Chapter 10

Conclusions and perspectives

This chapter concludes the dissertation by summarizing the main developments and results of this work. Possible directions for further research and indications for potential applications are given as well.

10.1 Summary

This thesis analyzes the problem of video compression with content-based functionalities in the framework of segmentation-based video coding systems. Two major problems have been considered. The first one is related with coding optimality in segmentation-based coding systems. Regarding this subject, the feasibility of a rate-distortion approach for a complete region-based coding system has been shown. The second one is how to address content-based functionalities in the coding system proposed as a solution of the first problem.

Optimality, as defined in the framework of rate-distortion theory, deals with obtaining a representation of the video sequence that leads to a minimum distortion of the coded signal for a given bit budget. In the case of segmentation-based coding systems this means to obtain an 'optimal' partition together with the best coding technique for each region of this partition so that the result is optimal in an operational rate-distortion sense. In Chapter 5 a formal description of the problem has been given for independent, non-scalable coding. An algorithm to solve the problem in the general case has been also proposed. The major concern of Chapter 6 is the application of this algorithm to a specific segmentation-based coding system, the so called SESAME. In SESAME, each frame is segmented into a set of regions, that are coded independently. Segmentation involves both spatial and motion homogeneity criteria. To exploit temporal redundancy, a prediction for both the partition and the texture of the current frame is created by using motion information. The time evolution of each region is defined along the sequence (time tracking). The results are optimal (or near-optimal) for the given framework in a rate-distortion sense. The definition of the coding strategy involves a global optimization of the partition as well as of the coding technique/quality level for each region. Experimental results have been presented for different test sequences, along with a comparison (in terms of compression performance) of SESAME with MPEG-4 and H.264.

In Chapter 7 the investigation has been extended to the problem of video coding optimization in the framework of a system that can address content-based functionalities. The focus has been in the various types of content-based scalability and object tracking. The generality of the problem has also been extended by trying to include the spatial and temporal dependencies between frames and scalability layers into the optimization schema. In this case the solution implies finding the optimal partition and set of quantizers for both the base and the enhancement layers. Due to the coding dependencies of the enhancement layer with respect to the base layer, the partition and the set of quantizers of the enhancement layer depend on the decisions made on the base layer. As this results in solutions that are extremely costly from the computational point of view, the study has been limited to sets of two frames or scalability layers. Also, a solution for the independent optimization problem (i.e. without tacking into account dependencies between different frames of scalability layers) has been proposed to reduce the computational complexity.

In Chapter 8 these solutions have been used to extend the SESAME coding system. The extended coding system, named XSESAME, supports different types of scalability (PSNR, Spatial and temporal) as well as content-based functionalities, such as content-based scalability and object tracking. The structure of XSESAME is an evolution of the structure of SESAME. Images are segmented into regions and prediction is used to exploit temporal redundancy as in the former coding system. In the enhancement layer, information can be refined by residual error coding or by using a finer partition. Additionally, methods to exploit spatial redundancy between base and enhancement scalability layers have been provided, as well as a new prediction mode to deal with temporal redundancy between enhancement layers. To properly address content-based functionalities, some modifications have been done in both the construction of the Projected Partition and the Partition Tree in order to ensure that the contours of the selected object are available at the Partition Tree. Two different operating modes for region selection in the enhancement layer have been presented: One (supervised) aimed at providing content-based functionalities at the enhancement layer and the other (unsupervised) aimed at coding efficiency, without content-based functionalities.

In order to improve the coding efficiency of partitions, the extension of the Multi-Grid Chain Code contour coding technique to inter-frame and scalable modes has been investigated. The results presented in this work show that, while the coding efficiency of MGCC (lossy) is superior to CC (lossless) in intra-mode coding, the lossy nature of MGCC introduces some problems that result in a performance degradation due to necessary side information that has to be sent to properly decode the contours. This overhead causes the inter and scalable MGCC technique to perform worse than inter and scalable CC.

Integration of object tracking into the segmentation-based coding system has been investigated in Chapter 8. In the general case, tracking is a very complex problem. If this capability has to be integrated into a coding system, additional problems arise due to conflicting requirements between coding efficiency and tracking accuracy. This has been solved by using a double partition approach, where pure spatial criteria are used to re-segment the partition used for coding. The projection of the re-segmented partition results in more precise adaption to object contours. A merging step is performed *a posteriori* to eliminate the excess of regions originated by the re-segmentation. Experimental results for different test sequences show a good performance, even in sequences with complex motion.

Experimental results for the different types of scalability have been provided in Chapter 8, demonstrating the different scalability modes. Comments on the operation of the algorithm can also be found in this chapter, such as the number of regions used or the types of texture coding techniques that are selected. A comparison with MPEG-4 is also provided. The operation of dependent and independent optimization modes have been compared. Dependent optimization allows to obtain a big gain in coding efficiency when comparing with the independent optimization algorithm.

