References

[1] T. Aach and A. Kaup. Statistical model-based change detection in moving video. *Signal Processing*, 31:165–180, 1993.

[2] T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, 7:147–160, 1995.

[3] T. Aach, A. Kaup, and R. Mester. Change detection in image sequences using gibbs random fields: A bayesian approach. In *International Workshop on Intelligent Signal Processing and Communication Systems, ISPACS*, pages 56–61, Oct. 1993.

[4] S. Abbasi and F. Mokhtarian. Shape similarity retrieval under affine transform: application to multi-view object representation and recognition. In *Proc. Seventh IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 450–455, Sept. 1999.

[5] L. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45(2):107–127, 2001.

[6] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora. Image sequence analysis for emerging interactive multimedia services - the european cost 211 framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(7):802–813, Nov. 1998.

[7] B. Appleton and C. Sun. Circular shortest paths by branch and bound. *Pattern Recognition*, 36(11):2513–2520, Nov. 2003.

[8] A. Bartoli, N. Dalal, B. Bose, and R. Horaud. From video sequences to motion panoramas. In *Proc. Workshop on Motion and Video Computing*, pages 201–207, Dec. 2002.

[9] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B 48*, pages 259–302, 1986.

[10] S. Birringer. Inexaktes Teil-Graph-Matching für die Suche von Videoobjekten mit Hilfe evolutionärer Algorithmen. Studienarbeit, Universität Mannheim, Mar. 2003.

[11] A. Blake and M. Isard. *Active Contours*. Springer Verlag, 1998.

[12] T. Brox, D. Farin, and P. H. N. de With. Multi-stage region merging for image segmentation. In *22nd Symposium on Information Theory in the Benelux*, pages 189–196, May 2001.

[13] L. Bruzzone and R. Cossu. An adaptive approach to reducing registration noise effects in unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 41:2455–2465, 2003.

[14] L. Bruzzone and D. Fernandez Prieto. An mrf approach to unsupervised change detection. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 143–147, 1999.

[15] L. Bruzzone and D. Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38:1171–1182, 2000.

[16] C. Calvo, A. Micarelli, and E. Sangineto. Automatic annotation of tennis video sequences. In *DAGM-Symposium*, pages 540–547. Springer, 2002.

[17] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *CVPR'97 Workshop on Content-Based Access of Image and Video Libraries*, pages 42–49, June 1997.

[18] A. Cavallaro and T. Ebrahimi. Video object extraction based on adaptive background and statistical change detection. In *Proc. of SPIE VCIP*, pages 465–475, Jan. 2000.

[19] A. Cavallaro and T. Ebrahimi. Classification of change detection algorithms for object-based applications. In *Proc. of Workshop on Image Analysis For Multimedia Interactive Services (WIAMIS-2003)*, Apr. 2003.

[20] A. Cavallaro, E. Salvador, and T. Ebrahimi. Shadow-aware object-based video processing. *IEE Vision, Image and Signal Processing*, to appear.

[21] I. Celasun and A. M. Tekalp. Optimal 2-d hierarchical content-based mesh design and update for object-based video. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(7):1135–1153, Oct. 2000.

[22] A. Chakraborty. Video structuring for multimedia applications. In *SPIE Proc. of Visual Communication and Image Processing*, pages 496–507, 2000.

[23] Y. Chen and J. Z. Wang. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1252–1267, 2002.

[24] S.-Y. Chien, C.-Y. Chen, W.-M. Chao, C.-W. Hsu, Y.-W. Huang, and L.-G. Chen. A fast and high subjective quality sprite generation algorithm with frame skipping and multiple sprites techniques. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 193–196, Sept. 2002.

[25] S.-Y. Chien, C.-Y. Chen, Y.-W. Huang, and L.-G. Chen. Multiple sprites and frame skipping techniques for sprite generation with high subjective quality and fast speed. In *Proc. IEEE International Conference Multimedia and Expo (ICME)*, pages 785–788, 2002.

[26] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 264–271. IEEE Computer Society, December 2001.

[27] Y.-Y. Chuang, D. B. Goldman, B. Curless, D. H. Salesin, and R. Szeliski. Shadow matting and compositing. *ACM Trans. Graph.*, 22(3):494–500, 2003.

[28] J. Clarke, S. Carlsson, and A. Zisserman. Detecting and tracking linear features efficiently. In R. B. Fisher and E. Trucco, editors, *Proc. 7th British Machine Vision Conf. (BMVA), Edinburgh*, pages 415–424, 1996.

[29] D. Connor and J. Limb. Properties of frame-difference signals generated by moving images. *IEEE Transactions on Communications*, 22:1564–1575, 1974.

[30] S. Coorg and S. Teller. Spherical mosaics with quaternions and dense correlation. *International Journal on Computer Vision*, 37(3):259–273, 2000.

[31] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Press, 1990.

[32] J. Costeira and T. Kanade. A multibody factorization method for independent moving objects. *International Journal on Computer Vision*, 29(3), September 1998.

[33] X. Dai and S. Khorram. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 36:1566–1577, 1998.

[34] J. Davis. Mosaics of scenes with moving objects. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 354–360, June 1998.

[35] G. de Haan and P. W. A. C. Biezen. An efficient true-motion estimator using candidate vectors from a parametric motion model. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(1):85–91, Feb. 1998.

[36] P. H. N. de With. A simple recursive motion estimation technique for compression of hdtv signals. In *Proc. International Conference on Image Processing and its Applications*, pages 417–420, Apr. 1992.

[37] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 11–20, New York, NY, USA, 1996. ACM Press.

[38] N. D. Doulamis, A. D. Doulamis, Y. S. Avrithis, and S. D. Kollias. Video content representation using optimal extraction of frames and scenes. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 875–879, 1998.

[39] F. Dufaux and J. Konrad. Efficient, robust, and fast global motion estimation for video coding. *IEEE Transactions on Image Processing*, 9(3), Mar. 2000.

[40] F. Dufaux and F. Moscheni. Background mosaicking for low bit rate video coding. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 673–676, 1996.

[41] E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra. *Proceedings of the IEEE*, 89:1368–1381, 2001.

[42] A. Ekin and A. M. Tekalp. Automatic soccer video analysis and summarization. In *SPIE Storage and Retrieval for Media Databases IV*, pages 339–350, Jan. 2003.

[43] D. Eppstein. Subgraph isomorphism in planar graphs and related problems. Technical report, Dept. of Information and Computer Science, University of California, May 1994.

[44] P. E. Eren and A. M. Tekalp. Bi-directional 2-D mesh representation for video object rendering, editing and superresolution in the presence of occlusion. *Signal Processing: Image Communication*, 18(5):321–336, May 2003.

[45] R. Fablet, P. Bouthemy, and M. Gelgon. Moving object detection in color image sequences using region-level graph labeling. In 6th *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 939–943, Oct. 1999.

[46] D. Farin and P. H. N. de With. Towards real-time MPEG-4 segmentation: A fast implementation of region-merging. In *21st Symposium on Information Theory in the Benelux*, pages 173–180, May 2000.

[47] D. Farin and P. H. N. de With. A new similarity measure for subpixel accurate motion analysis in object-based coding. In 5th *World Multi-Conference on Systemics, Cybernetics and Informatics (SCI)*, pages 244–249, July 2001.

[48] D. Farin and P. H. N. de With. Estimating physical camera parameters for 3DAV video coding. In 25th *Symposium on Information Theory in the Benelux*, pages 201–208, June 2004.

[49] D. Farin and P. H. N. de With. Estimating physical camera parameters based on multi-sprite motion estimation. In *SPIE Image and Video Communications and Processing, Vol. 5685*, volume 5685, pages 489–500, Jan. 2005.

[50] D. Farin and P. H. N. de With. Evaluation of a feature-based global-motion estimation system. In *SPIE Visual Communications and Image Processing*, pages 1331–1342, July 2005.

[51] D. Farin and P. H. N. de With. Misregistration errors in change detection algorithms and how to avoid them. In *Proc. IEEE International Conference on Image Processing ICIP*, volume 2, pages 438–441, Sept. 2005.

[52] D. Farin and P. H. N. de With. Reconstructing virtual rooms from panoramic images. In *26th Symposium on Information Theory in the Benelux*, pages 301–308, May 2005.

[53] D. Farin and P. H. N. de With. Automatic video-object segmentation employing multi-sprites with constrained delay. In *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2006.

[54] D. Farin and P. H. N. de With. Enabling arbitrary rotational camera-motion using multi-sprites with minimum coding-cost. *IEEE Transactions on Circuits and Systems for Video Technology*, accepted for publication.

[55] D. Farin, P. H. N. de With, and W. Effelsberg. Optimal partitioning of video sequences for MPEG-4 sprite encoding. In *24th Symposium on Information Theory in the Benelux*, pages 79–86, May 2003.

[56] D. Farin, P. H. N. de With, and W. Effelsberg. Recognition of user-defined video object models using weighted graph homomorphisms. In *SPIE Image and Video Communications and Processing (IVCP)*, volume 5022, pages 542–553, Jan. 2003.

[57] D. Farin, P. H. N. de With, and W. Effelsberg. Robust background estimation for complex video sequences. In *International Conference on Image Processing (ICIP)*, volume 1, pages 145–148, Sept. 2003.

[58] D. Farin, P. H. N. de With, and W. Effelsberg. A segmentation system with model assisted completion of video objects. In *SPIE Visual Communications and Image Processing (VCIP)*, volume 5150, pages 366–377, July 2003.

