Proc. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Madison, USA, June 2003.

Analyzing Appearance and Contour Based Methods for Object Categorization

Bastian Leibe and Bernt Schiele
Perceptual Computing and Computer Vision Group, ETH Zurich, Switzerland
{leibe,schiele}@inf.ethz.ch

Abstract

Object recognition has reached a level where we can identify a large number of previously seen and known objects. However, the more challenging and important task of categorizing previously unseen objects remains largely unsolved. Traditionally, contour and shape based methods are regarded most adequate for handling the generalization requirements needed for this task. Appearance based methods, on the other hand, have been successful in object identification and detection scenarios. Today little work is done to systematically compare existing methods and characterize their relative capabilities for categorizing objects.

In order to compare different methods we present a new database specifically tailored to the task of object categorization. It contains high-resolution color images of 80 objects from 8 different categories, for a total of 3280 images. It is used to analyze the performance of several appearance and contour based methods. The best categorization result is obtained by an appropriate combination of different methods.

1. Introduction

Even though generic object recognition and classification have been one of the goals of computer vision since its beginnings [1], we are still far from achieving this goal. On the other hand, the identification of known objects in different poses and under novel viewing conditions has made significant progress recently [2, 3, 4, 5, 6, 7, 8, 9]. At the same time, impressive results have been achieved for the detection of canonical views of individual categories, such as faces [10], cars [11, 12], pedestrians [13], and horses [14].

Still, little progress has been made for the more general task of multi-class object categorization, with some notable exceptions such as [15, 16]. Even more importantly, many recognition methods have not been tested on multi-class categorization, so that little is known about their respective capabilities to generalize beyond known and seen objects. Also, it is not clear what the role of different cues, such as contour, shape, color, and texture is for categorization. Traditionally, contour and shape based methods are considered most adequate for handling the generalization requirements needed for categorization tasks.

To address these issues, we have built a novel database specifically tailored to the task of object categorization. It contains 80 objects from 8 categories. Each object is represented by 41 views spaced evenly over the upper viewing hemisphere. This allows to analyze the performance of different recognition methods not only from a 1D circle or a few canonical viewpoints, but from multiple viewing positions. For each image a high-quality figure-ground segmentation mask is provided. This makes it possible to compare both appearance and contour based methods in the idealized setting of perfect segmentation. Even though any comparison on a particular database has its limitations, we strongly believe that databases such as the one we propose, as well as the comparison of different methods are important steps to enable progress in the area of object categorization. The database is made publicly available and other authors are invited to run and report experiments.

Section 2 casts the object categorization problem in a framework founded in Cognitive Psychology. This foundation motivates our object database, introduced in Section 3. Different contour and appearance-based methods are introduced in Section 4, and Section 5 presents experimental results comparing those methods as well as different cues for object categorization. As expected, different methods and cues have their respective strengths and weaknesses. Therefore, Section 6 proposes and discusses the combination of different methods.

2. Object Categorization

It is important to emphasize that the notion and the abstraction level of object classes is far from being uniquely and clearly defined. Notably, the question of how humans organize knowledge at different levels has received much attention in Cognitive Psychology [17]. Taking an example from Brown's work, a dog can not only be thought of as a *dog*, but also as a *boxer*, a *quadruped*, or in general an *animate being* [17]. Yet, *dog* is the term that comes to mind most easily, which is by no means accidental. Experiments show that there is a *basic level* in human categorization at which most knowledge is organized [18]. According to Rosch et al. [18, 19], this basic level is also

- the highest level at which category members have similar perceived shape.

- the highest level at which a single mental image can reflect the entire category.
- the highest level at which a person uses similar motor actions for interacting with category members.
- the level at which human subjects are usually fastest at identifying category members.
- the first level named and understood by children.

These points are the motivation for us to address multi-level object categorization rather than the less clearly defined problem of object classification. Basic level categorization is easiest for humans. At the next lower levels, subordinate categories and the exemplar level used in object identification can be found. The next higher level, superordinate categories, requires a higher degree of abstraction and world knowledge. It is thus useful to start the generic object recognition task in the framework of basic-level categories, which seem to be a good starting point for visual classification.