Finally, comments about all the work are provided in Section 10.2 and some possible lines for the future work are outlined in Section 10.3.

10.2 Conclusions

Video compression is a research field that has been very active in the last years, with a large number of people involved and with the approval of standards such as MPEG-4 and H.264. The improvements in raw coding efficiency have been immediately used in many applications, such as video streaming over the Internet or wireless channels. Content-based representation has been also a very active field. MPEG-4 has been the first video coding standard supporting content-based representation. However, MPEG-4 is a block-based coding system. Texture, motion and shape information is encoded and transmitted in a block by block basis. This mixed scheme may restrict separate access to the different information fields

The work that has been presented in this thesis shows that, by changing completely the coding paradigm from blocks to regions, content-based functionalities can be supported more easily by the coding system. It has been shown how the proposed segmentation-based coding system can address properly content-based functionalities such as object tracking and object-based scalability. Segmentation-based coding systems were introduced to overcome the problems of block-based systems in terms of coding efficiency. Until the moment, and except for very concrete cases, they can not match the performance of today's reference in coding efficiency, namely H.264. This is natural if the difference in the amount of research invested in both fields along the years is considered.

However, in the work presented here, this ability to represent content-based functionalities comes at the expense of increased complexity at the encoder. XSESAME encoder is significantly more computationally intensive than the MoMuSys encoder, even without the new prediction modes of the later.

10.2.1 Details of the encoder operation:

Important information about the operation of the encoder can be derived from the analysis of all the tests performed.

- *Number of regions*: Usually, partitions with a low number of regions are selected (even with only one region covering the entire frame), both in the base and in the enhancement layers. In many cases, the partition in the enhancement layer is the same as in the base layer and the improvement is achieved by refining the texture through residual error coding. In sequences where motion is important, more regions are used. This behavior is common to all types of scalability (PSNR, Spatial and Temporal).

  An increase in the bit-rate usually results in a decrease in the average number of regions per partition, specially in full frame mode. Using more regions require an increased percentage of the bit budget for coding the contours of these regions, at the expense of texture coding. This is less effective from an R-D point of view. In object functionalities mode, the number of regions is heavily influenced by the need to include all the contours of the object, so dependencies on bitrate are less evident.

- *Texture coding techniques*: For the base layer, the behavior is common to all scalability types and operation modes: Inter-frame techniques are selected more often than intra-frame techniques, specially at low bitrates. In sequences with a large motion or when the frame-rate is decreased, the use of intra-mode techniques becomes more noticeable.

  In the enhancement layer, the behavior depends on the scalability type and also on the operation mode: In PSNR mode, both inter-frame and layer intra techniques are widely used, with some advantage for inter-frame techniques. In Spatial scalability, layer intra techniques are less used because they require an interpolation step that reduces its performance. In object functionalities mode and for all types of scalability, inter-frame techniques are less used than in full frame scalability. The higher number of regions in object functionalities mode makes inter-frame techniques less efficient for small regions due to the cost of the motion vectors. Temporal scalability uses mainly inter-frame techniques.

- *Coding efficiency*: The analysis of the coding efficiency of the XSESAME coding system shows that there is much room for improvement in this point. In particular, some of the improved prediction techniques used in H.264 and MPEG-4 could be adopted.

In this thesis, the efforts to improve the coding efficiency have been devoted to three main points, using a formulation specific to segmentation-based coding systems:

- The use of a Rate-Distortion framework to drive the bit-allocation. By comparison with the region-based coding algorithm presented in [119] using a much simpler bit allocation rule, the current algorithm provides a gain of between 2 and 3 dB of PSNR for very low bit rates (30-40 kbps) [117].
- Improving partition coding: While prediction techniques and texture coding techniques have evolved enormously in the last years, the field of efficient partition coding in the general case (this is, with an arbitrary number of regions) is still open to many improvements. This was the motivation for concentrating our efforts in this field. In this thesis, the possibility to extend the Multi-Grid Chain Code (MGCC) contour coding technique to inter-frame mode and layer intra mode has been studied. Lossy techniques can provide better performance than lossless techniques but they must deal with the problem of preserving the partition topology. This is specially difficult in the case of generic partitions. The proposed approach to inter-frame and layer intra MGCC allows to solve this problem but the extra information necessary to preserve partition topology reduces the performance, specially for dense partitions.
- Study of a dependent optimization algorithm. Conclusions for this subject are given in Section 10.2.3.

