Bibliography

[1] F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954.

[2] P. Auer. On learning from multi-instance examples: Empirical evaluation of a theoretical approach. In *Proceedings of 14th International Conference on Machine Learning*, pages 21–29, 1997.

[3] M. Bar. Visual objects in context. *Nature Reviews Neuroscience*, 5(8):617–629, 2004.

[4] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan. Matching words and pictures. *JMLR*, 3:1107–1135, 2003.

[5] K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In *ICCV*, volume 2, pages 408–415, Vancouver, 2001.

[6] E. Bart, I. Porteous, P. Perona, and M. Welling. Unsupervised learning of visual taxonomies. In *CVPR*, pages 1–8, 2008.

[7] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 509–522, 2002.

[8] I. Biederman. On the semantics of a glance at a scene. *Perceptual organization*, pages 213–263, 1981.

[9] I. Biederman. Aspects and extension of a theory of human image understanding. *Computational processes in human vision: An interdisciplinary perspective*, pages Ablex Publishing Corporation, New Jersey, 1988.

[10] I. Biederman, RJ Mezzanotte, and JC Rabinowitz. Scene perception: detecting and judging objects undergoing relational violations. In *Cognitive Psychology*, volume 14, pages 143–77, 1982.

[11] C.M. Bishop. *Pattern recognition and machine learning*, volume 4. Springer New York, 2006.

[12] D. Blei and M. Jordan. Modeling annotated data. In *Proc. ACM SIGIR conf. on Research and development in information retrieval*, 2003.

[13] D.M. Blei and J. McAuliffe. Supervised topic models. *Advances in Neural Information Processing Systems*, 20:121–128, 2008.

[14] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[15] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. *Proc. CVPR*, 2008.

[16] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via plsa. In *9th European Conference on Computer Vision*, pages 517 – 30, Graz, Austria, 2006.

[17] A. Bosch, A. Zisserman, and X. Munoz. Scene classification using a hybrid generative/discriminative approach. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 30(4):712, 2008.

[18] S. Boughorbel, J.P. Tarel, and N. Boujemaa. Generalized histogram intersection kernel for image recognition. In *IEEE International Conference on Image Processing*, volume 3. IEEE, 2005.

[19] W.L. Buntine. Operations for learning with graphical models. *Arxiv preprint cs/9412102*, 1994.

[20] D. Cai, X. He, and J. Han. Efficient kernel discriminant analysis via spectral regression. *Proc. Int. Conf. on Data Mining*, pages 427–432, 2007.

[21] G. Carneiro, A. Chan, P. Moreno, and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. *IEEE PAMI*, 29(3):394–410, March, 2007.

[22] G. Carneiro and N. Vasconcelos. Formulating semantic image annotation as a supervised learning problem. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego*, 2005.

[23] AB Chan and N. Vasconcelos. Probabilistic kernels for the classification of auto-regressive visual processes. In *IEEE CVPR*, volume 1, 2005.

[24] H. Cheng, Z. Liu, and J. Yang. Sparsity induced similarity measure for label propagation. *ICCV*, 2009.

[25] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley, 1991.

[26] Ingemar J. Cox, Joumana Ghosn, Thomas V. Papathomas, and Peter N. Yianilos. Hidden annotation in content based image retrieval. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997.

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

[28] R. Datta, D. Joshi, J. Li, and J.Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 39:65, 2008.

[29] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[30] A. Dempster, N. Laird, and D. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *J. of the Royal Statistical Society*, B-39, 1977.

[31] L. Denoyer and P. Gallinari. Bayesian network model for semi-structured document classification. *Information Processing & Management*, 40(5):807–827, 2004.

[32] G. Doyle and C. Elkan. Accounting for word burstiness in topic models. In *Proceedings 26th International Conference on Machine Learning*, 2009.

[33] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, 2001.

[34] D. Dunn and W. Higgins. Optimal gabor filters for texture segmentation. *IEEE Trans. on Pattern. Analysis and Machine Intelligence*, 7(4), July 1995.

