Decision Forests

A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning

Decision Forests

A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning

---

Antonio Criminisi

Microsoft Research Ltd.
Cambridge, UK
antcrim@microsoft.com

Jamie Shotton

Microsoft Research Ltd.
Cambridge, UK
jamiesho@microsoft.com

Ender Konukoglu

Microsoft Research Ltd.
Cambridge, UK
enderk@microsoft.com

now
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 A. Criminisi, J. Shotton and E. Konukoglu, Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Foundations and Trends® in Computer Graphics and Vision, vol 7, nos 2–3, pp 81–227, 2011

ISBN: 978-1-60198-540-8

Foundations and Trends® in Computer Graphics and Vision
Volume 7 Issues 2–3, 2011

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, 2011, Volume 7, 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. 7, Nos. 2–3 (2011) 81–227

DOI: 10.1561/0600000035

Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning

Antonio Criminisi[1], Jamie Shotton[2], and Ender Konukoglu[3]

[1] *Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, antcrim@microsoft.com*
[2] *Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, jamiesho@microsoft.com*
[3] *Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, enderk@microsoft.com*

Abstract

This review presents a unified, efficient model of random decision forests which can be applied to a number of machine learning, computer vision, and medical image analysis tasks.

Our model extends existing forest-based techniques as it unifies classification, regression, density estimation, manifold learning, semi-supervised learning, and active learning under the same decision forest framework. This gives us the opportunity to write and optimize the core implementation only once, with application to many diverse tasks.

The proposed model may be used both in a discriminative or generative way and may be applied to discrete or continuous, labeled or unlabeled data.

The main contributions of this review are: (1) Proposing a unified, probabilistic and efficient model for a variety of learning tasks; (2) Demonstrating margin-maximizing properties of classification forests; (3) Discussing probabilistic regression forests in comparison with other nonlinear regression algorithms; (4) Introducing density forests for estimating probability density functions; (5) Proposing an efficient algorithm for sampling from a density forest; (6) Introducing manifold forests for nonlinear dimensionality reduction; (7) Proposing new algorithms for transductive learning and active learning. Finally, we discuss how alternatives such as random ferns and extremely randomized trees stem from our more general forest model.

This document is directed at both students who wish to learn the basics of decision forests, as well as researchers interested in the new contributions. It presents both fundamental and novel concepts in a structured way, with many illustrative examples and real-world applications. Thorough comparisons with state-of-the-art algorithms such as support vector machines, boosting and Gaussian processes are presented and relative advantages and disadvantages discussed. The many synthetic examples and existing commercial applications demonstrate the validity of the proposed model and its flexibility.

Contents

1

Overview and Scope

This review presents a unified, efficient model of random decision forests which can be used in a number of applications such as scene recognition from photographs, object recognition in images, automatic diagnosis from radiological scans and semantic text parsing. Such applications have traditionally been addressed by different, supervised or unsupervised machine learning techniques.

In this review, we formulate diverse learning tasks such as regression, classification and semi-supervised learning as instances of the same general decision forest model. The unified framework further extends to novel uses of forests in tasks such as density estimation and manifold learning. The underlying unified framework gives us the opportunity to implement and optimize the general algorithm for all these tasks only once, and then adapt it to individual applications with relatively small changes.

This review is directed at engineers and PhD students who wish to learn the basics of decision forests as well as more senior researchers interested in the new research contributions.

We begin by presenting a roughly chronological, non-exhaustive survey of decision trees and forests, and their use in the past two decades. Further references will be available in the relevant sections.

1.1 A Chronological Literature Review

One of the earlier works on decision trees is the seminal "Classification and Regression Trees (CART)" book by Breiman et al. [12], where the authors describe the basics of decision trees and their use for both classification and regression problems. Following that publication researchers then focused on algorithms for constructing (learning) optimal decision trees for different tasks using available training data. For this purpose, one of the most popular algorithms is "C4.5" of Quinlan [81]. Although, decision trees were proven to be useful, their application remained limited to relatively low dimensional data.