[59] D. Farin, P. H. N. de With, and W. Effelsberg. Minimizing MPEG-4 sprite coding-cost using multi-sprites. In *SPIE Visual Communications and Image Processing (VCIP)*, volume 5308, pages 234–245, Jan. 2004.

[60] D. Farin, P. H. N. de With, and W. Effelsberg. Video-object segmentation using multi-sprite background subtraction. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 1, pages 343–346, June 2004.

[61] D. Farin, W. Effelsberg, and P. H. N. de With. Robust clustering-based video-summarization with integration of domain-knowledge. In *International Conference on Multimedia and Expo (ICME)*, volume 1, pages 89–92, Aug. 2002.

[62] D. Farin, T. Haenselmann, S. Richter, G. Kühne, and W. Effelsberg. Segmentation and classification of moving video objects. In B. Furht and O. Marques, editors, *Handbook of Video Databases*, pages 561–591. CRC Press, Sept. 2003.

[63] D. Farin, J. Han, and P. H. N. de With. Fast camera calibration for the analysis of sport sequences. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, July 2005.

[64] D. Farin, M. Käsemann, P. H. N. de With, and W. Effelsberg. Rate-distortion optimal adaptive quantization and coefficient thresholding for MPEG coding. In *23rd Symposium on Information Theory in the Benelux*, pages 131–138, May 2002.

[65] D. Farin, S. Krabbe, W. Effelsberg, and P. H. N. de With. Robust camera calibration for sport videos using court models. In *SPIE Storage and Retrieval Methods and Applications for Multimedia*, volume 5307, pages 80–91, Jan. 2004.

[66] D. Farin, N. Mache, and P. H. N. de With. SAMPEG, a scene adaptive parallel MPEG-2 software encoder. In *SPIE Visual Communications and Image Processing (VCIP)*, volume 4310, pages 272–283, Jan. 2001.

[67] D. Farin, N. Mache, and P. H. N. de With. A software-based high-quality MPEG-2 encoder employing scene change detection and adaptive quantization. *IEEE Transactions on Consumer Electronics*, 48:887–897, Nov. 2002.

[68] D. Farin, M. Pfeffer, P. H. N. de With, and W. Effelsberg. Corridor scissors: A semi-automatic segmentation tool employing minimum-cost circular paths. In *International Conference on Image Processing (ICIP)*, pages 1177–1180, Oct. 2004.

[69] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

[70] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient matching of pictorial structures. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 66–73, June 2000.

[71] S. Finke. Robustes Tracking von kleinen Objekten unter Berücksichtigung von Überdeckungen. Diplomarbeit, Universität Mannheim, Jan. 2004.

[72] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV*, pages 823–836, London, UK, 2002. Springer-Verlag.

[73] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[74] D. Geiger, A. Gupta, L. A. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking, and matching deformable contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 17(3):294–302, 1995.

[75] T. Gleixner. Alpha-Kanal-Schätzung aus Einzelbildern. Studienarbeit, Universität Mannheim, Sept. 2003.

[76] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18:377–388, Apr. 1996.

[77] A. V. Goldberg and R. E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM*, 35(4):921–940, 1988.

[78] Y. Gong and X. Liu. Video summarization using singular value decomposition. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 174–180, June 2000.

[79] H. Greenspan, G. Dvir, and Y. Rubner. Region correspondence for image matching via EMD flow. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 27–31, June 2000.

[80] T. Haenselmann and W. Effelsberg. Wavelet based semi-automatic live-wire segmentation. In *SPIE Human Vision and Electronic Imaging VIII*, pages 260–269, January 2003.

[81] J. Han, D. Farin, and P. H. N. de With. Multi-level analysis of sports video sequences. In *Visual Communications and Image Processing (VCIP)*, Jan. 2006.

[82] J. Han, D. Farin, P. H. N. de With, and W. Lao. Automatic tracking method for sports video analysis. In *26th Symposium on Information Theory in the Benelux*, pages 309–316, May 2005.

[83] K. Haris, S. N. Efstratiadis, and N. Maglaveras. Watershed-based image segmentation with fast region merging. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 338–342, 1998.

[84] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of The Fourth Alvey Vision Conference, Manchester*, pages 147–151, 1988.

[85] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[86] R. I. Hartley. Self-calibration from multiple views with a rotating camera. In *ECCV '94: Proceedings of the third European conference on Computer vision (vol. 1)*, pages 471–478. Springer-Verlag New York, Inc., 1994.

[87] R. I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23, 1997.

[88] J. Hayet, J. Piater, and J. Verly. Robust incremental rectification of sports video sequences. In *Proc. British Machine Vision Conference (BMVC)*, Kingston (UK), 2004.

[89] J.-B. Hayet, J. H. Piater, and J. G. Verly. Fast 2D model-to-image registration using vanishing points for sports video analysis. In *Proc. IEEE International Conference on Image Processing ICIP*, volume 3, pages 417–420, Sept. 2005.

[90] P. S. Heckbert. Fundamentals of texture mapping and image warping. Master's thesis, Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720, June 1989.

[91] J. Hornegger and C. Tomasi. Representation issues in the ml estimation of camera motion. In *IEEE International Conference on Computer Vision (ICCV)*, pages 640–647, 1999.

[92] M. Irani and P. Anandan. All about direct methods. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and practice*. Springer-Verlag, 1999.

[93] ISO/IEC 14496-2, International Standard: Information technology - coding of audio-visual objects - part 2: visual.

[94] ISO/IEC 14496-5, International Standard: Information technology - coding of audio-visual objects - part 5: reference software.

[95] S. Iwase and H. Saito. Tracking soccer players based on homography among multiple views. In *Visual Communications and Image Processing (VCIP) 2003*, volume 5150, pages 283–292, July 2003.

[96] B. Jähne. *Digital Image Processing*. Springer Verlag, 2005.

[97] K. Jinzenji, H. Watanabe, S. Okada, and N. Kobayashi. MPEG-4 very low bit-rate video compression using sprite coding. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, page 2, Aug. 2001.

[98] S. Kamijo, K. Ikeuchi, and M. Sakauchi. Segmentations of spatio-temporal images by spatio-temporal markov random field model. In *Proc. of Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 298–313, 2001.

[99] K. Kanatani. Computational projective geometry. *CVGIP: Image Understanding*, 54(3):333–348, 1991.

[100] J. J. Kanski. *Klinische Ophthalmologie*. Urban and Fischer at Elsevier, 2005.

[101] T. Kasetkasem and P. Varshney. An image change detection algorithm based on markov random field models. *IEEE Transactions on Geoscience and Remote Sensing*, 40:1815–1823, 2002.

[102] H. Kim and K. Hong. Robust image mosaicing of soccer videos using self-calibration and line tracking. *Pattern Analysis & Applications*, 4(1):9–19, 2001.

[103] S. Kopf, T. Haenselmann, D. Farin, and W. Effelsberg. Automatic generation of summaries for the web. In *Proceedings of SPIE, Storage and Retrieval for Media Databases, Vol. 5307*, volume 5307, pages 417–428, Jan. 2004.

[104] S. Kopf, T. Haenselmann, D. Farin, and W. Effelsberg. Automatic generation of video summaries for historical films. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 3, pages 2067–2070, June 2004.

[105] S. Kopf, G. Kühne, and O. Schuster. Contour-based classification of video objects. In *Proceedings of SPIE, Storage and Retrieval for Media Databases*, pages 608–618, Jan. 2001.

[106] M. Kourogi, T. Kurata, J. Hoshino, and Y. Muraoka. Real-time image mosaicing from a video sequence. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 4, pages 133–137, Oct. 1999.

[107] S. Krabbe. Metadatenextraktion aus Videosequenzen innerhalb eines bekannten Weltmodells am Beispiel von Sportübertragungen. Diplomarbeit, Universität Mannheim, Dec. 2002.

[108] J. B. Kuipers. *Quaternions and Rotation Sequences*. Princeton University Press, 1998.

[109] T. Kurita. An efficient clustering algorithm for region merging. In *IEICE Trans. of Information and Systems*, volume E78-D, No. 12, 1995.

[110] R. Laganière and É. Vincent. Wedge-based corner model for widely separated views matching. In *IEEE International Conference on Pattern Recognition*, volume 3, pages 672–675, 2002.

[111] M. C. Lee, W. Chen, C. B. Lin, C. Gu, T. Markoc, S. I. Zabinsky, and R. Szeliski. A layered video object coding system using sprite and affine motion model. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):130–145, Feb. 1997.

[112] J. Li, J. Z. Wang, and G. Wiederhold. IRM: Integrated region matching for image retrieval. In *ACM Multimedia*, pages 147–156, 2000.

[113] L. Li and M. Leung. Integrating intensity and texture differences for robust change detection. *IEEE Transactions on Image Processing*, 11:105–112, 2002.

[114] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer, 1995.

[115] R. Lienhart, S. Pfeiffer, and W. Effelsberg. Video abstracting. In *Communications of the ACM*, volume 40, pages 55–62, 1997.

[116] Y. Liu, Q. Huang, Q. Ye, and W. Gao. A new method to calculate the camera focusing area and player position on playfield in soccer video. In *SPIE Visual Communications and Image Processing, 2005*, pages 1524–1533, July 2005.

[117] M. I. A. Lourakis and A. A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based

on the levenberg-marquardt algorithm. Technical report, Institute of Computer Science of the Foundation for Research and Technology - Hellas FORTH, Aug. 2004.