[35] P. Duygulu, K. Barnard, N. Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *European Conference on Computer Vision*, Copenhagen, Denmark, 2002.

[36] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In *Advances in Neural Information Processing Systems*, 2007.

[37] H.J. Escalante, C.A. Hérnadez, L.E. Sucar, and M. Montes. Late fusion of heterogeneous methods for multimedia image retrieval. In *Proceedings 1st ACM International Conference on Multimedia Information Retrieval*, pages 172–179. ACM, 2008.

[38] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[39] C. Fellbaum. *Wordnet: an electronic lexical database.* MIT Press, 1998.

[40] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 2009.

[41] S. Feng, R. Manmatha, and V. Lavrenko. Multiple bernoulli relevance models for image and video annotation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington DC, 2004.

[42] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, June 2003.

[43] M. Fink and P. Perona. Mutual boosting for contextual inference. *Neural Information Processing Systems*, 2004.

[44] J.W. Fisher, T. Darrell, W.T. Freeman, and P. Viola. Learning joint statistical models for audio-visual fusion and segregation. *Advances in Neural Information Processing Systems*, pages 772–778, 2001.

[45] D. Forsyth and M. Fleck. Body plans. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico*, pages 678–683, 1997.

[46] D. Gabor. Theory of communication. part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429–441, 1946.

[47] C. Galleguillos, A. Rabinovich, and S. Belongie. Object categorization using co-ocurrence, location and appearance. *IEEE Conference in Computer Vision and Pattern Recognition (CVPR) 2008, Anchorage, USA.*, 2008.

[48] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis.* Chapman Hall, 1995.

[49] T. Gevers and A. Smeulders. Picktoseek: Combining color and shape invariant features for image retrieval. *IEEE Trans. on Image Processing*, 9(1):102–119, January 2000.

[50] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice.* Chapman & Hall/CRC, 1996.

[51] Kristen Grauman and Trevor Darrell. The pyramid match kernel: Efficient learning with sets of features. *J. Mach. Learn. Res.*, 8:725–760, 2007.

[52] G. Griffin, A. Holub, and P. Perona. The caltech-256. Technical report, Caltech, 2006.

[53] N. Haering, Z. Myles, and N. Lobo. Locating dedicuous trees. In *Workshop in Content-based Access to Image and Video Libraries*, pages 18–25, 1997, San Juan, Puerto Rico.

[54] Junwei Han and Lei Guo. A new image retrieval system supporting query by semantics and example. In *ICIP (3)*, 2002.

[55] D.R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[56] X. He, O. King, W.Y. Ma, M. Li, and H.J. Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(1):39–48, 2003.

[57] G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. *10th European Conference on Computer Vision, Marseille, France*, page 30, 2008.

[58] T. Hofmann. Probabilistic latent semantic indexing. *ACM SIGIR*, pages 50–57, 1999.

[59] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[60] J. Iria, F. Ciravegna, and J. Magalhães. Web news categorization using a cross-media document graph. In *Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8. ACM, 2009.

[61] A. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition Journal*, 29, August 1996.

[62] N. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, 1984.

[63] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 119–126, 2003.

[64] IT Jolliffe. *Principal Component Analysis*. Springer, 2002.

[65] AJ Joshi, F. Porikli, and N. Papanikolopoulos. Multi-class active learning for image classification. In *CVPR*. IEEE, 2009.

[66] I. Khan, A. Saffari, and H. Bischof. Tvgraz: Multi-modal learning of object categories by combining textual and visual features. In *Proceedings 33rd Workshop of the Austrian Association for Pattern Recognition*, 2009.

[67] J.J. Kivinen, EB Sudderth, and MI Jordan. Learning multiscale representations of natural scenes using dirichlet processes. In *ICCV*. Citeseer, 2007.

[68] Yasushi Kiyoki, Takashi Kitagawa, and Takanari Hayama. A metadatabase system for semantic image search by a mathematical model of meaning. *SIGMOD Rec.*, 23(4):34–41, 1994.