In the nineties, researchers discovered how using ensembles of learners (e.g., generic "weak" classifiers) yields greater accuracy and generalization. This seems particulary true for high dimensional data, as often encountered in real life applications. One of the earliest references to ensemble methods is the boosting algorithm of Schapire [87], where the author discusses how iterative re-weighting of training data can be used to build "strong" classifiers as linear combination of many weak ones.

Combining the ideas of decision trees and ensemble methods gave rise to decision forests, that is, ensembles of randomly trained decision trees. The idea of constructing and using ensembles of trees with randomly generated node tests was introduced for the first time in the work of Amit and Geman [1, 2] for handwritten digit recognition. In that work the authors also propose using the mean of the tree probabilities as output of the tree ensemble.

In the subsequent work of Ho [47] tree training via randomized partitioning of the feature space is discussed further, and in [48] forests are shown to yield superior generalization to both boosting and pruned C4.5-trained trees, on some tasks. The author also shows comparisons between different split functions in the tree nodes.

Breiman's later work in [10, 11] further consolidated the role of random forests and popularized their use. There, the author introduces a different way of injecting randomness in the forest by randomly sampling the labeled training data (namely "bagging"). The author also describes techniques for predicting the forest test error based on measures of tree strength and correlation.

In computer vision, ensemble methods became popular with the seminal face and pedestrian detection papers of Viola and Jones [107, 108]. Random decision forests where used in [63] for image classification and in [60] for keypoint tracking in videos. Recent years have seen an explosion of forest-based techniques in the machine learning, vision and medical imaging literature [9, 15, 25, 31, 35, 37, 58, 59, 65, 67, 68, 69, 74, 79, 89, 92, 97, 110]. Decision forests compare favorably with respect to other techniques [15] and have lead to one of the biggest success stories of computer vision in recent years: the Microsoft Kinect for XBox 360 [39, 91, 66].

References

[1] Y. Amit and D. Geman, “Randomized inquiries about shape; an application to handwritten digit recognition,” Technical Report 401, Department of Statistics, University of Chicago, IL, 1994.

[2] Y. Amit and D. Geman., “Shape quantization and recognition with randomized trees,” *Neural Computation*, vol. 9, pp. 1545–1588, 1997.

[3] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, 2003.

[4] B. Benfold and I. Reid, “Unsupervised learning of a scene-specific coarse gaze estimator,” in *International Conference on Computer Vision*, Barcelona, Spain, 2011.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[6] C. M. Bishop, M. Svensen, and C. K. I. Williams, “GTM: The generative topographic mapping,” *Neural Computation*, 1998.

[7] A. Bjoerck, *Numerical Methods for Least Squares Problems*. SIAM, 1996.

[8] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011.

[9] A. Bosch, A. Zisserman, and X. Munoz, “Image classification using random forests and ferns,” in *Institute of Electrical and Electronics Engineers International Conference on Computer Vision*, 2007.

[10] L. Breiman, “Random forests,” Technical Report TR567, UC Berkeley, 1999.

[11] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[13] I. Budvytis, V. Badrinarayanan, and R. Cipolla, "Semi-supervised video segmentation using tree structured graphical models," in *Computer Vision and Pattern Recoginition*, 2011.

[14] A. Burr, "Active learning literature survey," Technical Report 2010-09-14, University Wisconsin Madison, Computer Sciences Technical Report, 2010.

[15] R. Caruana, N. Karampatziakis, and A. Yessenalina, "An empirical evaluation of supervised learning in high dimensions," in *International Conference on Machine Learning*, pp. 96–103, 2008.

[16] L. Cayton, "Algorithms for manifold learning," Technical Report CS2008-0923, UCSD, 2005.

[17] P. Chandna, S. Deswal, and M. Pal, "Semi-supervised learning based prediction of musculoskeletal disorder risk," *Journal of Industrial and Systems Engineering*, 2010.

[18] O. Chapelle, B. Schölkopf, and A. Zien, eds., *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006.