Another argument is that the distinction between object classes may be quite arbitrary when drawing strict borders between any two classes. In reality, some classes are inherently more similar than others (e.g. dogs and horses are more similar since they are quadrupeds than dogs and cars). Looking at multiple levels of object categorization rather than individual classes, it becomes a desired property that objects from the same superordinate category, such as quadrupeds, be classified as more similar than objects from different superordinate categories. If the object itself is not correctly recognized, then we want it to be assigned at least to a "similar" category (graceful degradation).

The experiments of this paper are restricted to basic level categories. In a first step, we explicitly do not want to model functional categories (e.g. *"things you can sit on"*) and ad-hoc categories (e.g. *"things you can find in an office environment"*) [20]. Even though those categories are important, they exist only on a higher level of abstraction and require a high degree of world knowledge and experience living in the real world.

It is important to note that categories do not exist per se in the world; they are a learned representation [18] and therefore depend on experience and education. So, it may not be possible to find *the* unique basic level for every object. However, there are objects that have become so much part of our daily life that their basic level is well-defined almost all over the world (e.g. apples, horses, cars, etc.). In the following section, we will introduce our evaluation database, which contains some of those categories.

3. The Database

Existing publicly available image databases, like the COIL [4], have been very influential. Most directly related to our endeavor, the RSORT database [15] contains full-sphere views, but only includes grayscale images and no segmentations. In this section, we present a new database for object categorization containing 80 objects from 8 carefully chosen categories, high-resolution color images, and segmentation masks for every image.

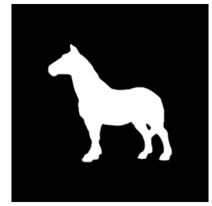
In our work, we want to explore categorization for both natural and artificial (human-made) objects. In particular, we include objects from the following basic areas: "fruits & vegetables"; "animals"; "human-made, small (graspable)"; and "human-made, big" (e.g. vehicles). Objects from these areas have different affordances, that is different ways of interacting with the environment, and different characteristics. For the first iteration of our database, we chose to include the following objects: apples, pears, and tomatoes for the "fruits & vegetables" area; cows, dogs, and horses for the "animals"; cups for the "graspable", and cars for the "vehicles" supercategory.

In principle, there are two ways how such a database can be built. A category can either be set up by a representative distribution of member objects reflecting their probabilities of occurrence in practice, or by a few prototypes that approximately span the category [21]. In light of the difficulty of establishing representative distributions and the effort involved in taking pictures of member objects, we resort to the second option. Figure 1 shows the current status of our database (in the following referred to as the ETH-80 database). For each category, we provide 10 objects that span large in-class variations while still clearly belonging to the category. Each object is represented by 41 images from viewpoints spaced equally over the upper viewing hemisphere (at distances of $22.5-26°$). The viewing positions were obtained by subdividing the faces of an octahedron to the third recursion level. For collecting the views, we employed an automated robot setup and a blue chromakeying background for easier segmentation. All images have been taken with a Sony DFW-X700 progressive scan digital camera with 1024×768 pixel resolution and a Tamron 6-12mm varifocal lens (F1.4). For every image, we provide a high-quality segmentation mask (Figure 1), so that shape and contour based methods can be easily applied. The full database is made available on our webpage[1].

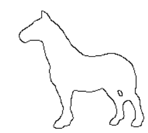
The intended test mode is leave-one-object-out crossvalidation. This means we train with 79 objects and test with the one unknown object. Recognition is considered successful if the correct category label is assigned. The results are averaged over all 80 possible test objects. We use the database for a best case analysis: categorization of unknown objects under the same viewing conditions, with a near-perfect figure-ground segmentation, and known scale. In a practical application, such perfect information is seldomly available. But if an algorithm does not work under

[1] http://www.vision.ethz.ch/pccv/

(a) Example Image



(b) Segmentation



(c) Extracted Contour

Figure 1: *The 8 categories of the ETH-80 database. Each category contains 10 objects with 41 views per object, spaced equally over the viewing hemisphere, for a total of 3280 images.*

these ideal conditions, it is likely to fail in practice.