[69] T. Kliegr, K. Chandramouli, J. Nemrava, V. Svatek, and E. Izquierdo. Combining image captions and visual analysis for image concept classification. In *Proceedings 9th International Workshop on Multimedia Data Mining at ACM SIG Knowledge Discovery and Data Mining*, pages 8–17. ACM New York, NY, USA, 2008.

[70] H. Kueck, P. Carbonetto, and N. Freitas. A constrained semi-supervised learning approach to data association. In *European Conference on Computer Vision*, Prague, Czech Republic, 2004.

[71] S. Lacoste-Julien, F. Sha, and M.I. Jordan. Disclda: Discriminative learning for dimensionality reduction and classification. *Advances in Neural Information Processing Systems 21 (NIPS08)*, 2008.

[72] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *NIPS, Vancouver*, 2003.

[73] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *Neural Information Processing Systems, Vancouver, Canada*, 2003.

[74] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proc. IEEE Conf. Comp. Vision Patt. Recog*, 2005.

[75] C. S. Lee, W.-Y. Ma, and H. Zhang. Information embedding based on user's relevance feedback for image retrieval. In *Proc. SPIE Vol. 3846, p. 294-304, Multimedia Storage and Archiving Systems IV*, pages 294–304, 1999.

[76] D. Li, N. Dimitrova, M. Li, and I.K. Sethi. Multimedia content processing through cross-modal association. In *Proceedings 11th ACM International Conference on Multimedia*, pages 604–611. ACM, 2003.

[77] Fei-Fei Li and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *IEEE CVPR*, pages 524–531, 2005.

[78] F.F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci US A*, 99(14):9596–601, 2002.

[79] L.J. Li and L. Fei-Fei. What, where and who? classifying event by scene and object recognition. *Proc. ICCV*, 2007.

[80] J.J. Lim, P. Arbeláez, C. Gu, and J. Malik. Context by region ancestry. In *ICCV*. Citeseer, 2010.

[81] Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, 28(1):84–95, 1980.

[82] J. Liu, Y. Yang, and M. Shah. Learning semantic visual vocabularies using diffusion distance. In *CVPR*. IEEE, 2009.

[83] Jingen Liu and Mubarak Shah. Scene modeling using co-clustering. *International Conference on Computer Vision*, 2007.

[84] B. Logan and A. Salomon. A music similarity function based on signal analysis. In *IEEE International Conference on Multimedia and Expo*, 2001.

[85] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[86] Jing Lu, Shao ping Ma, and Min Zhang. Automatic image annotation based-on model space. In *IEEE NLP-KE*, 2005.

[87] Ye Lu, HongJiang Zhang, Liu Wenyin, and Chunhui Hu. Joint semantics and feature based image retrieval using relevance feedback. *IEEE Transactions on Multimedia*, 5(3):339–347, 2003.

[88] D.J.C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge Univ Pr, 2003.

[89] M.I. Mandel and D.P.W. Ellis. Multiple-instance learning for music information retrieval. In *Proceedings of International Symposium of Music Information Retrieval*, 2008.

[90] R. Manmatha and S. Ravela. A syntatic characterization of appearence and its application to image retrieval. In *SPIE Conference on Human Vision and Electronic Imaging II*, volume 3016, 1997, San Jose, California.

[91] C.D. Manning, P. Raghavan, and H. Schütze. *An Introduction to Information Retrieval*. Cambridge University Press, 2008.

[92] O. Maron and T. Lozano-Perez. A framework for multiple instance learning. In *Neural Information Processing Systems, Denver, Colorado*, 1998.

[93] C.T. Meadow, B.R. Boyce, D.H. Kraft, and C.L. Barry. *Text Information Retrieval Systems.* Emerald Group Pub Ltd, 2007.

[94] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, pages 1615–1630, 2005.

[95] T.P. Minka. Estimating a dirichlet distribution. *http://research.microsoft.com/ minka/papers/dirichlet/*, 1:3, 2000.

[96] F. Monay and D. Gatica-Perez. Modeling semantic aspects for cross-media image indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1802–1817, 2007.