[19] Y. Chen, T.-K. Kim, and R. Cipolla, "Silhouette-based object phenotype recognition using 3D shape priors," in *International Conference on Computer Vision*, Barcelona, Spain, 2011.

[20] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[21] K. Crammer and Y. Singer, "On the algorithmic implementation of multi-class SVMs," *Journal of Machine Learning Research*, 2001.

[22] A. Criminisi, *Accurate Visual metrology from Single and Multiple Uncalibrated Images*. Distinguished dissertation series. Springer, 2001.

[23] A. Criminisi, T. Sharp, and A. Blake, "Geos: Geodesic image segmentation," in *Proceedings of European Conference on Computer Vision*, 2008.

[24] A. Criminisi, J. Shotton, and S. Bucciarelli, "Decision forests with long-range spatial context for organ localization in CT volumes," in *Medical Image Computing and Computer Assisted Intervention Workshop on Probabilistic Models for Medical Image Analysis*, 2009.

[25] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *Medical Image Computing and Computer Assisted Intervention Workshop on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging*, Beijing, 2010.

[26] J. De Porte, B. M. Herbst, W. Hereman, and S. J. van Der Walt, "An introduction to diffusion maps," *Techniques*, 2008.

[27] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.

[28] K. Driessens, P. Reutemann, B. Pfahringer, and C. Leschi, "Using weighted nearest neighbour to benefit from unlabelled data," in *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, 2010.

[29] N. Duchateau, M. De Craene, G. Piella, and A. F. Frangi, "Characterizing pathological deviations from normality using constrained manifold learning," in *Medical Image Computing and Computer Assisted Intervention*, 2011.

[30] M. Everingham, L. van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal on Computer Vision*, vol. 88, pp. 303–338, June 2010.

[31] G. Fanelli and J. Gall, "Real time head pose estimation with random regression forests," in *Institute of Electrical and Electronics Engineers Computer Vision and Pattern Recognition*, 2011.

[32] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the Association for Computing Machinery*, vol. 24, pp. 381–395, 1981.

[33] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma, "Learning the structure of manifolds using random projections," in *Neural Information Processing Systems*, 2007.

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

[35] J. Gall and V. Lempitsky, "Class-specific Hough forest for object detection," in *Institute of Electrical and Electronics Engineers Computer Vision and Pattern Recognition*, Miami, 2009.

[36] S. Gerber, T. Tasdizen, S. Joshi, and R. Whitaker, "On the manifold structure of the space of brain images," in *Medical Image Computing and Computer Assisted Intervention*, 2009.

[37] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance," *Neuroimage*, 2011.

[38] P. Geurts, "Extremely randomized trees," in *Machine Learning*, 2003.

[39] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in *Institute of Electrical and Electronics Engineers International Conference on Computer Vision*, 2011.

[40] R. Guerrero, R. Wolz, and D. Rueckert, "Laplacian eigenmaps manifold learning for landmark localization in brain MR images," in *Medical Image Computing and Computer Assisted Intervention*, 2011.

[41] S. S. Gupta, "Probability integrals of multivariate normal and multivariate t," *Annals of Mathematical Statistics*, vol. 34, no. 3, 1963.

[42] J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos, "GRAM: A framework for geodesic registration on anatomical manifolds," *Medical Image Analysis*, vol. 14, no. 5, 2010.

[43] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd Edition, 2003.

[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. 2001.

[45] D. Heath, S. Kasif, and S. Salzberg, "Induction of oblique decision trees," *Journal of Artificial Intelligence Research*, vol. 2, no. 2, pp. 1–32, 1993.

[46] C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Random projections for manifold learning — proofs and analysis," in *Neural Information Processing Systems*, 2007.