4. Recognition Methods

Using the database presented above, we want to compare different methods for multi-class object categorization. In particular, we want to address the question of what the role of color, texture, and shape is for this task. In this section, we introduce a selection of well-known recognition methods that are prototypical for these cues. Those methods serve as the basis for our experiments.

Color: One of the earliest appearance based recognition methods is recognition with color histograms [2]. Using this approach, we collect a global RGB histogram over all image pixels belonging to the object (as specified by the segmentation mask). Two histograms V and Q can be compared using the intersection measurement $\cap(Q, V) = \sum_i min(q_i, v_i)$ or the χ^2 divergence $\chi^2(Q, V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i}$. The test image is then assigned to the category containing the closest matching histogram. In our experiments, we obtained the best results with a histogram resolution of 16-16-16 for the color channels and using the χ^2 measurement.

Texture: For the texture cue, we use a generalization of the color histogram approach to histograms of local gray-value derivatives at multiple scales [9]. In our experiments, we compare two versions of this approach. The first is a rotation-variant descriptor and uses only first derivatives in x and y direction over 3 different scales. The second uses rotation invariant features, namely the gradient magnitude and the Laplacian, again over 3 scales. Both the $D_x D_y$ and the Mag-Lap version have been applied to the COIL database in the past with 100% recognition rate [9]. In our experiments, we obtained best results with the scales set to $\sigma_{1,2,3} = (1, 2, 4)$, 16 histogram bins per dimension, and the χ^2 measurement for histogram comparison. As shown in [9], histogram based approaches can also be used locally to recognize objects from a small set of sample points taken from the test image. In this paper, however, we use only the simpler alternative of matching histograms.

Global Shape: For the shape cues, we make a difference between global and local shape. As representatives for global shape, we use PCA-based methods [22, 4]. There are two principal ways of using PCA for recognition. In the traditional method [4], one single global eigenspace for all categories is built and the training images are projected into this space. Recognition then becomes a nearest-neighbor search in the eigenspace for the closest training example. The other approach is to build separate eigenspaces for each category and measure the reconstruction error ("distance from feature space" [22]), that is the quality by which the class-specific eigenspace can represent the test image. This approach can be generalized even further towards view-specific eigenspaces [23], which we will leave for future experiments.

The class-specific approach has the advantage that it can be extended easily to a larger number of categories – only the eigenspaces for the new classes have to be recomputed – but it is not yet known how it scales. We have made experiments with both approaches and found no significant differences in their recognition performance. Since our experiments require the recalculation of the eigenspace for every object, and the global eigenspace version takes an order of magnitude longer to compute, we only report results on the version with class-specific eigenspaces.

In two separate experiments, we apply PCA to the raw segmentation masks ("pure" global shape) and to the segmented grayvalue images. For the segmentation masks, the best recognition performance was achieved using only the first 30 eigenvectors; for grayvalue images, best results were obtained using the first 40 eigenvectors. For all PCA experiments, the images are downscaled to a size of 128×128 pixels. In contrast to [4], we do not adapt the scale for individual views of an object such that its bounding box always fills the whole image. In our experience, the varying scales distort the eigenspace and could potentially hurt recognition performance.

Local Shape: We have chosen contours as a representative feature for local shape. Over the years, numerous methods have been developed for contour-based recognition, e.g. deformable prototypes [21] or shock graphs [24], to name but a few. We pick out a method based on the Shape Context proposed by Belongie [25], which has achieved excellent results, for example for handwritten digit recognition.

In this approach, an object view is represented by a discrete set of points sampled regularly along the internal or external contours. For every point, a log-polar histogram, the Shape Context, is computed that approximates the distribution of adjacent point locations relative to the reference point. In order to achieve scale invariance, the outer radius for the histograms is typically set to the mean distance between all point pairs.