10.2.2 Object tracking

Experimental results show that a good performance can be achieved in most sequences. Objects can be tracked with accuracy even in sequences with complex motion. However, in some cases, regions from the background are mistakenly included in the selected object. To avoid this kind of errors, any high level information about the object (color, shape, ...) could be used to help the decision on whether a regions belongs to the object or to the background.

10.2.3 Dependent optimization

Dependent optimization improve significantly the coding efficiency (between 0.7 and 2 dB), depending on the scalability type (PSNR, spatial or temporal) and also on the operation mode (full frame or object functionalities). Larger gains are obtained for object functionalities mode due to the larger number of regions in this mode.

In PSNR scalability, the main difference between the dependent and the independent optimization algorithms is that dependent optimization allocates more bits to the enhancement layer at the expense of the base layer. This is because the goal is to maximize the quality of the enhancement layer. In Temporal scalability, the bit budget is shared almost equitatively between both layers because the optimization algorithm is oriented towards maximizing the PSNR averaged between the base and enhancement layers. Spatial scalability is where the gains are less evident because, being the base layer encoded at a lower resolution, a lower amount of the bit budget is usually allocated to this layer. The contribution of the base layer to the quality of the enhancement layer is lower than in the PSNR case, for example. Therefore, the number of choices of the dependent optimization algorithm are reduced and smaller gains are obtained.

The dependent optimization algorithm usually selects partitions with less regions than the independent optimization algorithm, specially in full frame mode. In object functionalities mode, the number of regions is heavily conditioned by the need to include all the contours of the object. This reduces the number of choices of the algorithm and the reduction of the number of regions is less noticeable.

The problem with the dependent optimization approach is the high computational cost. Even with reduced sets of two frames/enhancement layers, the encoding time increases approximately 20 times with respect to the independent approach. This larger complexity currently limits the usefulness of dependent optimization.

10.3 Possible extensions

In this section, some extensions to the work presented in this thesis are given. The extensions can be grouped into three categories:

- Improvement of coding efficiency: While this coding system is more biased to applications relying in content-based functionalities than to raw compression efficiency, improving coding efficiency would increase the usability of the system, for example if these applications are to be used over low bandwidth channels such as UMTS. This can be done by improving temporal and spatial prediction or using better coding techniques. Tools similar to the ones already used in MPEG-4 or H.264 could be adapted to this coding system, namely improved motion estimation, use of B-frames, AC/DC prediction, new wavelet texture coding ...
- Modifications on the structure of the Partition Tree: Currently, the Partition Tree is structured as a multiway tree (that is, a tree with any number of children for each node). The number of children for each node is given by the segmentation algorithm, that decides how to split a given regions according to homogeneity (size or contrast) criteria. In some situations, a given region can be segmented in many components. In this case, the high cost of the contours of the children nodes makes almost impossible for the optimization algorithm to select them. One possible solution would be to mod-

ify the segmentation algorithm to limit the number of components resulting from the segmentation of any given region. Additionally, other data structures to represent the Partition Tree could be investigated. One of the most promising are Binary Partition Trees (BPT). This kind of region-oriented image representation has successfully been used for segmentation applications [114].

- Improvement of existing content-based functionalities: Tracking could be improved by refining the criterion to include/reject regions in the borders of the object. For instance, model-based criteria could be used on these regions. An improvement in the motion estimation algorithm would possibly help improving the tracking capability.

  Extension of the criteria for the construction of Partition Tree. Currently, the definition of the levels of the Partition Tree is based on spatial homogeneity (splitting levels) and motion similarity (merging levels). Model-based algorithms or image analysis tools could be used to drive or influence these processes (for instance, identify regions from human faces and do not allow these regions to be merged with non-face regions).

  User interactivity is a key point in content-based functionalities. This interactivity could materialize in the possibility to drive the construction of the Partition Tree (for instance, by forcing some regions to merge or preventing the merging of other regions) or the construction of the enhancement layer by modifying the selection mask. Another example would be the use of model-based techniques as described in the previous paragraph where the model can be refined by user interaction and perhaps learn from this interaction.

- New functionalities: New scalability modes based on combinations of the existing modes (for instance, spatio-temporal scalability) could be introduced.

  Integration with MPEG-7 [61, 4]. The tracking capabilities of XSESAME can allow its use as a description generation tool. For instance, the tracking capabilities allow to follow a given object so that information about the presence and position of this object along the sequence can be obtained.

  The use of a separate coding syntax facilitate such application. On the other hand, it would be worth investigating how MPEG-7 descriptions could be used at the decoder for video post-processing applications or to improve the coding efficiency of the video coder [111].

Bibliography

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(4):384–401, July 1985.

