Acknowledgements

All contributors to this book would like to acknowledge the financial support by the Network of Excellence MUSCLE within the 6th Framework Program for Research of the European Commision, under contract no. FP6-507752. In addition, individual chapter acknowledgements are as follows:

Chapter 4 The authors would like to thank A. Potamianos for providing the initial experimental setup for AV-ASR, I. Kokkinos for visual front-end discussions and face detection code, K. Murphy for using his HMM toolkit, and J.N. Gowdy for the use of the CUAVE database. They are also grateful to G. Gravier for his extensive feedback on the work.

Chapter 5 The authors' work has also been supported by various agencies since 2001, including the European Union Marie Curie Programme (MOUMIR HP-99-108), Enterprise Ireland (projects MUSE-DTV, CASMS, DumpingDetective, Detection of Illicit content), Science Foundation Ireland Research Frontiers and Eureka Programmes, Adobe Systems Inc. and The Irish Research Council for Science, Engineering & Technology.

Chapter 6 The work was supported in part by the Scientific and Technical Research Council of Turkey (TUBITAK) under Grant no. EEEAG-105E065 and Ministry of Industry and Trade of Turkey under Grant no. SANTEZ-105E121.

Chapter 7 The research was co-funded by the European Union and the Hellenic Ministry of Education in the framework of program Pythagoras II of the Operational Program for Education and Initial Vocational Training within the 3rd Community Support Program.

Chapter 8 The authors would like to thank C. Kotropoulos and his group from Aristotle University of Thessaloniki for providing the annotated movie video database, A. Zlatintsi at the National Technical University of Athens (NTUA) for the annotation of the movie clips and the design and coordination of the evaluation of the movie summarizations, and all the members of the CVSP Lab at NTUA for helping as evaluators of the movie summaries.

Chapter 14 The authors acknowledge the support by EPSRC, British Telecommunications Plc, SIRA, and the Imaging Faraday Partnership for their support in this work.

References

1. "Apple Human Interface Guidelines," http://developer.apple.com/documentation/UserExperience/Conceptual/OSXHIGuidelines.
2. "Apple iPhone Human Interface Guidelines," http://developer.apple.com/documentation/iPhone/Conceptual/iPhoneHIG/iPhoneHIG.pdf.
3. "HCI term definition," http://en.wikipedia.org/wiki/Human-Computer_Interaction.
4. "IBM WebSphere Voice Server," http://www-306.ibm.com/software/pervasive/voice_server.
5. "J. Blat, A first view of multimedia standards," http://www.iua.upf.es/~jblat/material/doctorat/multimedia_standards.html.
6. "SALT forum," http://www.saltforum.org/.
7. "The NICE (Natural Interactive Communication for Edutainment) project," http://www.niceproject.com/.
8. "W3C Extensible MultiModal Annotation markup language (EMMA)," http://www.w3.org/TR/emma/.
9. "W3C Ink Markup Language," http://www.w3.org/TR/InkML/.
10. "W3C Mobile Web Best Practices," http://www.w3.org/TR/mobile-bp/.
11. "W3C Multimodal Architecture and Interfaces," http://www.w3.org/TR/mmi-arch/.
12. "W3C MultiModal Interaction Requirements," http://www.w3.org/TR/mmi-reqs/.
13. "W3C MultiModal Interaction Use Cases," http://www.w3.org/TR/mmi-use-cases/.
14. "W3C MultiModal Interaction Working Group: Multimodal Interaction Framework," http://www.w3.org/TR/mmi-framework/.
15. "W3C MultiModal Interaction Working Group: System and Environment Framework," http://www.w3.org/TR/sysenv/.
16. "Mpeg-7 overview (version 9)," ISO/IEC JTC1/SC29/WG11 N5525, Tech. Rep., March 2003.
17. B. Adams et al., "IBM research TREC-2002 video retrieval system," in *Proc. Text Retrieval Conference*, 2002.
18. H. Ailisto, I. Korhonen, T. Tuomisto, S. Siltanen, E. Strömmer, L. Pohjanheimo, J. Hyväkkä, P. Välkkynen, A. Ylisaukko-oja, and H. Keränen,

"Communications technologies. VTT's research programme 2002-2006. physical browsing for ambient intelligence (pb-ami). VTT publications 629," http://www.vtt.fi/inf/pdf/publications/2007/P629.pdf, pp. 284 – 308, 2007.
19. A. A. Alatan, "Automatic multi-modal dialogue scene indexing," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 2001, pp. 374–377.
20. A. A. Alatan and A. N. Akansu, "Multi-modal dialog scene detection using hidden-markov models for content-based multimedia indexing," *Multimedia Tools and Applications*, vol. 14, pp. 137–151, 2001.
21. A. A. Alatan, A. N. Akansu, and W. Wolf, "Comparative analysis of hidden markov models for multi-modal dialogue scene indexing," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, vol. IV, 2000, pp. 2401–2404.
22. P. Aleksic and A. Katsaggelos, "Audio-visual biometrics," *Proc. IEEE*, vol. 11, pp. 2025–2044, 2006.
23. A. Allauzen and J.-L. Gauvain, "Open vocabulary ASR for audiovisual document indexation," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2005.
24. B. Alp, P. Haavisto, T. Jarske, K. Oistamo, and Y. Neuvo, "Median based algorithms for image sequence processing," in *SPIE Conf. on Visual Communications and Image Processing*, 1990, pp. 122–134.
25. E. Ammicht, E. Fosler-Lussier, and A. Potamianos, "Information seeking spoken dialogue systems - Part I: Semantics and pragmatics," *IEEE Trans. Multimedia*, vol. 9, no. 3, pp. 532–549, 2007.
26. L. Amsaleg and P. Gros, "Content-based retrieval using local descriptors: Problems and issues from a database perspective," *Pattern Analysis & Applications*, vol. 2001, no. 4, pp. 108–124, 2001.
27. E. André and T. Rist, "Presenting through performing: on the use of multiple lifelike characters in knowledge-based presentation systems," in *Proc. Int'l Conf. on Intelligent User Interfaces*, New Orleans, Louisiana, 2000, pp. 1–8.
28. D. Androutsos, L. Guan, and A. V. G. Editors), "Special issue on semantic retrieval of multimedia," *IEEE Signal Process. Mag.*, vol. 23, no. 2, March 2006.
29. H. Aoki, "High–speed dialog detection for automatic segmentation of recorded tv program," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, 2005, pp. 49–58.
30. H. Aoki, S. Shimotsuji, and O. Hori, "A shot classification method of selecting effective key-frames for video browsing," in *Proc. ACM Int'l Conference on Multimedia*, 1996, pp. 1–10.
31. A. Arampatzis, T. Van Der Weide, C. Koster, and P. Van Bommel, *Linguistically Motivated Information Retrieval*. New York, New York, Etats-Unis: M. Dekker, 2000, vol. 69, pp. 201–222.
32. A. Arasu, J. Cho, H. Garcia-Molina, A. Paepke, and S. Raghavan, "Searching the web," *ACM Transactions on Internet Technology*, vol. 1, no. 1, pp. 2–43, Aug. 2001.
33. W. Arentz and B. Olstad, "Classifying offensive sites based on image content," *Computer Vision and Image Understanding*, vol. 94, pp. 295–310, 2004.
34. S. Aukstakalnis and D. Blatner, *Silicon Mirage; The Art and Science of Virtual Reality*. Berkeley, CA, USA: Peachpit Press, 1992.
35. Y. Avrithis, A. Doulamis, N. Doulamis, and S. Kollias, "Summarization of videotaped presentations: automatic analysis of motion and gesture," *Computer Vision and Image Understanding*, vol. 75, no. 12, pp. 3–24, 1998.

36. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, 2001.
37. J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. C. Jain, and C. F. Shu, "Virage image search engine: an open framework for image management," in *Proc. of SPIE*, vol. 2670. SPIE, 1996, pp. 76–87.
38. T. M. Bae, S. H. Jin, and Y. M. Ro, "Video segmentation using hidden Markov model with multimodal features," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, 2004, pp. 401–409.
39. R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg, *Readings in Human-Computer Interaction: Toward the year 2000.* Morgan Kaufmann Publishers, 1995.
40. R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan, "The smart phone: a ubiquitous input device," *IEEE Pervasive Comput.*, vol. Volume 5, no. 1, pp. 70– 77, Jan-March 2006.
41. R. Ballagas, M. Rohs, and J. G. Sheridan, "Sweep and point and shoot: phonecam-based interactions for large public displays," in *Proc. ACM-SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press, 2005, pp. 1200–1203.
42. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int'l J. of Comp. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.
43. Z. Barzelay and Y. Y. Schechner, "Harmony in motion," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2007.
44. P. W. Battaglia, R. A. Jacobs, and R. N. Aslin, "Bayesian integration of visual and auditory signals for spatial localization," *J. of the Opt. Soc. Am. (A)*, vol. 20, no. 7, pp. 1391–1397, July 2003.
45. S. Bayer, "Building a standards and research community with the Galaxy Communicator software infrastructure," in *Practical Spoken Dialog Systems*, D. A. Dahl, Ed. Kluwer Academic Publishers, 2004, pp. 167–196.
46. F. Beaver, *Dictionary of Film Terms.* Twayne Publishing, New York, 1994.
47. F. Béchet, A. L. Gorin, J. H. Wright, and D. Hakkani-Tür, "Detecting and extracting named entities from spontaneous speech in a mixed initiative spoken dialogue context: How may I help you?" *Speech Communication*, vol. 42, no. 2, pp. 207–225, 2004.
48. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "The R*-tree: An efficient and robust access method for points and rectangles," in *ACM SIGMOD International Conference on Management of Data*, Atlantic City, New Jersey, Etats-Unis, 23-25 May 1990, pp. 322–331.
49. J. R. Bellegarda, "Large vocabulary speech recognition with multispan statistical language models," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 1, pp. 76–84, 2000.
50. ——, "Statistical language model adaptation: review and perspectives," *Speech Communication*, vol. 42, no. 1, pp. 93–108, 2004.
51. T. Bellman and I. S. MacKenzie, "A probabilistic character layout strategy for mobile text entry," in *Proc. Graphics Interface*, 1998, pp. 168–176.
52. S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

53. S. Bengio, "An asynchronous hidden Markov model for audio-visual speech recognition," in *Proc. Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds. MIT Press, 2003, pp. 1237–1244.
54. C. Benoit, J. C. Martin, C. Pelachaud, L. Schomaker, and B. Suhm, "Audio-visual and multimodal speech systems," in *Handbook of Standards and Resources for Spoken Language Systems*, D. Gibbon, I. Mertins, and R. Moore, Eds. Kluwer Academic Publishers, 2000.
55. J. L. Bentley, "Multidimensional binary search trees in database applications," *IEEE Trans. Softw. Eng.*, vol. 5, no. 4, pp. 333–340, 1979.
56. S. Berchtold, D. A. Keim, and H.-P. Kriegel, "The X-tree : An index structure for high-dimensional data," in *22th International Conference on Very Large Data Bases*, Mumbai (Bombay), Inde, 3-6 Sep. 1996, pp. 28–39.
57. C. Berg, J. P. R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups.* Springer-Verlag, 1984.
58. N. O. Bernsen and L. Dybkjaer, "Is speech the right thing for your application?" in *Proc. Int'l Conf. on Spoken Language Processing*, Sydney, Australia, 1998, pp. 3209–3212.
59. P. Bertelson and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," *Percept. Psychophysics*, vol. 29, pp. 578–584, 1981.
60. M. Betser and G. Gravier, "Multiple events tracking in sound tracks," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, 2004.
61. K. Bharat and M. R. Henzinger, "Improved algorithms for topic distillation in a hyperlinked environment," in *Proc. of the ACM SIGIR Conference*, Melbourne, 1998, pp. 104–111.
62. B. Bigi, Y. Huang, and R. De Mori, "Vocabulary and language model adaptation using information retrieval," in *Proc. Int'l Conf. on Spoken Language Processing*, 2004.
63. V. Bilici, E. Krahmer, S. te Riele, and R. Veldhuis, "Preferred modalities in dialogue systems," in *Proc. Int'l Conf. on Spoken Language Processing*, Beijing, China, 2000.
64. O. Bimber and R. Raskar, "Modern approaches to augmented reality," in *Proc. ACM Int'l conference on Computer Graphics and Interactive Techniques*, 2005.
65. C. Bishop, "Training with noise is equivalent to Tikhonov regularization," *Neural Computation*, vol. 7, no. 1, pp. 108–116, 1995.
66. C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.
67. H. M. Blanken, A. P. de Vries, H. E. Blok, and L. F. (Editors), *Multimedia Retrieval (Data-Centric Systems and Applications).* Springer, 2007.
68. D. Bohus and A. I. Rudnicky, "Error handling in the RavenClaw dialog management framework," in *Human Language Technology Conference*, Vancouver, Canada, 2005, pp. 225–232.
69. R. A. Bolt, "Put-That-There : Voice and gesture at the graphics interface," in *Proc. ACM Int'l conference on Computer Graphics and Interactive Techniques.* ACM Press New York, NY, USA, 1980, pp. 262–270.
70. D. Bordwell and K. Thompson, *Film Art: An Introduction.* McGraw–Hill, Inc., 4th ed., New York, 1993.
71. J. Boreczky and L. Wilcox, "A hidden Markov model framework for video segmentation using audio and image features," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1998, pp. 3741–3744.

