Object Categorization

Object Categorization

Axel Pinz

Graz University of Technology
Austria

axel.pinz@tugraz.at

the essence of knowledge

Boston – Delft

**Foundations and Trends® in
Computer Graphics and Vision**

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

The preferred citation for this publication is A. Pinz, Object Categorization, Foundation and Trends® in Computer Graphics and Vision, vol 1, no 4, pp 255–353, 2005

Printed on acid-free paper

ISBN: 1-933019-91-3

Foundations and Trends® in Computer Graphics and Vision
Volume 1 Issue 4, 2005

Editorial Scope

Foundations and Trends® in Computer Graphics and Vision will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

Information for Librarians
Foundations and Trends® in Computer Graphics and Vision, 2005, Volume 1, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Foundations and Trends® in
Computer Graphics and Vision
Vol. 1, No 4 (2005) 255–353

DOI: 10.1561/0600000003

Object Categorization

Axel Pinz[1]

[1] *Graz University of Technology, Austria, axel.pinz@tugraz.at*

Abstract

This article presents foundations, original research and trends in the field of object categorization by computer vision methods. The research goals in object categorization are to detect objects in images and to determine the object's categories. Categorization aims for the recognition of generic classes of objects, and thus has also been termed 'generic object recognition'. This is in contrast to the recognition of specific, individual objects. While humans are usually better in generic than in specific recognition, categorization is much harder to achieve for today's computer architectures and algorithms. Major problems are related to the concept of a 'visual category', where a successful recognition algorithm has to manage large intra-class variabilities versus sometimes marginal inter-class differences. It turns out that several techniques which are useful for specific recognition can also be adapted to categorization, but there are also a number of recent developments in learning, representation and detection that are especially tailored to categorization.

Recent results have established various categorization methods that are based on local salient structures in the images. Some of these methods use just a 'bag of keypoints' model. Others include a certain amount of geometric modeling of 2D spatial relations between parts, or 'constellations' of parts. There is now a certain maturity in these approaches and they achieve excellent recognition results on rather complex image databases. Further work focused on the description of shape and object contour for categorization is only just emerging. However, there remain a number of important open questions, which also define current and future research directions. These issues include localization abilities, required supervision, the handling of many categories, online and incremental learning, and the use of a 'visual alphabet', to name a few. These aspects are illustrated by the discussion of several current approaches, including our own patch-based system and our boundary fragment-model. The article closes with a summary and a discussion of promising future research directions.

Contents

1
Introduction

This article provides a review of existing representations, algorithms, systems and databases for visual object categorization. It describes the state of the art in this field, which has been a long standing goal, and is still a mainly unsolved problem in computer vision research. The time chosen for writing is motivated by recent success in recognition from local, salient parts, which can be considered a significant step towards object categorization.

Who are the supposed readers of this document, and what potential benefits are there for them? Students and graduate students in computer vision will get a thorough review of the state of the art in visual object categorization. Researchers in computer vision might benefit from a more complete point of view, including a number of approaches which they have not focused on within the scope of their own research. Researchers in related fields should find this article a valuable reference.

But the article goes beyond a pure review of the state of the art. It includes original research in categorization, presents a prototype system for categorization, discusses our databases and provides experimental results on object categorization and localization in still images.

1.1 Problem statement

We can define visual object categorization as the process of assigning a specific object to a certain category. This process has also been termed 'generic object recognition' (generic OR), and it is in contrast to 'specific OR', which deals with the recognition of a specific, individual object. Examples of categories in generic OR are people, children, dogs, cars, bikes or dishes, while specific OR might aim at recognizing a certain individual, like Albert Einstein, or a specific object like my car. An individual object might also be termed a specific *instance* of a more generic category. Categories can also be organized in hierarchies (child – human being – mammal), and categories might overlap – a tall glass might be used as a vase. Throughout the remainder of this article, we will use the terms *'categorization'* for visual object categorization or generic OR, and *'specific OR'* for the recognition of individual objects.

Looking at humans, and comparing their recognition performance with artificial systems, it turns out that humans are much better in categorization than machines, but specific OR can often be handled more efficiently, reliably or simply faster by an artificial vision system. VanRullen and Thorpe [199] point out that humans can perform ultra-rapid categorization tasks. They can decide whether a briefly flashed image belongs to a certain category in less than 150ms, and they provide experimental evidence for the two categories 'animal' and 'means of transport'. On the other hand, there are numerous solutions to industrial inspection, which recognize and localize specific objects much faster and much more reliably than humans can do (see [31] for an example of such an industrial product, and [76] for the underlying theoretical foundations). A further aspect of categorization is the sheer number of visual categories. There is evidence from cognitive psychology that humans deal with about 30,000 different categories (see Biederman [21]). This would require solving currently intractable computational complexity.