[2] J. K. Aggarwal and W. N. Martin. Analyzing dynamic scenes containing multiple moving objects. In T. S. Huang, editor, *Image sequence analysis*, pages 355–380. Springer-Verlag, New York, 1981.

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

[4] O. Avaro and P. Salembier. MPEG-7 systems: Overview. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6):760–764, June 2001.

[5] H.J. Barnard. *Image and Video Coding Using a Wavelet decomposition.* PhD thesis, Technische Universiteit Delft, Faculteit der Elektrotechniek, Vakgroep Informatietheorie, 1994.

[6] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, 1957.

[7] T. Berger. *Rate-Distortion Theory. A Mathematical Theory for data compression.* Prentice-Hall, 1971.

[8] M. Bertero, T.A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–887, 1988.

[9] S. Beucher and F. Meyer. The morphological approach to segmentation: the watershed transformation. In E. Dougherty, editor, *Mathematical morphology in image processing*, chapter 12, pages 433–481. Marcel Dekker, 1993.

[10] P. Bouthemy and E. Franois. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Internationa Journal on Computer Vision*, 10(2):157–182, 1993.

[11] P. J. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

[12] R. Castagno, T. Ebrahimi, and M. Kunt. Video segmentation based on multiple features for interactive multimedia applications. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):562–571, September 1998.

[13] E. Chang and A. Zakhor. Variable bit rate MPEG video storage on parallel disk arrays. In *First International Workshop on Community Networking Integrated Multimedia Services to the Home*, pages 127–137, San Francisco, USA, July 1994.

[14] L. Chiariglione. MPEG and multimedia communications. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):5–18, February 1997.

[15] H.264/AVC Software Coordination. http://bs.hhi.de/ suehring/tml/.

[16] I. Corset, L. Bouchard, S. Jeannin, P. Salembier, F. Marqués, M. Pardàs, R. Morros, F. Meyer, and B. Marcotegui. Segmentation-based coding system allowing the manipulation of objects (SESAME). Technical Report ISO/IECJTC1/SC29/WG11/MPEG95/408, LEP, UPC, CMM, November 1995.

[17] G. Côté, B. Erol, M. Gallant, and F. Kossentini. H.263+: Video coding at low bit rates. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(7):849–866, November 1998.

[18] Y. Deng and B.S. Manjunath Netra-v. Toward an object-based video representation. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):616–627, September 1998.

[19] M. Dudon, O. Avaro, and C. Roux. Triangular active mesh for motion estimation. *Signal Processing: Image Communication*, 10:21–41, 1997.

[20] F. Dufaux and F. Moscheni. Spatio-temporal segmentation based on motion and static segmentation. In *IEEE International Conference on Image Processing, ICIP'95*, Washington, DC, October 1995.

[21] F. Dufaux and F. Moscheni. Segmentation-based motion estimation for second generation video coding techniques. In L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*, pages 79–124. Kluwer Academic Publishers, 1996. ISBN: 0 7923 9680 4.

[22] J.L. Dugelay and H. Sanson. Differential methods for the identification of 2D and 3D motion models in image sequences. *Signal Processing: Image Communication*, 7:105–127, 1995.

[23] J.G. Dunham. Optimum unifirm piecewise linear approximation of planar curves. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8:67–75, January 1986.

[24] T. Ebrahimi and M. Kunt. Visual data compression for multimedia applications. *Proceedings of the IEEE*, 86(6), June 1998.

[25] Michelle Effros. Optimal modeling for complex system design. *IEEE Signal Processing Magazine*, 15(6):51–73, November 1998.

[26] Z. Li et al. Adaptive Basic Unit Layer Rate Control for JVT. Technical report, Pattaya, Thailand, March 2003.

[27] Z. Li et al. Proposed draft of adaptive rate control. Technical report, Geneva, May 2003.

[28] H. Everett. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11:399–417, March 1963.

[29] G.D. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61:268–278, March 1973.

[30] H. Freeman. On the coding of arbitrary geometric configurations. *IRE Trans. Electronic Comp.*, EC(10):260–268, June 1961.

[31] N. García, F. Jaureguizar, and J.I. Ronda. Pixel-based video compression schemes. In L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*, pages 31–78. Kluwer Academic Publishers, 1996. ISBN: 0 7923 9680 4.

[32] A. Gasull, F. Marqués, and J. A. García. Lossy image contour coding with multiple grid chain code. In *Workshop on Image Analysis and Synthesis in Image Coding94, WIASIC'94*, pages B4.1–B4.4, Berlin, Germany, October 1994.

[33] P. Gerken. Object-based analysis-synthesis coding of image sequences at very low bit rates. *IEEE Trans. on Circuits and Systems for Video Technology*, 4(3):228–235, June 1994.