Point correspondences between different shapes can be found by matching the log-polar histograms. In their original implementation, Belongie et al. match shapes by iteratively deforming one contour using thin plate splines [25]. Here, we compare two simpler approaches. In the first method, we search a continuous path around the main object contour using a dynamic programming approach (similar to Dynamic Time Warping). We allow that adjacent points on one contour be matched to the same point on the other contour, and that a mismatching point be skipped, but every point on one of the contours must be matched and the overall matching order must be kept. The final score is the sum over all individual matching costs. The second approach is just a one-to-one matching between contour points using a greedy strategy. Here, the matching score is also the sum over all individual matching costs. In both cases, best results were obtained using 100 points on the contour, 5 radius and 12 sector bins, and the intersection measurement for comparing shape context histograms.

5. Results

In this section, the methods described above are applied to the object categorization task. As all methods depend on a set of parameters, we made a series of preliminary experiments to determine the optimal parameter settings for every method. In the following, we report only the best results.

5.1. Global Recognition Rates

Table 1 shows the recognition results for the different methods, both averaged over the whole database and broken up per category. As already mentioned in Section 3, the test mode is leave-one-object-out crossvalidation. So, the results always show the performance for the categorization of unknown objects. As can be seen, the contour-based methods perform best with 86.4% recognition rate. Next best are the global-shape based PCA variations with 83.41% and 82.99%, respectively. The texture histograms are only slighly behind with 82.23% for the rotation-invariant case, and 79.79% for the rotation-variant one. With only 64.85% recognition rate, color performs worst.

Globally, there is only a slight difference between the two PCA methods. However, on the category level significant differences become apparent. For the apple and tomato categories, the version with grayvalue images outperforms the mask-based version. Here, the global shape is similar for both categories, but the objects in both classes have a characteristic, class-specific texture. As a result, shape ambiguities between the categories can be resolved by additional information from the grayvalue images. For the cow, dog and horse categories, on the other hand, the mask-based version shows better performance. Here, the global shape is again similar for all three categories. However, the ambiguities cannot be resolved by resorting to the grayvalue information encoded in the eigenspaces, because there is no characteristic texture for those categories. On the contrary, the in-class variation for texture is so high that using localized grayvalue information actually hurts performance. The behavior of both contour based methods is similar to the one for PCA on mask images, only on a globally higher level. Between the two contour-based methods, there is no significant difference.

For the texture histograms, the rotation invariant version has a better global performance than the rotation variant

	Color	D_xD_y	Mag-Lap	PCA Masks	PCA Gray	Cont. Greedy	Cont. DynProg	Avg.
apple	57.56%	**85.37%**	80.24%	78.78%	**88.29%**	77.07%	76.34%	77.66%
pear	66.10%	90.00%	85.37%	**99.51%**	**99.76%**	90.73%	91.71%	89.03%
tomato	**98.54%**	94.63%	**97.07%**	67.80%	76.59%	70.73%	70.24%	82.23%
cow	86.59%	82.68%	**94.39%**	75.12%	62.44%	86.83%	86.34%	82.06%
dog	34.63%	62.44%	74.39%	72.20%	66.34%	**81.95%**	**82.93%**	67.84%
horse	32.68%	58.78%	70.98%	77.80%	77.32%	**84.63%**	**84.63%**	69.55%
cup	79.76%	66.10%	77.80%	**96.10%**	**96.10%**	**99.76%**	**99.02%**	87.81%
car	62.93%	**98.29%**	77.56%	**100.0%**	**97.07%**	**99.51%**	**100.0%**	90.77%
total	64.85%	79.79%	82.23%	83.41%	82.99%	86.40%	86.40%	80.87%

Table 1: *Recognition Results for the categorization of unknown objects.*

Category	Primary feature(s)	Secondary feature(s)
apple	PCA Gray	Texture D_xD_y
pear	PCA Gray / Masks	
tomato	Color	Texture Mag-Lap
cow	Texture Mag-Lap	Contour / Color
dog	Contour	
horse	Contour	
cup	Contour	PCA Gray / Masks
car	PCA Masks / Contour	Texture D_xD_y

Table 2: *Best primary and secondary features for our categories, as derived from the recognition results.*

one. On the per-category level, however, the methods show more distinct behaviors. Rotation variant features seem to be significantly better for the apple, pear, and car categories, that is for those objects where the relative orientation of texture elements or lines is important for recognition. For those categories that contain mainly circular texture elements (like the specularities on most of the tomatoes), or where the relative number of edge pixels on its own is a characteristic feature (as seems to be the case for the animals and cups), the rotation invariant texture descriptor gives the better results.

