Kernel Methods in Computer Vision

Kernel Methods in Computer Vision

Christoph H. Lampert

Max Planck Institute for Biological Cybernetics
72076 Tübingen
Germany
http://www.christoph-lampert.de
chl@tuebingen.mpg.de

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

The preferred citation for this publication is C. Lampert, Kernel Methods in Computer Vision, Foundations and Trends® in Computer Graphics and Vision, vol 4, no 3, pp 193–285, 2008

ISBN: 978-1-60198-268-1

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, 2008, Volume 4, 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. 4, No. 3 (2008) 193–285

DOI: 10.1561/0600000027

Kernel Methods in Computer Vision

Christoph H. Lampert

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, http://www.christoph-lampert.de, chl@tuebingen.mpg.de

Abstract

Over the last years, *kernel methods* have established themselves as powerful tools for computer vision researchers as well as for practitioners. In this tutorial, we give an introduction to kernel methods in computer vision from a geometric perspective, introducing not only the ubiquitous support vector machines, but also less known techniques for regression, dimensionality reduction, outlier detection, and clustering. Additionally, we give an outlook on very recent, non-classical techniques for the prediction of structure data, for the estimation of statistical dependency, and for learning the kernel function itself. All methods are illustrated with examples of successful application from the recent computer vision research literature.

Contents

1

Overview

Computer vision has established itself as a broad subfield of computer science. It spans all areas for building automatic systems that extract information from images, covering a range of applications, from the organization of visual information, over control and monitoring tasks, to interactive and real-time systems for human–computer interaction. Despite this variability, some principled algorithms have emerged over the last years and decades that are useful in many different scenarios and thereby transcend the boundaries of specific applications. One recently very successful class of such algorithms are *kernel methods*. Based on the fundamental concept of defining *similarities* between objects they allow, for example, the prediction of properties of new objects based on the properties of known ones (*classification, regression*), or the identification of common subspaces or subgroups in otherwise unstructured data collections (*dimensionality reduction, clustering*).

1.1 The Goals of This Tutorial

With this tutorial, we aim at giving an introduction to kernel methods with emphasis on their use in computer vision. In the chapter

"Introduction to Kernel Methods" we use the problem of binary classification with *support vector machines* as introductory example in order to motivate and explain the fundamental concepts underlying all kernel methods. Subsequently, *"Kernels for Computer Vision"* gives an overview of the kernel functions that have been used in the area of computer vision. It also introduces the most important concepts one needs to know for the design of new kernel functions. Although support vector machines (SVMs) are the most popular examples of kernel methods, they are by far not the only useful ones. In the rest of this tutorial, we cover a variety of kernel methods that go beyond binary classification, namely algorithms for *"Multiclass Classification"*, *"Outlier Detection"*, *"Regression"*, *"Dimensionality Reduction"*, and *"Clustering"*. We also include some recent non-standard techniques, namely *"Structured Prediction"*, *"Dependency Estimation"*, and techniques for *"Learning the Kernel"* from data. In each case, after introducing the underlying idea and mathematical concepts, we give examples from the computer vision research literature where the methods have been applied successfully. It is our hope that this double-tracked approach will give pointers into both directions, theory and application, for the common benefit of researchers as well as practitioners.

1.2 What This Tutorial Is Not

This work is not meant to replace an introduction into *machine learning* or generic *kernel methods*. There are excellent textbooks for this purpose, e.g., [88] and [90]. In contrast to a formal introduction, we will sometimes take shortcuts and appeal to the reader's geometric intuition. This is not out of disrespect for the underlying theoretical concepts, which are in fact one of the main reasons why kernel methods have become so successful. It is rather because otherwise we would not be able to achieve our main goal: to give a concise overview of the plethora of kernel methods and to show how one can use them for tackling many interesting computer vision problems.