[34] M. Gilge. Region-oriented texture coding. In L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*, pages 171–218. Kluwer Academic Publishers, 1996. ISBN: 0 7923 9680 4.

[35] M. Gilge, T. Engelhardt, and Mehlan R. Coding of arbitrarely shaped image segments based on a generalized orthogonal transform. *EURASIP, Image Communications*, 1(2):153–180, October 1989.

[36] C. Gomila. Codificació de particions mitjanant Multi-Grid Chain Code. Master's thesis, ETSET Barcelona, UPC, Barcelona,Spain, February 1997.

[37] R.M. Gray. *Source Coding Theory.* Kluwer Academic Publishers, Norwell, MA, 1990.

[38] C. Gu. 3D contour image coding by morphological filters and motion estimation. In IEEE, editor, *International Conference on Acoustics, Speech and Signal Processing, ICASSP'94*, Australia, April 1994.

[39] M. Hotter and R. Thoma. Image segmentation based on object-oriented mapping parameter estimation. *Signal Processing*, 15(3):315–334, October 1988.

[40] C.L. Huang and C.Y. Hsu. A new motion compensation method for image sequence coding using hierarchical grid interpolation. *IEEE Trans. on Circuits and Systems for Video Technology*, 4(1):42–52, February 1994.

[41] ISO-IEC/SC29/WG11/NO400. *Test Model 5*, April 1993. Document AVC-491b, Document 2.

[42] ITU-T. Recommendation H.261. Video codec for audiovisual services at px64 kbit/s. March 1993.

[43] ITU-T. Recommendation H.263. Video coding for low bitrate communication. March 1996.

[44] ITU-T. Recommendation H.263 v2 (H.263+). Video coding for low bitrate communication. February 1998.

[45] ITU-T. Recommendation H.263 v3 (H.263++). Video coding for low bitrate communication. November 2000.

[46] E. Jensen, K. Rijkse, I. Lagendijk, and P. van Beek. Coding of arbitrarily sahaped image segments. In *Proceedings of Workshop on Image Analysis and Synthesis in Image Coding*, Berlin, October 1994.

[47] ISO/IEC JTC1/SC29/WG11. MPEG-4 Video Verification Model version 18.0. January 2001.

[48] MPEG-4 Committee Draft ISO/IEC JTC1/SC29/WG11. Coding of moving pictures and audio. ISO/IEC 14496-2. October 1998.

[49] G. Karlsson and M. Vetterli. Three dimensional subband coding of video. In *International Conference on Acoustics, Speech and Signal Processing*, pages 1100–1103, New York, April 1988.

[50] H. Katata, N. Ito, T.Aono, and H.Kusao. Object Wavelet Transform for coding of arbitrarily-shaped image segments. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):234–237, February 1997.

[51] A. Kaup. Object-based texture coding of moving video in MPEG-4. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(1):5–15, February 1999.

[52] M. Kim, J.G. Choi, D. Kim, M.H. Lee, C. Ahn, and Y.S. Ho. A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(8):1216–1226, December 1999.

[53] R. Kresch and D. Malah. Quadtree and bitplane decompositions as particular cases of the generalized morphological skeleton. In IEEE, editor, *1995 IEEE Workshop on Nonlinear Signal and Image Processing*, Halkidiki, Greece, June 1995.

[54] S. Kruse, A. Graffunder, and S. Askar. A new tracking scheme for semi-automatic video object segmentation. In *Proceedings of the Workshop on Image Analysis for Multimedia Interactive Services*, pages 93–96, Berlin, Germany, May 1999.

[55] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second generation image coding techniques. *Proceedings of the IEEE*, 73(4):549–575, April 1985.

[56] Hung-Ju Lee, T. Chiang, and Ya-Qin Zhang. Scalable rate control for MPEG-4 video. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(6):878–894, September 2000.

[57] J. Lee. Rate-distortion optimization of parametrized quantization matrix for MPEG-2 encoding. In *IEEE International Conference on Image Processing, ICIP'98*, Chicago, USA, October 1998.

[58] J. Lee and B.W. Dickinson. Joint optimization of frame type selection and bit allocation for MPEG video coders. In *IEEE International Conference on Image Processing, ICIP'94*, pages 962–966, Texas, U.S.A., November 1994.

[59] A.S. Lewis and G. Knowles. Video compression using 3D wavelet transform. *IEE Electronic Letters*, 26(6):396–398, March 1990.

[60] Liang-Jin Ling, A. Ortega, and C.-C. Jay Kuo. A gradient-based rate control algorithm with applications to MPEG video. In *IEEE International Conference on Image Processing, ICIP'95*, pages 392–395, Washington, DC, October 1995.

[61] B. Manjunath, P. Salembier, and T. Sikora, editors. *Introduction to MPEG-7*. Wiley, 2002.