This article sets out to answer the following questions: How can artificial systems perform categorization? What are the key building blocks that are required to build a categorization system? What are the main challenges? What are the bottlenecks and unsolved problems? This will also shed light on the more general question: Why is categorization simpler for humans than for machines and why is specific OR simpler for machines than for humans?

1.2 Historical development

In the following, I give a very brief sketch of some major landmarks in the history of object recognition research. This is not meant to be a complete review, but rather some useful information to pave the way for later discussion. One of the major early landmarks is certainly the work of David Marr [126], who proposed viewer-centered and object-centered representational levels (image – primal sketch – 2-1/2D sketch – 3D object model), as well as visual modules which can be used to generate these descriptions (e.g. 'shape from X' to produce a 2-1/2D sketch). Marr's ideas influenced at least a decade of research, and have led to a so-called 'reconstruction school' which advocates that 3D reconstruction and 3D modeling of a scene (and thus of the objects in the scene) are necessary for further reasoning.

On the other hand, there is the 'recognition school' which favors working in the 2D domain, with 2D images, features and descriptors which are extracted from these images. Their pattern classification [43] or pattern recognition [145] approach is fundamentally different from the reconstructionist paradigm. Much of the content of this article actually is in the spirit of a 'recognition school' approach. We will discuss, for example, the 'bag of keypoints' approach, in which salient points are extracted from images, and descriptors are calculated to form feature vectors. These feature vectors can be used to learn a discriminative model from training images, and to recognize (categorize) test images. But we will also present the generative 'constellation' model, which employs a 'light' 2D geometry in terms of spatial 2D relations between key parts of the object model.

Only very recently have we seen efforts to combine discriminative and generative approaches in categorization research. This confluence of recognition and reconstruction schools has already been predicted by Aloimonos and Shulman [4] in 1989.

There are further milestones which should be mentioned. Biederman [20, 21] proposed his 'recognition by components' (RBC) theory. Volumetric primitives, so-called 'geons' can be used to recognize objects in a qualitative (and thus generic) way. While this theory is quite elegant, its implementations (see [18, 39]) lacked due to low level vision problems, so that geon-based recognition has not been applied to real-world categorization problems. Research in perceptual grouping proceeds in a similar manner [167]. Low-level 2D primitives are grouped to build object descriptions, either in a pure bottom-up (data driven) manner, or top-down, including prior knowledge (models) about the expected image content. At the other end of the spectrum of potential solutions to OR, we find the idea of purely image- or 'appearance'-based recognition, for instance in parametric eigenspace [143]. This idea has triggered a vast number of extremely successful appearance-based approaches to specific OR.

In general, there has been more research in specific OR than in categorization in the past. Success in specific OR has influenced a number of approaches to categorization, although most of the developed algorithms for specific OR are not directly applicable to categorization. There is a paradigm of specific OR by alignment, in which spatial correspondence between groups of image features and model features is found by searching for the geometric transformation that aligns these features best. This includes affine transformations for planar objects [84] and 3D model to 2D image feature matching [121]. Another way to compare image and model features is to extract features which are invariant against certain geometric [142] or radiometric distortions [3]. Efficient indexing is needed, when a database of potentially many object models has to be matched against features extracted from a query image that contains a certain specific object. This can, for instance, be done by geometric hashing [207], a technique which is robust against partial occlusion and geometric transformations. Finally, the success in global appearance-based recognition [143] has motivated research in local appearance-based methods for specific OR [122]. At this point, we can observe that techniques for specific OR and for categorization meet.

In categorization, these various ideas have led to the development of a number of recent approaches which try to:

- model appearance more locally,
- group simple geometric primitives, and
- use learning algorithms to find common patterns that can be shared over many individuals of a category.

Within the past 5 years, we have seen a rapid development and rise in the success of object categorization in increasingly difficult, cluttered, and realistic scenes[1]. We can also observe a number of contributions from related fields as machine learning, neurosciences and cognitive psychology.