[97] Y. Mori, H. Takahashi, and R. Oka. Automatic word assignment to images based on image division and vector quantization. In *Proceedings of Recherche d'Information Assistée par Ordinateur (RIAO)*. Citeseer, 2000.

[98] Henning Muller, Stephane Marchand-Maillet, and Thierry Pun. The truth about corel - evaluation in image retrieval. In *CIVR '02: Proceedings of the International Conference on Image and Video Retrieval*, pages 38–49, 2002.

[99] S. Nakamura. Statistical multimodal integration for audio-visual speech processing. *IEEE Transactions on Neural Networks*, 13(4):854–866, 2002.

[100] D. Navon. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3):353–383, 1977.

[101] W. Niblack and et al. The qbic project: Querying images by content using color, texture, and shape. In *Storage and Retrieval for Image and Video Databases*, pages 173–181, SPIE, Feb. 1993, San Jose, California.

[102] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. *Proc. ECCV*, 4:490–503, 2006.

[103] A. Oliva and P.G. Schyns. Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41(2):176–210, 2000.

[104] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[105] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Visual Perception*, 2006.

[106] M. Paramita, M. Sanderson, and P. Clough. Diversity in photo retrieval: Overview of the ImageCLEF 2009 photo task. *CLEF working notes*, 2009.

[107] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *Int. Journal of Computer Vision*, Vol. 18(3):233–254, June 1996.

[108] T.T. Pham, N.E. Maillot, J.H. Lim, and J.P. Chevallet. Latent semantic fusion model for image retrieval and annotation. In *Proceedings 16th ACM International Conference on Information and Knowledge Management*, pages 439–444. ACM, 2007.

[109] R. Picard. Digital libraries: Meeting place for high-level and low-level vision. In *Proc. Asian Conf. on Computer Vision*, December 1995, Singapore, USA.

[110] J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 6174, 1999.

[111] M. Porat and Y. Zeevi. Localized texture processing in vision: Analysis and synthesis in the gaborian space. *IEEE Trans. on Biomedical Engineering*, 36(1):115–129, January 1989.

[112] D. Putthividhya, HT Attias, and SS Nagarajan. Supervised topic model for automatic image annotation. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010.

[113] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool. Modeling scenes with local descriptors and latent aspects. *Proceedings. Tenth IEEE International Conference on Computer Vision*, Vol. 1:883 – 90, 2005.

[114] P. Quelhas, F. Monay, J.M. Odobez, D. Gatica-Perez, and T. Tuytelaars. A thousand words in a scene. *IEEE transactions on pattern analysis and machine intelligence*, 29(9):1575–1589, 2007.

[115] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. *Computer Vision, IEEE 11th International Conference on*, pages 1–8, 2007.

[116] D. Ramage, D. Hall, R. Nallapati, and C.D. Manning. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 248–256. Association for Computational Linguistics, 2009.

[117] T. Randen and J.H. Husoy. Filtering for texture classification: A comparative study. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(4):291–310, 1999.

[118] K. Rao and P. Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications.* Academic Press, 1990.

[119] N. Rasiwasia, PL Moreno, and N. Vasconcelos. Bridging the gap: Query by semantic example. *Multimedia, IEEE Transactions on*, 9(5):923–938, 2007.

[120] N. Rasiwasia and N. Vasconcelos. Scene classification with low-dimensional semantic spaces and weak supervision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[121] L.W. Renninger and J. Malik. When is scene identification just texture recognition? *Vision Research*, 44(19):2301–2311, 2004.

[122] HA Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 203–208, 1996.

[123] Y. Rui and T. Huang. Optimizing learning in image retrieval. *IEEE CVPR*, 2000.

[124] G. Salton. *The SMART Retrieval System - Experiments in Automatic Document Processing.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1971.

[125] G. Salton and J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill, New York, 1983.

[126] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, 1975.

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

[128] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling appearance, shape and context. *International Journal of Computer Vision*, pages 1–22, 2007.

[129] J. Sivic, B.C. Russell, A. Efros, A. Zisserman, and W.T. Freeman. Discovering object categories in image collections. *Proc. ICCV*, 1:65, 2005.