72. A. Bosson, G. Cawley, Y. Chan, and R. Harvey, "Nonretrieval: blocking pornographic images," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, 2002, pp. 50–60.
73. M. M. Bouamrane and S. Luz, "Meeting browsing: State-of-the-art review," *Multimedia Systems*, vol. 12, pp. 439–457, 2007.
74. N. Boujemaa, S. Boughorbel, and C. Vertan, "Soft color signatures for image retrieval by content," in *Eusflat'2001*, vol. 2, 2001, pp. 394–401.
75. N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. L. Saux, and H. Sahbi, "IKONA: Interactive generic and specific image retrieval," in *Proc. Int'l Workshop on Multimedia Content-Based Indexing and Retrieval*, 2001.
76. N. Boujemaa, J. Fauqueur, and V. Gouet, "What's beyond query by example?" in *Trends and Advances in Content-Based Image and Video Retrieval*, R. V. L. Shapiro, H.P. Kriegel, Ed. Springer Verlag, 2004.
77. H. Bourlard and S. Dupont, "A new ASR approach based on independent processing and recombination of partial frequency bands," in *Proc. Int'l Conf. on Spoken Language Processing*, 1996, pp. 426–429.
78. A. Bovik, P. Maragos, and T. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3245–3265, Dec 1993.
79. A. Briassouli and N. Ahuja, "Extraction and analysis of multiple periodic motions in video sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1244–1261, Jul. 2007.
80. P. F. Brown, V. J. D. Pietra, P. V. de Souza, J. C. Lai, and R. L. Mercer, "Class-based n-gram models of natural language," *Computational Linguistics*, vol. 18, no. 4, pp. 467–479, 1992.
81. A. Budanitsky and G. Hirst, "Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures," in *Proceedings of the Workshop on WordNet and Other Lexical Resources NAACL 2001*, 2001.
82. M. D. Byrne, "Cognitive architectures," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 2003, pp. 97–117.
83. S. K. Card, J. D. Mackinlay, and G. G. Robertson, "A morphological analysis of the design space of input devices," *ACM Transactions on Information Systems*, vol. 9, no. 2, pp. 99–122, 1991.
84. S. K. Card, A. Newell, and T. P. Moran, *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1983.
85. B. Carpenter, *The Logic of Typed Feature Structures*. Cambridge University Press New York, NY, USA, 1992.
86. J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan, "Embodiment in conversational interfaces: Rea," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*. ACM Press New York, NY, USA, 1999, pp. 520–527.
87. L. Chaisorn, T.-S. Chua, and C.-H. Lee, "A multi-modal approach to story segmentation for news video," *World Wide Web*, vol. 6, pp. 187–208, 2003.
88. Y. Chan, R. Harvey, and J. A. Bangham, "Using colour features to block dubious images," in *Proc. European Signal Processing Conference*, 2000.
89. E. Y. Chang, B. Li, G. Wu, and K. Goh, "Statistical learning for effective visual image retrieval," in *Proc. IEEE Int'l Conf. on Image Processing*, September 2003, pp. 609–612.

90. P. Chang, M. Han, and Y. Gong, "Extract highlights from baseball game video with Hidden Markov Models," in *Proc. IEEE Int'l Conf. on Image Processing*, September 2002, pp. 609–612.
91. C. Chelba and F. Jelinek, "Structured language modeling," *Computer Speech and Language*, vol. 14, no. 4, pp. 283–332, 2000.
92. G. Chen and D. Kotz, "A survey of context-aware mobile computing research," Dept. of Computer Science, Dartmouth College, Tech. Rep. TR2000-381, November 2000.
93. L. Chen, J.-L. Gauvain, L. Lamel, and G. Adda, "Dynamic language modeling for broadcast news," in *Proc. Int'l Conf. on Spoken Language Processing*, 2004.
94. L. Chen and M. T. Özsu, "Rule-based scene extraction from video," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 2, 2002, pp. 737–740.
95. L. Chen, S. Rizvi, and M. T. Özsu, "Incorporating audio cues into dialog and action scene extraction," in *Proc. IS&T/SPIE's 15th Annual Symp. Electronic Imaging - Storage and Retrieval for Media Databases*, 2003, pp. 252–264.
96. Z. Chen, L. Wenyin, F. Zhang, M. Li, and H.-J. Zhang, "Web mining for web image retrieval," *Journal of the American Society of Information Science*, vol. 52, no. 10, pp. 831–839, 2001.
97. S. F. Cheng, W. Chen, and H. Sundaram, "Semantic visual templates: linking visual features to semantics," in *Proc. IEEE Int'l Conf. on Image Processing*, Chicago, Illinois, 1998, pp. 531–535.
98. W. Chou and B. H. Juang, *Pattern Recognition in Speech and Language Processing*. CRC Press, Boca Raton, FL, USA, 2002.
99. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing*. Kluwer Academic Publ., 1990.
100. V. Claveau, P. Sbillot, C. Fabre, and P. Bouillon, "Learning semantic lexicons from a part-of-speech and semantically tagged corpus using inductive logic programming," *Journal of Machine Learning Research*, vol. 4, pp. 493–525, Aug. 2003.
101. P. R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow, "Quickset: multimodal interaction for distributed applications," in *Proc. ACM Int'l Conference on Multimedia*. ACM Press New York, NY, USA, 1997, pp. 31–40.
102. P. Cohen and S. Oviatt, "The role of voice in human-machine communication," in *Voice communication between humans and machines*, D. B. Roe and J. G. Wilpon, Eds. Washington, DC, USA: National Academy Press, 1994, pp. 34–75.
103. R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, "A system for video surveillance and monitoring," Robotics Institute, Carnegie Mellon University, Technical report CMU-RI-TR-00-12, May 2000.
104. G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309–347, 1992.
105. J. Coopersmith, "Pornography, videotape, and the internet," *IEEE Technol. Soc. Mag.*, pp. 27–34, Spring 2000.
106. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
107. T. Cootes, G. Edwards, and T. C.J., "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, 2001.

108. F. L. Corno F. and S. I., "A cost effective solution for eye-gaze assistive technology," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, 2002.
109. N. Corporate, "T9 text input solutions," http://www.t9.com/.
110. H. Corporation, "Twiddler2," http://www.handykey.com/.
111. C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation: A review," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 28–37, 2006.
112. R. Coudray and B. Besserer, "Global motion estimation for MPEG-encoded streams," in *Proc. IEEE Int'l Conf. on Image Processing*, 2004, pp. 3411–3414.
113. ——, "Motion based segmentation using MPEG streams and watershed method," in *International Symposium on Visual Computing*, 2005, pp. 729–736.
114. H. D. Crane and C. M. Steele, "Accurate three-dimensional eye tracker," *J. of the Opt. Soc. Am.*, vol. 17, pp. 691–705, 1978, 17, 691-705.
115. M. Cristani, M. Bicego, and V. Murino, "On-line adaptive background modelling for audio surveillance," in *Proc. Int'l Conf. on Pattern Recognition*, 2004.
116. ——, "Audio-visual event recognition in surveillance video sequences," *IEEE Trans. Multimedia*, vol. 9, pp. 257–267, Feb. 2007.
117. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, 1998.
118. S. Cunningham, N. Reeves, and M. Britland, "An ethnographic study of music information seeking: implications for the design of a music digital library," in *Proc. ACM/IEEE Joint Conference on Digital Libraries (JCDL'03).* Houston, Texas, US: IEEE Computer Society, 2003, pp. 5–16.
119. R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 781–796, August 2000.
120. R. Dahyot, A. C. Kokaram, N. Rea, and H. Denman, "Joint audio visual retrieval for tennis broadcasts," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, April 2003.
121. R. Dahyot, N. Rea, A. Kokaram, and N. Kingsbury, "Inlier modeling for multimedia data analysis," in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, Siena Italy, September 2004, pp. 482–485.
122. T. Darrell, J. Fisher, P. Viola, and B. Freeman, "Audio-visual segmentation and the cocktail party effect," in *Proc. Int'l Conf. on Multimodal Interfaces*, 2000.
123. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 39, 2007.
124. A. Davison, Y. Cid, and N. Kita, "Real-time 3D SLAM with wide-angle vision," in *Proceedings of the 5th IFAC/EURON Sumbosiom on Intelligent Autonomous Vehicles*, Lissabon, Portugal, July 2004.
125. "Standard international iso/iec 15836,the dublin core metadata element set," Nov. 2003.
126. Y. Dedeoglu, "Moving object detection, tracking and classification for smart video surveillance," Master's thesis, Bilkent University, Department of Computer Engineering, 2004.
127. M. Delakis, G. Gravier, and P. Gros, "Score oriented Viterbi search in sport video structuring using HMM and segment models," in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, 2006.

128. ——, "Audiovisual integration with segment models for tennis video parsing," *Computer Vision and Image Understanding*, 2008, in press.
129. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," *J. of Royal Stat. Soc. (Series B)*, vol. 39, no. 1, pp. 1–38, 1977.
130. L. Deng, J. Dropo, and A. Acero, "Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 3, pp. 412–421, 2005.
131. H. Denman, N. Rea, and A. Kokaram, "Content-based analysis for video from snooker broadcasts," *Computer Vision and Image Understanding*, vol. 92, pp. 141–306, November/December 2003.
132. M. L. Dertouzos, *The Unfinished Revolution: Making Computers Human-Centric*. HarperCollins Publishers, 2001.
133. A. Desolneux, L. Moisan, and J.-M. Morel, "Edge detection by Helmholtz principle," *J. Math. Imaging and Vision*, vol. 14, pp. 271–284, 2001.
134. A. Dielmann and S. Renals, "Automatic meeting segmentation using dynamic Bayesian networks," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 25–36, January 2007.
135. V. Digalakis, J. Rohlicek, and M. Ostendorf, "ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition," *IEEE Trans. Speech Audio Process.*, pp. 431–442, 1993.
136. V. V. Digalakis, "Segment-based stochastic models of spectral dynamics for continuous speech recognition," Ph.D. dissertation, Speech Processing and Interpretation Laboratory, Universit de Boston, 1992.
137. D. Dimitriadis, P. Maragos, and A. Potamianos, "Auditory Teager energy cepstrum coefficients for robust speech recognition," in *Proc. Int'l Conf. on Speech Communication and Technology*, Lisboa, Portugal, Sep 2005.
138. N. Dimitrova, L. Agnihorti, and G. Wei, "Video classification based on HMM using text and faces," in *Proc. European Signal Processing Conference*, 2000.
139. A. Dix, J. Finlay, G. Abowd, and R. Beale, *Human-Computer Interaction*. Prentice Hall, 2004.
140. A. Doulamis, N. Doulamis, Y. Avrithis, and S. Kollias, "A fuzzy video content representation for video summarization and content-based," *Signal Processing*, vol. 80, no. 6, pp. 1049–1067, Jun 2000.
141. J. Downie, *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, 2003, vol. 37, ch. Music Information Retrieval, pp. 295–340.
142. S. A. Drab and N. M. Artner, "Motion detection as interaction technique for games & applications on mobile devices," in *Pervasive Mobile Interaction Devices (PERMID 2005) Workshop at the Pervasive 2005*, Munich, DR, May 2005.
143. J. Driver, "Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading," *Nature*, vol. 381, pp. 66–68, May 1996.
144. A. T. Duchowski, "A breadth-first survey of eye tracking applications," *Behavior Research Methods, Instruments, and Computers*, vol. 34, no. 4, pp. 455–470, 2002.
145. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. J. Wiley & Sons, 2001.

146. F. Dufaux and J. Konrad, "Efficient, robust and fast global motion estimation for video coding," *IEEE Trans. Image Process.*, vol. 9, pp. 497–501, 2000.
147. S. Dupont and J. Luettin, "Audio-visual speech modeling for continuous speech recognition," *IEEE Trans. Multimedia*, vol. 2, no. 3, pp. 141–151, 2000.
148. S. Eickeler and S. Muller, "Content-based video indexing of TV broadcast news using hidden Markov models," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1999, pp. 2997–3000.
149. A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, July 2003.
150. C. Elting, S. Rapp, G. Möhler, and M. Strube, "Architecture and implementation of multimodal plug and play," in *Proc. Int'l Conf. on Multimodal Interfaces*. ACM Press New York, NY, USA, 2003, pp. 93–100.
151. R. Engel and N. Pfleger, "Modality fusion," in *SmartKom: Foundations of Multimodal Dialogue Systems*, W. Wahlster, Ed. Springer-Verlag, New York, NY, 2006, pp. 223–236.
152. E. Erzin, A. Cetin, and Y. Yardimci, "Subband analysis for robust speech recognition in the presence of car noise," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1995.
153. G. Evangelopoulos and P. Maragos, "Multiband modulation energy tracking for noisy speech detection," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2024–2038, Nov 2006.
154. R. Fagin, R. Kumar, and D. Sivakumar, "Efficient similarity search and classification via rank aggregation," in *ACM SIGMOD International Conference on Management of Data*, San Diego, Californie, Etats-Unis, 9-12 Jun. 2003, pp. 301–312.
155. X. Fan, X. Xie, W.-Y. Ma, H.-J. Zhang, and H.-Q. Zhou, "Visual attention based image browsing on mobile devices," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, vol. I, 2003, pp. 53–56.
156. C. Fangxiang, W. Christmas, and J. Kittler, "Periodic human motion description for sports video databases," in *Proc. Int'l Conf. on Pattern Recognition*, vol. 3, 2004, pp. 870 – 873.
157. C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.
158. P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *Int'l J. of Comp. Vis.*, vol. 61, no. 1, pp. 55–79, 2005.
159. M. Ferecatu, "Image retrieval with active relevance feedback using both visual and keyword-based descriptors," Ph.D. dissertation, INRIA—University of Versailles Saint Quentin-en-Yvelines, France, 2005.
160. M. Ferecatu, M. Crucianu, and N. Boujemaa, "Retrieval of difficult image classes using svm-based relevance feedback," in *Proc. ACM Int'l Workshop on Multimedia Information Retrieval*, October 2004, pp. 23 – 30.
161. M. Ferman and A. M. Tekalp, "Probabilistic analysis and extraction of video content," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 2, 1999, pp. 91–95.
162. K. P. Fishkin, A. Gujar, B. L. Harrison, T. P. Moran, and R. Want, "Embodied user interfaces for really direct manipulation," *Communications of the ACM*, vol. 43, no. 9, pp. 74–80, 2000.
163. M. M. Fleck, D. A. Forsyth, and C. Bregler, "Finding Naked People," in *Proc. European Conf. on Computer Vision*, vol. 2, 1996, pp. 593–602.