1.3 Potential applications

There are a number of obvious applications of categorization to image database annotation, image retrieval and video annotation. But potential applications of categorization go far beyond that. Reliable categorization in real-time will open up applications in surveillance, driver assistance, autonomous robots, interactive games, virtual and augmented reality and telecommunications. A more general view might include systems for 'cognitive personal assistance' with many potential aspects, ranging from user support in complex environments to very basic support capabilities for elderly or disabled people.

1.4 Outline of this review article

The article is structured in three major parts (Sections 2 – 4). I start with an in-depth analysis of major issues related to solving the problem of categorization mentioned in Section 2. This analysis provides at the same time an introduction to the main topics, which are then discussed in detail in Section 3 which presents the major building blocks for

[1] This may partly be related to recent European research initiatives. There has been substantial funding of basic research in 'Cognitive Vision' within the 5th framework program of the European Union, with an even broader perspective of 'Cognitive Systems' in the current, 6th framework program. There has been strong support of categorization research within these programs.

categorization systems. Finally, Section 4 presents two major aspects of our own research in categorization: a region-based approach, and categorization with a boundary-fragment-model.

You will probably recognize that the subject is quite broad and heterogeneous (ranging from the representation of scale in images, over machine learning, to 2D spatial models for categorization). Thus, there is no isolated section on the 'state of the art' and related work. I prefer, rather, to cite relevant publications throughout the whole article, which is hopefully more useful to you, the potential reader.

Finally, there is a common thread, which should provide some extra value for those who manage to read the complete article sequentially. However, many sections stand on their own and may also be consulted individually.

References

[1] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2005.

[2] S. Agarwal and D. Roth, "Learning a sparse representation for object recognition," in *Proc. ECCV*, pp. 113–130, 2002.

[3] R. O. Alferez and Y. F. Wang, "Geometric and illumination invariants for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 505–536, 1999.

[4] Y. Aloimonos and D. Shulman, *Integration of visual modules: An extension of the Marr paradigm*, Academic Press, 1989.

[5] J. Amores, N. Sebe, and P. Radeva, "Fast spatial pattern discovery integrating Boosting with constellations of contextual descriptions," in *Proc. CVPR*, pp. 769–774, 2005.

[6] C. H. Anderson, P. J. Burt, and G. S. van der Wal, "Change detection and tracking using pyramid transform techniques," in *Proc. SPIE Intelligent Robots and Computer Vision*, pp. 72–78, 1985.

[7] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 26–33, 1986.

[8] A. Bar-Hillel, T. Hertz, and D. Weinshall, "Object class recognition by boosting a part-based model," in *Proc. CVPR*, 2005.

[9] K. Barnard, V. Cardei, and B. Funt, "A comparison of color constancy algorithms. Part one: Methodology and experiments with synthesized data," *IEEE Transactions in Image Processing*, vol. 11, no. 9, pp. 972–984, 2002.

[10] K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of color constancy algorithms. Part two. Experiments with image data," *IEEE Transactions in Image Processing*, vol. 11, no. 9, pp. 985–996, 2002.

[11] E. Bart and S. Ullman, "Cross-generalization: Learning novel classes from a single example by feature replacement," in *Proc. CVPR*, 2005.

[12] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. CVPR*, pp. 774–781, 2000.

[13] P. R. Beaudet, "Rotational invariant image operators," in *Proc. ICPR*, pp. 579–583, 1978.

[14] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proc. CVPR*, pp. 1000–1006, 1997.

[15] J. S. Beis and D. G. Lowe, "Indexing without invariants in 3D object recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1000–1015, 1999.

[16] H. Bekel, I. Bax, G. Heidemann, and H. Ritter, "Adaptive computer vision: Online learning for object recognition," in *Proc. DAGM 2004*, (C. E. R. et al., ed.), pp. 447–454, Springer, 2004.

[17] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[18] R. Bergevin and M. D. Levine, "Generic object recognition: Building and matching coarse descriptions from line drawings," *PAMI*, vol. 15, no. 1, pp. 19–36, 1993.

[19] E. J. Bernstein and Y. Amit, "Part-based statistical models for object classification and detection," in *Proc. CVPR*, 2005.

[20] I. Biederman, "Human image understanding: Recent research and a theory," *CVGIP*, vol. 32, pp. 29–73, 1985.

[21] I. Biederman, "Visual object recognition," in *Chapter 4 of: An invitation to cognitive science, vol. 2, Visual Cognition*, (S. F. Kosslyn and D. N. Osherson, eds.), pp. 121–165, MIT press, 1995.