[130] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477, 2003.

[131] M. Slaney. Semantic-audio retrieval. In *IEEE International Conference on Acoustics Speech and Signal Processing*, volume 4, pages 4108–4111. IEEE, 2002.

[132] Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and trecvid. In *Proceedings 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[133] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval: the end of the early years. *IEEE Trans. on Pattern. Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[134] J. Smith and S. Chang. Visualseek: a fully automated content-based image query system. In *ACM Multimedia, Boston, Massachussetts*, pages 87–98, 1996.

[135] J. R. Smith, C.-Y. Lin, M. R. Naphade, A. Natsev, and B. L. Tseng. Validity-weighted model vector-based retrieval of video. In *Proceedings of the SPIE, Volume 5307, pp. 271-279 (2003).*, pages 271–279, 2003.

[136] JR Smith. Image retrieval evaluation. *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998.

[137] J.R. Smith, M. Naphade, and A. Natsev. Multimedia semantic indexing using model vectors. *ICME*, pages 445–448, 2003.

[138] C.G.M. Snoek and M. Worring. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1):5–35, 2005.

[139] M. Steyvers and T. Griffiths. Probabilistic topic models. *Handbook of latent semantic analysis*, 427, 2007.

[140] EB Sudderth, A. Torralba, WT Freeman, and AS Willsky. Learning hierarchical models of scenes, objects, and parts. *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005*, 2, 2005.

[141] M. Swain and D. Ballard. Color indexing. *Int. Journal of Computer Vision*, Vol. 7(1):11–32, 1991.

[142] Martin Szummer and Rosalind Picard. Indoor-outdoor image classification. In *Workshop in Content-based Access to Image and Video Databases*, 1998, Bombay, India.

[143] S. M. M. Tahaghoghi, James A. Thom, and Hugh E. Williams. Are two pictures better than one? In *ADC '01: Proceedings of the 12th Australasian database conference*, pages 138–144, Washington, DC, USA, 2001. IEEE Computer Society.

[144] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[145] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, pages 169–191, 2003.

[146] A. Torralba, K.P. Murphy, and W.T. Freeman. Contextual models for object detection using boosted random fields. *Advances in Neural Information Processing Systems*, 2004.

[147] L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient object category recognition using classemes. *ECCV*, pages 776–789, 2010.

[148] T. Tsikrika and J. Kludas. Overview of the wikipedia multimedia task at ImageCLEF 2009. In *Working Notes for the CLEF Workshop*, 2009.

[149] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2):467–476, February 2008.

[150] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, 2008.

[151] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[152] A. Vailaya, A. Jain, and H. Zhang. On image classification: City vs. landscape. *Pattern Recognition*, 31:1921–1936, December 1998.

[153] K.E.A. Van De Sande, T. Gevers, and C.G.M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1582–1596, 2009.

[154] J.C. van Gemert, J.M. Geusebroek, C.J. Veenman, and A.W.M. Smeulders. Kernel codebooks for scene categorization. *Proc. ECCV*, pages 696–709, 2008.

[155] M. Vasconcelos, N. Vasconcelos, and G. Carneiro. Weakly supervised top-down image segmentation. *CVPR*, pages 1001–1006, 2006.

[156] N. Vasconcelos. Minimum probability of error image retrieval. *IEEE Trans. on Signal Processing*, August 2004.

[157] N. Vasconcelos. Minimum probability of error image retrieval. *IEEE Trans. on Signal Processing*, 52(8), August 2004.

[158] N. Vasconcelos. A unified view of image similarity. In *Proc. Int. Conf. Pattern Recognition*, Barcelona, Spain, 2000.

[159] N. Vasconcelos. Image indexing with mixture hierarchies. In *Proc. IEEE CVPR*, Kawai, Hawaii, 2001.

[160] N. Vasconcelos and M. Kunt. Content-based retrieval from image databases: Current solutions and future directions. In *Proc. Int. Conf. Image Processing*, Thessaloniki, Greece, 2001.