[118] Y. Lu, W. Gao, and F. Wu. Sprite generation for frame-based video coding. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 473–476, 2001.

[119] Y. Lu, W. Gao, and F. Wu. Efficient background video coding with static sprite generation and arbitrary-shape spatial prediction techniques. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(5):394–405, 2003.

[120] H. Luo and A. Eleftheriadis. Rubberband: An improved graph search algorithm for interactive object segmentation. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 101–104, 2002.

[121] J. Maciel and J. P. Costeira. A global solution to sparse correspondence problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(2):187–199, Feb. 2003.

[122] S. Mann and R. W. Picard. Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing*, 6(9), Sept. 1999.

[123] M. Massey and W. Bender. Salient stills: Process and practice. *IBM Systems Journal*, 35(3&4):557–573, 1996.

[124] R. Mech and M. Wollborn. A noise robust method for segmentation of moving objects in video sequences. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2657–2660, Apr. 1997.

[125] B. T. Messmer and H. Bunke. A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 20(5):493–504, May 1998.

[126] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *European Conference on Computer Vision (ECCV)*, pages 128–142. Springer, 2002. Copenhagen.

[127] R. Mohr and B. Triggs. Projective geometry for image analysis; a tutorial given at ISPRS, Vienna, Sept. 1996.

[128] F. Mokhtarian and A. K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14:789–805, Aug. 1992.

[129] H. Moravec. Visual mapping by a robot rover. In *Proceedings of the 6th International Joint Conference on Artificial Intelligence*, pages 599–601, August 1979.

[130] E. N. Mortensen and W. A. Barrett. Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing*, 60:349–384, 1998.

[131] Y. Morvan, D. Farin, and P. H. N. de With. Matching-pursuit dictionary pruning for MPEG-4 video object coding. In *Internet and multimedia systems and applications*, volume 1, pages 476–481, Feb. 2005.

[132] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications in Pure and Applied Mathematics*, 42(5):577–685, 1989.

[133] J. Nesvadba, P. Fonseca, A. Sinitsyn, F. de Lange, M. Thijssen, P. van Kaam, H. Liu, R. van Leeuwen, J. Lukkien, A. Korostelev, J. Ypma, B. Kroon, H. Celik, A. Hanjalic, U. Naci, J. Benois-Pineau, P. de With, and J. Han. Real-time and distributed AV content analysis system for consumer electronics networks. In *IEEE International Conference on Multimedia and Expo (ICME)*, July 2005.

[134] C.-W. Ngo, T.-C. Pong, and H.-J. Zhang. On clustering and retrieval of video shots. In *ACM Multimedia*, pages 51–60, 2001.

[135] H. Nicolas. Optimal criterion for dynamic mosaicking. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 4, pages 133–137, Oct. 1999.

[136] P. Nunes and F. M. Pereira. Scene level rate control algorithm for MPEG-4 video coding. In *SPIE Visual Communications and Image Processing (VCIP)*, pages 194–205, 2001.

[137] Y. Ohno, J. Miura, and Y. Shirai. Tracking players and estimation of the 3D position of a ball in soccer games. In *"Proc. International Conference on Pattern Recognition (ICPR)*, volume 1, pages 145–148, Sept. 2000.

[138] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22:266–280, Mar. 2000.

[139] M. Pastrnak, D. Farin, and P. H. N. de With. Adaptive decoding of MPEG-4 sprites for memory-constrained embedded systems. In *26th Symposium on Information Theory in the Benelux*, pages 137–144, May 2005.

[140] M. Pastrnak, P. Poplavko, P. H. N. de With, and D. Farin. Data-flow timing models of dynamic multimedia applications for multiprocessor systems. In *4th IEEE International Workshop on System-on-Chip for Real-Time Applications (SoCRT)*, pages 206–209, July 2004.

[141] I. Patras, E. A. Hendriks, and R. L. Lagendijk. Video segmentation by MAP labeling of watershed segments. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(3):326–332, 2001.

[142] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert. The diver project: Interactive digital video repurposing. *IEEE Multimedia*, 11(11):54–61, 2004.

[143] H. Peinsipp. Implementation of a Java applet for demonstration of block-matching motion-estimation algorithms. Studienarbeit, Universität Mannheim, Oct. 2003.

[144] M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. Technical report, Yale University, Center for Computational Vision & Control, Nov. 1997.

[145] M. Pelillo, K. Siddiqi, and S. W. Zucker. Continuous-based heuristics for graph and tree isomorphisms, with application to computer vision. In *NIPS 99 Workshop on Complexity and Neural Computation*, Dec. 1999.

[146] M. Pfeffer. Entwicklung eines Algorithmus zur benutzerunterstützten Segmentierung mehrerer unabhängiger Videoobjekte. Diplomarbeit, Universität Kaiserslautern and Universität Mannheim, Aug. 2003.

[147] M. Pilu. A direct method for stereo correspondence based on singular value decomposition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 261–266, June 1997.

[148] P. Piscaglia, A. Cavallaro, M. Bonnet, and D. Douxchamps. High level description of video surveillance sequences. In *Proc. of EC-MAST*, pages 316–331, 1999.

[149] M. Pollefeys, R. Koch, M. Vergauwen, B. Deknuydt, and L. V. Gool. Three-dimensional scene reconstruction from images. In *SPIE Electronic Imaging, Three-Dimensional Image Capture and Applications III*, volume 3958, pages 215–226, 2000.

[150] H. V. Poor. *An Introduction to Signal Detection and Estimation, 2nd ed.* Springer-Verlag, 1994.

[151] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes in C.* Cambridge Univ. Press, 1988.

[152] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing*, to appear.

[153] K. Ratakonda, M. I. Sezan, and R. Crinon. Hierarchical video summarization. In *SPIE Proc. Visual Communications and Image Processing (VCIP)*, pages 1531–1541, 1999.

[154] J. M. Rehg and T. Kanade. Visual tracking of high DOF articulated structures: an application to human hand tracking. In *European Conference on Computer Vision (2)*, pages 35–46, 1994.

[155] I. D. Reid and A. Zisserman. Goal-directed video metrology. In *Proc. European Conference on Computer Vision (ECCV)*, pages 647–658, 1996.

[156] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using Kalman-filtering. In *Proc. of ICRAM*, pages 193–199, 1995.

[157] P. Rosin. Thresholding for change detection. In *Computer Vision, 1998. Sixth International Conference on*, pages 274–279, 1998.

[158] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graphics (special issue, Proc. of SIGGRAPH 2004)*, 23(3):309–314, 2004.

[159] P. J. Rousseeuw and K. Van Driessen. Computing LTS regression for large data sets. *Institute of Mathematical Statistics Bulletin*, 27(6), 1998.

[160] Y. Rubner. *Perceptual Metrics for Image Database Navigation.* PhD thesis, Stanford University, 1999.

[161] M. Ruzon and C. Tomasi. Alpha estimation in natural images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 18–25, June 2000.

[162] E. Salvador, A. Cavallaro, and T. Ebrahimi. Shadow identification and classification using invariant color models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1545–1548, May 2001.

[163] C. Schellewald and C. Schnörr. Subgraph matching with semidefinite programming. In V. D. G. Alberto Del Lungo and A. Kuba, editors, *Electronic Notes in Discrete Mathematics*, volume 12. Elsevier Science Publishers, 2003.

[164] J. Schmidt and H. Niemann. Using quaternions for parametrizing 3-d rotations in unconstrained nonlinear optimization. In *Vision, Modeling, and Visualization*, pages 399–406, Nov. 2001.

[165] S. Seedorf. Implementierung eines Java-Applets zur Visualisierung von Geometrietransformationen für Image-Mosaicing. Studienarbeit, Universität Mannheim, July 2003.

[166] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry.* Oxford University Press, 1952.

[167] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 593–600, 1994.

[168] H.-Y. Shum, M. Han, and R. Szeliski. Interactive construction of 3D models from panoramic mosaics. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–433, June 1998.

[169] M. A. Smith and T. Kanade. Video skimming and characterization through the combination of image and language understanding techniques. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 775–781, 1997.

[170] S. M. Smith and J. M. Brady. SUSAN — a new approach to low level image processing. *International Journal of Computer Vision (IJCV)*, 23(1):45–78, May 1997.

[171] A. Smolic and J. Ohm. Robust global motion estimation using a simplified M-estimator approach. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 868–871, Sept. 2000.

[172] A. Smolic, T. Sikora, and J.-R. Ohm. Direct estimation of long-term global motion parameters using affine and higher order polynomial models. In *Proc. Picture Coding Symposium (PCS)*, pages 239–242, Apr. 1999.

[173] A. Smolic, T. Sikora, and J.-R. Ohm. Direct estimation of long-term global motion parameters using affine and higher order polynomial models. In *Proc. Picture Coding Symposium (PCS)*, Apr. 1999.

[174] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 246–252, 1999.

[175] C. V. Stewart. MINPRAN: a new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(10):925–938, Oct. 1995.

[176] G. Sudhir, J. C. M. Lee, and A. K. Jain. Automatic classification of tennis video for high-level content-based retrieval. In *IEEE International Workshop on Content Based Access of Image and Video Databases*, pages 81–90, 1998.

[177] H. Suesse and W. Ortmann. Robust matching of affinely transformed objects. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 375–378, Sept. 2003.