[47] T. K. Ho, "Random decision forests," in *International Conference on Document Analysis and Recognition*, pp. 278–282, 1995.
[48] T. K. Ho, "The random subspace method for constructing decision forests," *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
[49] http://research.microsoft.com/groups/vision/decisionforests.aspx.
[50] http://research.microsoft.com/projects/medicalimageanalysis/.
[51] T. Joachims, "Advances in kernel methods — support vector learning," chapter *Making Large-Scale SVM Learning Practical*, The MIT Press, 1999.
[52] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
[53] B. M. Kelm, S. Mittal, Y. Zheng, A. Tsymbal, D. Bernhardt, F. Vega-Higuera, K. S. Zhou, P. Meer, and D. Comaniciu, "Detection, grading and classification of coronary stenoses in computed tomography angiography," in *Medical Image Computing and Computer Assisted Intervention*, 2011.
[54] P. Kontschieder, S. Rota Buló, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *International Conference on Computer Vision*, Barcelona, Spain, 2011.
[55] E. Konukoglu, A. Criminisi, S. Pathak, D. Robertson, S. White, D. Haynor, and K. Siddiqui, "Robust linear registration of CT images using random regression forests," in *Society of Photographic Instrumentation Engineers PIE Medical Imaging*, 2011.
[56] M. Kristan, D. Skocaj, and A. Leonardis, "Incremental learning with gaussian mixture models," in *Computer Vision Winter Workshop (CVWW)*, Moravske Toplice, Slovenia, 2008.
[57] C. H. Lampert, "Kernel methods in computer vision," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 3, 2008.
[58] C. Leistner, A. Saffari, J. Santner, and H. Bischoff, "Semi-supervised random forests," in *International Conference on Computer Vision*, 2009.
[59] V. Lempitsky, M. Verhoek, A. Noble, and A. Blake, "Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography," in *Functional Imaging and Modelling of the Heart (FIMH)*, 2009.
[60] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, 2006.
[61] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
[62] T. Malisievicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-svms for object detection and beyond," in *International Conference on Computer Vision*, Barcelona, Spain, 2011.
[63] R. Maree, P. Geurts, J. Piater, and L. Wehenkel, "Random subwindows for robust image classification," in *Proceedings of Computer Vision and Pattern Recognition*, 2005.
[64] N. Meinshausen, "Node harvest," *The Annals of Applied Statistics*, vol. 4, no. 4, 2010.

[65] B. Menze, B. M. Kelm, D. N. Splitthoff, U. Koethe, and F. A. Hamprecht, "On oblique random forests," in *Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database*, 2011.

[66] Microsoft Corp. Redmond WA, Kinect for XBox 360.

[67] A. Montillo and H. Ling, "Age regression from faces using random forests," in *International Conference on Image Processing*, 2009.

[68] A. Montillo, J. Shotton, J. E. Winn, D. Metaxas, E. Iglesias, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of CT images," in *Information Processing in Medical Imaging (IPMI)*, 2011.

[69] F. Moosman, B. Triggs, and F. Jurie, "Fast discriminative visual codebooks using randomized clustering forests," in *Proceedings of Neural Information Processing Systems*, 2006.

[70] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and eigenfunctions of Fokker-Plank operators," in *Proceedings of Neural Information Processing Systems*, 2005.

[71] R. Navaratnam, A. W. Fitzgibbon, and R. Cipolla, "The joint manifold model for semi-supervised multi-valued regression," in *International Conference on Computer Vision*, 2007.

[72] R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, pp. 125–139, 2001.

[73] S. Nowozin and C. H. Lampert, "Structured learning and prediction in computer vision," *Foundations and Trends in Computer Graphics and Vision*, vol. 6, no. 3, 4, 2010.

[74] S. Nowozin, C. Rother, S. Bagon, T. Sharp, B. Yao, and P. Kohli, "Decision tree fields," in *International Conference on Computer Vision*, 2011.

[75] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns," *Institute of Electrical and Electronics Engineers Transactions on Pattern Aanalysi and Machine Intelligence*, vol. 32, no. 3, 2010.

[76] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

[77] O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. Martinez Moller, S. Nekolla, and N. Navab, "Fast multiple organs detection and localization in whole-body MR Dixon sequences," in *Medical Image Computing and Computer Assisted Intervention*, Toronto, 2011.