164. F. Fleuret and H. Sahbi, "Scale-invariance of support vector machines based on the triangular kernel," in *3rd International Workshop on Statistical and Computational Theories of Vision*, October 2003.
165. M. S. Flickner, H. Niblack, W. Ashley, J. Q. H. Dom, B. Gorkani, M. Hafner, J. Lee, D. Petkovic, D. Steele, and D. Yanker, "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.
166. F. Flippo, A. Krebs, and I. Marsic, "A framework for rapid development of multimodal interfaces," in *Proc. Int'l Conf. on Multimodal Interfaces*. ACM Press New York, NY, USA, 2003, pp. 109–116.
167. J. D. Foley, V. L. Wallace, and P. Chan, "The human factors of computer graphics interaction techniques," *IEEE Comput. Graph. Appl.*, vol. 4, no. 11, pp. 13–48, 1984.
168. J. Foote, "An overview of audio information retrieval," *Multimedia Systems*, vol. 7, no. 1, pp. 2–10, 1999.
169. G. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–277, 1973.
170. W. T. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 9, pp. 891–906, 1991.
171. B. Frey, T. Kristjansson, L. Deng, and A. Acero, "Learning dynamic noise models from noisy speech for robust speech recognition," in *Proc. Advances in Neural Information Processing Systems*, vol. 8, 2001, pp. 472–478.
172. W. A. Fuller, *Measurement Error Models*. Wiley, 1987.
173. B. Furht and O. M. (Editors), *Handbook of Video Databases: Design and Applications*. CRC Press, Boca Raton, FL, 2003.
174. D. Gabor, "Theory of communication," *Journal Inst. of Elec. Eng. London*, vol. 93, no. III, pp. 429–457, 1946.
175. W. O. Galitz, *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. John Wiley & Sons, 2002.
176. S. Galliano, E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre, and G. Gravier, "The ESTER phase II evaluation campaign for the rich transcription of French broadcast news," in *Proc. Int'l Conf. on Speech Communication and Technology*, 2005.
177. E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman Publishing Co., Boston, MA, USA, 1995.
178. C. Garcia and M. Delakis, "Convolutional face finder: A neural architecture for fast and robust face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1408–1423, 2004.
179. S. Gauch, J. Wang, and S. Rachakonda, "A Corpus Analysis Approach for Automatic Query Expansion and its Extension to Multiple Databases," *ACM Transactions on Information Systems*, vol. 17, no. 3, pp. 250–269, 1999.
180. S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, 1984.
181. T. Gevers and A. W. M. Smeulders, "Content-based image retrieval: An overview," in *Emerging Topics in Computer Vision*, G. Medioni and S. B. Kang, Eds. Prentice Hall, 2004, ch. 8.
182. D. Gildea and T. Hofmann, "Topic-based language models using EM," in *Proc. European Conf. on Speech Communication and Technology*, 1999.

183. A. Girgensohn, J. Boreczky, and L. Wilcox, "Keyframe-based user interfaces for digital video," *IEEE Computer*, vol. 34, no. 9, pp. 61–67, Sep 2001.
184. H. Glotin, D. Vergyri, C. Neti, G. Potamianos, and J. Luettin, "Weighting schemes for audio-visual fusion in speech recognition," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2001.
185. E. B. Goldstein, *Sensation and Perception.* California: Wadsworth Publ. Co., 1984.
186. Y. Gong, "A method of joint compensation of additive and convolutive distortions for speaker-independent speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 975–983, 2005.
187. Y. Gong, L. T. Sin, C. H. Chuan, H. Zhang, and M. Sakauchi, "Automatic parsing of TV soccer programs," in *Proc. of IEEE Int'l Conference on Multimedia Computing and Systems*, vol. 7, May 1995, pp. 167–174.
188. G. Gravier and D. Moraru, "Towards phonetically-driven hidden Markov models: Can we incorporate phonetic landmarks in HMM-based ASR?" in *Proc. ISCA Tutorial and Research Workshop on Non Linear Speech Processing*, ser. Lecture Notes in Artificial Intelligence, M. C. et al., Ed., vol. 4885. Springer Verlag, 2007, pp. 161–168.
189. U. Grenander, *Elements of Pattern Theory.* The Jonhs Hopkins Univ. Press, 1996.
190. W. G. Griswold, P. Shanahan, S. W. Brown, R. Boyer, M. Ratto, R. B. Shapiro, and T. M. Truong, "ActiveCampus: Experiments in community-oriented ubiquitous computing," *Communications of the ACM*, vol. 43, no. 9, pp. 74–80, 2000.
191. J. Gustafson, "Developing multimodal spoken dialogue systems. empirical studies of human-computer interaction," Ph.D. dissertation, Department of Speech, Music and Hearing, KTH, 2002.
192. A. Guttman, "R-trees: A dynamic index structure for spatial searching," in *ACM SIGMOD International Conference on Management od Data*, Boston, Massachusetts, Etats-Unis, 18-21 Jun. 1984, pp. 47–57.
193. M. Hakkarainen and C. Woodward, "Symball - camera driven table tennis for mobile phones," in *ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE 2005*, 2005, pp. 15 – 17.
194. M. Haller, M. Billinghurst, and B. H. Thomas, *Emerging Technologies of Augmented Reality.* Hershey, PA, USA: IGI Publishing, 2006.
195. J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection," in *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology.* New York, NY, USA: ACM Press, 2005, pp. 115–118.
196. J. Hannuksela, P. Sangi, and J. Heikkilä, "A vision-based approach for controlling user interfaces of mobile devices," in *Proc. of IEEE Workshop on Vision for Human-Computer Interaction (V4HCI)*, 2005.
197. J. P. Hansen, A. W. Anderson, and P. Roed, *Symbiosis of Human and Artifact.* Elsevier Science, 1995, vol. 20A, ch. Eye gaze control of multimedia systems, pp. 37–42.
198. B. L. Harrison, K. P. Fishkin, A. Gujar, C. Mochon, and R. Want, "Squeeze me, hold me, tilt me! an exploration of manipulative user interfaces," in *Proc. ACM-SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1998, pp. 17–24.

199. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. New York, NY, USA: Cambridge University Press, 2003.
200. A. Hauptmann, "Lessons for the future from a decade of Informedia video analysis research," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, vol. 3568, 2005, pp. 1–10.
201. A. Hauptmann, R. Yan, T. Ng, W. Lin, R. Jin, D. Christel, M. Chen, and R. Baron, "Video classification and retrieval with the Informedia digital video library system," in *Proc. Text Retrieval Conference*, Gaithersburg, MD, USA, Nov 2002.
202. P. A. Heeman, "POS tags and decision trees for language modeling," in *Proc. the Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1999.
203. F. Heijden, *Image Based Measurement Systems: Object Recognition and Parameter Estimation*. WILEY, Jan. 1996.
204. H. Helmholtz, *Physiological Optics, Vol.III: The Perceptions of Vision (J. P. Southall, Trans.)*. Rochester, NY: Optical Soc. Amer., 1910, 1925.
205. M. Hennecke, D. Stork, and K. Prasad, "Visionary speech: Looking ahead to practical speechreading systems," in *Speechreading by Humans and Machines*, D. Stork and M. Hennecke, Eds. Berlin, Germany: Springer, 1996, pp. 331–349.
206. A. Henrich, "The lsd^{h}-tree: An access structure for feature vectors," in *14th International Conference on Data Engineering*, Orlando, Floride, Etats-Unis, 23-27 Feb. 1998, pp. 362–369.
207. A. Henrysson, M. Billinghurst, and M. Ollila, "Virtual object manipulation using a mobile phone," in *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*. New York, NY, USA: ACM Press, 2005, pp. 164–171.
208. H. Hermansky, M. Pavel, and S. Tibrewala, "Towards ASR using partially corrupted speech," in *Proc. Int'l Conf. on Spoken Language Processing*, Oct. 1996, pp. 458–461.
209. J. Hershey and J. Movellan, "Audio-vision: Using audio-visual synchrony to locate sounds," in *Proc. Advances in Neural Information Processing Systems*, 1999.
210. J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy, "Combining sensory information: Mandatory fusion within, but not between, senses," *Science*, vol. 298, pp. 1627–1630, 2002.
211. K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz, "Sensing techniques for mobile interaction," in *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM Press, 2000, pp. 91–100.
212. A. Hliaoutakis, G. Varelas, E. Voutsakis, E. G. Petrakis, and E. Milios, "Information retrieval by semantic similarity," *Intern. Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 3, no. 3, pp. 55–73, July/Sept. 2006, special Issue of Multimedia Semantics.
213. P. Honkamaa, J. Jäppinen, and C. Woodward, "Interactive outdoor mobile augmentation using markerless tracking and gps," in *Proc. Mobile Ubiquitous Multimedia (MUM2007), Oulu, Finland*, December 2007.
214. P. Honkamaa, S. Siltanen, J. Jäppinen, C. Woodward, and O. Korkalo, "A lightweight approach for augmented reality on camera phones using 2d images

to simulate 3d," in *Proc. Virtual Reality International Conference (VRIC), Laval, France*, April 2007, pp. 285–288.
215. B. K. Horn, *Robot Vision*. Cambridge, Massachusetts: MIT Press, 1986.
216. E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumiere Project: Bayesian user modeling for inferring the goals and needs of software users," in *Proc. Int'l Conf. on Uncertainty in Artificial Intelligence*, Madison, Wisconsin, 1998, pp. 256–265.
217. X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.
218. S. Huet, "Informations morpho-syntaxiques et adaptation thématique pour améliorer la reconnaissance de la parole," Ph.D. dissertation, Université de Rennes 1, Rennes, France, dec 2007.
219. S. Huet, G. Gravier, and P. Sébillot, "Morphosyntactic processing of N-best lists for improved recognition and confidence measure computation," in *Proc. Int'l Conf. on Speech Communication and Technology*, 2007.
220. Q. Huo and C. Lee, "A bayesian predictive approach to robust speech recognition," *IEEE Trans. Speech Audio Process.*, pp. 200–204, 2000.
221. I. K. Ibrahim, *Handbook of Research on Mobile Multimedia*. Hershey, PA, USA: IGI Publishing, 2006.
222. F. Idris and S. Panchanathan, "Review of image and video indexing techniques," *Journal of Visual Communication and Image Representation*, vol. 8, no. 2, pp. 146–166, 1997.
223. C. Inferense Group, Cavendish Laboratory, "Dasher developments," http://www.inference.phy.cam.ac.uk/dasher/development/.
224. Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Query databases through multiple examples," in *Proc. Int'l Conf. on Very Large Data Bases*, New York. USA, 1998, pp. 218–227.
225. U. Iurgel, R. Meermeier, S. Eickeler, and G. Rigoll, "New approaches to audiovisual segmentation of TV news for automatic topic retrieval," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2001, pp. 1397–1400.
226. H. Iwata, "Haptic interfaces," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 2003, pp. 206–219.
227. R. Iyer and M. Ostendorf, "Modeling long distance dependence in language: Topic mixtures *versus* dynamic cache models," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 1, pp. 30–39, 1999.
228. J.-Hu and A. Bagga, "Identifying story and preview images in news web pages," in *Proc. 7th Int'l Conf. on Document Analysis and Recognition (ICDAR'2003)*, Edinburgh, Scotland, Aug. 2003, pp. 640–644.
229. F. Jabloun and A. E. Cetin, "The teager energy based feature parameters for robust speech recognition in car noise," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*. Washington, DC, USA: IEEE Computer Society, 1999, pp. 273–276.
230. R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: what you look at is what you get," *ACM Transactions on Information Systems*, vol. 9, no. 2, pp. 152–169, 1991.
231. R. Jacob, "Eye movement-based human-computer interaction techniques: Toward non-command interfaces," *Advances in Human-Computer Interaction*, vol. 4, pp. 150–190, 1993.