[178] C. Sun and S. Pallottino. Circular shortest path in images. *Pattern Recognition*, 36(3):711–721, Mar. 2003.

[179] R. Szeliski. Image mosaicing for tele-reality applications. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 44–53, Dec. 1994.

[180] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258. ACM Press/Addison-Wesley Publishing Co., 1997.

[181] C. Thiel. Entwicklung einer skriptgesteuerten Videoanalyse basierend auf MPEG-7 Deskriptoren. Diplomarbeit, Universität Mannheim, Dec. 2003.

[182] P. H. S. Torr and C. Davidson. IMPSAC: synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(3):354–364, Mar. 2003.

[183] P. H. S. Torr and A. Zisserman. MLESAC: a new robust estimator with application to estimating image geometry. *Compututer Vision and Image Understanding*, 78(1):138–156, 2000.

[184] A. Torsello and E. R. Hancock. Efficiently computing weighted tree edit distance using relaxation labeling. In *Energy Minimization Methods in Computer Vision and Pattern Recognition, Third International Workshop, EMMCVPR 2001, France*, volume 2134 of *Lecture Notes in Computer Science*, pages 438–453. Springer, Sept. 2001.

[185] J. Townshend, C. Justice, C. Gurney, and J. McManus. The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30:1054–1060, 1992.

[186] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *International Conference on Computer Vision*, page 255, 1999.

[187] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment – A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.

[188] A. Vetro, H. Sun, and Y. Wang. MPEG-4 rate control for multiple video objects. *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, 9(1):186–199, Feb. 1999.

[189] L. Vincent and P. Soile. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(6):583–597, 1991.

[190] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision - to appear*, 2002.

[191] J. Vogel, J. Widmer, D. Farin, M. Mauve, and W. Effelsberg. Priority-based distribution trees for application-level multicast. In *Proceedings of the 2nd Workshop on Network and System Support for Games (ACM NETGAMES 2003)*, pages 148–157, May 2003.

[192] J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *The IEEE Transactions on Image Processing Special Issue: Image Sequence Compression*, 3(5):625–638, September 1994.

[193] J. H. Ward. Hierarchical grouping to optimize an objective function. *J. American Stat. Assoc.*, 58:236–245, 1963.

[194] H. Watanabe and K. Jinzenji. Sprite coding in object-based video coding standard: MPEG-4. In *Proc. World Multiconf. on SCI 2001*, volume XIII, pages 420–425, 2001.

[195] T. Watanabe, M. Haseyama, and H. Kitajima. A soccer field tracking method with wire frame model from TV images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 1633–1636, Oct. 2004.

[196] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):780–785, July 1997.

[197] N. Xu and N. Ahuja. Object contour tracking using graph cuts based active contours. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 277–280, Sept. 2002.

[198] A. Yamada, Y. Shirai, and J. Miura. Tracking players and a ball in video image sequence and estimating camera parameters for 3d interpretation of soccer games. In *Proc. 16th Int. Conf. on Pattern Recognition*, pages 303–306, Aug. 2002.

[199] X. Yu and D. Farin. Current and emerging topics in sports video processing. In *IEEE International Conference on Multimedia and Expo (ICME)*, July 2005.

[200] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. Technical Report RR-2676, INRIA, Oct. 1995.

[201] Z. Zhang. Determining the epipolar geometry and its uncertainty - a review. *International Journal of Computer Vision*, 27(2):161–195, Mar. 1998.

[202] F. Ziliani. An image segmentation procedure based on statistical change detection. Technical Report LTS 98.02, Ecole Polytechnique Federale de Lausanne (EPFL), May 1998.

Summary

Practically established video compression and storage techniques still process video sequences as rectangular images without further semantic structure. However, humans watching a video sequence immediately recognize acting objects as semantic units. This semantic object separation is currently not reflected in the technical system, making it difficult to manipulate the video at the object level. The realization of object-based manipulation will introduce many new possibilities for working with videos like composing new scenes from pre-existing video objects or enabling user-interaction with the scene.

Moreover, object-based video compression, as defined in the MPEG-4 standard, can provide high compression ratios because the foreground objects can be sent independently from the background. In the case that the scene background is static, the background views can even be combined into a large panoramic *sprite* image, from which the current camera view is extracted. This results in a higher compression ratio since the sprite image for each scene only has to be sent once.

A prerequisite for employing object-based video processing is automatic (or at least user-assisted semi-automatic) segmentation of the input video into semantic units, the video objects. This segmentation is a difficult problem because the computer does not have the vast amount of pre-knowledge that humans subconsciously use for object detection. Thus, even the simple definition of the desired output of a segmentation system is difficult. The subject of this thesis is to provide algorithms for segmentation that are applicable to common video material and that are computationally efficient.

The thesis is conceptually separated into three parts. In Part I, an automatic segmentation system for general video content is described in detail. Part II introduces object models as a tool to incorporate user-defined knowledge about the objects to be extracted into the segmentation process. Part III concentrates on the modeling of camera motion in order to relate the observed camera motion to real-world camera parameters.

The segmentation system that is described in Part I is based on a background-subtraction technique. The pure background image that is required for this technique is synthesized from the input video itself. Sequences that contain rotational camera motion can also be processed since the camera motion is estimated and the input images are aligned into a panoramic scene-background. This approach is fully compatible to the MPEG-4 video-encoding framework, such that the segmentation system can be easily combined with an object-based MPEG-4 video codec.

After an introduction to the theory of projective geometry in Chapter 2, which is required for the derivation of camera-motion models, the estimation of camera motion is discussed in Chapters 3 and 4. It is important that the camera-motion estimation is not influenced by foreground object motion. At the same time, the estimation should provide accurate motion parameters such that all input frames can be combined seamlessly into a background image. The core motion estimation is based on a feature-based approach where the motion parameters are determined with a robust-estimation algorithm (RANSAC) in order to distinguish the camera motion from simultaneously visible object motion. Our experiments showed that the robustness of the original RANSAC algorithm in practice does not reach the theoretically predicted performance. An analysis of the problem has revealed that this is caused by numerical instabilities that can be significantly reduced by a modification that we describe in Chapter 4.

The synthetization of static-background images is discussed in Chapter 5. In particular, we present a new algorithm for the removal of the foreground objects from the background image such that a pure scene background remains. The proposed algorithm is optimized to synthesize the background even for difficult scenes in which the background is only visible for short periods of time. The problem is solved by clustering the image content for each region over time, such that each cluster comprises static content. Furthermore, it is exploited that the times, in which foreground objects appear in an image region, are similar to the corresponding times of neighboring image areas.

The reconstructed background could be used directly as the sprite image in an MPEG-4 video coder. However, we have discovered that the counterintuitive approach of splitting the background into several independent parts can reduce the overall amount of data. In the case of general camera motion, the construction of a single sprite image is even impossible. In Chapter 6, a *multi-sprite partitioning* algorithm is presented, which separates the video sequence into a number of segments, for which independent sprites are synthesized. The partitioning is computed in such a way that the total area of the resulting sprites is minimized, while simultaneously

satisfying additional constraints. These include a limited sprite-buffer size at the decoder, and the restriction that the image resolution in the sprite should never fall below the input-image resolution. The described multi-sprite approach is fully compatible to the MPEG-4 standard, but provides three advantages. First, any arbitrary rotational camera motion can be processed. Second, the coding-cost for transmitting the sprite images is lower, and finally, the quality of the decoded sprite images is better than in previously proposed sprite-generation algorithms.

Segmentation masks for the foreground objects are computed with a change-detection algorithm that compares the pure background image with the input images. A special effect that occurs in the change detection is the problem of image misregistration. Since the change detection compares co-located image pixels in the camera-motion compensated images, a small error in the motion estimation can introduce segmentation errors because non-corresponding pixels are compared. We approach this problem in Chapter 7 by integrating *risk-maps* into the segmentation algorithm that identify pixels for which misregistration would probably result in errors. For these image areas, the change-detection algorithm is modified to disregard the difference values for the pixels marked in the risk-map. This modification significantly reduces the number of false object detections in fine-textured image areas.

The algorithmic building-blocks described above can be combined into a segmentation system in various ways, depending on whether camera motion has to be considered or whether real-time execution is required. These different systems and example applications are discussed in Chapter 8.

Part II of the thesis extends the described segmentation system to consider object models in the analysis. Object models allow the user to specify which objects should be extracted from the video. In Chapters 9 and 10, a graph-based object model is presented in which the features of the main object regions are summarized in the graph nodes, and the spatial relations between these regions are expressed with the graph edges. The segmentation algorithm is extended by an object-detection algorithm that searches the input image for the user-defined object model. We provide two object-detection algorithms. The first one is specific for cartoon sequences and uses an efficient sub-graph matching algorithm, whereas the second processes natural video sequences. With the object-model extension, the segmentation system can be controlled to extract individual objects, even if the input sequence comprises many objects.

Chapter 11 proposes an alternative approach to incorporate object models into a segmentation algorithm. The chapter describes a semi-automatic segmentation algorithm, in which the user coarsely marks the object and

the computer refines this to the exact object boundary. Afterwards, the object is tracked automatically through the sequence. In this algorithm, the object model is defined as the texture along the object contour. This texture is extracted in the first frame and then used during the object tracking to localize the original object. The core of the algorithm uses a graph representation of the image and a newly developed algorithm for computing shortest circular-paths in planar graphs. The proposed algorithm is faster than the currently known algorithms for this problem, and it can also be applied to many alternative problems like shape matching.