The limited space that is available in a text like this has another unfortunate consequence: we have to omit a lot of technical details that most textbooks on kernel method spend many pages on. In particular, we will not cover: *probabilistic foundations*, such as the statistical assumptions on how the data we work with was generated; *statistical learning theory*, including the highly elegant PAC theory and generalization bounds; *optimization theory*, such as dualization and convexity; and *numerics*, for example the many methods developed to solve the SVMs and related training problems. All good textbooks on support vector machines and kernel methods cover at least some of these topics, and it is our hope that after reading this introduction into *Kernel Method for Computer Vision*, your interest will be aroused to do further background reading.

References

[1] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng, "Discriminative learning of markov random fields for segmentation of 3D scan data," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 169–176, 2005.

[2] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *International Conference on Machine Learning (ICML)*, 2004.

[3] H. S. Baird, "Document image defect models and their uses," in *International Conference on Document Analysis and Recognition (ICDAR)*, 1993.

[4] G. H. Bakır, M. Wu, and J. Eichhorn, "Maximum-margin feature combination for detection and categorization," Technical Report, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 2005.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *European Conference on Computer Vision (ECCV)*, pp. 404–417, May 2006.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 19, no. 7, pp. 711–720, 1997.

[7] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[8] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of Machine Learning Research (JMLR)*, vol. 2, pp. 125–137, 2002.

[9] K. P. Bennett and E. J. Bredensteiner, "Duality and geometry in SVM classifiers," in *International Conference on Machine Learning (ICML)*, 2000.

[10] C. M. Bishop, *Neural Networks for Pattern Recognition.* USA: Oxford University Press, 1995.

[11] M. B. Blaschko and C. H. Lampert, "Correlational spectral clustering," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[12] M. B. Blaschko and C. H. Lampert, "Learning to localize objects with structured output regression," in *European Conference on Computer Vision (ECCV)*, 2008.

[13] M. B. Blaschko and C. H. Lampert, "Object localization with global and local context kernels," in *British Machine Vision Conference (BMVC)*, 2009.

[14] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *International Conference on Image and Video Retrieval (CIVR)*, pp. 401–408, 2007.

[15] S. P. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[16] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, no. 9, pp. 1124–1137, 2004.

[17] L. Breiman, *Classification and Regression Trees.* Chapman & Hall/CRC, 1998.

[18] M. Bădoiu, S. Har-Peled, and P. Indyk, "Approximate clustering via core-sets," in *ACM Symposium on Theory of Computing*, pp. 250–257, 2002.

[19] O. Chapelle, "Training a support vector machine in the primal," *Neural Computation*, vol. 19, no. 5, pp. 1155–1178, 2007.

[20] O. Chapelle, "Training a support vector machine in the primal," *Neural Computation*, vol. 19, no. 5, pp. 1155–1178, 2007.

[21] M. Charikar, S. Khuller, D. Mount, and G. Narasimhan, "Algorithms for facility location problems with outliers," in *ACM SIAM Symposium on Discrete Algorithms (SODA)*, pp. 642–651, 2001.

[22] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *IEEE International Conference on Image Processing (ICIP)*, pp. 34–37, 2001.

[23] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *Journal of Machine Learning Research (JMLR)*, vol. 2, pp. 265–292, 2001.

[24] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. Kandola, "On kernel-target alignment," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 15, pp. 367–373, 2001.

[25] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *ECCV Workshop on Statistical Learning in Computer Vision*, 2004.

[26] D. Decoste and B. Schölkopf, "Training invariant support vector machines," *Machine Learning*, vol. 46, no. 1, pp. 161–190, 2002.

[27] T. Deselaers, D. Keysers, and H. Ney, "Improving a discriminative approach to object recognition using image patches," in *Symposium of the German Pattern Recognition Society (DAGM)*, pp. 326–335, 2005.

[28] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.

[29] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pp. 65–72, 2005.

[30] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 10, pp. 155–161, 1996.

[31] B. Efron and G. Gong, "A leisurely look at the bootstrap, the jackknife, and cross-validation," *American Statistician*, vol. 37, no. 1, pp. 36–48, 1983.

[32] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[33] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Transactions on Computers*, vol. 100, no. 22, pp. 67–92, 1973.