[161] N. Vasconcelos and A. Lippman. Learning over multiple temporal scales in image databases. In *Proc. European Conference on Computer Vision, Dublin, Ireland*, 2000.

[162] Nuno Vasconcelos. *Bayesian models for visual information retrieval*. PhD thesis, Massachusetts Institute of Technology, 2000.

[163] A. Vinokourov, D.R. Hardoon, and J. Shawe-Taylor. Learning the semantics of multimedia content with application to web image retrieval and classification. In *4th International Symposium on Independent Component Analysis and Blind Source Separation*, 2003.

[164] A. Vinokourov, J. Shawe-Taylor, and N. Cristianini. Inferring a semantic representation of text via cross-language correlation analysis. *Advances in Neural Information Processing Systems*, pages 1497–1504, 2003.

[165] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 1(2), 2002.

[166] J. Vogel and B. Schiele. A semantic typicality measure for natural scene categorization. *DAGM.04 Annual Pattern Recognition Symposium*, 2004.

[167] C. Wang, D. Blei, and F.F. Li. Simultaneous image classification and annotation. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009.

[168] G. Wang, D. Hoiem, and D. Forsyth. Building text features for object image classification. In *Proceedings of 19th International Conference on Pattern Recognition*, 2009.

[169] G. Wang, D. Hoiem, and D. Forsyth. Learning image similarity from flickr groups using stochastic intersection kernel machines. In *IEEE International Conference on Computer Vision*, pages 428–435, 2009.

[170] Y. Wang, P. Sabzmeydani, and G. Mori. Semi-latent dirichlet allocation: A hierarchical model for human action recognition. *Human Motion–Understanding, Modeling, Capture and Animation*, pages 240–254, 2007.

[171] T. Westerveld. Image retrieval: Content versus context. *Content-Based Multimedia Information Access*, pages 276–284, 2000.

[172] T. Westerveld and A P. de Vries. Experimental evaluation of a generative probabilistic image retrieval model on 'easy' data. In *In Proceedings of the Multimedia Information Retrieval Workshop*, Toronto, Canada, August 2003.

[173] T. Westerveld and A.P. de Vries. Experimental evaluation of a generative probabilistic image retrieval model on.easy.data. In *Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval, Multimedia Information Retrieval Workshop*. Citeseer, 2003.

[174] L. Wolf and S. Bileschi. A critical view of context. *International Journal of Computer Vision*, 69(2):251–261, 2006.

[175] Y. Yang, D. Xu, F. Nie, J. Luo, and Y. Zhuang. Ranking with local regression and global alignment for cross media retrieval. In *Proceedings 17th ACM International Conference on Multimedia*, pages 175–184. ACM, 2009.

[176] Y. Yang, Y.T. Zhuang, F. Wu, and Y.H. Pan. Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval. *IEEE Transactions on Multimedia*, 10(3):437–446, 2008.

[177] H. Zhang, A. Berg, M. Maire, and J. Malik. Svm-knn: Discriminative nearest neighbor classification for visual category recognition. *Proc. CVPR*, 2:2126–2136, 2006.

[178] H. Zhang, Y. Zhuang, and F. Wu. Cross-modal correlation learning for clustering on image-audio dataset. In *Proceedings 15th ACM International Conference on Multimedia*, page 276. ACM, 2007.

[179] X. Zhou, N. Cui, Z. Li, F. Liang, and T.S. Huang. Hierarchical gaussianization for image classification. In *IEEE 12th International Conference on Computer Vision*, pages 1971–1977. IEEE, 2009.

[180] J. Zhu, A. Ahmed, and E.P. Xing. Medlda: maximum margin supervised topic models for regression and classification. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1257–1264. ACM, 2009.

[181] Y. Zhuang, Y. Yang, F. Wu, and Y. Pan. Manifold learning based cross-media retrieval: a solution to media object complementary nature. *Journal of VLSI Signal Processing*, 46(2):153–164, 2007.

[182] Y.T. Zhuang, Y. Yang, and F. Wu. Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval. *IEEE Transactions on Multimedia*, 10(2):221–229, 2008.