Part III of the thesis elaborates on different techniques to derive information about the physical 3-D world from the camera motion. In the segmentation system, we employ camera-motion estimation, but the obtained parameters have no direct physical meaning. Chapter 12 discusses an extension to the camera-motion estimation to factorize the motion parameters into physically meaningful parameters (rotation angles, focal-length) using camera autocalibration techniques. The speciality of the algorithm is that it can process camera motion that spans several sprites by employing the above multi-sprite technique. Consequently, the algorithm can be applied to arbitrary rotational camera motion.

For the analysis of video sequences, it is often required to determine and follow the position of the objects. Clearly, the object position in image coordinates provides little information if the viewing direction of the camera is not known. Chapter 13 provides a new algorithm to deduce the transformation between the image coordinates and the real-world coordinates for the special application of sport-video analysis. In sport videos, the camera view can be derived from markings on the playing field. For this reason, we employ a model of the playing field that describes the arrangement of lines. After detecting significant lines in the input image, a combinatorial search is carried out to establish correspondences between lines in the input image and lines in the model. The algorithm requires no information about the specific color of the playing field and it is very robust to occlusions or poor lighting conditions. Moreover, the algorithm is generic in the sense that it can be applied to any type of sport by simply exchanging the model of the playing field.

In Chapter 14, we again consider panoramic background images and particularly focus ib their visualization. Apart from the planar background-sprites discussed previously, a frequently-used visualization technique for panoramic images are projections onto a cylinder surface which is unwrapped into a rectangular image. However, the disadvantage of this approach is that the viewer has no good orientation in the panoramic image because he looks into all directions at the same time. In order to provide

a more intuitive presentation of wide-angle views, we have developed a visualization technique specialized for the case of indoor environments. We present an algorithm to determine the 3-D shape of the room in which the image was captured, or, more generally, to compute a complete floor plan if several panoramic images captured in each of the rooms are provided. Based on the obtained 3-D geometry, a graphical model of the rooms is constructed, where the walls are displayed with textures that are extracted from the panoramic images. This representation enables to conduct virtual walk-throughs in the reconstructed room and therefore, provides a better orientation for the user.

Summarizing, we can conclude that all segmentation techniques employ some definition of foreground objects. These definitions are either explicit, using object models like in Part II of this thesis, or they are implicitly defined like in the background synthetization in Part I. The results of this thesis show that implicit descriptions, which extract their definition from video content, work well when the sequence is long enough to extract this information reliably. However, high-level semantics are difficult to integrate into the segmentation approaches that are based on implicit models. Intead, those semantics should be added as postprocessing steps. On the other hand, explicit object models apply semantic pre-knowledge at early stages of the segmentation. Moreover, they can be applied to short video sequences or even still pictures since no background model has to be extracted from the video. The definition of a general object-modeling technique that is widely applicable and that also enables an accurate segmentation remains an important yet challenging problem for further research.

Samenvatting

De huidige praktisch bewezen videocompressie- en opslagtechnieken bewerken videosequenties nog steeds als rechthoekige beelden zonder enige semantische structuur. Echter, mensen die naar sequenties van videobeelden kijken, nemen onmiddellijk de daarin optredende objecten als semantisch relevante eenheden waar. Deze semantische objectherkenning wordt niet gereflecteerd in de technische implementatie, zodat het moeilijk is om videobeelden te manipuleren op objectniveau. De realisatie van objectmanipulatie zal veel nieuwe mogelijkheden introduceren om met videobeelden te werken, zoals het samenstellen van nieuwe scènes van reeds bestaande video-objecten en speciale gebruikersinteractie met de gerepresenteerde scène.

Daarnaast kan objectgebaseerde videocompressie zoals gedefinieerd in de MPEG-4 standaard, tot hoge compressiefactoren leiden, omdat de objecten op de voorgrond onafhankelijk van de achtergrond kunnen worden verzonden. In het geval van een statische achtergrond in de scène kunnen de verschillende achtergrondbeelden worden gecombineerd in een groot panoramisch beeld, genaamd *sprite*-beeld, waarvan het actuele camerablikveld kan worden geëxtraheerd. Dit concept resulteert in een hogere compressiefactor, omdat het sprite-beeld voor elke scène slechts eenmaal hoeft te worden verzonden.

Een voorwaarde voor het gebruiken van objectgebaseerde videobewerking is automatische (of op zijn minst met hulp van de gebruiker semi-automatische) segmentatie van de videobeelden aan de ingang in semantische eenheden, ook wel video-objecten genoemd. Deze segmentatie is een complex probleem, omdat een computer niet de enorme voorkennis heeft, die mensen onbewust gebruiken voor het detecteren van objecten. Zelfs een eenvoudige definitie van het gewenste uitgangsresultaat van het segmentatiesysteem is moeilijk. Het onderwerp van dit proefschrift is om algoritmen te ontwikkelen voor segmentatie die toepasbaar zijn voor gebruikelijk videomateriaal en die rekenkundig gezien efficiënt zijn.

Het proefschrift is conceptueel gesplitst in drie delen. In Deel I wordt

een automatisch segmentatiesysteem voor generieke beeldinhoud in detail beschreven. Deel II introduceert objectmodellen als gereedschap om voorkennis van de gebruiker toe te voegen over de te extraheren objecten in het segmentatieproces. Deel III concentreert zich op het modelleren van camerabeweging om de geobserveerde camerabeweging te relateren aan de werkelijke, fysische cameraparameters.

Het segmentatiesysteem dat wordt beschreven in Deel I is gebaseerd op een techniek met achtergrond-subtractie. Het pure achtergrondbeeld dat nodig is voor deze techniek, is gesynthetiseerd van het ingangsvideosignaal zelf. Sequenties van videobeelden die een draaiende camerabeweging bevatten kunnen ook worden bewerkt, omdat de camerabeweging wordt geschat en de ingangsbeelden in een panoramische achtergrond van de scène worden samengesteld. Deze benadering is volledig compatibel met de MPEG-4 videocodering, zodat het segmentatiesysteem probleemloos kan worden gecombineerd met een objectgebaseerde MPEG-4 videocoder.

Na een introductie in de theorie van projectieve geometrie in Hoofdstuk 2, die nodig is voor de afleiding van camerabewegingsmodellen, wordt de schatting van camerabeweging besproken in de Hoofdstukken 3 en 4. Het is belangrijk dat de schatting van de camerabeweging niet wordt beïnvloed door de objectbeweging op de voorgrond van de scène. Tegelijkertijd moet de schatting tot nauwkeurige bewegingsparameters leiden, zodanig dat alle ingangsbeelden naadloos kunnen worden samengevoegd in het achtergrondsbeeld. De bewegingsschatting is in de kern een *feature*-gebaseerde benadering, waarin de bewegingsparameters worden bepaald met een robuust schattingsalgoritme (RANSAC), om een onderscheid te maken tussen camerabeweging en de gelijktijdig zichtbare objectbeweging. Experimenten hebben aangetoond dat het originele RANSAC-algoritme de theoretische voorspelde robuustheid in de praktijk niet realiseert. Een analyse van dit probleem heeft opgeleverd dat dit door numerieke instabiliteiten wordt veroorzaakt. Deze kunnen significant worden gereduceerd door een algoritmemodificatie die in Hoofdstuk 4 wordt beschreven.

De synthetisatie van beelden met statische achtergrond wordt beschreven in Hoofdstuk 5. Een bijzondere bijdrage is een nieuw algoritme voor het verwijderen van objecten op de voorgrond in het achtergrondbeeld, zodat een pure scène-achtergrond overblijft. Het voorgestelde algoritme is geoptimaliseerd om de achtergrond reconstrueren, zelfs voor moeilijke scènes waarin de achtergrond slechts korte tijd zichtbaar is. Dit probleem is opgelost door de beeldinhoud van een gebied temporeel zodanig te klusteren dat elk kluster een statische beeldinhoud heeft. Tevens wordt benut dat de tijden waarin voorgrondobjecten in een gebied zichtbaar zijn gelijkwaardig zijn aan de corresponderende tijden van naburige beeldgebieden.

De gereconstrueerde achtergrond zou direct kunnen worden gebruikt als sprite beeld in een MPEG-4 videocoder. Het is echter een interessante ontdekking dat een anti-intuïtieve benadering, om de achtergrond te splitsen in verscheidene onafhankelijke delen, de totale hoeveelheid beelddata kan verminderen. In het geval van generieke camerabeweging is de constructie van een enkel sprite-beeld zelfs onmogelijk. In Hoofdstuk 6 wordt een *multi-sprite* partitioneringsalgoritme gepresenteerd, dat de sequentie van videobeelden verdeeld in een aantal segmenten waarvoor onafhankelijke sprites worden opgebouwd. De partitionering wordt zodanig berekend, dat de totale oppervlakte van de resulterende sprites wordt geminimaliseerd, terwijl gelijktijdig extra voorwaarden worden gerealiseerd. Deze voorwaarden zijn een beperkte sprite buffergrootte in de decoder en de beperking dat de beeldresolutie in de sprite nooit lager mag zijn dan de ingangsbeeldresolutie. De beschreven multi-sprite benadering is volledig compatibel met de MPEG-4 standaard, maar heeft desondanks drie voordelen. Ten eerste kan elke draaiende camerabeweging worden gebruikt. Ten tweede zijn de coderingskosten voor het overdragen van de sprite-beelden lager en ten derde, is de kwaliteit van de gedecodeerde sprite-beelden beter dan dat van eerdere algoritmen voor spritegeneratie.