[22] M. Bister, J. Cornelis, and A. Rosenfeld, "A critical view of pyramid segmentation algorithms," *Pattern Recognition Letters*, vol. 11, no. 9, pp. 605–617, 1990.

[23] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification via pLSA," in *Proc. ECCV*, pp. 517–530, 2006.

[24] G. Bouchard and B. Triggs, "Hierarchical part-based visual object categorization," in *Proc. CVPR*, 2005.

[25] R. A. Brooks and L. A. Stein, "Building brains for bodies," *Autonomous Robots*, vol. 1, pp. 7–25, 1994.

[26] M. C. Burl, M. Weber, and P. Perona, "A probabilistic approach to object recognition using local photometry and global geometry," in *Proc. ECCV*, pp. 628–641, 1998.

[27] P. J. Burt, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[28] B. Caputo, C. Wallraven, and M. E. Nilsback, "Object categorization via local kernels," in *Proc. ICPR*, pp. 132–135, 2004.

[29] P. Carbonetto, G. Dorko, and C. Schmid, "Bayesian learning for weakly supervised object classification," Tech. Rep., INRIA Rhone-Alpes, Grenoble, France, August 2004.

[30] F. Chabat, G. Z. Yang, and D. M. Hansell, "A corner orientation detector," *Image and Vision Computing*, vol. 17, pp. 761–769, 1999.

[31] Cognex Corporation, "http://www.cognex.com/products/visiontools/patmax.asp," page visited April 26, 2005.

[32] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, 2002.

[33] D. Crandall, P. Felzenszwalb, and D. Huttenlocher, "Spatial priors for part-based recognition using statistical models," in *Proc. CVPR*, 2005.

[34] D. Crandall and D. Huttenlocher, "Weakly supervised learning of part-based spatial models for visual recognition," in *Proc. ECCV*, pp. 16–29, 2006.

[35] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in *ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–22, 2004.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, 2005.

[37] A. Demiriz, K. Bennett, and J. Shawe-Taylor, "Linear programming boosting via column generation," *Machine Learning*, vol. 46, no. 1-3, pp. 225–254, 2002.

[38] T. Deselaers, D. Keysers, and H. Ney, "Discriminative training for object recognition using image patches," in *Proc. CVPR*, pp. 157–162, 2005.

[39] S. Dickinson, A. Pentland, and A. Rosenfeld, "3-D Shape recovery using distributed aspect matching," *PAMI*, vol. 14, no. 2, pp. 174–198, 1992.

[40] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *In Proc. International Conference on Computer Vision*, 2003.

[41] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman, "The schema system," *Int. J. Computer Vision*, vol. 2, pp. 209–250, 1989.

[42] L. Dreschler and H. H. Nagel, "Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene," *Computer Graphics and Image Processing*, vol. 20, pp. 199–228, 1982.

[43] R. O. Duda, P. E. Hart, and D. C. Stork, *Pattern classification*, Wiley, 2001.

[44] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, "Object recognition as machine translation; Learning a lexicon for a fixed image vocabulary," in *Proc. ECCV*, pp. 97–112, 2002.

[45] I. R. Fasel, M. S. Bartlett, and J. R. Movellan, "A comparison of Gabor filter methods for automatic detection of facial landmarks," in *Proc. 5th Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[46] L. Fei-Fei, R. Fergus, and P. Perona, "A Bayesian approach to unsupervised one-shot learning of object categories," in *In Proc. International Conference on Computer Vision*, pp. 1134–1141, 2003.

[47] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. CVPR*, 2004.

[48] L. Fei-fei, R. Fergus, and A. Torralba, "Recognizing and learning object categories," http://people.csail.mit.edu/torralba/iccv2005, Tutorial presented at ICCV 2005 pages visited Feb. 7, 2006.

[49] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. CVPR*, 2005.

[50] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[51] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *Int. J. Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[52] P. Felzenszwalb and D. Huttenlocher, "Spatial priors for part-based recognition using statistical models," in *Proc. CVPR*, 2005.

[53] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *In Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR*, 2003.

[54] R. Fergus, P. Perona, and A. Zisserman, "A visual category filter for Google images," in *Proc. ECCV*, pp. 242–256, 2004.

[55] R. Fergus, P. Perona, and A. Zisserman, "Learning object categories from Google's image search," in *Proc. ICCV*, 2005.

[56] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," in *Proc. CVPR*, 2005.