[62] B. Marcotegui, P. Correia, F. Marqués, R. Mech, R. Rosa, M. Wollborn, and F. Zanoguera. A video generation tool allowing friendly user interactions. In *IEEE International Conference on Image Processing, ICIP'99*, page 26AP4.10, Kobe, Japan, October 1999.

[63] B. Marcotegui, F. Marqués, and F. Meyer. Allowing content-based functionalities in segmentation-based coding schemes. *Annales des Télécomunications*, 52(7-8):380–387, July-August 1997.

[64] B. Marcotegui and F. Meyer. Morphological segmentation of image sequences. In J. Serra and P. Soille, editors, *Mathematical morphology and its applications to image processing*, pages 101–108. Kluwer Academic Publishers, 1994.

[65] F. Marqués. Motion stability in image sequence segmentation using the watershed algorithm. In P. Maragos, R.W. Schafer, and M.A. Butt, editors, *Third workshop on Mathematical morphology and its applications to image processing*, pages 321–328. Kluwer Academic Publishers, Atlanta, USA, May 1996.

[66] F. Marqués, S. Fioravanti, and P. Brigger. Coding of image partitions by morphological skeletons using overlapping structuring elements. In IEEE, editor, *1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pages 250–253, Halkidiki, Greece, June 20-22 1995.

[67] F. Marqués and A. Gasull. Partition coding using multigrid chain code and motion compensation. In *IEEE International Conference on Image Processing, ICIP'96*, pages II.935–II.938, Lausanne, Switzerland, September 1996.

[68] F. Marqués, C. Gomila, and A. Gasull. Partition coding method and device. International Patent 99.400436.4 SPID:PHF 99511 EP-P CLA/CLA/LAND-GG, February 1999.

[69] F. Marqués, C. Gomila, and A. Gasull. Partition decoding method and device. International Patent 99.401761.4 SPID:PHF 99561 EP-P CLA/CLA/LAND-GG, February 1999.

[70] F. Marqués and J. Llach. Tracking of generic objects for video object generation. In *IEEE International Conference on Image Processing, ICIP'98*, pages 628–632, Chicago, USA, October 1998.

[71] F. Marqués and C. Molina. Object tracking for content-based functionalities. In *SPIE Visual Communications and Image Processing VCIP'97*, San Jose (CA), USA, February 1997.

[72] F. Marqués, M. Pardàs, and P. Salembier. Coding-oriented segmentation of video sequences. In L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*, pages 79–124. Kluwer Academic Publishers, 1996. ISBN: 0 7923 9680 4.

[73] F. Marqués, P. Salembier, M. Pardàs, R. Morros, I. Corset, S. Jeannin, B. Marcotegui, and F. Meyer. A segmentation-based coding system allowing manipulation of objects

(sesame). In *IEEE International Conference on Image Processing, ICIP'96*, Lausanne, Switzerland, September 1996. (Invited paper).

[74] F. Marqués, J. Sauleda, and T. Gasull. Shape and location coding for contour images. In *Picture Coding Symposium*, pages 18.6.1–18.6.2, Lausanne, Switzerland, March 1993.

[75] F. Marqués, V. Vera, and A. Gasull. Recursive image sequence segmentation by hierarchical models. In *Proc. of the 12th International Conference on Pattern Recognition*, pages 523–525, Oct 1994.

[76] R. Mech and M. Wollworn. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera. *EURASIP, Signal Processing*, 66(2):203–217, April 1998.

[77] T. Meier and K.N. Ngan. Automatic segmentation of moving objects for video object plane generation. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):525–538, September 1998.

[78] T. Meier and K.N. Ngan. Video segmentation for content-based coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(8):1190–1203, December 1999.

[79] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21–46, September 1990.

[80] T. Minami and K. Shinohara. Encoding of line drawings with multiple grid chain code. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8:265–276, March 1986.

[81] J.L. Mitchell, W.B. Pennebaker, C. Fogg, and D.J. LeGall. *MPEG Video Compression Standard.* Chapman and Hall, New York, USA, 1997.

[82] J.R. Morros, F. Marqués, M. Pardàs, and P. Salembier. Video sequence segmentation based on rate-distortion theory. In *SPIE Visual Communications and Image Processing VCIP'96*, volume 2727, pages 1185–1196, Orlando (FL), USA, March 1996.

[83] R. Morros. WWW, 2004.

[84] F. Moscheni, S. Bhattacharjee, and M. Kunt. Spatiotemporal segmentation based on region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:897–915, September 1998.

[85] MPEG-4 Reference Software FDIS. Document ISO/IEC JTC1/SC29/WG11/N2805 48th MPEG meeting, Vancouver, July 1999.