Segmentatiemaskers voor de voorgrondobjecten worden bepaald met een algoritme voor het detecteren van veranderingen, dat het pure achtergrondbeeld vergelijkt met de ingangsbeelden. Een speciaal effect dat optreedt in de veranderingsdetectie is het probleem van foutieve beeldpositionering. Omdat de veranderingsdetectie overeenkomstige beeldelementen in de camerabewegingsgecompenseerde beelden vergelijkt, kan een kleine fout in de bewegingsschatting leiden tot segmentatiefouten. De reden hiervoor is dat niet-corresponderende beeldelementen worden vergeleken. In Hoofdstuk 7 wordt dit opgelost door zogenaamde *risicomaskers* in het segmentatie-algoritme te integreren, die beeldelementen identificeren waarvoor foutieve beeldpositionering waarschijnlijk zal resulteren in fouten. Voor deze beeldgebieden is het veranderingsdetectie-algoritme gemodificeerd zodanig dat de beeldverschillen van deze beeldelementen niet worden gebruikt. Deze modificatie vermindert het aantal verkeerde objectdetecties aanzienlijk in beeldgebieden met veel detailinformatie.

De hierboven beschreven algoritmemodules kunnen op verschillende manieren in het segmentatiesysteem worden gecombineerd, afhankelijk van of camerabeweging moet worden geïntegreerd of wanneer real-time executie noodzakelijk is. Deze verschillende systemen en voorbeeldtoepassingen worden bediscussiëerd in Hoofdstuk 8.

Deel II van het proefschrift verbreedt het segmentatiesysteem door objectmodellen mede in de analyse te betrekken. Objectmodellen maken

het mogelijk voor de gebruiker om te specificeren welke objecten uit het videosignaal moeten worden onttrokken. In de Hoofdstukken 9 en 10 wordt een graafgebaseerd objectmodel gepresenteerd, waarin de eigenschappen van de belangrjkste objectgebieden worden samengevat in de knooppunten van de graaf en de spatiële relaties tussen deze gebieden worden uitgedrukt door de verbindingen van de graaf. Het segmentatie-algoritme is uitgebreid met een objectdetectie-algoritme dat in het ingangsbeeld zoekt naar het door de gebruiker gedefiniëerde objectmodel. Twee algoritmen voor objectdetectie zijn ontwikkeld. Het eerste is specifiek geschikt voor tekenfilmbeelden en gebruikt een efficiënt deelgraaf-zoekalgoritme, het tweede objectdetectie-algoritme kan daarentegen algemene videobeelden bewerken. Door de uitbreiding met het objectmodel kan het segmentatiesysteem worden gecontroleerd om individuele objecten te extraheren, zelfs wanneer de ingangsbeeldsequentie veel objecten bevat.

Hoofdstuk 11 stelt een alternatieve benadering voor om objectmodellen te integreren in een segmentatie-algoritme. Dit hoofdstuk beschrijft een semi-automatisch segmentatie-algoritme, waarin de gebruiker globaal het object markeert en de computer dit verfijnt tot de exacte objectcontour. Hierna wordt het object gevolgd gedurende de beeldsequentie. In dit algoritme is het objectmodel gedefinieerd als de beeldstructuur (textuur) langs de objectcontour. Deze textuur wordt onttrokken in het eerste beeld en dan gebruikt gedurende het volgen van het object om het originele object te localiseren. De kern van het algoritme gebruikt een graafrepresentatie van het beeld en een nieuw ontwikkeld algoritme voor het berekenen van de kortste rondgaande paden (cykels) in planaire graven. Het voorgestelde algoritme is sneller dan de algemeen bekende algoritmen voor dit probleem en het kan ook worden toegepast voor veel alternatieve problemen, zoals het vergelijken van objectvormen.

Deel III van het proefschrift gaat dieper in op verschillende technieken om informatie over de fysische 3-D wereld af te leiden van de camerabeweging. In het segmentatiesysteem gebruiken we camerabewegingsschatting, maar de verkregen parameters hebben geen directe fysische betekenis. Hoofdstuk 12 behandelt een uitbreiding naar camerabewegingsschatting om de bewegingsparameters te factoriseren naar fysisch zinvolle parameters (draaihoek, brandpuntsafstand), die zijn gebaseerd op zelf-calibratie. Het speciale element in het algoritme is dat sequenties kunnen worden bewerkt met een camerabeweging die zich over verscheidene sprites uitstrekt, wanneer de eerder genoemde multi-sprite techniek wordt gebruikt. De consequentie is dat het algoritme kan worden toegepast voor generiek draaiende camerabewegingen.

Voor de analyse van videosequenties is het vaak nodig om de object-

posities te bepalen en te volgen. Het is duidelijk dat de objectpositie in beeldcoördinaten weinig informatie geeft wanneer het blikveld van de camera onbekend is. Hoofdstuk 13 bespreekt een nieuw algoritme om de transformatie tussen beeldcoördinaten en de wereldcoördinaten af te leiden voor de speciale toepassing van video-analyse van sportwedstrijden. In sportbeelden kan het blikveld van de camera worden bepaald via markeringen op het speelveld. Om deze reden is een model van het speelveld gebruikt, dat de inrichting van de veldlijnen beschrijft. Nadat de significante lijnen in het beeld zijn gedetecteerd, wordt een combinatorische zoekstrategie uitgevoerd om overeenkomsten tussen lijnen in het beeld en veldlijnen in het model te vinden. Het algoritme heeft geen informatie nodig over de specifieke kleur van het speelveld en het is zeer robuust tegen afdekkingen van objecten of slechte belichtingscondities. Bovendien is het algoritme generiek toepasbaar voor elke andere sport door eenvoudigweg het speelveldmodel te verwisselen.

In Hoofdstuk 14 beschouwen we opnieuw panoramische achtergrondbeelden en focusseren in het bijzonder op hun visualisatie. Behalve de eerder besproken vlakke achtergrond sprites, is het projecteren op een cilinderoppervlak een gebruikelijke visualisatietechniek, waarbij het oppervlak wordt afgerold tot een vlak rechthoekig beeld. Het nadeel van deze techniek is echter dat de kijker geen goede oriëntatie heeft in het panoramische beeld, omdat hij alle richtingen tegelijk observeert. Om te kunnen voorzien in een meer gebruikersvriendelijke visualisatie van panoramische beelden, is een techniek ontwikkeld die speciaal geschikt is voor inpandige ruimtes. We presenteren een algoritme om de 3-D vorm van de kamer waar het beeld was opgenomen te bepalen, of meer algemeen, het berekenen van het complete vloerplan wanneer panoramische beelden van elke ruimte ter beschikking staan. Gebaseerd op de verkregen 3-D geometrie wordt een grafisch model geconstrueerd, waarbij de muren worden getoond met de beeldstructuur die is geëxtraheerd van de panoramische beelden. Deze visualisatie maakt het mogelijk om virtuele wandelingen in de gereconstrueerde kamer te maken en voorziet daardoor in een betere oriëntatie voor de gebruiker.

Samenvattend kan worden geconcludeerd dat alle segmentatietechnieken een zekere definitie van voorgrondobjecten toepassen. Deze definities zijn ofwel expliciet, gebruik makend van objectmodellen zoals in Deel II van dit proefschrift of zij zijn impliciet gedefiniëerd, zoals bijvoorbeeld de achtergrondsynthetisatie in Deel I. De resultaten van dit proefschrift tonen aan dat impliciete modellen die hun definitie onttrekken aan de video-inhoud, goed werken wanneer de beeldsequentie lang genoeg is om deze informatie betrouwbaar te extraheren. Semantiek op hoog niveau is echter moeilijk te integreren in segmentatiebenaderingen die gebaseerd zijn op impliciete

modellen. In plaats daarvan moet deze semantiek in nabewerkingsstappen worden toegevoegd. Expliciete objectmodellen passen daarentegen semantische voorkennis toe in de aanvangsstappen van de segmentatie. Bovendien kunnen deze modellen worden toegepast voor korte videosequenties of zelfs individuele beelden, omdat geen achtergrondmodel hoeft te worden geëxtraheerd van het videosignaal. De definitie van een algemene objectmodelleringstechniek die breed toepasbaar is en die ook een nauwkeurige segmentatie mogelijk maakt, blijft een belangrijk doch uitdagend probleem voor verder onderzoek.

Zusammenfassung

Die gegenwärtig in der Praxis verwendeten Videokompressions- und Speichertechniken verarbeiten die Videosequenzen nach wie vor als rechteckige Bilder ohne weitere semantische Struktur. Andererseits nehmen wir als Menschen sofort die agierenden Objekte als semantische Einheiten wahr. Diese Zerlegung in semantische Objekte wird momentan auf der technischen Seite nicht durchgeführt, was die Manipulation des Videos auf Objektebene erschwert. Die Realisierung objektbasierter Manipulation wird neue Möglichkeiten für die Verarbeitung von Videos erlauben, wie z.B. das Zusammensetzen neuer Szenen aus vorgefertigten Videoobjekten oder die Interaktion des Benutzers mit der dargestellten Szene.

