visual object categorization will enable humanoid robots to deal with unknown objects and consequently with more general situations in real application scenarios.

VI. Acknowledgments

This work was partially conducted within the EU Cognitive Systems project GRASP (FP7-215821) funded by the European Commission and the German Humanoid Research project SFB588 funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft).

References

[1] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[2] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2002.

[3] Rosch E. Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7(4):573–605, 1975.

[4] P.M. Roth and M. Winter. Survey of appearance-based methods for object recognition. *Inst. for Computer Graphics and Vision, Graz University of Technology, Austria, Tech. Rep. ICG-TR-01/08*, 2008.

[5] P. Azad, T. Asfour, and R. Dillmann. Combining appearance-based and model-based methods for real-time object recognition and 6d localization. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 5339–5344, 2006.

[6] K. Okada, M. Kojima, S. Tokutsu, T. Maki, Y. Mori, and M. Inaba. Multi-cue 3D object recognition in knowledge-based vision-guided humanoid robot system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3217–3222, 2007.

[7] M. Ulrich, C. Wiedemann, and C. Steger. Cad-based recognition of 3d objects in monocular images. In *International Conference on Robotics and Automation*, pages 1191–1198, 2009.

[8] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, volume 1, page 22, 2004.

[9] R. Fergus, P. Perona, A. Zisserman, et al. Object class recognition by unsupervised scale-invariant learning. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 2003.

[10] R.B. Rusu, A. Holzbach, M. Beetz, and G. Bradski. Detecting and Segmenting Objects for Mobile Manipulation. In *Proceedings of IEEE Workshop on Search in 3D and Video (S3DV), Kyoto-Japan*, 2009.

[11] Radu Bogdan Rusu, Michael Beetz, Andreas Holzbach, Rosen Diankov, and Gary Bradski. Perception for mobile manipulation and grasping using active stereo. In *Humanoid Robots, 2009 9th IEEE-RAS International Conference on*, Paris, 12/2009 2009.

[12] A. Johnson. Spin-images: a representation for 3-D surface matching. *Robotics Institute. Pittsburgh, Pennsylvania: Carnegie Mellon University*, 1998.

[13] A.E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.

[14] O. Stasse, S. Dupitier, and K. Yokoi. 3d object recognition using spin-images for a humanoid stereoscopic vision system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Beijing, China*, pages 2955–2960, 2006.

[15] N. Brusco, M. Andreetto, A. Giorgi, and G.M. Cortelazzo. 3D registration by textured spin-images. In *3D Digital Imaging and Modeling*, pages 262–269, 2005.

[16] P.A. De Alarcón, A.D. Pascual-Montano, and J.M. Carazo. Spin images and neural networks for efficient content-based retrieval in 3d object databases. *Lecture notes in computer science*, pages 225–234, 2002.

[17] D. Huber, A. Kapuria, R. Donamukkala, and M. Hebert. Parts-based 3d object classification. In *IEEE Computer Society Conference On Computer Vision and Pattern Recognition*, volume 2. IEEE Computer Society; 1999, 2004.

[18] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi. Humanoid robot HRP-2. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1083–1090, 2004.

[19] AS Mian, M. Bennamoun, and R. Owens. Three-dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1584–1601, 2006.

[20] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, page 164. Eurographics Association, 2003.

[21] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.

[22] J. Corney, H. Rea, D. Clark, J. Pritchard, M. Breaks, and R. MacLeod. Coarse filters for shape matching. *IEEE Computer Graphics and Applications*, pages 65–74, 2002.

[23] M. Ben-Chen and C. Gotsman. Characterizing shape using conformal factors. In *Proceedings of Eurographics Workshop on Shape Retrieval*, 2008.

[24] J. Bohg, C. Barck-Holst, K. Huebner, M. Ralph, B. Rasolzadeh, D. Song, and D. Kragic. Towards Grasp-Oriented Visual Perception for Humanoid Robots. *In International Journal of Humanoid Robotics, Special Issue on Active Vision of Humanoids*, 6(3):387–434, 2009.

[25] N. Amenta, S. Choi, and R.K. Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2-3):127–153, 2001.

[26] T.K. Dey and S. Goswami. Tight cocone: a water-tight surface reconstructor. *Journal of Computing and Information Science in Engineering*, 3:302, 2003.

[27] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The princeton shape benchmark. *Shape Modeling International, Genova, Italy*, 2004.

[28] The KIT ObjectModels Web Database. `http://i61p109.itec.uni-karlsruhe.de/ObjectModelsWebUI`, September 2010.

[29] D. Gonzalez-Aguirre, T. Asfour, and R. Dillmann. Eccentricity Edge-Graphs from HDR Images for Object Recognition by Humanoid Robots. In *Humanoid Robots, 2010 10th IEEE-RAS International Conference on*, 2010.

[30] Nicole Atzpadin, Peter Kauff, and Oliver Schreer. Stereo analysis by hybrid recursive matching for real-time immersive video conferencing. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3), 2004.

[31] J.W.H. Tangelder and R.C. Veltkamp. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, 39(3):441–471, 2008.

[32] Hart P. Duda R. and Stork D. *Pattern Classification*. A Wiley publication. Wiley, 2 edition, 2001.

[33] C. Staelin. Parameter selection for support vector machines. *Hewlett-Packard Company, Tech. Rep. HPL-2002-354R1*, 2003.