232. A. Jaimes, J. B. Pelz, T. Grabowski, J. Babcock, and S. F. Chang, "Using human observers' eye movements in automatic image classifiers," in *Proceedings of SPIE Human Vision and Electronic Imaging VI, San Jose, CA*, 2001.
233. A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, December 2005.
234. A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, pp. 34–40, January 2004.
235. A. Jameson, "Adaptive interfaces and agents," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 2003, pp. 305–330.
236. F. Jensen, S. Lauritzen, and K. Olsen, "Bayesian updating in recursive graphical models by local computations," *Computational Statistics Quarterly*, vol. 4, pp. 269–282, 1990.
237. M. Johnston, "Unification-based multimodal parsing," in *Proc. of the 36th annual meeting on Association for Computational Linguistics*, Montreal, Canada, 1998, pp. 624–630.
238. M. Johnston and S. Bangalore, "Finite-state multimodal integration and understanding," *Natural Language Engineering*, vol. 11, no. 2, pp. 159–187, 2005.
239. M. Johnston, P. R. Cohen, D. McGee, S. L. Oviatt, J. A. Pittman, and I. Smith, "Unification-based multimodal integration," in *Proc. of the 8th conference of European chapter of the Association for Computational Linguistics*, Madrid, Spain, 1997, pp. 281–288.
240. M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection." *Int'l J. of Comp. Vis.*, vol. 46, no. 1, pp. 81–96, 2002.
241. E. Jonietz, "Augmented reality: Special issue 10 emerging technologies 2007, MIT technology review," 2007.
242. S. X. Ju, M. J. Black, S. Minneman, and D. Kimber, "Summarization of videotaped presentations: automatic analysis of motion and gesture," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 686–696, 1998.
243. O. O. K. and S. F. W. M., "Perceptual image retrieval using eye movements," in *Proceedings of International Workshop on Intelligent Computing in Pattern Analysis/Synthesis*, 2007, xi'an, China, 26-27 August.
244. J. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, Albuquerque N.M., Apr 1990, pp. 381–384.
245. ——, "Construction and evaluation of a robust multifeature speech/music discriminator," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1997, pp. 1331–1334.
246. E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*. Stamford, Connecticut: McGraw-Hill, 4 edition, 2000.
247. A. Katsamanis, G. Papandreou, and P. Maragos, "Audiovisual-to-articulatory speech inversion using active appearance models for the face and hidden markov models for the dynamics," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2008.
248. A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, "Multimodal fusion by adaptive compensation for feature uncertainty with application to audiovisual speech recognition," in *Proc. European Signal Processing Conference*, 2006.

249. M. Kay, "Functional grammar," in *Proc. of the 5th Annual Meeting of the Berkeley Linguistics Society*, 1979, pp. 142–158.
250. C. Kayser, C. Petkov, M. Lippert, and N. Logothetis, "Mechanisms for allocating auditory attention: an auditory saliency map," *Current Biology*, vol. 15, no. 21, pp. 1943–1947, 2005.
251. R. Keiller, "Using VoiceXMl 2.0 in the VxOne unified messaging application," in *Practical Spoken Dialog Systems*, D. A. Dahl, Ed. Kluwer Academic Publishers, 2004, pp. 143–163.
252. H. Keränen, L. Pohjanheimo, and H. Ailisto, "Tag manager: a mobile phone platform for physical selection services," in *IEEE International Conference on Pervasive Services 2005 (ICPS'05)*, 2005, pp. 405 – 412.
253. A. Kerne, "Collage machine: an interactive agent of web recombination," *Leonardo*, vol. 33, no. 5, pp. 347–350, 2000.
254. M. Kherfi, D. Ziou, and A. Bernardi, "Image retrieval from the world wide web: Issues, techniques, and systems," *ACM Computing Surveys*, vol. 36, no. 1, pp. 35–67, March 2004.
255. E. Kidron, Y. Y. Schechner, and M. Elad, "Cross-modal localization via sparsity," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1390–1404, Apr. 2007.
256. E. Kijak, G. Gravier, P. Gros, L. Oisel, and F. Bimbot, "Hmm based structuring of tennis videos using visual and audio cues," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, vol. 3, July 2003, pp. 309–312.
257. E. Kijak, G. Gravier, L. Oisel, and P. Gros, "Audiovisual integration for sport broadcast structuring," *Multimedia Tools and Applications*, vol. 30, pp. 289–312, 2006.
258. A. Kilgarriff and M. Palmer, "Special Issue on Senseval," *Computers and the Humanities*, vol. 34, no. 1/2, Apr. 2000.
259. C. W. Kim, R. Ansari, and A. E. Cetin, "A class of linear-phase regular biorthogonal wavelets," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1992, pp. 673–676.
260. J. Kim and J. Peal, "A computational model for causal and diagnostic reasoning in inference systems," in *Proc. Int'l Joint Conf. on Artificial Intel.*, 1983, pp. 190–193.
261. M. Kipp, "ANVIL - A generic annotation tool for multimodal dialogue," in *Proc. European Conf. on Speech Communication and Technology*, 2001, pp. 1367–1370.
262. J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
263. D. Klakow, "Selecting articles from the language model training corpus," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2000.
264. J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.
265. P. Knees, M. Schedl, T. Pohle, and G. Widmer, "An innovative three-dimensional user interface for exploring music collections enriched with meta-information from the web," in *Proc. ACM Int'l Conference on Multimedia*, Santa Barbara, California, USA, October 23-26 2006, pp. 17–24.
266. P. Knees, M. Schedl, and G. Widmer, "Multiple lyrics alignment: Automatic retrieval of song lyrics," in *Proc. Int'l Conf. on Music Information Retrieval*, London, UK, September 11-15 2005, pp. 564–569.

267. D. C. Knill, D. Kersten, and A. L. Yuille, *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996, ch. Introduction: A Bayesian Formulation of Visual Perception, pp. 1–21.
268. C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, Jun 1985.
269. K. Koffka, *Principles of Gestalt Psychology*. Routledge, 1935, 1999.
270. W. Köhler, *Gestalt Psychology*. New York: Liveright Publish. Corp., 1947, 1970.
271. T. Kohonen, *Self-Organizing Maps*, 3rd ed., ser. Springer Series in Information Sciences. Berlin: Springer, 2001, vol. 30.
272. T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proc. IEEE*, vol. 84, no. 10, pp. 1358–1384, October 1996.
273. A. Kokaram and P. Delacourt, "On the motion-based diagnosis of video from cricket broadcasts," in *Irish Signals and Systems Conference*, June 2002.
274. A. Kokaram, N. Rea, R. Dahyot, A. M. Tekalp, P. Bouthemy, P. Gros, and I. Sezan, "Browsing sports video," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 47–58, March 2006.
275. A. C. Kokaram, *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*. Springer Verlag, 1998.
276. M. Kotti, E. Benetos, C. Kotropoulos, and I. Pitas, "A neural network approach to audio-assisted movie dialogue detection," *Neurocomputating*, vol. 71, pp. 157–166, 2007.
277. M. Kotti, C. Kotropoulos, B. Ziolko, I. Pitas, and V. Moschou, "A framework for dialogue detection in movies," *Lecture Notes in Computer Science*, vol. 4105, pp. 371–378, 2006.
278. W. Kraaij and R. Pohlmann, "Comparing the Effect of Syntactic vs. Statistical Phrase Indexing Strategies for Dutch," in *2nd European Conference on Research and Advanced Technology for Digital Libraries*, C. Nicolaou and C. Stephanides, Eds. Lecture Notes in Computer Science, Springer Verlag, 1998, vol. 1513, pp. 605–614.
279. P. Král, C. Cerisara, and J. Klečková, "Automatic dialog acts recognition based on sentence structure," in *Proc. Int'l Conf. on Speech Communication and Technology*, 2005, pp. 825–828.
280. B. Kroon, J. Nesvadba, and A. Hanjalic, "Dialog detection in narrative video by shot and face analysis," in *Proc. IS&T/SPIE's 19th Annual Symp. Electronic Imaging -Multimedia Content Access: Algorithms and Systems*, vol. 6506, 2007, pp. 315–325.
281. S. Kumar and P. R. Cohen, "Towards a fault-tolerant multi-agent system architecture," in *Proc. International Conference on Autonomous Agents*. ACM Press New York, NY, USA, 2000, pp. 459–466.
282. M. La Cascia, S. Sethi, and S. Sclaroff, "Combining textual and visual cues for content-based image retrieval on the world wide web," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998, pp. 24–28.
283. J. Lai and N. Yankelovich, "Conversational speech interfaces," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 2003, pp. 698–713.

284. M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: in defense of weak fusion," *Vision Research*, vol. 35, no. 3, pp. 389–412, 1995.
285. C. Leacock, M. Chodorow, and G. A. Miller, "Using corpus statistics and WordNet relations for sense identification," *Computational Linguistics*, vol. 24, no. 1, pp. 147–165, 1998.
286. J. J. Lee, J. Kim, and J. H. Kim, "Data-driven design of HMM topology for on-line handwriting recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 1, 2001.
287. B. Lehane, N. O'Connor, and N. Murphy, "Action sequence detection in motion pictures," in *Proc. European Workshop Integration of Knowledge, Semantics and Digital Media Technology*, 2004.
288. ——, "Dialogue scene detection in movies using low and mid-level visual features," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, 2004, pp. 286–296.
289. H. Lejsek, F. H. Asmundsson, B. Thor-Jonsson, and L. Amsaleg, "Scalability of local image descriptors: A comparative study," in *Proc. ACM Int'l Conference on Multimedia*, Oct. 2006.
290. R. Lempel and A. Soffer, "PicASHOW: Pictorial authority search by hyperlinks on the web," *ACM Transactions on Information Systems*, vol. 20, no. 1, pp. 1–24, Jan. 2002.
291. D. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.
292. D. Lennon, N. Harte, and A. Kokaram, "A HMM framework for motion based parsing for video from observational psychology," in *Irish Machine Vision and Image Processing Conference*, DCU, Dublin, Ireland, August 2006, pp. 110–117.
293. D. Lennon, "Motion based parsing," Master's thesis, Trinity College Dublin, 2007.
294. R. Leonardi, P. Migliorati, and M. Prandini, "Semantic indexing of soccer audio-visual sequences: A multimodal approach based on controlled markov chains," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, May 2004.
295. M. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State-of-the-art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 2, no. 1, pp. 1–19, 2006.
296. Y. Li, Z. A. Bandar, and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 4, pp. 871–882, July/Aug. 2003.
297. Y. Li and C. C. J. Kuo, "Real-time segmentation and annotation of MPEG video based on multimodal content analysis I & II," Univ. Southern California, Los Angeles, Technical Report, Tech. Rep., 2000.
298. ——, *Video Content Analysis Using Multimodal Information*. Springer, 2003.
299. Y. Li, S. Narayanan, and C. C. J. Kuo, "Identification of speakers in movie dialogues using audiovisual cues," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, vol. 2, 2002, pp. 2093–2096.
300. ——, "Content-based movie analysis and indexing based on audiovisual cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1073–1085, 2004.

301. T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *Proc. Int'l Conf. on Music Information Retrieval*, London, UK, September 11-15 2005, pp. 34–41.
302. D. Lin, "An information-theoretic definition of similarity," in *Proc. Int'l Conf. on Machine Learning*, 1998, pp. 296–304.
303. D. J. Litman and S. Pan, "Predicting and adapting to poor speech recognition in a spoken dialogue system," in *Proc. National Conference on Artificial Intelligence and Conference on Innovative Applications of Artificial Intelligence*. AAAI Press / The MIT Press, 2000, pp. 722–728.
304. C.-B. Liu and N. Ahuja, "Motion based retrieval of dynamic objects in videos," in *Proc. ACM Int'l Conference on Multimedia*, 2004, pp. 288–291.
305. H. Liu and P. Singh, "Conceptnet: a practical commonsense reasoning toolkit," *BT Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.
306. Y. Liu, E. Shriberg, A. Stolcke, and M. P. Harper, "Using machine learning to cope with imbalanced classes in natural speech: Evidence from sentence boundary and disfluency detection," in *Proc. Int'l Conf. on Spoken Language Processing*, 2004.
307. B. Logan, A. Kositsky, and P. Moreno, "Semantic analysis of song lyrics," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*. Taipei, Taiwan: IEEE Computer Society, June 27-30 2004, pp. 827–830.
308. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l J. of Comp. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
309. P. P. G. P. Ltd, "Virtual laser keyboard," http://www.virtual-laser-keyboard.com/.
310. L. Lu, H. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 7, pp. 504–516, 2002.
311. Y. Lu, C. Hu, X. Zhu, H.-J. Zhang, and Q. Yang, "A unified framework for semantic and feature based relevance feedback in image retrieval systems," in *Proc. ACM Int'l Conference on Multimedia*, Los Angeles CA, USA, 2000, pp. 31–37.
312. J. Luettin, G. Potamianos, and C. Neti, "Asynchronous stream modeling for large vocabulary audio-visual speech recognition," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2001.
313. Y. Ma, L. Lu, H. Zhang, and M. Li, "A user attention model for video summarization," in *Proc. ACM Int'l Conference on Multimedia*, 2002.
314. Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, "A generic framework of user attention model and its application in video summarization," *IEEE Trans. Multimedia*, vol. 7, pp. 907–919, 2005.
315. I. S. MacKenzie and R. W. Soukoreff, "Text entry for mobile computing: models and methods, theory and practice," *Human-Computer Interaction*, vol. 17, no. 2, pp. 147–198, 2002.
316. J. P. G. Mahedero, Á. Martínez, P. Cano, M. Koppenberger, and F. Gouyon, "Natural language processing of lyrics," in *Proc. ACM Int'l Conference on Multimedia*. New York, NY, USA: ACM Press, 2005, pp. 475–478.
317. G. Maltese and F. Mancini, "An automatic technique to include grammatical and morphological information in a trigram-based statistical language model," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1992.