In general, it becomes clear that no single method is superior for all categories. Interestingly, though, almost all of the above methods are the best choice for at least one category. For example, the global color distribution, which is in general not a characteristic feature for many basic-level categories, still performs well for cows and tomatoes. From this we can conclude that for multi-class object categorization, we need multiple features and different combinations of features. Table 2 shows a list of the most discriminative primary and secondary features for our categories (achieving best and second best recognition results).

5.2. Confusions

In Section 2, we have stated the need for graceful degradation of an object categorization system. We therefore want to evaluate which objects are treated as similar or are confused by the different methods. We hope this can shed more light onto how the methods perform and how they may generalize to larger tasks with more categories.

In order to examine this more closely, we look at the confusion matrix for each method. By iteratively grouping together those categories that are confused most often, we obtain a hierarchy of groupings. Figure 2 shows the grouping hierarchies for color, rotation invariant texture, PCA on segmentation masks, and contours. As can be seen from these diagrams, the contour based method results in the most intuitive hierarchy, grouping together both the fruits and the animals. Both PCA and texture succeed in grouping together the animals, but manage only two of the three fruit categories. Interestingly, those groupings are different for the two cues: apples and tomatoes are treated as similar in terms of global shape; apples and pears in terms of texture. As could be expected, color again performs worst.

The out-of-class confusions that occurred most often in our experiments are cows with cars for the shape and contour cues, and apples with cups for texture. These are mainly degenerate views from above, where a cow has a roughly rectangular outline, or from a medium height, where the cup handle is not visible and only an ambiguous shape remains. In real-world situations and with unconstrained viewpoints, such confusions are likely to appear.

Interestingly, rotation-invariant texture is the cue that best groups the animal categories together. When taken for a single class, this cue can recognize them with 99.59% accuracy – significantly better than it is possible with global shape or contours. It only fails when trying to distinguish the individual types of animals.

6. Multi-Cue Combination

The results from our experiments stress the need for multi-cue combination. In the following, we examine how recognition performance can be improved by applying a decision tree [26] that at each level bases its decisions on one cue only. Starting again from the confusion matrices, we seek an optimal partition of the categories that minimizes the number of misclassifications. We then make our decision