[78] O. Pauly, D. Mateus, and N. Navab, "STARS: A new ensemble partitioning approach," in *International Conference on Computer Vision Workshop on Information Theory In Computer Vision and Pattern Recognition*, 2011.

[79] N. Payet and S. Todorovic, "$(rf)^2$ random forest random field," in *Neural Information Processing Systems*, 2010.

[80] R. L. Plackett, "A reduction formlula for normal multivariate integrals," *Biometrika*, vol. 41, 1954.

[81] J. R. Quinlan, *C4.5: Programs for Machine Learning*. 1993.

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

[83] G. Rogez, J. Rihan, S. Ramalingam, P. Orrite, and C. Torr, "Randomized trees for human pose detection," in *Computer Vision and Pattern Recognition*, 2008.

[84] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Institute of Electrical and Electronics Engineers Workshop on Applications of Computer Vision*, 2005.

[85] M. R. Sabuncu and K. V. Leemput, "The relevance voxel machine (RVoxM): A bayesian method for image-based prediction," in *Medical Image Computing and Computer Assisted Intervention*, 2011.

[86] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischoff, "On-line random forests," in *International Conference on Computer Vision workshop on On-Line learning for Computer Vision*, 2009.

[87] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.

[88] G. A. F. Seber and C. J. Wild, *Non Linear Regression*. New York: John Wiley and Sons, 1989.

[89] T. Sharp, "Implementing decision trees and forests on a GPU," in *European Conference on Computer Vision*, 2008.

[90] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *Proceedings of Computer Vision and Pattern Recognition (Computer Vision and Pattern Recognition)*, Washington, DC, USA, 1997.

[91] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from a single depth image," in *Institute of Electrical and Electronics Engineers Computer Vision and Pattern Recognition*, 2011.

[92] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *Institute of Electrical and Electronics Engineers Computer Vision and Pattern Recognition*, 2008.

[93] B. W. Silverman, *Density Estimation*. London: Chapman and Hall, 1986.

[94] J. Skilling, "Bayesian inference and maximum entropy methods in science and engineering," in *American Institute of Physics*, 2004.

[95] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," Technical Report, 1998.

[96] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *Journal of Machine Learning Research*, vol. 7, July 2006.

[97] A. Statnikov, L. Wang, and C. A. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC Bioinformatics*, 2008.

[98] G. Stewart and J. Sun, *Matrix Perturbation Theory*. Elsevier, 1990.

[99] G. J. Szekely and M. L. Rizzo, "Testing for equal distributions in high dimensions," *Interstat*, 2004.

[100] M. Szummer and T. Jaakkola, "Partially labelled classification with markov random walks," in *Neural Information Processing Systems*, 2001.

[101] J. B. Tenenbaum, V. deSilva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, 2000.

[102] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[103] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, 2007.

[104] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Proceedings of the Institute of Electrical and Electronics Engineers International Conference on Computer Vision (International Conference on Computer Vision)*, pp. 1589–1596, 2005.

[105] Z. Tu, "Auto-context and its application to high-level vision tasks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[106] V. Vapnik, *The nature of statistical learning theory.* Springer Verlag, 2000.

[107] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Conference Vision*, 2004.

[108] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *International Conference on Computer Vision*, 2003.

[109] J. Wang, "On transductive support vector machines," in *Prediction and Discovery*, American Mathematical Society, 2007.

[110] P. Yin, A. Criminisi, I. Essa, and J. Winn, "Tree-based classifiers for bilayer video segmentation," in *Computer Vision and Pattern Recognition*, 2007.

[111] Q. Zhang, R. Souvenir, and R. Pless, "On manifold structure of cardiac MRI data: Application to segmentation," in *Institute of Electrical and Electronics Engineers Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, 2006.

[112] X. Zhu and A. Goldberg, "Introduction to semi-supervised learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, 2009.

[113] A. Zien and C. S. Ong, "Multiclass multiple kernel learning," in *International Conference on Machine Learning*, 2007.