318. L. Mangu, E. Brill, and A. Stolcke, "Finding consensus in speech recognition: Word error minimization and other applications of confusion networks," *Computer Speech and Language*, vol. 14, no. 4, pp. 373–400, 2000.
319. B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley, 2002.
320. P. Maragos, J. Kaiser, and T. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3024–3051, Oct 1993.
321. D. Marcu, "The rhetorical parsing of unrestricted texts: A surface-based approach," *Computational Linguistics*, vol. 26, no. 3, pp. 395–448, 2000.
322. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Acad. Press, 1979.
323. J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solutions of ill-posed problems in computational vision," *J. of the Amer. Stat. Assoc.*, vol. 82, no. 37, pp. 76–89, March 1987.
324. I. Marsic, A. Medl, and J. Flanagan, "Natural communication with information systems," *Proc. IEEE*, vol. 88, no. 8, pp. 1354–1366, 2000.
325. D. L. Martin, "The Open Agent Architecture: A framework for building distributed software system," *Applied Artificial Intelligence*, vol. 13, no. 1, pp. 91–128, 1999.
326. J. Martinez, "Standards - mpeg-7 overview of mpeg-7 description tools, part 2," *IEEE Multimedia*, vol. 9, no. 3, pp. 83–93, Jul-Sep 2002.
327. R. Mason, R. Gunst, and J. Hess, *Statistical Design and Analysis of Experiments*. Wiley, 1989.
328. D. Massaro and D. Stork, "Speech recognition and sensory integration," *American Scientist*, vol. 86, no. 3, pp. 236–244, 1998.
329. I. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 198–213, 2002.
330. R. Mayer, T. A. Aziz, and A. Rauber, "Visualising class distribution on self-organising maps," in *Proc. Int'l Conf. on Artificial Neural Networks*. Porto, Portugal: Springer, September 9 - 13 2007, pp. 359–368.
331. S. McAdams, "Recognition of auditory sound sources and events," in *Thinking in Sound:The Cognitive Psychology of Human Audition*. Oxford University Press, 1993.
332. J. McCarthy, M. A. Sasse, and R. J., "Could I have the menu please?: An eye tracking study of design conventions," in *Proc. Annual Conf. on Human-Computer Interaction*, 2003, 8-12 Sep 2003, Bath, UK.
333. I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang, "Automatic analysis of multimodal group actions in meetings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 305–317, 2005.
334. H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, pp. 746–748, 1976.
335. M. F. McTear, "Spoken dialogue technology: enabling the conversational user interface," *ACM Computing Surveys*, vol. 34, no. 1, pp. 90–169, 2002.
336. B. Merialdo, "Tagging English text with a probabilistic model," *Computational Linguistics*, vol. 20, no. 2, pp. 155–171, 1994.
337. Microsoft, "Microsoft surface," http://www.microsoft.com/surface/.
338. MicroVision, "Microvision, inc." http://www.microvision.com/.

339. P. Milgram and F. Kishino, "Taxonomy of mixed reality virtual displays," *Institute of Electronics, Information, and Communication Engineers Trans. Information and Systems*, vol. E77-D, no. 12, pp. 1321–1239, 1994.
340. G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for information processing," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
341. M. Minksy, "A framework for representing knowledge," in *The Psychology of Computer Vision*, P. Winston, Ed. McGraw-Hill, 1977, pp. 211–277.
342. G. Monaci and P. Vandergheynst, "Audiovisual Gestalts," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition Workshop*. New York, NY: IEEE Computer Society, 2006, p. 200.
343. F. Mörchen, A. Ultsch, M. Nöcker, and C. Stamm, "Databionic visualization of music collections according to perceptual distance," in *Proc. Int'l Conf. on Music Information Retrieval*, London, UK, September 11-15 2005, pp. 396–403.
344. A. Morris, A. Hagen, H. Glotin, and H. Bourlard, "Multi-stream adaptive evidence combination for noise robust ASR," *Speech Communication*, vol. 34, pp. 25–40, 2001.
345. "Standard international iso/iec 21000 information technology – "multimedia framework"."
346. "MPEG-7 requirements document v.15, iso/iec jtc1/sc29/wg11, mpeg01/n4320," Jul. 2001.
347. S. V. Mulken, E. André, and J. Müller, "The persona effect: How substantial is it?" in *Proc. of HCI on People and Computers*. Springer-Verlag, London, UK, 1998, pp. 53–66.
348. D. Mumford, *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996, ch. Pattern Theory: A unifying perspective, pp. 25–61.
349. D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure & Appl. Math.*, vol. XLII, no. 4, 1989.
350. K. Murphy, "Dynamic Bayesian networks: Representation, inference and learning," Ph.D. dissertation, Univ. of California, Berkeley, 2002.
351. J. Myers and A. Well, *Research Design and Statistical Analysis*. Lawrence Erlbaum Associates, 2003.
352. MyVu, "Myvu corp." http://www.myvu.com/.
353. C. Naas and L. Gong, "Ten principles for designing Human-Computer dialog systems," in *Practical Spoken Dialog Systems*, D. A. Dahl, Ed. Kluwer Academic Publishers, 2004, pp. 25–40.
354. J. Nam, M. Alghoniemy, and A. Tewfik, "Audio-visual content-based violent scene characterization," in *Proc. IEEE Int'l Conf. on Image Processing*, 1998, pp. 353–357.
355. J. Nam, A. E. Çetin, and A. H. Tewfik, "Speaker identification and video analysis for hierarchical video shot classification." in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 2, 1997, pp. 550–553.
356. K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio based biometric score fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, February 2008.
357. M. R. Naphade and T. S. Huang, "Extracting semantics from audio-visual content: the final frontier in multimedia retrieval," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 793–810, 2002.

358. A. Nasr, Y. Estève, F. Béchet, T. Spriet, and R. de Mori, "A language model combining N-grams and stochastic finite state automata," in *Proc. European Conf. on Speech Communication and Technology*, 1999.
359. X. Naturel, G. Gravier, and P. Gros, "Fast structuring of large television streams using program guides," in *Proceedings of the 4th International Workshop on Adaptive Multimedia Retrieval (AMR), Geneva, Switzerland*, ser. Lecture Notes in Computer Science, vol. 4398, Jul. 2006, pp. 223–232.
360. J. G. Neal and S. C. Shapiro, "Intelligent multi-media interface technology," *ACM-SIGCHI Bulletin*, vol. 20, no. 1, pp. 75–76, 1988.
361. A. V. Nefian, "Coupled hidden markov model for audiovisual speech recognition," *US Patent No. 7,165,029*, Jan. 2007.
362. A. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, "Dynamic bayesian networks for audio-visual speech recognition," *EURASIP Journal on Applied Signal Processing*, vol. 11, pp. 1–15, 2002.
363. R. Neumayer, M. Dittenbach, and A. Rauber, "PlaySOM and PocketSOMPlayer: Alternative interfaces to large music collections," in *Proc. Int'l Conf. on Music Information Retrieval*. London, UK: Queen Mary, University of London, September 11-15 2005, pp. 618–623.
364. R. Neumayer and A. Rauber, "Integration of text and audio features for genre classification in music information retrieval," in *Proc. European Conf. on Information Retrieval*, Rome, Italy, April 2-5 2007, pp. 724–727.
365. ——, "Multi-modal music information retrieval - visualisation and evaluation of clusterings by both audio and lyrics," in *Proceedings of the 8th Conference Recherche d'Information Assistée par Ordinateur (RIAO'07)*. Pittsburgh, PA, USA: ACM, May 29th - June 1 2007.
366. J. Nielsen, *Usability Engineering*. Academic Press, 1993.
367. ——, "Alert Box: F-Shaped pattern for reading web content," http://www.useit.com/alertbox/reading_pattern.html, 2006.
368. J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*. ACM Press New York, NY, USA, 1990, pp. 249–256.
369. L. Nigay and J. Coutaz, "A design space for multimodal systems: concurrent processing and data fusion," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*. ACM, New York, NY, USA, 1993, pp. 172–178.
370. S. G. Nikolov, D. R. Bull, and I. D. Glichrist, "Gaze-contingent multi-modality displays of multi-layered geographical maps," in *Proc. of the 5th Intl. Conf. on Numerical Methods and Applications (NM&A02)*, 2002, symposium on Numerical Methods for Sensor Data Processing, Borovetz, Bulgaria.
371. Nokia, "Mara - the mobile augmented reality applications project," http://research.nokia.com/research/projects/mara/index.html.
372. T. Numajiri, A. Nakamura, and Y. Kuno, "Speed browser controlled by eye movements," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, 2002, august 26-29, Lausanne, 2002.
373. J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, December 1995.
374. U. of South Australia, "Tinmith ar system," http://www.tinmith.net/.
375. K. Ohtsuki, K. Bessho, Y. Matsuo, S. Matsunaga, and Y. Hayashi, "Automatic multimedia indexing," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 69–78, 2006.

376. N. Oliver, A. Garg, and E. Horvitz, "Layered representations for learning and inferring office activity from multiple sensory channels," *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 163–180, 2004.
377. N. Orio, "Music retrieval: A tutorial and review," *Foundations and Trends in Information Retrieval*, vol. 1, no. 1, pp. 1–90, September 2006.
378. M. Ostendorf, V. Digalakis, and O. Kimball, "From HMMs to segment models: A unified view of stochastic modeling for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 5, pp. 360–378, 1996.
379. M. Ostendorf, "From HMMs to Segment Models," in *Automatic Speech and Speaker Recognition - Advanced Topics*. Kluwer Academic Publishers, 1996, ch. 8.
380. S. Oviatt, "Mutual disambiguation of recognition errors in a multimodal architecture," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, USA, 1999, pp. 576–583.
381. ——, "Ten myths of multimodal interaction," *Communications of the ACM*, vol. 42, no. 11, pp. 74–81, 1999.
382. ——, "Multimodal interfaces," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 2003, pp. 286–304.
383. S. Oviatt, R. Coulston, S. Tomko, B. Xiao, R. Lunsford, M. Wesson, and L. Carmichael, "Toward a theory of organized multimodal integration patterns during human-computer interaction," in *Proc. Int'l Conf. on Multimodal Interfaces*. ACM Press, New York, NY, USA, 2003, pp. 44–51.
384. O. K. Oyekoya and F. W. M. Stentiford, "Exploring human eye behaviour using a model of visual attention," in *Proc. Int'l Conf. on Pattern Recognition*, 2004, cambridge UK, August.
385. ——, "A performance comparison of eye tracking and mouse interfaces in a target image identification task," in *2nd European Workshop on the Integration of Knowledge, Semantics & Digital Media Technology (EWIMT)*, 2005, london, 30th Nov - 1st Dec, 2005.
386. O. Oyekoya, "Eye tracking: A perceptual interface for content based image retrieval," Ph.D. dissertation, University College London, UK, 2007.
387. L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," Computer Systems Laboratory, Stanford Univ., CA, Tech. Rep., 1998.
388. E. Pampalk, A. Rauber, and D. Merkl, "Content-based Organization and Visualization of Music Archives," in *Proc. ACM Int'l Conference on Multimedia*. Juan les Pins, France: ACM, December 1-6 2002, pp. 570–579.
389. G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos, "Multimodal fusion and learning with uncertain features applied to audiovisual speech recognition," in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, 2007, pp. 264–267.
390. W. Pasman and C. Woodward, "Implementation of an augmented reality system on a pda," in *Proc. IEEE/ACM Int'l Symposium on Mixed and Augmented Reality*, October 2003, pp. 276–277.
391. E. K. Patterson, S. Gurbuz, Z. Tufekci, and J. N. Gowdy, "CUAVE: A new audio-visual database for multimodal human-computer interface research," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2002.

392. J. Peng and D. R. Heisterkamp, "Kernel indexing for relevance feedback image retrieval," in *Proc. IEEE Int'l Conf. on Image Processing*, Barcelona, Spain, 2003.
393. A. Pentland, B. Moghaddam, and T. Starner, "View–based and modular eigenspaces for face recognition," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 1994, pp. 84–91.
394. A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *Int'l J. of Comp. Vis.*, vol. 18, no. 3, pp. 233–254, 1996.
395. M. Perakakis and A. Potamianos, "The effect of input mode on inactivity and interaction times of multimodal systems," in *Proc. Int'l Conf. on Multimodal Interfaces*. ACM New York, NY, USA, 2007, pp. 102–109.
396. ——, "A study in efficiency and modality usage in multimodal form filling systems," *IEEE Trans. Audio, Speech, and Language Processing*, 2008, to appear.
397. K. Perlin, "Quikwriting: continuous stylus-based text entry," in *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM Press, 1998, pp. 215–216.
398. E. Petajan, "Automatic lipreading to enhance speech recognition," Ph.D. dissertation, Univ. of Illinois, Urbana-Campaign, 1984.
399. E. G. Petrakis, K. Kontis, E. Voutsakis, and E. Milios, "Relevance feedback methods for logo and trademark image retrieval on the web," in *ACM Symposium on Applied Computing (ACM SAC'2006)*, Dijon, France, April 23-27 2006, pp. 1084–1088, special Track on Information Access and Retrieval (IAR).
400. E. G. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou, "X-similarity: Computing semantic similarity between concepts from different ontologies," *Journal of Digital Information Management*, vol. 4, no. 4, pp. 233–238, December 2006.
401. S. Pfeiffer, R. Lienhart, and W. Effelsberg, "Scene determination based on video and audio features," *Multimedia Tools and Applications*, vol. 15, pp. 59–81, 2001.
402. A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "A dynamic programming approach to speech/music discrimination of radio recordings," in *Proc. European Signal Processing Conference*, 2007.
403. F. Pitié, S.-A. Berrani, R. Dahyot, and A. Kokaram, "Off-line multiple object tracking using candidate selection and the viterbi algorithm," in *Proc. IEEE Int'l Conf. on Image Processing*, Genoa, Italy, 2005.
404. V. Pitsikalis, A. Katsamanis, G. Papandreou, and P. Maragos, "Adaptive multimodal fusion by uncertainty compensation," in *Proc. Int'l Conf. on Spoken Language Processing*, 2006, pp. 2458–2461.
405. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.
406. L. Pohjanheimo, H. Keränen, and H. Ailisto, "Implementing touchme paradigm with a mobile phone," in *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence*. New York, NY, USA: ACM Press, 2005, pp. 87–92.
407. J.-P. Poli and J. Carrive, "Modeling television schedules for television stream structuring, Singapour," in *Proceedings of ACM MultiMedia Modeling*, Jan. 2007, pp. 680–689.