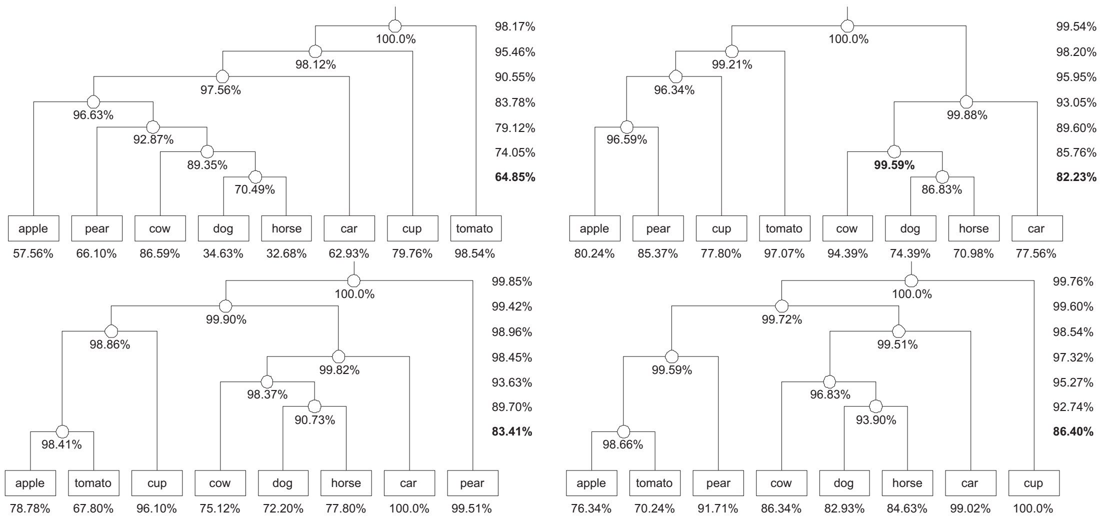


Figure 2: *Grouping hierarchies for four different cues: color (top left), rotation invariant texture (top right), PCA on masks (bottom left), and contours (bottom right). The diagrams show, from bottom to top, the best groupings for each cue. At each node the local recognition rate for this grouping is displayed. The numbers to the right show the global recognition rate after the groups are split.*

based on the cue that produces the best partition and iteratively refine the resulting group of categories. For this, we have to recompute the confusion matrices for all cues while leaving out those views that have already been misclassified. In this example, we stop at the category level, but we expect that the results can be improved when the approach is pursued down to a view or aspect level.

Figure 3 shows the resulting optimal decision trees for the case where all cues are available, and for the case where local shape is not. The performance for the first case is clearly better, with 93.02% recognition rate compared to 89.97% for the second case. However, both versions are comparable up to the point where the individual animal categories need to be distinguished. Here, the main difference occurs, and 3% performance is lost because the other cues are not as good at separating the animals. Using only color and texture and no shape information at all, the performance is significantly worse with only 86.4% combined recognition rate (not shown). This confirms that both global and local shape are important cues for object categorization.

7. Discussion & Conclusion

In this work, we have analyzed the performance of several state-of-the-art appearance- and contour-based recognition methods for the more general task of multi-class object categorization. As basis for our analysis, we have introduced a new database containing several categories and both object appearances and segmentation masks. We hope it will serve to bring together the communities of appearance and contour based recognition. That there is a potential for mutual benefit can be seen from our results. Contours proved to be the best single cue for the categories in our database, followed by global shape and (rotation invariant) texture descriptors. What is even more important, though, is that every cue we tested turned out to be the best choice for at least one category. This shows that there is significant potential for improvement by using multiple cues.

In the second part of our analysis, we have demonstrated how this potential can be used in the form of a multi-cue decision tree. Using all available cues, we were thus able to improve the global recognition rate from 86.4% to 93%. Contours again played an important role in this improvement. Without them, the recognition rate could only be increased to about 90%, mostly because the remaining cues were not able to distinguish the different animal categories. Without both contours and global shape, recognition performance could only be increased from 83.4% to 86.4% – a performance the contour-based methods achieved on their own. This emphasizes the importance of shape-based cues for object categorization.

It is important to bear in mind that this work shows a best case analysis. Transferring methods from a lab setting to the real world is not a trivial task, and it may well be that some necessary features cannot be extracted in sufficient quality for a particular method to work. What we can deduct from the experiments is an opposite argument: if a method does not achieve good results under our idealized conditions, it is likely to fail in practice. In that respect, our finding that no single method achieved over 87% recognition rate is an even stronger argument for the necessity of multiple cues.

The size of the database will be increased in the future, with more objects per category and a larger number of categories. However, the ultimate test case is the real world.