[34] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *Journal of Machine Learning Research (JMLR)*, vol. 3, pp. 1289–1305, 2003.

[35] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf, "Kernel measures of conditional dependence," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 22, pp. 489–496, 2008.

[36] P. Gehler and S. Nowozin, "Feature combination methods for kernel classifiers," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[37] P. Gehler and S. Nowozin, "Let the kernel figure it out; principled learning of pre-processing for kernel classifiers," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[38] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 780–784, 2002.

[39] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 1458–1465, 2005.

[40] D. Gusfield, *Algorithms on Strings, Trees, and Sequences.* New York: Cambridge University Press, 1997.

[41] B. Haasdonk and H. Burkhardt, "Invariant kernel functions for pattern analysis and machine learning," *Machine Learning*, vol. 68, no. 1, pp. 35–61, 2007.

[42] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, pp. 147–151, 1988.

[43] M. Hein and O. Bousquet, "Hilbertian metrics and positive definite kernels on probability measures," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2005.

[44] D. Hiemstra, "A probabilistic justification for using tf×idf term weighting in information retrieval," *International Journal on Digital Libraries*, vol. 3, no. 2, pp. 131–139, 2000.

[45] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[46] R. Horst and N. V. Thoai, "DC programming: Overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, 1999.

[47] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3–4, pp. 321–377, 1936.

[48] C. Hsu and C. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[49] Y. G. Jiang, C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *International Conference on Image and Video Retrieval (CIVR)*, pp. 494–501, 2007.

[50] T. Joachims, "Training linear SVMs in linear time," in *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 217–226, New York, NY, USA: ACM, 2006.

[51] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer, 1986.

[52] M. Kloft, U. Brefeld, P. Laskov, and S. Sonnenburg, "Non-sparse multiple kernel learning," in *NIPS Workshop on Kernel Learning: Automatic Selection of Optimal Kernels*, 2008.

[53] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1137–1145, 1995.

[54] A. Kumar and C. Sminchisescu, "Support kernel machines for object recognition," in *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[55] J. T. Y. Kwok and I. W. H. Tsang, "The pre-image problem in kernel methods," *IEEE Transactions on Neural Networks*, vol. 15, no. 6, pp. 1517–1525, 2004.

[56] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *International Conference on Machine Learning (ICML)*, pp. 282–289, 2001.

[57] P. L. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," *International Journal of Neural Systems*, vol. 10, no. 5, pp. 365–378, 2000.

[58] C. H. Lampert and M. B. Blaschko, "A multiple kernel learning approach to joint multi-class object detection," in *Symposium of the German Pattern Recognition Society (DAGM)*, pp. 31–40, 2008.

[59] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Beyond sliding windows: Object localization by efficient subwindow search," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[60] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Efficient subwindow search: A branch and bound framework for object localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2009.

[61] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research (JMLR)*, vol. 5, pp. 27–72, 2004.

[62] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2169–2178, 2006.

[63] Y. Li, S. Gong, J. Sherrah, and H. Liddell, "Support vector machine based multi-view face detection and recognition," *Image and Vision Computing*, vol. 22, no. 5, pp. 413–427, 2004.

[64] A. Linden and J. Kindermann, "Inversion of multilayer nets," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 425–430, 1989.

[65] E. Loupias, N. Sebe, S. Bres, and J. M. Jolion, "Wavelet-based salient points for image retrieval," in *IEEE International Conference on Image Processing (ICIP)*, pp. 518–521, 2000.

[66] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[67] S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines," in *SPIE Symposium on Electronic Imaging*, pp. 35–45, San Jose, CA, 2004.

[68] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, pp. 281–296, 1967.

[69] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. Wiley New York, 1992.

[70] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K. R. Müller, "Fisher discriminant analysis with kernels," in *IEEE Workshop on Neural Networks for Signal Processing*, pp. 41–48, 1999.

[71] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 1, pp. 63–86, 2004.

[72] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*. Cambridge, MA, USA: MIT Press, 1992.