[57] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation by image exploration," in *Proc. ECCV*, pp. 40–54, 2004.

[58] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Object detection by contour segment networks," in *Proc. ECCV*, pp. 14–28, 2006.

[59] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Trans. Computers*, vol. 22, no. 1, pp. 67–92, 1973.

[60] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Linear scale space," *Journal of Mathematical Imaging and Vision*, vol. 4, pp. 325–351, 1994.

[61] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, 1993.

[62] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Intercommission conference on Fast Processing of Photogrammetric Data*, pp. 281–305, 1987.

[63] D. Forsyth and J. Ponce, *Computer vision, a modern approach*, Prentice Hall, 2003.

[64] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[65] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995.

[66] Y. Freund and R. Shapire, "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[67] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," 1998.

[68] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Software*, vol. 3, pp. 209–226, 1977.

[69] M. Fussenegger, A. Opelt, A. Pinz, and P. Auer, "Object recognition using segmentation for feature detection," in *Proc. ICPR'04*, pp. 41–44, 2004.

[70] A. Ghosh and N. Petkov, "Robustness of shape descriptors to incomplete contour representations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1793–1804, 2005.

[71] J. Gibson, *The ecological approach to visual perception*, Lawrence Erlbaum, 1979.

[72] Z. Gigus and J. Malik, "Computing the aspect graph for line drawings of polyhedral objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 2, pp. 113–122, 1990.

[73] S. Gilles, *Robust description and matching of images*, Ph.D. thesis, University of Oxford, 1998.

[74] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, 2nd ed., 2002.

[75] G. Granlund, "Organization of architectures for cognitive vision systems," in *Proceedings of Workshop on Cognitive Vision*, (Schloss Dagstuhl, Germany), October 2003.

[76] W. E. L. Grimson, *Object recognition by computer: The role of geometric constraints*, MIT Press, 1990.

[77] M. Hanheide, C. Bauckhage, and G. Sagerer, "Memory consistency validation in a cognitive vision system," in *Proc. ICPR*, pp. 459–462, 2004.

[78] R. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[79] R. M. Haralick and L. G. Shapiro, *Computer and robot vision*, Vol. II, Addison-Wesley, 1993.

[80] C. Harris and M. Stevens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conference*, pp. 147–151, 1988.

[81] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, Springer, 2001.

[82] A. Holub and P. Perona, "A discriminative framework for modelling object classes," in *In Proc. Computer Vision and Pattern Recognition, CVPR*, 2005.

[83] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. IT-8, pp. 179–187, 1962.

[84] D. P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment with an image," *Int. J. Computer Vision*, vol. 5, no. 2, pp. 195–212, 1990.

[85] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley, 2001.

[86] J. Illingworth and J. Kittler, "A survey of the Hough transform," *CVGIP*, vol. 44, pp. 87–116, 1988.

[87] F. Jurie and C. Schmid, "Scale-invariant shape features for recognition of object categories," in *Proc. CVPR*, 2004.

[88] T. Kadir and M. Brady, "Scale, saliency and image description," *Int. J. Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[89] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *ECCV (1)*, pp. 228–241, 2004.

[90] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*, Wiley, 1990.

[91] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image Descriptors," in *Proc. CVPR*, pp. 506–513, 2004.

[92] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," *Pattern Recognition Letters*, vol. 1, pp. 95–102, 1982.

[93] J. J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.

[94] J. J. Koenderink, "Scale-time," *Biol. Cybern.*, vol. 58, pp. 159–162, 1988.

[95] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biol. Cybern.*, vol. 55, pp. 367–375, 1987.

[96] M. P. Kumar, P. H. S. Torr, and A. Zisserman, "Extending pictorial structures for object recognition," in *Proc. BMVC*, 2004.

[97] M. P. Kumar, P. H. S. Torr, and A. Zisserman, "Obj cut," in *Proc. CVPR*, 2005.

[98] l. Van Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns," in *Proc. ECCV*, pp. 642–651, 1996.

[99] R. Laganiére, "Morphological corner detection," in *Proc. 6th ICCV*, pp. 280–285, 1999.

[100] I. Laptev, "On space-time interest points," *Int. J. Computer Vision*, vol. 64, no. 2/3, pp. 107–123, 2005.

[101] I. Laptev and T. Lindeberg, "Space-time interest points," in *Proc. ICCV*, 2003.