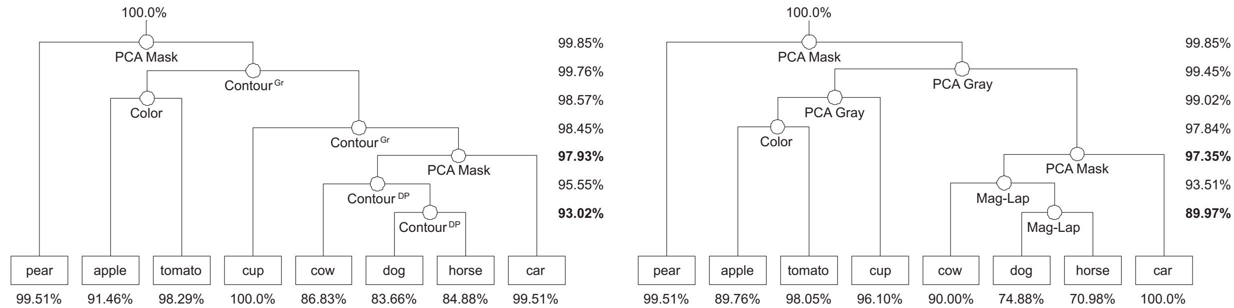


Figure 3: *Optimal multi-cue decision trees when all cues are available (left) and when local shape is not (right). The numbers to the right of each tree show the global recognition rate after each split. Note that the performance of both trees differs significantly only for distinguishing the animal categories.*

Thus, our long-term vision is to use this database as a training set and test on pictures taken under more realistic and less controlled viewing conditions. For this goal, we will produce a series of test sets with increasing difficulty, with objects placed in the real world including cluttered settings, occlusions, and different lighting conditions.

With the exception of the contour-based approaches, all methods analyzed in this paper have been global. It would be interesting to compare also local, part- or region-based approaches, such as [8, 12, 13, 27]. This work provides a framework in which they can be tested.

Acknowledgments: This work is part of the CogVis project, funded in part by the Comission of the European Union (IST-2000-29375), and the Swiss Federal Office for Education and Science (BBW 00.0617). The authors would like to thank Markus Graf for his expertise on categorization and Markus Kaeppeli for his contribution to the code.

References

[1] D. Marr. *Vision*. W.H. Freeman, 1982.

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

[3] R.P.N. Rao and D.H. Ballard. Object indexing using an iconic sparse distributed memory. In *ICCV'95*, 1995.

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

[5] B.W. Mel. Seemore: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. In *ICPR'96*, 1996.

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

[7] R.C. Nelson and A. Selinger. A cubist approach to object recognition. In *ICCV'98*, pages 614–621, 1998.

[8] D. Lowe. Object recognition from local scale invariant features. In *ICCV'99*, 1999.

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

[10] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *Trans. PAMI*, 20(1):23–38, 1998.

[11] H. Schneiderman and T. Kanade. A statistical method of 3d object detection applied to faces and cars. In *CVPR'00*, 2000.

[12] M. Weber, M. Welling, and P. Perona. Unsupervised learning of object models for recognition. In *ECCV'00*, 2000.

[13] C. Papageorgiou and T. Poggio. A trainable system for object detection. *IJCV*, 38(1):15–33, 2000.

[14] D. Forsyth and M. Fleck. Body plans. In *CVPR'97*, 1997.

[15] R.C. Nelson and A. Selinger. Large-scale tests of a keyed, appearance-based 3-d object recognition system. *Vision Research*, 38(15), 1998.

[16] Y. Keselman and S. Dickinson. Generic model abstraction from examples. In *CVPR'01*, 2001.

[17] R. Brown. How shall a thing be called? *Psychological Review*, 65:14–21, 1958.

[18] E. Rosch, C. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.

[19] G. Lakoff. *Women, Fire, and Dangerous Things–What Categories Reveal about the Mind*. Univ. of Chicago Press, 1987.

[20] L.W. Barsalou. Ad-hoc categories. *Memory and Cognition*, 11:211–227, 1983.

[21] S. Sclaroff. Deformable prototypes for encoding shape categories in image databases. *Pattern Recognition*, 30(4), 1997.

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cog. Neurosci.*, 3:71–86, 1991.

[23] A. Leonardis, H. Bischof, and J. Maver. Multiple eigenspaces. *Pattern Recognition*, 35(11):2613–2627, 2002.

[24] D. Macrini, A. Shokoufandeh, S. Dickinson, K. Siddiqi, and S. Zucker. View-based 3-d object recognition using shock graphs. In *ICPR'02*, 2002.

[25] S. Belongie, J. Malik, and J. Puchiza. Matching shapes. In *ICCV'01*, 2001.

[26] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, New York, 2nd edition, 2001.

[27] S. Agarwal and D. Roth. Learning a sparse representation for object detection. In *ECCV'02*, 2002.