Desweiteren kann objektbasierte Videokompression, wie sie im MPEG-4-Standard definiert wurde, hohe Kompressionsfaktoren erreichen, da die Objekte im Vordergrund unabhängig vom Hintergrund übertragen werden können. Für den Fall dass der Szenenhintergrund statisch ist, können die Hintergrundansichten sogar in ein großes Panoramabild (*Sprite*) zusammengefügt werden, vom dem die aktuelle Kameraansicht wieder extrahiert wird. Dies resultiert in einem erhöhten Kompressionsfaktor, da das Sprite-Bild für jede Szene nur einmal gesendet werden muss.

Eine Voraussetzung für objektbasierte Videoverarbeitung ist die automatische (oder zumindest benutzerunterstützte, halbautomatische) Segmentierung des Eingabevideos in semantische Einheiten; den Videoobjekten. Diese Segmentierung ist ein schwieriges Problem, da der Computer nicht das unermessliche Vorwissen zur Verfügung hat, das Menschen unterbewusst für die Objekterkennung benutzen. Daher ist schon eine einfache Definition der erwünschten Ausgabe eines Segmentierungssystems schwierig. Das Thema dieser Dissertation ist es, Algorithmen für die Segmentierung zu entwickeln, die für gewöhnliches Videomaterial geeignet und effizient in der Berechnung sind.

Diese Dissertation ist konzeptuell in drei Teile gegliedert. In Teil I wird ein automatischen Segmentierungssystem für allgemeine Videoinhalte detailliert beschrieben. Teil II führt Objektmodelle als ein Werkzeug ein,

um benutzerdefiniertes Wissen über die zu extrahierenden Objekte in den Segmentierungsprozess einfließen zu lassen. Teil III beschreibt die Modellierung der Kamerabewegung, um die beobachtete Kamerabewegung mit den realen physischen Kameraparametern in Zusammenhang zu bringen.

Das in Teil I beschriebene Segmentierungssystem basiert auf der Technik der Hintergrundssubtraktion. Das reine Hintergrundbild, welches für diese Technik benötigt wird, wird aus dem Eingabevideo selbst synthetisiert. Sequenzen, in denen drehende Kamerabewegungen enthalten sind, können auch verarbeitet werden, da die Kamerabewegung geschätzt wird und die Eingabebilder in ein Panoramabild des Szenenhintergrunds zusammengesetzt werden. Dieser Ansatz ist voll kompatible zum MPEG-4 Videokompressionsverfahren, so dass das Segmentierungssystem problemlos mit einem objektbasierten MPEG-4 Videocodec kombiniert werden kann.

Nach einer Einführung in die Theorie der projektiven Geometrie in Kapitel 2, was für die Herleitung der Kamerabewegungsmodelle benötigt wird, wird die Schätzung der Kamerabewegung in den Kapiteln 3 und 3 diskutiert. Es ist wichtig, dass die Schätzung der Kamerabewegung nicht durch gleichzeitig vorhandene Bewegungen von Vordergrundobjekten beeinflußt wird. Andererseits sollte sie präzise Bewegungsparameter bestimmen, so dass alle Eingabebilder nahtlos in ein Hintergrundbild zusammengefügt werden können. Der Kern der Bewegungsschätzung verwendet einen featurebasierten Ansatz, bei dem die Bewegungsparameter mit einem robusten Schätzalgorithmus (RANSAC) bestimmt werden, um die Kamerabewegung von gleichzeitig sichtbarer Objektbewegung unterscheiden zu können. Unsere Experimente zeigten, dass die Robustheit des ursprünglichen RANSAC-Algorithmus in der Praxis nicht die theoretisch vorausgesagte Leistung erreicht. Eine Analyse des Problems ergab, dass dies in numerischen Instabilitäten begründet liegt, die durch eine in Kapitel 4 beschriebene Modifikation des Algorithmus erheblich reduziert werden können.

Die Synthese statischer Hintergrundbilder wird in Kapitel 5 diskutiert. Dabei präsentieren wir im speziellen einen neuen Algorithmus für das Entfernen von Vordergrundobjekten aus dem Hintergrundbild, so dass der reine Szenenhintergrund verbleibt. Der vorgeschlagene Algorithmus ist daraufhin optimiert, den Hintergrund auch in schwierigen Szenen rekonstruieren zu können, in denen er nur für kurze Zeiträume sichtbar ist. Das Problem wird gelöst, indem die Bildinhalte einer Region zeitlich so zu Clustern gruppiert werden, dass die Cluster jeweils einem statischen Bildinhalt entsprechen. Desweiteren wird ausgenutzt, dass die Zeiten, in denen in einer Region Vordergrundobjekte sichtbar sind, ähnlich sind wie die entsprechenden Zeiten der benachbarten Bildregionen.

Der rekonstruierte Hintergrund könnte direkt als Sprite-Bild in einem MPEG-4 Videocoder verwendet werden. Allerdings haben wir herausgefunden, dass der unintuitive Ansatz, den Hintergrund in mehrere unabhängige Teile zu zerlegen, die gesamte Datenmenge reduzieren kann. Im allgemeinen Fall unbeschränkter Kamerabewegung ist die Konstruktion eines einzelnen Sprite-Bildes sogar unmöglich. In Kapitel 6 wird ein Algorithmus zur *Multi-Sprite* Zerlegung präsentiert, welcher die Videosequenz in eine Anzahl Segmente unterteilt, für die dann unabhängige Sprites erstellt werden. Die Zerlegung wird so bestimmt, dass die Gesamtfläche des resultierenden Sprites minimiert wird, während gleichzeitig zusätzliche Nebenbedingungen erfüllt werden müssen. Dazu zählt eine Limitierung der Größe des Sprite-Bildspeichers im Decoder und die Einschränkung, dass die Bildauflösung im Sprite niemals unter die Auflösung des Eingabebildes sinken darf. Der beschriebene Multi-Sprite Ansatz ist vollständig kompatibel zum MPEG-4 Standard, aber bietet drei Vorteile. Erstens erlaubt er die Verarbeitung beliebiger drehender Kamerabewegungen. Zweitens sind die Kodierungskosten für die Übertragung des Sprite-Bildes geringer, und schliesslich ist die Qualität des dekodierten Sprite-Bildes besser als in früheren Algorithmen zur Spritegenerierung.

Die Segmentierungsmasken der Vordergrundobjekte werden mit einem Algorithmus zur Detektion von Änderungen zwischen dem reinen Hintergrundbild und den Eingabebildern bestimmt. Ein spezieller Effekt, der in der Änderungsdetektion auftritt, ist das Problem der Fehlausrichtung der Bilder. Da die Änderungsdetektion Bildpunkte an korrespondierenden Bildpositionen vergleicht, kann ein kleiner Fehler in der Bewegungsschätzung zu Segmentierungsfehlern führen, falls Pixel verglichen werden, die nicht korrespondieren. Wir gehen dieses Problem in Kapitel 7 dadurch an, Risikomasken in den Segmentierungsalgorithmus einzuführen, welche diejenigen Bildpunkte markieren, für welche eine Fehlausrichtung der Bilder wahrscheinlich zu Fehlern führen würde. Für diese Bildbereiche wird der Algorithmus zur Änderungsdetektion so modifiziert, dass er die Bilddifferenzen für diese Bildpunkte nicht beachtet. Diese Modifikation reduziert die Anzahl der Fehldetektionen von Objekten in feintexturierten Bildbereichen erheblich.

Die oben beschriebenen Algorithmenmodule können auf verschiedene Weise in ein Segmentierungssystem kombiniert werden, abhängig davon, ob ggf. Kamerabewegungen beachtet werden müssen oder ob eine Ausführung in Echtzeit benötigt wird. Diese unterschiedlichen Systeme und Beispielanwendungen werden in Kapitel 8 diskutiert.

Teil II der Arbeit erweitert das beschriebene Segmentierungssystem so, dass Objektmodelle in die Analyse einbezogen werden. Objektmodelle erlauben es dem Benutzer, die Objekte, die aus dem Video extrahiert werden sollen, zu spezifizieren. In den Kapiteln 9 und 10 wird ein graphenbasiertes Objektmodell präsentiert, in dem die Eigenschaften der elementaren Objektregionen in den Knoten des Graphen zusammengefasst sind und die räumlichen Beziehungen zwischen den Regionen mit Kanten im Graph repräsentiert werden. Der Segmentierungsalgorithmus wird mit einer Objektdetektion erweitert, welche im Eingabebild nach dem benutzerdefinierten Objektmodell sucht. Wir präsentieren zwei Algorithmen zur Objektdetektion. Der erste ist spezialisiert auf Zeichentricksequenzen und benutzt einen Algorithmus zur effizienten Suche von Teilgraphen, wohingegen der zweite reale Videosequenzen verarbeitet. Mit der Erweiterung um Objektmodelle kann das Segmentierungssystem so kontrolliert werden, dass es individuelle Objekte extrahiert, selbst wenn die Eingabesequenz mehrere Objekte enthält.

Kapitel 11 schlägt einen alternativen Ansatz vor um Objektmodelle in einen Segmentierungsalgorithm zu integrieren. Das Kapitel beschreibt einen halbautomatischen Segmentierungsalgorithmus, bei dem der Benutzer das Objekt grob markiert und der Computer dies zur exakten Objektkontur verfeinert. Anschliessend wird das Objekt automatisch durch die Sequenz verfolgt. In diesem Algorithmus wird das Objektmodell als die Textur entlang der Objektkontur definiert. Diese Textur wird im ersten Bild extrahiert und dann während der Objektverfolgung benutzt, um das ursprüngliche Objekt wiederzufinden. Der Kern des Algorithmus benutzt eine Graphdarstellung des Bildes und einen neu entwickelten Algorithmus zur Berechnung kürzester zirkulärer Pfade in planaren Graphen. Der vorgeschlagene Algorithmus ist schneller als die derzeit bekannten Algorithmus für dieses Problem und er kann ebenso für viele andere Probleme benutzt werden, wie z.B. dem Vergleich von Objektformen.