[102] I. Laptev and T. Lindeberg, "Local descriptors for spatio-temporal recognition," in *Proc. ECCV Workshop on Spatial coherernce for visual motion analysis*, 2004.

[103] I. Laptev and T. Lindeberg, "Velocity adaptation of space-time interest points," in *Proc. ICPR*, 2004.

[104] L. J. Latecki, R. Lakämper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with single closed contour," in *Proc. CVPR*, pp. 424–429, 2000.

[105] S. Lazebnik, C. Schmid, and J. Ponce, "Semi-local affine parts for object recognition," in *Proc. BMVC*, 2004.

[106] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," in *Proc. ECCV Workshop on Statistical Learning in Computer Vision*, 2004.

[107] B. Leibe and B. Schiele, "Scale-invariant object categorization using a scale-adaptive mean-shift search," in *Proc. DAGM Pattern Recognition Symposium*, 2004.

[108] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Proc. CVPR*, 2005.

[109] A. Leonardis and H. Bischof, "Robust recognition using eigenimages," *Computer Vision and Image Understanding: CVIU*, vol. 78, no. 1, pp. 99–118, 2000.

[110] T. K. Leung, M. C. Burl, and P. Perona, "Probabilistic affine invariants for recognition," in *Proc. CVPR*, pp. 678–684, 1998.

[111] T. Lindeberg, "Scale-space for discrete signals," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234–254, 1990.

[112] T. Lindeberg, *Scale space theory in computer vision*, Kluwer, 1994.

[113] T. Lindeberg, "Linear spatio-temporal scale-space," in *Proc Scale-Space'97: Scale-Space Theories in Computer Vision*, pp. 113–127, Springer, 1997.

[114] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

[115] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 77–116, 1998.

[116] T. Lindeberg, "Time-recursive velocity-adapted spatio-temporal scale-space filters," in *Proc. ECCV*, pp. 52–67, Springer, 2002.

[117] T. Lindeberg, A. Akbarzadeh, and I. Laptev, "Galilean-diagonalized spactio-temporal interest points," in *Proc. ICPR*, 2004.

[118] T. Lindeberg and J.-O. Eklundh, "Scale-space primal sketch: Construction and experiments," *Image and Vision Computing*, vol. 10, no. 1, pp. 3–18, 1992.

[119] N. K. Logothetis and D. L. Sheinberg, "Visual object recognition," *Annu. Rev. Neurosci.*, vol. 19, pp. 577–621, 1996.

[120] D. G. Lowe, *Perceptual organization and visual cognition*, Kluwer, 1985.

[121] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, no. 3, pp. 355–395, 1987.

[122] D. G. Lowe, "Object recognition from local scale-invariant features," in *In Proc. International Conference on Computer Vision*, pp. 1150–1157, 1999.

[123] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, no. 2, pp. 91–110, 2004.

[124] S. Maitra, "Moment invariants," *Proceedings of the IEEE*, vol. 67, no. 4, pp. 697–699, 1979.

[125] R. Marée, P. Geurts, J. Piater, and L. Wehenkel, "Random subwindows for robust image classification," in *Proc. CVPR*, 2005.

[126] D. Marr, *Vision: A computational investigation into the human representation and processing of visual information.*, W. H. Freeman, 1982.

[127] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society London B*, vol. 207, pp. 187–217, 1980.

[128] A. M. Martinez and A. C. Kak, "PCA versus LDA," *PAMI*, vol. 23, no. 2, pp. 228–233, 2001.

[129] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. 13th BMVC*, pp. 384–393, 2002.