408. P. Poller and V. Tschernomas, "Multimodal fission and media design," in *SmartKom: Foundations of Multimodal Dialogue Systems*, W. Wahlster, Ed. Springer-Verlag, New York, NY, 2006, pp. 379–400.
409. A. Potamianos, E. Fosler-Lussier, E. Ammicht, and M. Perakakis, "Information seeking spoken dialogue systems - Part II: Multimodal dialogue," *IEEE Trans. Multimedia*, vol. 9, no. 3, pp. 550–566, 2007.
410. A. Potamianos, H. Kuo, A. Pargellis, A. Saad, and Q. Zhou, "Design principles and tools for multimodal dialog systems," in *Proc. ESCA Workshop Interact. Dialog. Multi-Modal Syst.*, Kloster Irsee, Germany, Jun. 1999.
411. A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," *J. of the Acous. Soc. Am.*, vol. 99, no. 6, pp. 3795–3806, Jun 1996.
412. A. Potamianos, E. Sanchez-Soto, and K. Daoudi, "Stream weight computation for multi-stream classifiers," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2006.
413. G. Potamianos, C. Neti, G. Gravier, and A. Garg, "Automatic recognition of audio-visual speech: Recent progress and challenges," *Proc. IEEE*, vol. 91, no. 9, pp. 1306–1326, 2003.
414. G. Potamianos, C. Neti, G. Gravier, A. Garg, and A. W. Senior, "Recent advances in the automatic recognition of audio-visual speech," *Proc. IEEE*, vol. 91, no. 9, pp. 1–18, 2003.
415. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes*. Cambridge Univ. Press, 1992.
416. C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions of interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–982, 2000.
417. S. Quackenbush and A. Lindsay, "Overview of MPEG-7 audio," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 725–729, 2001.
418. E. R. Baeza-Yates, *Modern Information Retrieval*. Addison Wesley, 1999.
419. S. Raaijmakers, J. Den Hartog, and J. Baan, "Multimodal topic segmentation and classification of news video," in *Proc. Text Retrieval Conference*, vol. 2, 2002, pp. 33–36.
420. L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
421. L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. NJ, USA: Prentice-Hall, 1993.
422. K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis, and S. Kollias, *Signal Processing: Image Communication*, 2007, submitted for publication.
423. K. Rapantzikos and M. Zervakis, "Robust optical flow estimation in MPEG sequences," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, Mar 2005.
424. Z. Rasheed and M. Shah, "Scene detection in hollywood movies and tv shows," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 343–348.
425. K. Ratakonda, M. Sezan, and R. Crinon, "Hierarchical video summarization," in *Proc. SPIE, Visual Comm. and Image Proc.*, vol. 3653, Dec 1998, pp. 1531–1541.
426. A. Rauber and M. Frühwirth, "Automatically analyzing and organizing music archives," in *Proceedings of the 5th European Conference on Research and Ad-*

vanced Technology for Digital Libraries (ECDL'01), ser. LNCS. Darmstadt, Germany: Springer, September 4-8 2001, pp. 402–414.
427. A. Rauber, E. Pampalk, and D. Merkl, "Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles," in *Proc. Int'l Conf. on Music Information Retrieval*, Paris, France, October 13-17 2002, pp. 71–80.
428. M. Rautiainen et al., "TREC 2002 video track experiments at MediaTeam Oulu and VTT," in *Proc. Text Retrieval Conference*, 2002.
429. N. Rea, R. Dahyot, and A. Kokaram, "Semantic event detection in sports through motion understanding," in *Proc. ACM Int'l Conference on Image and Video Retrieval*, Dublin, Ireland, July 2004.
430. ——, "Classification and representation of semantic content in broadcast tennis videos," in *Proc. IEEE Int'l Conf. on Image Processing*, Genoa, Italy, 2005.
431. N. Rea, C. Lambe, G. Lacey, and R. Dahyot, "Multimodal periodicity analysis for illicit content detection in videos," in *IET European Conference on Visual Media Production (CVMP)*, London, UK, November 2006, pp. 106–114.
432. N. Rea, "High-level event detection in broadcast sports video," Ph.D. dissertation, Trinity College Dublin, 2005.
433. L. M. Reeves, J. C. Martin, M. McTear, T. V. Raman, K. M. Stanney, H. Su, Q. Y. Wang, J. Lai, J. A. Larson, and S. Oviatt, "Guidelines for multimodal user interface design," *Communications of the ACM*, vol. 47, no. 1, pp. 57–59, 2004.
434. G. Reitmayr and T. Drummond, "Going out: robust, model-based tracking for outdoor augmented reality," in *Proc. IEEE/ACM Int'l Symposium on Mixed and Augmented Reality*, 2006.
435. J. Rekimoto and M. Saitoh, "Augmented surfaces: a spatially continuous work space for hybrid computing environments," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*, 1999, pp. 378–385.
436. P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proc. Int'l Joint Conf. on Artificial Intel.*, C. S. Mellish, Ed. San Mateo: Morgan Kaufmann, Aug. 1995, pp. 448–453.
437. D. A. Robinson, "A method of measuring eye movement using a scleral search coil in a magnetic field," *IEEE Trans. Bio-Med. Electron.*, vol. BME-10, pp. 137–145, 1963.
438. J. T. Robinson, "The K-D-B-tree: A search structure for large multidimensional dynamic indexes," in *ACM SIGMOD International Conference on Management of Data*, Ann Arbor, Michigan, Etats-Unis, 29 Apr. - 1 May 1981, pp. 10–18.
439. C. Rocchi, O. Stock, M. Zancanaro, M. Kruppa, and A. Krüger, "The museum visit: generating seamless personalized presentations on multiple devices," in *Proc. Int'l Conf. on Intelligent User Interfaces*. New York, NY, USA: ACM Press, 2004, pp. 316–318.
440. J. Rocchio, "Relevance feedback in information retrieval," in *The SMART Retrieval System - Experiments in Automatic Document Processing*, G. Salton, Ed. Prentice Hall, Englewood Cliffs, 1971, pp. 313–323.
441. M. Rohs, "Marker-based embodied interaction for handheld augmented reality games," *Journal of Virtual Reality and Broadcasting*, vol. 4, no. 5, Mar. 2007, `urn:nbn:de:0009-6-7939`, ISSN 1860-2037.

442. R. C. Rose, E. M. Hofstetter, and D. A. Reynolds, "Integrated models of signal and background with application to speaker identification in noise," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 245–257, 1994.
443. A. Rosenfeld, D. Doermann, and D. D. (Editors), *Video Mining*. Springer, 2003.
444. A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, September 2003.
445. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer-Verlag, 2006.
446. L. Rothrock, R. Koubek, F. Fuchs, M. Haas, and G. Salvendy, "Review and reappraisal of adaptive interfaces: toward biologically inspired paradigms," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 1, pp. 47–84, 2002.
447. V. Roto, "Best practices and future visions for search user interfaces: A workshop," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*, 2003.
448. H. A. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes," *Proc. Advances in Neural Information Processing Systems*, vol. 8, pp. 875–881, 1996.
449. B. Rueber, "Obtaining confidence measures from sentence probabilities," in *Proc. European Conf. on Speech Communication and Technology*, 1997.
450. Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39–62, 1999.
451. Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 2, Washington, DC, 1997.
452. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, 1998.
453. B. Russell, *A History of Western Philosophy*. New York: Simon & Schuster, 1945.
454. S. Sagayama, K. Shinoda, M. Nakai, and H. Shimodaira, "Analytic methods for acoustic model adaptation: a review," in *Proc. of ISCA Workshop on Adaptation Methods*, Sophia-Antipolis France, 2001, pp. 67–76.
455. P. Salembier and J. R. Smith, "MPEG-7 multimedia description schemes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 748–759, 2001.
456. G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.
457. G. Salton and M. J. McGill, "The SMART and SIRE experimental retrieval systems," in *Readings in Information Retrieval*, K. S. Jones and P. Willett, Eds. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1997, pp. 381–399.
458. G. Salton, *Automatic Text Processing*. Addison-Wesley, 1989.
459. G. Salton, C. Yang, and C. Yu, "A Theory of Term Importance in Automatic Text Analysis," *Journal of the American Society for Information Science*, vol. 26, no. 1, pp. 33–44, 1975.
460. G. Salvendy, *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, New York, NY, USA, 2005.

461. D. D. Salvucci and G. J. H., "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium*. New York: ACM Press, 2000, pp. 71–78.
462. M. Sargin, Y. Yemez, E. Erzin, and A. Tekalp, "Audiovisual synchronization and fusion using canonical correlation analysis," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1396–1403, Nov. 2007.
463. E. Saykol, U. Gudukbay, and O. Ulusoy, "A histogram-based approach for object-based query-by-shape-and-color in multimedia databases," Bilkent University, Technical Report BUCE-0201, 2002.
464. L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," vol. 46, no. 3, pp. 647–654, 1998.
465. M. Schedl, P. Knees, and G. Widmer, "Discovering and visualizing prototypical artists by web-based co-occurrence analysis," in *Proc. Int'l Conf. on Music Information Retrieval*, London, UK, September 11-15 2005, pp. 21–28.
466. B. Schilit and M. Theimer, "Disseminating active map information to mobile hosts," *IEEE Netw.*, vol. 8, no. 5, pp. 22–32, 1994.
467. T. Schnell, T. Wu, "Applying eye tracking as alternative approach for activation of controls and functions in aircraft," in *Proceedings of the 5th International Conference On Human Interaction with Complex Systems (HICS)*, 2000, p. 113.
468. B. Schölkopf, "The kernel trick for distances," in *Proc. Advances in Neural Information Processing Systems*, vol. 12. MIT Press, 2000, pp. 301–307.
469. B. Schölkopf and A. Smola, *Learning with Kernels*. MIT Press, 2002.
470. R. B. Segal and J. O. Kephart, "Swiftfile: An intelligent assistant for organizing e-mail," in *Proc. of AAAI 2000 Spring Symposium on Adaptive User Interfaces*, 2000, pp. 107–112.
471. T. K. Sellis, N. Roussopoulos, and C. Faloutsos, "The R+-tree: A dynamic index for multi-dimensional objects," in *Proc. Int'l Conf. on Very Large Data Bases*, Brighton, Royaume-Uni, 1-4 Sep. 1987, pp. 507–518.
472. S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue, "GALAXY-II: A reference architecture for conversational system development," in *Proc. Int'l Conf. on Spoken Language Processing*, vol. 3, Sydney, Australia, 1998, pp. 931–934.
473. S. Seneff, R. Lau, and J. Polifroni, "Organization, communication, and control in the GALAXY-II conversational system," in *Proc. European Conf. on Speech Communication and Technology*, Budapest, Hungary, 1999, pp. 1271–1274.
474. S. Seneff, M. McCandless, and V. Zue, "Integrating natural language into the word graph search for simultaneous speech recognition and understanding," in *Proc. European Conf. on Speech Communication and Technology*, 1995.
475. SensAble, "Sensable technologies," http://www.sensable.com/.
476. W. A. Sethares and T. W. Staley, "Periodicity transforms," *IEEE Trans. Signal Process.*, vol. 47, no. 11, November 1999.
477. F. Seydoux and J.-C. Chappelier, "Semantic indexing using minimum redundancy cut in ontologies," in *Proc. of International Conference on Recent Advances in Natural Language Processing (RANLP 2005)*, September 2005, pp. 486–492.
478. A. Shaikh, S. Juth, A. Medl, I. Marsic, C. Kulikowski, and J. Flanagan, "An architecture for multimodal information fusion," in *Proc. of the Workshop on Perceptual User Interfaces*, Banf, Canada, 1997, pp. 91–93.

479. R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, 1998.
480. H.-T. Shen, B.-C. Ooi, and K.-L. Tan, "Giving meanings to www images," in *Proc. ACM Int'l Conference on Multimedia*, Marina del Rey, CA, 2000, pp. 39–47.
481. S. Sherr, *Input Devices*. Academic Press, Orlando, FL, USA, 1990.
482. J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 21-23 June 1994, pp. 593–600.
483. B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Longman Publishing Co., Boston, MA, USA, 1997.
484. ——, *Leonardo's Laptop: Human Needs and the New Computing Technologies*. Cambridge, MA, USA: MIT Press, 2002.
485. E. H. Shortliffe, *Computer-based medical consultation: MYCIN*. New York, NY: Elsevier, 1976.
486. C. L. Sidner, "Building spoken-language collaborative interface agents," in *Practical Spoken Dialog Systems*, D. A. Dahl, Ed. Kluwer Academic Publishers, 2004, pp. 197–226.
487. J. Sietsma and R. Dow, "Creating artificial neural networks that generalize," *Neural Networks*, vol. 4, pp. 67–79, 1991.
488. T. Sikora, "The MPEG-7 visual standard for content description – an overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, 2001.
489. M. Silfverberg, I. S. MacKenzie, and T. Kauppinen, "An isometric joystick as a pointing device for handheld information terminals," in *Proc. of Graphics Interface*, B. Watson and J. W. Buchanan, Eds., 2001, pp. 119–126.
490. S. Siltanen, M. Hakkarainen, O. Korkalo, T. Salonen, J. Sääski, C. Woodward, T. Kannetis, M. Perakakis, and A. Potamianos, "Multimodal user interface for augmented assembly," in *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, 2007.
491. S. Siltanen and J. Hyväkkä, "Implementing a natural user interface for camera phones using visual tags," in *Proceedings of the 7th Australasian User Interface Conference (AUIC2006)*, 2006, pp. 113 – 116.
492. J. Siroux, M. Guyomard, F. Multon, and C. Remondeau, "Oral and gestural activities of the users in the Georal system," in *International Conference on Cooperative Multimodal Communication*, vol. 2, Eindhoven, The Netherlands, 1995, pp. 287–298.
493. J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int'l Conf. on Computer Vision*, vol. 2, Oct. 2003, pp. 1470–1477.
494. M. Slaney and M. Covell, "FaceSync: A linear operator for measuring synchronization of video facial images and audio tracks," in *Proc. Advances in Neural Information Processing Systems*, 2001.
495. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.
496. J. R. Smith and S. F. Chang, "Visualseek: a fully automated content-based image query system," in *Proc. ACM Int'l Conference on Multimedia*. ACM Press New York, NY, USA, 1997, pp. 87–98.