Teil III der Arbeit widmet sich verschiedenen Techniken um Informationen über die physische 3-D Welt aus der Kamerabewegung abzuleiten. Im Segmentierungssystem haben wir die Bewegung der Kamera geschätzt, allerdings hatten die berechneten Parameter keine direkte physikalische Bedeutung. Kapitel 12 diskutiert eine Erweiterung für die Schätzung der Kamerabewegung, um die Bewegungsparameter mit Techniken der Selbstkalibrierung in physikalisch bedeutungsvolle Parameter (wie Drehwinkel oder Brennweite) zu faktorisieren. Die Spezialität des Algorithmus ist, dass er mit Hilfe der Multi-Sprite-Technik Kamerabewegungen verarbeiten kann, die sich über mehrere Sprites erstrecken. Folglich kann der Algorithmus für beliebige drehende Kamerabewegungen angewendet werden.

Für die Analyse von Videosequenzen ist es oft erforderlich, die Position von Objekten zu bestimmen und zu verfolgen. Natürlich liefert die Objektposition in Bildkoordinaten wenig Informationen falls die Blickrichtung der Kamera unbekannt ist. Kapitel 13 beschreibt einen neuen Algorithmus um die Transformation zwischen Bildkoordinaten und Weltkoordinaten für die Spezialanwendung der Sportvideoanalyse zu bestimmen. In Sportvideos kann die Kameraansicht von Markierungen auf dem Spielfeld abgeleitet werden. In diesem Sinne benutzen wir ein Modell des Spielfeldes, welches die Anordnung der Linien beschreibt. Nach der Extraktion der wesentlichen Linien im Eingabebild wird eine kombinatorische Suche durchgeführt um Korrespondenzen zwischen den Linien im Eingabebild und den Linien im Modell herzustellen. Der Algorithmus benötigt keine Information über die spezifische Spielfeldfarbe und ist sehr robust gegenüber Verdeckungen oder ungünstigen Beleuchtungsverhältnissen. Des weiteren ist der Algorithmus generisch in dem Sinne, dass er an jede Sportart angepasst werden kann, indem lediglich das Spielfeldmodell ausgetauscht wird.

In Kapitel 14 betrachten wir wieder Hintergrundpanoramas and konzentrieren uns dabei speziell auf deren Visualisierung. Abgesehen von den ebenen Hintergrund-Sprites, die oben diskutiert wurden, sind Projektionen auf Zylinderoberflächen, die danach zu einem rechteckigen Bild ausgerollt werden, eine gebräuchliche Darstellungstechnik. Der Nachteil dieses Ansatzes ist jedoch, dass der Betrachter sich im Panoramabild nicht gut orientieren kann, da er gleichzeitig in alle Richtungen schaut. Um eine intuitivere Darstellung für weitwinklige Ansichten bereitzustellen, haben wir eine Darstellungstechnik entwickelt, die für Innenraumansichten spezialisiert ist. Wir präsentieren einen Algorithmus, um die 3-D Form des Raumes zu bestimmen, in dem das Bild aufgenommen wurde, oder, allgemeiner, um den kompletten Grundriss zu berechnen, falls Panoramabilder von jedem der Räume zur Verfügung stehen. Basierend auf der ermittelten 3-D Geometrie wird ein graphisches Modell des Raumes erstellt, wobei die Wände Texturen aus den Panoramabildern zugewiesen bekommen. Diese Darstellung erlaubt es, virtuelle Begehungen im rekonstruierten Raum durchzuführen und ermöglicht dadurch dem Betrachter eine verbesserte Orientierung.

Zusammenfassend können wir feststellen, dass sämtliche Segmentierungstechniken eine gewisse Definition für Vordergrundobjekte benutzen. Diese Definitionen sind entweder explizit durch Objektmodelle gegeben, wie in Teil II der Arbeit, oder sie sind implizit definiert, wie z.B. durch die Hintergrundssynthese aus Teil I. Die Ergebnisse dieser Arbeit zeigen, dass implizite Beschreibungen, die ihre Definition aus dem Videoinhalt selbst ableiten, gut funktionieren, wenn die Sequenz lang genug ist, diese Information zuverlässig zu extrahieren. Es ist jedoch schwierig, höhere Se-

mantik in Segmentierungsansätze zu integrieren, die auf impliziten Modellen aufbauen. In diesem Fall sollte die Semantik stattdessen in Nachverarbeitungsschritten hinzugefügt werden. Explizite Objektmodelle bringen dagegen das semantische Vorwissen früh in den Segmentierungsvorgang ein. Desweiteren können sie auf kurze Videosequenzen oder sogar Standbilder angewendet werden, da kein Hintergrundmodell aus dem Video extrahiert werden muss. Die Definition einer allgemeinen Objektmodellierungstechnik, die breit anwendbar ist und die auch eine genaue Segmentierung ermöglicht, bleibt ein wichtiges aber anspruchsvolles Problem für die weitere Forschung.

Acknowledgments

The research for this thesis was carried out in three research groups, starting in Mannheim at the Circuitry and Simulation group, headed by Prof. Peter de With. After Prof. de With changed his position to the Technical University of Eindhoven, I could continue my research at the Computer Science IV group in Mannheim, headed by Prof. Wolfgang Effelsberg. Finally, I finished the research at the Video Coding and Architectures group at the Technical University of Eindhoven.

However, seen in a more global picture, the story begins much earlier. First, I want to thank my parents for always supporting me and opening the possibility to receive a good education. The basis for my technical education was provided by the University of Stuttgart, where I appreciate especially the time I spent at the Image Understanding group of Prof. Levy for writing my study thesis and master thesis. In particular, I also thank Niels Mache, who supervised these works and with whom I still have a continuing friendship and exchange of ideas.

After the studies in Stuttgart, I joined the Circuitry and Simulation group at the University of Mannheim. I am thankful to Peter de With, who opened the opportunity for me to work towards a PhD on an exciting topic. I also thank Wolfgang Effelsberg for inviting me to join his group and continue my research there without having to change the topic. In both groups I enjoyed to have the full freedom of research and to be allowed to explore any topic that I was interested in. Special thanks go to Peter de With for the detailed reviewing of every paper that we have published, for his endless support and confidence. Finally, I want to express thanks to all my former colleagues in Mannheim and my current colleagues in Eindhoven for the very friendly working environment.

In 2003, Wolfgang Effelsberg and Roy Pea from the Stanford University made it possible that I could spend some time at the Stanford Center for Innovations in Learning (SCIL) to work on panoramic-video processing. I thank both of them for this unique opportunity. This stay has been a valuable experience for me, and I enjoyed the warm working atmosphere

in the project team, which included Roy Pea, Mike Mills, Joe Rosen, and many others. The ideas developed there have lead to Chapter 14.

In the past years, numerous students worked on their study and master thesis under my supervision. Thereby, they helped substantially in the implementation and development of new ideas. The valuable student work is too numerous to be listed exhaustively, but in particular, I want to thank (in temporal order) Alexander Staller, Thomas Brox [12], Michael Käsemann [64], Susanne Krabbe [107], Stefan Birringer [10], Stefan Seedorf [165], Tobias Gleixner [75], Holger Peinsipp [143], Magnus Pfeffer [146], Christian Thiel [181], and Sascha Finke [71].

It is also noteworthy that all research and writing of this thesis was carried out exclusively with open-source software. I am indebted to everyone working on these wonderful software projects, especially the programs gcc, emacs, the Linux kernel, LaTeX, and tgif.

Finally, I want to thank the promotion committee for reviewing the thesis. In particular, the very detailed reading and endless flow of comments provided by Peter de With and Wolfgang Effelsberg were very helpful. I take full responsibility for any remaining non-conform hyphenation and misplaced comma.

Biography

Dirk Farin was born in Tübingen, Germany in 1973. He graduated in computer science and electrical engineering from the University of Stuttgart, Germany, in 1999. Subsequently, he became research assistant at the Department of Circuitry and Simulation at the University of Mannheim, where he started his research on video-object segmentation. He joined the Department of Computer Science IV at the University of Mannheim in 2001. Since 2004, he has a post-doc position in the Video Coding and Architecture group at the Technical University of Eindhoven, Netherlands. Apart from video-object segmentation, his research interests include video compression, content analysis, and 3-D reconstruction. Currently, he is involved in a joint project of Philips and the Technical University of Eindhoven about the development of video capturing and compression systems for 3-D television. He received a best student paper award at the SPIE Visual Communications and Image Processing conference in 2004 for his work on multi-sprites, and two best student paper awards at the Symposium on Information Theory in the Benelux in 2001 and 2003. He is member of the program committee of the IEEE International Conference on Image Processing and reviewer for several journals including IEEE Multimedia and IEEE Circuits and Systems for Video Technology. In 2005, he organized a special session about sports-video analysis at the IEEE International Conference on Multimedia and Expo. Mr. Farin developed popular open-source and commercial software including an MPEG-2 decoder, two MPEG-2 encoders, libraries with computer-vision algorithms, and image-format conversion software.