[86] H.G. Musmann, M. Hotter, and J. Ostermann. Object-oriented analysis-synthesis coding of moving images. *Signal Processing: Image Communication*, 1(2):117–138, October 1989.

[87] H.G. Musmann, P. Pirsch, and H.J. Grallert. Advances in picture coding. *Proceedings of the IEEE*, 73(4):523–548, April 1985.

[88] A.N. Netravali and J.O. Limb. Picture coding: a review. *Proceedings of the IEEE*, 68(3):366–406, March 1980.

[89] H. Nicolas, S. Pateux, and D. Le Guen. Minimum description length criterion and segmentation map coding for region-based video compression. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(2):184–198, February 2001.

[90] J.M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *EURASIP, Signal Processing*, 66(2):143–156, April 1998.

[91] A. Ortega and K. Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50, November 1998.

[92] P. Van Otterloo. *A contour-oriented approach for shape analysis.* Prentice Hall International (UK), 1991.

[93] M. Pardàs and P. Salembier. 3D morphological segmentation and motion estimation for image sequences. *EURASIP Signal Processing*, 38(2):31–43, September 1994.

[94] M. Pardàs and P. Salembier. Joint region and motion estimation with morphological tools. In J. Serra and P. Soille, editors, *Second Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 93–100, Fontainebleau, France., September 1994. Kluwer Academic Press.

[95] M. Pardàs and P. Salembier. Time-recursive segmentation of image sequences. In EURASIP, editor, *EUSIPCO 94, VII European Signal Processing Conference*, pages 18–21, Edinburgh, U.K., September 13-16 1994.

[96] M. Pardàs, P. Salembier, and B. Gonzalez. Motion and region overlapping estimation for segmentation-based video coding. In *IEEE International Conference on Image Processing*, volume II, pages 428–432, Austin, Texas, November 1994.

[97] M. Pardàs, P. Salembier, F. Marqués, and R. Morros. Partition tree for segmentation-based video coding. In *IEEE International Conference on Acoustics, Speech & Signal Processing, ICASSP'96*, pages 1982–1985, Atlanta (GA), USA, May 1996.

[98] S. Pateux. *Segmentation spatio-temporelle et codage orienté-régions de séquences vidéo basés sur le formalisme MDL.* PhD thesis, Université de Rennes 1, 1998.

[99] Ioannis Patras, E.A. Hendriks, and R.L. Lagendijk. Video segmentation by map labeling of watershed segments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(3):326–332, March 2001.

[100] F. Pereira and T. Ebrahimi. *The MPEG-4 Book.* Prentice Hall PTR, Upper Saddle River, NJ, 2002.

[101] A.J. Pinho. A method for encoding region boundaries based on transition points. *Image and Vision Computing*, 16(3):213–218, March 1998.

[102] A. Puri and Wong. Spatial domain resolution scalable video coding. In SPIE, editor, *SPIE Visual Communications and Image Processing VCIP'93*, Boston (MA), November 1993.

[103] S. Rajala, M. Civanlar, and W. Lee. Video data compression using three-dimensional segmentation based on HVS properties. In IEEE, editor, *International Conference on Acoustics, Speech and Signal Processing*, pages 1092–1905, New York (NY), USA, 1988.

[104] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Trans. on Image Processing*, 3(5):533–545, September 1994.

[105] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distorsion sense. *IEEE Trans. on Image Processing*, 2(2):160–175, April 1993.

[106] K. Ramchandran and M. Vetterli. Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility. *IEEE Trans. on Image Processing*, 3(5):533–545, September 1994.

[107] E. Reusens. Joint optimization of representation model and frame segmentation for generic video compression. *Signal Processing*, 46:105–117, September 1995.

[108] J. Rissanen. Modeling by the shortest data description. *Automatica*, 14:465–471, 1978.

[109] J. Roese, W. Pratt, and G. Robinson. Interframe cosine transform image coding. *IEEE Transactions on Communications*, 25(11):1329–1339, November 1977.

[110] J.I. Ronda, M. Eckert, F. Jaureguizar, and N. García. Rate control and bit allocation for MPEG-4. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(8):1243–1258, December 1999.

[111] J. Ruíz and P. Salembier. Metadata-based coding tools for hybrid video codecs. In *23rd Picture Coding Symposium, PCS'2003*, pages 473–477, Saint-Malo, France, April 2003.

[112] P. Salembier. Morphological multiscale segmentation for image coding. *EURASIP Signal Processing*, 38(3):359–386, September 1994.

[113] P. Salembier. Motion compensated partition coding. In SPIE, editor, *SPIE Visual Communications and Image Processing VCIP'96*, volume 2727, Orlando, USA, March 1996.