497. J. R. Smith, S. Basu, C.-Y. Lin, M. R. Naphade, and B. Tseng, "Integrating features, models and semantics for content-based retrieval," in *Proc. Int'l Workshop on Multimedia Content-Based Indexing and Retrieval*, September 2001, pp. 95–98.
498. M. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding techniques," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 1997, p. 775.
499. R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 450–455, March 2005.
500. C. G. M. Snoek, M. Worring, J.-M. Geusebroek, D. C. Koelma, F. J. Seinstra, and A. W. M. Smeulders, "The semantic pathfinder: Using an authoring metaphor for generic multimedia indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1678–1689, October 2006.
501. C. G. Snoek and M. Worring, "Multimodal video indexing: A review of the state-of-the-art," *Multimedia Tools and Applications*, vol. 25, no. 1, pp. 5–35, January 2005.
502. M. Sonka, V. Hlavec, and R. Boyle, *Image Processing Analysis, and Machine Vision*. PWS Publishing, 1999, ch. 6 & 14.
503. K. Spärck Jones, S. Walker, and S. E. Robertson, "A Probabilistic Model of Information Retrieval: Development and Comparative Experiments - Part 1 and 2," *Information Processing and Management*, vol. 36, no. 6, pp. 779–840, 2000.
504. K. Stanney, S. Samman, L. Reeves, K. Hale, W. Buff, C. Bowers, B. Goldiez, D. Nicholson, and S. Lackey, "A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations," *International Journal of Human-Computer Interaction*, vol. 17, no. 2, pp. 229–257, 2004.
505. C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.
506. B. E. Stein and M. A. Meredith, *The Merging of the Senses*. MIT Press Cambridge, MA, 1993.
507. F. W. M. Stentiford, "Attention based similarity," *Pattern Recognition*, vol. 40, no. 3, pp. 771–783, 2006.
508. R. J. Sternberg, *Cognitive Psychology*, 4th ed. Thomson Wadsworth, 2006.
509. A. Stolcke, Y. König, and M. Weintraub, "Explicit word error minimization in N-best list rescoring," in *Proc. European Conf. on Speech Communication and Technology*, 1997.
510. D. Stork and M. Hennecke, Eds., *Speechreading by Humans and Machines*. Berlin, Germany: Springer, 1996.
511. J. Sturm, B. Cranen, F. Wang, J. Terken, and I. Bakx, "Effects of prolonged use on the usability of a multimodal form-filling interface," in *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*. Springer, The Netherlands, 2004, pp. 329–348.
512. B. Suhm and A. Waibel, "Towards better language models for spontaneous speech," in *Proc. Int'l Conf. on Spoken Language Processing*, 1994.
513. X. Sun and M. Kankanhalli, "Video summarization using R-sequences," *Real-time imaging*, vol. 6, no. 6, pp. 449–459, Dec 2000.

514. H. Sundaram and S. F. Chang, "Computable scenes and structures in films," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 482–491, 2002.
515. M. Suzuki, Y. Kajiura, A. Ito, and S. Makino, "Unsupervised language model adaptation based on automatic text collection from WWW," in *Proc. Int'l Conf. on Speech Communication and Technology*, 2006.
516. D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, no. 11-13, pp. 1191–1199, 1999.
517. L. Taycher, L. Cascia, and S. Sclaroff, "Image digestion and relevance feedback in the imagerover www search engine," in *Int'l Conf. on Visual Information Systems*, San Diego, Dec. 1997, pp. 85–94.
518. H. Teager and S. Teager, "Evidence of nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modelling*. Kluwer Academic, 1990, pp. 241–261.
519. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd ed. Acad. Press, 2006.
520. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*. Washington DC: Winston & Sons, 1977.
521. C. Tillmann and H. Ney, "Word triggers and the EM algorithm," in *Proc. of the Workshop Computational Natural Language Learning (CoNLL)*, 1997, pp. 117–124.
522. A. Tomkins, "Social and semantic structures in web search," in *Proc. IEEE/WIC/ACM International Conference on Web Intelligence*, Silicon Valley, CA, 2007.
523. S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. ACM Int'l Conference on Multimedia*. Ottawa, Canada: ACM Press, 2001, pp. 107–118.
524. S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," in *Proc. Int'l Conf. on Machine Learning*. Morgan Kaufmann, 2000, pp. 999–1006.
525. B. U. Töreyin, Y. Dedeoglu, and A. E. Çetin, "Hmm based falling person detection using both audio and video." in *ICCV Workshop on HCI*, 2005, pp. 211–220.
526. B. U. Toreyin, E. B. Soyer, I. Onaran, and A. E. Cetin, "Falling person detection using multi-sensor signal processing," *EURASIP Journal on Advances in Signal Processing*, 2007.
527. A. Treisman and G. Gelade, "A feature integration theory of attention," *Cognit. Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
528. B. Truong, C. Dorai, and S. Venkatesh, "New enhancements to cut, fade, and dissolve detection processes in video segmentation," in *Proc. ACM Int'l Conference on Multimedia*, 2000, pp. 219–227.
529. N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," in *Proc. ACM Int'l conference on Computer Graphics and Interactive Techniques*, 2004.
530. G. Tür, D. Hakkani-Tür, A. Stolcke, and E. Shriberg, "Integrating prosodic and lexical cues for automatic topic segmentation," *Computational Linguistics*, vol. 21, no. 1, pp. 31–57, 2001.
531. M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

532. M. Turunen, "Jaspis - a spoken dialogue architecture and its applications," Ph.D. dissertation, University of Tampere, Department of Information Studies, 2004.
533. M. Turunen and J. Hakulinen, "Jaspis 2 - an architecture for supporting distributed spoken dialogues," in *Proc. European Conf. on Speech Communication and Technology*, Geneva, Switzerland, 2003, pp. 1913–1916.
534. G. Tzanetakis and P. Cook, "MARSYAS: A framework for audio analysis," *Organized Sound*, vol. 4, no. 30, pp. 169–175, 2000.
535. ——, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, July 2002.
536. S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video manga: generating semantically meaningful video summaries," in *Proc. ACM Int'l Conference on Multimedia*, 1999, pp. 383–392.
537. M. Utiyama and H. Isahara, "A statistical model for domain-independent text segmentation," in *Proc. Annual Meeting of the Association for Computational Linguistics*, 2001, pp. 499–506.
538. A. Valli and J. Véronis, "Étiquetage grammatical de corpus oraux : problèmes et perspectives," *Revue française de linguistique appliquée*, vol. 4, no. 2, pp. 113–133, 1999.
539. J. R. Vallino, "Interactive Augmented Reality," Ph.D. dissertation, Department of Computer Science, University of Rochester, 1998.
540. V. Vapnik, *Statistical Learning Theory.* New York: Wiley-Interscience, 1998.
541. G. B. Varile and A. Zampolli, *Survey of the State of the Art in Human Language Technology.* Cambridge University Press, 1997.
542. A. Vassiliou, A. Salway, and D. Pitt, "Formalizing stories: sequences of events and state changes," in *Proc. of IEEE Int'l Conference on Multimedia and Expo*, vol. 1, 2004, pp. 587–590.
543. D. Vaufreydaz, M. Akbar, and J. Rouillard, "Internet documents: A rich source for spoken language modeling," in *Proc. IEEE Workshop Automatic Speech Recognition and Understanding*, 1999.
544. D. Vergyri, K. Kirchhoff, K. Duh, and A. Stolcke, "Morphology-based language modeling for arabic speech recognition," in *Proc. Int'l Conf. on Spoken Language Processing*, 2004.
545. C. Vertan and N. Boujemaa, "Upgrading color distributions for image retrieval: can we do better?" in *International Conference on Visual Information Systems (Visual2000)*, November 2000.
546. F. Vignoli and S. Pauws, "A music retrieval system based on user-driven similarity and its evaluation," in *Proc. Int'l Conf. on Music Information Retrieval*, London, UK, September 11-15 2005, pp. 272–279.
547. P. Viola and M. Jones, "Robust real-time face detection," *Int'l J. of Comp. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
548. M. Vlachos, P. Yu, and V. Castelli, "On periodicity detection and structural periodic similarity," in *SIAM International Conference on Data Mining*, 2005.
549. E. Voorhees, "Using WORDNET for Text Retrieval," in WORDNET*: An Electronic Lexical Database*, C. Fellbaum, Ed. The MIT Press, 1998.
550. E. Voutsakis, E. G. Petrakis, and E. Milios, "Intellisearch: Intelligent search for images and text on the web," in *Image Analysis and Recognition (ICIAR 2006)*, Povoa de Varzim, Portugal, Sept. 18-20 2006, pp. 697–708.

551. E. Voutsakis, E. Petrakis, and E. Milios, "Weighted link analysis for logo and trademark image retrieval on the web," in *Proc. of IEEE/WIC/ACM Int'l Conference on Web Intelligence*, Compiegne, France, Sept. 2005, pp. 581–585.
552. VTT, "Phonemouse demo software," http://www.vtt.fi/multimedia/.
553. W. Wahlster, "Towards symmetric multimodality: Fusion and fission of speech, gesture, and facial expression," in *KI : Advances in Artificial Intelligence: 26th Annual German Conference on AI.* Springer, 2003, pp. 1–18.
554. ——, *SmartKom: Foundations of Multimodal Dialogue Systems.* Springer-Verlag, New York, Secaucus, NJ, USA, 2006.
555. M. T. Wallace, G. E. Roberson, W. D. Hairston, B. E. Stein, J. W. Vaughan, and J. A. Schirillo, "Unifying multisensory signals across time and space," *Exp. Brain Research*, vol. 158, pp. 252–258, 2004.
556. F. Wang, Y.-F. Ma, H.-J. Zhang, and J.-T. Li, "A generic framework for semantic sports video analysis using dynamic Bayesian networks," in *International Multimedia Modelling Conference*, 2005.
557. J. Z. Wang, J. Li, G. Wiederhold, and O. Firschein, "System for screening objectionable images using Daubechies' wavelets and color histograms," in *International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*, 1997, pp. 20–30.
558. J. Wang, S. Zhai, and J. Canny, "Camera phone based motion sensing: interaction techniques, applications and performance study," in *Proc. ACM Symposium on User Interface Software and Technology*, 2006.
559. Y. Wang, Z. Liu, and J.-C. Huang, "Multimedia content analysis-using both audio and visual clues," *IEEE Signal Process. Mag.*, vol. 17, no. 6, pp. 12–36, Nov. 2000.
560. D. J. Ward and D. J. C. MacKay, "Fast hands-free writing by gaze direction," *Nature*, vol. 418, p. 838, 2002.
561. D. J. Ward, "Adaptive computer interfaces," Ph.D. dissertation, University of Cambridge, 2001.
562. C. Ware and H. Mikaelian, "An evaluation of an eye tracker as a device for computer input," in *Proc. ACM-SIGCHI conference on Human factors in computing systems*, 1987.
563. M. Weintraub, Y. Aksu, S. Dharanipragada, S. Khudanpur, H. Ney, J. Prange, A. Stolcke, F. Jelinek, and E. Shriberg, "LM95 project report: Fast training and portability," Center for Language and Speech Processing, Johns Hopkins University, Tech. Rep., 1996.
564. D. A. White and R. Jain, "Similarity indexing with the SS-tree," in *12th International Conference on Data Engineering*, 26 Feb. - 1 Mar. 1996, pp. 516–523.
565. C. D. Wickens and H. J. G., *Engineering Psychology and Human Performance.* Prentice Hall, NJ, 2000.
566. J. Wilkinson and B. Devlin, "The material exchange format (mxf) and its application," *SMPTE journal*, vol. 111, no. 9, pp. 378–384, 2002.
567. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann, Academic Press, 2000, ch. 4.
568. W. Wolf, "Hidden markov model parsing of video programs," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 1997, pp. 2609–2611.
569. L. Wu, K. Sycara, T. Payne, and C. Faloutsos, "FALCON: Feedback adaptive loop for content-based retrieval," in *Proc. Int'l Conf. on Very Large Data Bases*, Cairo, Egypt, Sept. 2000, pp. 297–306.