[130] T. Matsuyama and V. Hwang, "SIGMA: A framework for image understanding – integration of bottom-up and top-down analyses," in *Proc. 9th IJCAI*, pp. 908–915, 1985.
[131] P. Meer, "Stochastic image pyramids," *Computer Graphics and Image Processing*, vol. 45, pp. 269–294, 1989.
[132] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Proc. ICCV*, pp. 1792–1799, 2005.
[133] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proc. ICCV*, pp. 525–531, 2001.
[134] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *ECCV (1)*, pp. 128–142, 2002.
[135] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proc. CVPR*, pp. 257–263, 2003.
[136] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
[137] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
[138] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1/2, pp. 43–72, 2005.
[139] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Computer Vision and Image Understanding*, vol. 94, no. 1–3, pp. 3–27, 2004.
[140] A. Montanvert, P. Meer, and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 307–316, 1991.
[141] P. Moreels and P. Perona, "Common-frame model for object recognition," in *Proc. NIPS*, 2004.
[142] J. Mundy and A. Zisserman, eds., *Geometric invariance in computer vision*, MIT Press, 1992.
[143] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *Int.J. Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995.
[144] K. Murphy, A. Torralba, and W. T. Freeman, "Using the forest to see the trees: A graphical model relating features, objects, and scenes," in *Proc. NIPS*, 2003.
[145] M. Nadler and E. P. Smith, *Pattern recognition engineering*, Wiley, 1993.
[146] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. ECCV*, pp. 490–503, 2006.
[147] S. Obdržálek and J. Matas, "Object recognition using local affine frames on distinghuished regions," in *Proc. 13th BMVC*, pp. 113–122, 2002.
[148] S. Obdržálek and J. Matas, "Image retrieval using local compact DCT-based representation," in *Proc. 25th DAGM*, pp. 490–497, 2003.
[149] S. Obdržálek and J. Matas, "Sub-linear indexing for large scale object recognition," in *Proc. BMVC*, pp. 1–10, 2005.
[150] S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert, "Learning and inference in parametric switching linear dynamic systems," in *Proc. ICCV*, 2005.

[151] K. Ohba and K. Ikeuchi, "Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, pp. 1043–1047, 1997.

[152] A. Opelt, *Generic object recognition*, Ph.D. thesis, Graz University of Technology, 2006.

[153] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, "Weak hypotheses and boosting for generic object detection and recognition," in *ECCV'04*, (T. Pajdla and J. Matas, eds.), pp. 71–84, Springer, 2004.

[154] A. Opelt and A. Pinz, "Object localization with boosting and weak supervision for generic object recognition," in *Proc. SCIA*, 2005.

[155] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with Boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 416–431, 2006.

[156] A. Opelt, A. Pinz, and A. Zisserman, "A boundary-fragment-model for object detection," in *Proc. ECCV*, pp. 575–588, 2006.

[157] A. Opelt, A. Pinz, and A. Zisserman, "Incremental learning of object detectors using a visual shape alphabet," in *Proc. CVPR*, 2006. Best paper prize – runner up.

[158] S. Osher and N. Paragios, eds., *Geometric level set methods in imageing vision and graphics*, Springer, 2003.

[159] N. C. Oza and S. Russell, "Online bagging and boosting," in *Proc. Workshop on Artificial Intelligence and Statistics*, 2001.

[160] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 266–280, 2000.

[161] P. Perona, "A concise taxonomy of visual recognition," http://vasc.ri.cmu.edu/~hebert/04workshop/presentations/Perona-Sicily-Oct04.pdf, presented at the International Object Recognition Workshop, Sicily, October 1004, http://www.pascal-network.org/Workshops/IOR04/Programme/, pages visited May 10, 2005.

[162] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[163] A. Pinz and J.-P. Andreu, "Qualitative spatial reasoning to infer the camera position in generic object recognition," in *Proceedings ICPR'98*, pp. 770–773, 1998.

[164] A. Rosenfeld, *Multiresolution image processing and analysis*, Springer, 1984.

[165] P. M. Roth, H. Grabner, D. Skocaj, H. Bischof, and A. Leonardis, "Online conservative learning for person detection," in *Proc. VS-PETS Workshop at ICCV*, 2005.

[166] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *Proc. CVPR*, 2004.

[167] S. Sarkar and K. Bowyer, *Computing perceptual organization in computer vision*, World Scientific, 1994.

[168] F. Scalzo and J. H. Piater, "Statistical learning of visual feature hierarchies," in *Proc. CVPR*, 2005.

[169] F. Schaffalitzky and A. Zisserman, "Multi-view matching of unordered image sets, or 'How do I organize my holiday snaps?'," in *ECCV (1)*, pp. 414–431, 2002.

[170] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 19, no. 5, pp. 530–535, 1997.

[171] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[172] H. Schneiderman and T. Kanade, "Object detection using the statistics of parts," *Int. J. Computer Vision*, vol. 56, no. 3, pp. 151–177, 2004.

[173] G. Schweighofer, A. Opelt, and A. Pinz, "Improved object categorization by unsupervised object localization," in *Proc. Int. Workshop on Learning for Adaptable Visual Systems LAVS'04*, (St. Catherine's College, Cambridge), 2004.

[174] E. Seemann, B. Leibe, K. Mikolajczyk, and B. Schiele, "An evaluation of local shape-based features for pedestrian detection," in *Proc. BMVC*, 2005.