[114] P. Salembier and L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Trans. on Image Processing*, 9(4):561–576, April 2000.

[115] P. Salembier and F. Marqués. Region-based representation of image and video: Segmentation tools for multimedia services. *IEEE Trans. on Circuit and Systems for Video Technology*, 9(8):1147–1169, December 1999.

[116] P. Salembier, F. Marqués, and A. Gasull. Coding of partition sequences. In L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*. Kluwer, 1996. ISBN: 0 7923 9680 4.

[117] P. Salembier, F. Marqués, M. Pardàs, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer, and B. Marcotegui. Segmentation-based video coding system allowing the manipulation of objects. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):60–73, February 1997.

[118] P. Salembier and M. Pardàs. Hierarchical morphological segmentation for image sequence coding. *IEEE Trans. on Image Processing*, 3(5):639–651, September 1994.

[119] P. Salembier, L. Torres, F. Meyer, and C. Gu. Region-based video coding using mathematical morphology. *Proceedings of IEEE (Invited Paper)*, 83(6):843–857, June 1995.

[120] H. Sanson. Toward a robust parametric identification of motion on regions of arbitrary shape by non-linear optimization. In *IEEE International Conference on Image Processing, ICIP'95*, pages 203–206, October 1995.

[121] L.L. Schumaker. *Spline functions: basic theory.* Wiley-Interscience, New York, 1981.

[122] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, 1982.

[123] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. Journal*, 27:379–423, 1948.

[124] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. Acoustics, Speech and Signal Processing*, 36(9):1445–1453, September 1988.

[125] T. Sikora. The MPEG-4 video standard verification model. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(2):19–31, February 1997.

[126] T. Sikora. MPEG digital video coding standards. *IEEE Signal Processing Magazine*, 14(5):82–100, September 1997.

[127] T. Sikora, S. Bauer, and B. Makai. Efficiency of shape-adaptive 2-d transforms for coding of arbitrarily shaped image segments. *IEEE Trans. on Circuits and Systems for Video Technology*, 1995.

[128] C. Stiller. Object-based estimation of dense motion fields. *IEEE Trans. on Image Processing*, 6(2):234–250, April 1997.

[129] C. Stiller and J. Konrad. Estimating motion in image sequences: A tutorial on modelling and computation of 2D motion. *IEEE Signal Processing Magazine*, 16(4):70–91, July 1999.

[130] G. Sullivan, T. Wiegand, and K.P. Lim. Joint Model reference encoding methods and decoding concealment methods; Section 2.6: Rate Control. Technical report, San Diego, September 2003.

[131] Gary J. Sullivan and Thomas Wiegand. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 15(6):74–90, November 1998.

[132] D.S. Taubman and M.W. Marcellin. *JPEG2000. Image compression fundamentals, standards and practice.* Kluwer Academic Publishers, Boston, 2002.

[133] ISO-IEC CD 13818 Information Technology. Generic coding of moving pictures and associated audio (MPEG-2). Technical report, Motion Picture Expert Group, November 1993.

[134] L. Torres and M. Kunt. *Video Coding: The second generation approach.* Kluwer Academic Publishers, Boston, 1996.

[135] G. Tziritas and C. Labit. *Motion Analysis for Image Sequence Coding.* Elsevier Science B. V., The Netherlands, 1994.

[136] Peter van Beek, A.M. Tekalp, N. Zhuang, I. Celasun, and M. Xia. Hierarchical 2-d mesh representation, tracking and compression for object-based video. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(2):353–369, March 1999.

[137] A. Vetro, H. Sun, and Y. Wang. MPEG-4 Rate Control for Multiple Video Objects. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(1):186–199, February 1999.

[138] T. Wiegand, A. Schwarz, H.and Joch, F. Kossentini, and G.J. Sullivan. Rate-constrained coder control and comparison of video coding standards. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(7):688– 703, July 2003.

[139] T. Wiegand and G. Sullivan. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10

AVC). Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Pattaya, Thailand, March 2003.

[140] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video Coding Standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(7):560–576, July 2003.

[141] Thomas Wiegand and Bern Girod. *Multi-Frame Motion-Compensated Prediction for Video Transmission*. Kluwer Academic Publishers, Boston, 2001.

[142] P. Willemin, T. Reed, and M. Kunt. Image sequence coding by split and merge. *IEEE Transactions on Communications*, 39(12):1845–1855, 1991.

[143] C.T. Zahn and R.Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. on Computers*, 21(3):269–281, March 1972.

[144] F. Zanoguera, B. Marcotegui, and F. Meyer. A toolbox for interactive segmentation based on nested partitions. In *IEEE International Conference on Image Processing, ICIP'99*, Kobe, Japan, October 1999.