570. Z. Wu and M. Palmer, "Verb semantics and lexical selection," in *Proc. Annual Meeting of the Association for Computational Linguistics*, New Mexico State University, Las Cruces, New Mexico, 1994, pp. 133–138.
571. L. Xie, P. Xu, S.-F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with domain knowledge and hidden Markov models," *Pattern Recognition Letters*, vol. 25, no. 7, pp. 767–775, 2004.
572. Y. Yaşaroğlu and A. A. Alatan, "Summarizing video: content, features & hmm topologies," in *Proc. Int. Workshop Very Low Bitrate Video Coding*, 2003, pp. 101–110.
573. M. M. Yeung and B.-L. Yeo, "Video visualization for compact presentation and fast browsing of picturial content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 771–785, 1997.
574. L. Ying, S.-H. Lee, C.-H. Yeh, and C.-C. Kuo, "Techniques for movie content analysis and skimming," in *IEEE Signal Process. Mag.*, vol. 23, no. 2, Mar 2006, pp. 79–89.
575. L. Ying, S. Narayanan, and C. Kuo, "Content-based movie analysis and indexing based on audiovisual cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1073–1085, Aug. 2004.
576. M. Ylikerälä and H. Kuukkanen, "Pluggable 3D stereographics," in *Articles on Experiences 4 - Digital Media & Game. Kylänen, Mika (Ed.). Lapland Centre of Expertise for the Experience Industry (LCEEI).*, 2006, pp. 168 – 176.
577. N. B. Yoma, F. McInnes, and M. Jack, "Weighted matching algorithms and reliability in noise canceling by spectral subtraction," in *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, vol. 2, 1997, pp. 1171–1174.
578. N. Yoma and M. Villar, "Speaker verification in noise using a stochastic version of the weighted viterbi algorithm," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 3, pp. 158–166, 2002.
579. S. Young, "A review of large-vocabulary continuous-speech," *IEEE Signal Process. Mag.*, vol. 13, no. 5, pp. 45–57, 1996.
580. S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, "The HTK book (for HTK version 3.2)," Cambridge University Engineering Department, Tech. Rep., 2002.
581. S. Young, N. Russell, and J. Thornton, "Token passing: A simple conceptual model for connected speech recognition systems," Cambridge University Engineering Dept, Tech. Rep. CUED/F-INFENG/TR38, 1989.
582. A. L. Yuille, "Energy functions for early vision and analog networks," *Biological Cybernetics*, vol. 61, pp. 115–123, 1989.
583. A. L. Yuille and H. H. Bülthoff, *Perception as Bayesian Inference*. Cambridge University Press, 1996, ch. Bayesian Decision Theory and Psychophysics, pp. 123–161.
584. J. Zacks and B. Tversky, "Event structure in perception and conception," *Psychological Bulletin*, no. 127, pp. 3–21, 2001.
585. J. M. Zacks, T. S. Braver, M. A. Sheridan, D. I. Donaldson, A. Z. Snyder, J. M. Ollinger, R. L. Buckner, and M. E. Raichle, "Human brain activity time-locked to perceptual event boundaries," *Nature Neuroscience*, vol. 4, no. 6, pp. 651–655, June 2001.
586. S. Zhai and P. Milgram, *Quantifying Coordination in Multiple DOF Movement and its Application to Evaluating 6 DOF Input Devices*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1998.

587. Y. Zhai, Z. Rasheed, and M. Shah, "Semantic classification of movie scenes using finite state machines," *IEE Proc. - Vision, Image, and Signal Processing*, vol. 152, no. 6, pp. 896–901, 2005.
588. D. Zhang, D. Gatica-Perez, S. Bengio, I. McCowan, and G. Lathoud, "Modeling individual and group actions in meetings: A two-layer HMM framework," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition Workshop*, vol. 7, 2004, pp. 117–124.
589. H.-J. Zhang, Z. Chen, W.-Y. Liu, and M. Li, "Relevance feedback in content-based image search," in *Proc. 12th Int'l Conf. on New Information Technology (NIT)*, Beijing, China, Aug. 2003, pp. 29–31, (invited keynote).
590. H.-J. Zhang and Z. Su, "Improving CBIR by semantic propagation and cross-mode query expansion," in *Proc. Int'l Workshop on Multimedia Content-Based Indexing and Retrieval*, September 2001, pp. 83–86.
591. T. Zhang and C. C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 4, pp. 441–457, 2001.
592. R. Zhao and W. Grosky, "Narrowing the semantic gapimproved text-based web document retrieval using visual features," *IEEE Trans. Multimedia*, vol. 4, no. 2, pp. 189–200, June 2002.
593. X. S. Zhou and T. S. Huang, "Unifying keywords and visual contents in image retrieval," *IEEE Multimedia*, vol. 9, no. 2, pp. 23–33, 2002.
594. ——, "Relevance feedback for image retrieval: a comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, 2003.
595. Z. Zhu and T. S. Huang, *Multimodal Surveillance: Sensors, Algorithms, and Systems*. Artech House Publishers, Jul. 2007.
596. Y. Zhuang, Y. Rui, T. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering," in *Proc. IEEE Int'l Conf. on Image Processing*, 1998, pp. 866–870.
597. X. Zou and B. Bhanu, "Tracking humans using multi-modal fusion," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition Workshop*. Washington, DC, USA: IEEE Computer Society, 2005, p. 4.
598. E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, 2nd ed., ser. Series of Information Sciences. Berlin: Springer, 1999, vol. 22.

Index

1D projection 137
3D interfaces 56

augmented reality 61

acoustic model 57
  adaptation 71
acoustic models 59
action classification 135, 136
action recognition 127, 128, 142
action spotting 142
Active Appearance Models 120
active learning 229
adaptable interfaces 53, 69, 70
adaptation 68, 295
adaptive interfaces 68
adaptive multimedia 80
ambient intelligence 68
angular kernel 231
annotated image 261
Applications 43
  Audio-Visual ASR 43
  Biometrics 46
  Broadcast News 45
  images 46
  meetings 47
  sports 44
  TV Structuring 45
AROnSite 324
asycnrhonous HMM 92
Audio 132
audio descriptors *see* audio feature extraction
audio feature extraction 245
  critical bands 245
  Rhythm Patterns 245
  Statistical Spectrum Descriptors 245
audio features 23, 179
  generic/specific 23
  MFCC 24
  pitch 24
  short-term energy 24
  short/long-term 23
  speaking rate 25
  ZCR 24
audio processing 179
audio similarity perception 242
audio test collections 246
audiovisual 3, 18
audiovisual automatic speech recognition 111
audiovisual fusion 179
audiovisual integration 91, 97
audiovisual saliency 179
augmented reality 311, 314, 320, 322
authorities 265
autocorrelation 132
automatic speech recognition 201
Automatic Speech Recognition (ASR) *see* speech recognition

background subtraction 145
Baum-Welch 136
Baum-Welsh 139
Bayesian 3, 137
Bayesian Belief Networks (BBN) 70
Bayesian decision theory 8

Bayesian estimation 8
Bayesian inference 10
biometric identification 68

CBIR 222
Christmas Carols 242
classification feature 148
click to talk 288
clustering 325
co-citation analysis 265
cognition 7
cognitive psychology 8
collaborative
  authoring 80
  filtering 72, 78
collection analysis module 273
color histogram 130
common meaning representation 65
compositionality 288
computational attention 179
conceptual content 221
conceptual feature vector 223
conceptual similarity 224
confidence measures 202, 211
consensus decoding 209
consistency 286, 295
content based
  image retrieval 78
Content Based Image Retrieval 222
content-based image retrieval 261
content-case analysis 128
context aware 318
context awareness 74
context-free grammar 142
crawler module 272
cricket 130
cross-modal integration 3
curl 138

desktop metaphor 55
Dialogue density 163
dialogue manager 282
Dialogue Manager (DM) 57
Dialogue scene 159
Dialogue velocity 163
direct manipulation 55
direct parsing 128
discourse markers 213
discrete cosine transform 148
dynamic bayesian networks 107
dysvideo 136

electronic tag 319
energy spectrum 263
Euclidean distance 244
event detection 179
Expectation-Maximization 118
eye tracking 60

fast fourier transform 148
feature extraction 148
feature structures 64
features 5
fight detection 149
Film Syntax 158
Finite State Machine (FSM) 57
form-filling 288
fusion 3, 19, 111, 153

gestalt psychology 8
gesture based application 322
gesture based interaction 316, 317, 322
gesture interfaces 60
global motion estimation 130
GOMS model 52
Graphical User Interfaces *see* GUI
GUI 50, 55, 58, 62, 81, 83
  toolkits 83

handwriting recognition 135
haptic 316
haptic interfaces 61
head mounted display 314
Head Mounted Displays (HMD) 61
hidden markov model 127
Hidden Markov Model (HMM) 57
hierarchical HMM 103
HITS 260
HMD *see* head mounted display
HMM 127, 134–136, 139
  continuous 135, 137
  discrete 135
  topology 140
Hough transform 129
hubs 265
Human Computer Interaction (HCI) 50
Human Model Processor (HMP) 51

hybrid descriptor 221
hybrid image search 223
hybrid retrieval systems 260
hypernym graph 225
hypernymy 224

illicit content 133
illicit content analysis 128
image analysis 261
image content 260, 261
image retrieval 259
image retrieval on the Web 260
information retrieval 212, 259
intensity histogram 262
interactive image retrieval 221
Interactive Voice Response (IVR) 59
interface guidelines 83
intermodal 5, 16
intramodal 5, 16

key concepts 223
key-frame selection 179

language model 57, 202
  adaptation 71
language model adaptation 212, 214, 215
language model interpolation 214
Laplace kernel 231
layered HMM 92
lexical cohesion 213
link analysis 260
Logical story unit 159
logo 261
logo-trademark detection 263
logo-trademark similarity 263

map based user interfaces 244
Markov Decision Process (MDP) 73
Maximum Likelihood 12
Maximum-A-Posteriori 12
McGurk effect 14
measurement noise compensation 111
mixed-initiative 284
mobile augmented reality *see* augmented reality
mobile interfaces 74
modality efficiency 291
modality synergy 280, 293
modality-selection 288
model based parsing 128, 134
Model-View-Controller *see* MVC
moment invariants 263
monomodal 3
morpho-syntactic knowledge 206
morpho-syntactic rescoring 208
morphology 205
most ambiguous and orthogonal examples 232
most ambiguous examples 232
motion estimation 315, 317
Motion Field 131
motion trajectories 131
MPEG 131
MPEG-7 23
MPEG-7 Standard 157
mulitstream models 92
multi-agent architectures 81
multi-touch 56
multi-touch screen 313, 316
multicue 5
multimedia 3
multimedia fission 62
multimedia maps 79
multimedia retrieval 70
multimedia skimming 79
multimedia summarization 79
multimedia systems
  adaptive 70
multimodal 3, 5
  architectures 81
  frameworks 82
  integration pattern 67
  mutual disambiguation 67
  synergistic error correction 67
multimodal dialogue systems 279
multimodal fusion 62, 63, 92
multimodal interfaces 62, 279
multimodal processing 179
multimodal synergy 62
multistream HMM 92
music information retrieval 243
  content-based 243
  cultural data 243
  song lyrics 242, 243
    lyrics fetching 246
music map 325
music maps 326

MVC 52, 72
  paradigm 53, 54, 66, 80
MVC paradigm 281

n-class model 202
n-gram model 202
Natural Language Generation (NLG) 57
natural language processing 201
Natural Language Understanding (NLU) 57
NFC 319

object actions 131
object classification 148
ontology 224
open-mike 288

PageRank 260
parsing 135
pca 130
perception 3, 7
periodicity 132
philosophy 8
PhoneMouse 322
physical browsing 319
PicASHOW 261
PlaySOM 325
PocketSOM 325
pornographic 127, 132
pragmatics 205
Principal Component Analysis 130

queries by example 260
queries by example image 260
query expansion 269
query focused graph 266
query point movement 269
query uncertainty 260

RBF kernel 231
relevance feedback 71, 73, 229, 260, 268
retrieval by image annotations 260
retrieval by image content 260
retrieval module 273
RFID 319
rotation 138

Scenes 158
score-oriented Viterbi search 104
segment models 91, 98
segmentation 145
Self-Organizing Map 243
  adaptation function 244
  BMU selection 244
  multimodal 249
  quantitative evaluation 249
  visualization 244
semantic
  annotation 78
  fusion 63, 64
  gap 78
  multimedia understanding 78
  parsing
    robust 57
  unification 65
  web 82
semantic gap 222, 242
semantic persistence 286
semantic projection 228
semantic similarity 224, 261
Semantic Similarity Retrieval Model 264
semantics 205
sensation 7
Shot accuracy 161
Shots 158
shout detection 150
silhouette 148
similarity adaptation 269
similarity measure 225
skin 131, 137
smart mobile terminal 313
snooker 129, 130, 135, 141
SOM based user interfaces *see* map based user interfaces
speech dictation 59
speech interfaces 58
speech recognition 56, 57
speech synthesis 57
Spoken Dialogue Systems (SDS) 57, 59
sport action tracking 130
sport analysis 128
sport media analysis 127
SSRM 264
state machine 135
stereo display 314

Stereogames 323
storage module 273
Story unit 159
stream weights 112, 115
strong fusion 22
SVM scale invariance 231
SymBall 322
symmetric multimodality 287
synergistic error correction 288
syntactics 205

tactile 316
tagging, ASR transcripts 207
tagging, part-of-speech 206
talking heads 68
tennis 129, 141
TEOCEP 150
term re-weighting 269
text features 246
  $tf \times idf$ model 246
  bag-of-words models 246
  term weighing 246
text segmentation 213, 214
text-to-speech synthesis (TTS) 58
topic adaptation 212
topic segmentation 107
touchscreen 326
trademark 261
trajectory 135
turn-taking 285

unification integration 64
usability 51, 52
usability principles 52
  consistency 53
  familiarity 53
  predictability 53
  transparency 53
user model 69
  adaptation 69
  application 69
user-initiative 284

Vector Space Model 264
video annotation 179
video processing 179
Video shot string 165
video structuring 91
video summarization 179
view classification 129
virtual reality 61
visemes 68
visual dominance effect 51
visual marker 321
visual tag 319
Viterbi 130, 137–139
voice browser 84
VoiceXML 83

wavelet 148
weak fusion 20
Weighted PicASHOW 267
WIMP interface 55
word error minimization 209
WordNet 264
Wordnet 224
World Wide Web 259
WPicASHOW 267
WWW 259
WYSIWYG 55

Zooming User Interfaces (ZUI) 56