[175] A. Selinger and R. C. Nelson, "Improving appearance-based object recognition in cluttered background," in *Proc. ICPR*, pp. 1–8, 2000.

[176] M. E. Sereno, T. Trinath, M. Augath, and N. K. Logothetis, "Three-dimensional shape representation in monkey cortex," *Neuron*, vol. 36, pp. 635–652, 2002.

[177] T. Serre, L. Wolf, and T. Poggio, "A new biologically motivated framework for robust object recognition," in *In Proc. Computer Vision and Pattern Recognition, CVPR*, 2005.

[178] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, pp. 10–21, 1949.

[179] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[180] J. Shi and C. Tomasi, "Good features to track," in *Proc. CVPR*, pp. 593–600, 1994.

[181] J. Shotton, A. Blake, and R. Cipolla, "Contour-based learning for object detection," in *Proc. ICCV*, 2005.

[182] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost: Joint appearance, shape and context modeling for multi-class object recognitoin and segmentation," in *Proc. 9th ECCV*, pp. 1–15, 2006.

[183] A. Siebert, "Retrieval of gamma corrected images," *Pattern Recognition Letters*, vol. 22, no. 2, pp. 249–256, 2001.

[184] J. Sivic, B. C. Russell, A. A. Elfros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in *Proc. ICCV*, 2005.

[185] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proceedings of the International Conference on Computer Vision*, pp. 1470–1477, Oct. 2003.

[186] S. M. Smith and J. M. Brady, "SUSAN – a new approach to low level image processing," *Int. J. Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[187] L. Stark and K. Bowyer, "Generic recognition through qualitative reasoning about 3-d shape and object function," in *Proc. CVPR*, pp. 251–256, 1991.

[188] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Learning hierarchical models of scenes, objects, and parts," in *Proc. ICCV*, 2005.

[189] B. M. ter Haar Romeny, ed., *Geometry-driven diffusion in computer vision*, Kluwer, 1994.

[190] J. Thureson and S. Carlsson, "Appearance based qualitative image description for object class recognition," in *Proc. ECCV*, pp. 518–529, 2004.

[191] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection," in *Proc. CVPR*, 2004.

[192] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Proc. ICCV*, 2005.

[193] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[194] T. Tuytelaars, *Local, invariant features for registration and recognition*, Ph.D. thesis, K.U. Leuven, 2000.

[195] T. Tuytelaars and L. Van Gool, "Content-based image retrieval based on local affinely invariant regions," in *Proc. ICVS*, pp. 493–500, 1999.

[196] T. Tuytelaars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *Proc. 11th BMVC*, pp. 412–425, 2000.

[197] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[198] I. Ulusoy and C. M. Bishop, "Generative versus discriminative methods for object recognition," in *Proc. CVPR*, 2005.

[199] R. VanRullen and S. J. Thorpe, "Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects," *Perception*, vol. 30, pp. 655–668, 2001.

[200] V. N. Vapnik, *The nature of statistical learning theory*, Springer, 1999.

[201] M. M. Veloso, P. E. Rybski, and F. von Hundelshausen, "FOCUS: A generalized method for object discovery for robots that observe and interact with humans," in *Proc. HRI'06*, ACM, 2006.

[202] D. Vernon, "Cognitive vision – the development of a discipline," http://europa.eu.int/information_society/istevent/2004/cf/document.cfm?doc_id=568, page visited August 1st, 2006.

[203] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[204] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *ECCV (1)*, pp. 18–32, 2000.

[205] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. ICCV*, 2005.

[206] A. Witkin, "Scale-space filtering," in *Proc. 8th IJCAI*, pp. 1019–1022, 1983.

[207] H. J. Wolfson and I. Rigoutsos, "Geometric hashing: An overview," *IEEE Computational Science & Engineering*, vol. 4, no. 4, pp. 10–21, 1997.

[208] R. A. Young, "The Gaussian derivative model for spatial vision: I. retinal mechanisms," *Spatial Vision*, vol. 2, pp. 273–293, 1987.

[209] A. L. Yuille and T. A. Poggio, "Scaling theorems for zero-crosssings," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 15–25, 1986.

[210] W. Zhang, B. Yu, G. J. Zelinsky, and D. Samaras, "Object class recognition using multiple layer boosting with heterogeneous features," in *Proc. CVPR*, pp. 323–330, 2005.