[73] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia Object Image Library (COIL-100)," Technical Report, Department of Computer Science Columbia University, New York, 1996.

[74] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *European Conference on Computer Vision (ECCV)*, pp. 490–503, 2006.

[75] S. Nowozin, G. Bakır, and K. Tsuda, "Discriminative subsequence mining for action classification," in *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[76] F. Odone, A. Barla, and A. Verri, "Building kernels from binary strings for image matching," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 169–180, 2005.

[77] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Learning*, pp. 185–208, MIT Press, 1999.

[78] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multiclass classification," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 13, pp. 547–553, 1999.

[79] N. Quadrianto, L. Song, and A. J. Smola, "Kernelized sorting," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 22, 2008.

[80] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[81] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[82] A. Rényi, "On measures of dependence," *Acta Mathematica Hungarica*, vol. 10, no. 3, pp. 441–451, 1959.

[83] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of Machine Learning Research (JMLR)*, vol. 5, pp. 101–141, 2004.

[84] C. Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables," in *International Conference on Machine Learning (ICML)*, pp. 515–521, 1998.

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

[86] B. Schölkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," in *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 252–257, 1995.

[87] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[88] B. Schölkopf and A. J. Smola, *Learning with Kernels*. MIT Press, 2002.

[89] S. Shalev-Shwartz, Y. Singer, and N. Srebro, "Pegasos: Primal estimated sub-GrAdient SOlver for SVM," in *International Conference on Machine Learning (ICML)*, pp. 807–814, 2007.

[90] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[91] J. Sietsma and R. J. F. Dow, "Creating artificial neural networks that generalize," *Neural Networks*, vol. 4, no. 1, pp. 67–79, 1991.

[92] P. Simard, Y. LeCun, J. S. Denker, and B. Victorri, "Transformation invariance in pattern recognition — Tangent distance and tangent propagation," in *Neural Networks: Tricks of the Trade*, pp. 239–274, Springer, 1999.

[93] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision*, pp. 1470–1477, 2003.

[94] J. R. Smith and S. F. Chang, "Tools and techniques for color image retrieval," in *Storage and Retrieval for Image and Video Databases (SPIE)*, pp. 426–437, 1996.

[95] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *The Journal of Machine Learning Research*, vol. 7, pp. 1531–1565, 2006.

[96] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision (IJCV)*, vol. 7, no. 1, pp. 11–32, 1991.

[97] M. Szummer, P. Kohli, and D. Hoiem, "Learning CRFs using graph cuts," in *European Conference on Computer Vision (ECCV)*, pp. 582–595, 2008.

[98] B. Taskar, C. Guestrin, and D. Koller, "Max-margin markov networks," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 17, 2003.

[99] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[100] M. Tipping and B. Schölkopf, "A kernel approach for vector quantization with guaranteed distortion bounds," in *Artificial Intelligence and Statistics (AISTATS)*, pp. 129–134, 2001.

[101] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," *Journal of Machine Learning Research (JMLR)*, vol. 6, pp. 1453–1484, 2005.

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

[103] T. Tuytelaars, C. Lampert, M. Blaschko, and W. Buntine, "Unsupervised object discovery: A comparison," *International Journal of Computer Vision (IJCV)*, 2009.

[104] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[105] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[106] M. Varma and D. Ray, "Learning the discriminative power-invariance tradeoff," in *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[107] C. Wallraven, B. Caputo, and A. Graf, "Recognition with local features: The kernel recipe," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 257–264, 2003.

[108] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping, "Use of the zero norm with linear models and kernel methods," *Journal of Machine Learning Research (JMLR)*, vol. 3, pp. 1439–1461, 2003.

[109] J. Wood, "Invariant pattern recognition: A review," *Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996.

[110] M. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215–220, 2002.

[111] R. A. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spatial Vision*, vol. 2, no. 4, pp. 273–293, 1987.

[112] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision (IJCV)*, vol. 73, no. 2, pp. 213–238, 2007.