BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATICS
DEPARTMENT OF BROADBAND INFOCOMMUNICATIONS AND ELECTROMAGNETIC THEORY

Wavelet based efficient video compression methods

Ph.D. Thesis

Written by:

Lajos Konyha
2008.

Supervisors:
Dr. Kálmán Fazekas
Dr. István Erényi

The subject and aim of dissertation

My line in this dissertation is lossy compression of moving pictures. I am going to introduce my results about image processing based on wavelet transform. Lossy compression methods are qualified upon two parameters: the compression rate and the quality of decoded image. These two numeric values are evaluated geared to each other during the evaluation of compression methods. The aim is high quality with low data rate. Nowadays the most efficient image compression methods are based on wavelet transform and on SPIHT procedure coding wavelet coefficients. My results will be presented in three groups. These groups cover the various separable areas of image coding.

In the first part I am using three-dimensional dyadic wavelet transformation to exploit spatial and temporal redundancy and a modified SPIHT procedure to achieve efficient compression. In order to process the great amount of data I suggest a new method which grants an efficient real-time image coding.

In the second part I manage the coding of two-dimensional objects of arbitrary shapes. I have modified the SPIHT procedure so that it is able to code the coefficients made by wavelet transform chosen according to the shape efficiently.

In the third part I present my achievements in the area of motion compensation. Block based motion compensation cannot be used efficiently for motion of arbitrary shapes or for complex motion. In these cases mesh based motion compensation is effective. One important task is to create the mesh that covers the whole object. For this task I suggest a new method that produces the nodes and structure of the mesh in an effective way.

Applied experimental and inspection methods

I have used the following inspection methods during research:

- examination of video and image compression standards, and the specified and suggested methods inside
- examination of linear, non-linear, reversible, irreversible transformations
- examination of two- and three-dimensional transformations
- examination of multiresolution signal decomposition
- examination of one- and more dimensional filters
- examination of linear, non-linear and adaptive quantization
- examination of entropy coding methods
- research in libraries (articles of reviews and conferences), research in the Internet (in professional databases, like IEEE Xplore)

I have used the following experimental methods during research:

- test of algorithms via computer simulation (C, C++, MATLAB)
- computer implementations of image processing methods, their evaluation (C,

[6] **G. Xing, J. Li, S. Li, Y. Zhang**: Arbitrarily Shaped Video Object Coding by Wavelet, *EEE Transactions on Circuits and Systems for Video Technology, Volume: 11 Issue: 10, 2001, pp. 1135 – 1139*

[7] **Murat Tekalp**: Digital Video Processing (Paperback), *Prentice Hall PTR, 1995*

Referenced publications

Location of reference	Referenced publication
Jurij F. Tasič, Urban Burnik: **The Challenge of Personalisation and Mobility: COST #276**, *EUROCON 2003, Computer as a Tool*, vol II. pp.411-515, Ljubljana, Slovenia, September 22-24, 2003	[S1] [S17]
Kuan-Liang Chen: **Wafer ID Recognition Technology Research**, Thesis Id: 093NKIT5650013, Advisor: Wei-Chih Hsu, 2004	[S12]
M.Mitrea, F.Prêteux, M.Petrescu, A.Vlad: **The StirMark Watermarking Attack in the DWT Domain**, *Proceedings 12th IEEE International Conference on Systems, Signals and Image Processing (IWSSIP'05), Halkida, Greece*, Vol. 2, September 2005, p. 5-9.	[S9]
Iveta Gladišová: **Simple Image Indexing and Retrieval Technique Based on Dominant Colours**, *13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006)*, pp.345-347, Budapest, Hungary, 21-23 September 2006	[S17] [S34]
Suresh, K.V.; Kumar, G.M.; Rajagopalan, A.N.; **Superresolution of License Plates in Real Traffic Videos,** Intelligent Transportation Systems, IEEE Transactions on, Volume 8, Issue 2, June 2007 Page(s):321 - 331	[S12]
Chien-Kuo Li: **Using Omni-Directional and PTZ Cameras to Implement Real-Time Tracking of Moving Objects on a DSP Board**, *Thesis*, Chang Yuan Christian University, Taiwan, 2006	[S12]
Sheng-Long Lee: **Using Motion Information and Multiple Classifiers in a License Plate Recognition System for Moving Vehicles on the Road**, *Thesis*, Chang Yuan Christian University, Taiwan, 2004	[S12]

References

[1] **C. Fogg, D. J. LeGall, J. L. Mitchell, W. B. Pennebaker**: MPEG Video Compression Standard, *Springer, 1996*

[2] **J. M. Shapiro**: Embedded Image Coding Using Zerotrees of Wavelet Coefficients, *IEEE Transactions on Signal Processing, Vol. 41, No. 12 (1993), p. 3445-3462*

[3] **A. Said, W. A. Pearlman**: A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees, *IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, pp. 243-250, June 1996.*

[4] **J. Van Overloop, W. Van De Sype, K. N. Denecker, P. De Neve, E. Sundermann, I. L. Lemahieu**: Experimental Comparison of 2D and 3D Wavelet Image Compression Methods for Medical Image Sets, *Medical Imaging 1998: Image Display, June 1998, pp. 348-358*

[5] **E. Moyano, F. J. Quiles, A. Garrido, J. Duato, L. Orozco-Barbosa**: Efficient 3D Wavelet Transform Decomposition for Video Compression, *Digital and Computational Video, 2001. Proceedings. Second International Workshop on 8-9 Feb. 2001 pp. 118 – 125*

C++, MATLAB, assembly)
- optimization of procedures by hardware level programming
- valuation of implementations through results of standard tests
- computerized data processing and display

New scientific results of the dissertation

1st thesis group: Three-dimensional dyadic wavelet transformation for real-time video coding.

During video compression the current methods use transform coding to exploit spatial redundancy. 8X8, 16x16 or more adjacent pixels are used at once. Contrarily simple difference images are used to exploit temporal redundancy. Maximum 3 images are used (current frame, backward predictor, forward predictor) **[1]**.

Human visual system (HVS) has the ability that it can adapt itself to sudden changes with a little delay. So it can detect quick changes weakly (also it can detect quick spatial changes weakly). In natural video contents the temporal redundancy usually appears in more subsequent frames (same frames, or slow, continual transition), except some special cases (change of scenes, disappearing or reappearing parts due to motion), when there is a sudden change in two subsequent frames. In order to exploit spatial redundancy wavelet transformation is the most efficient method. It is practical to use them to exploit temporal redundancy, too **[2, 3]**.

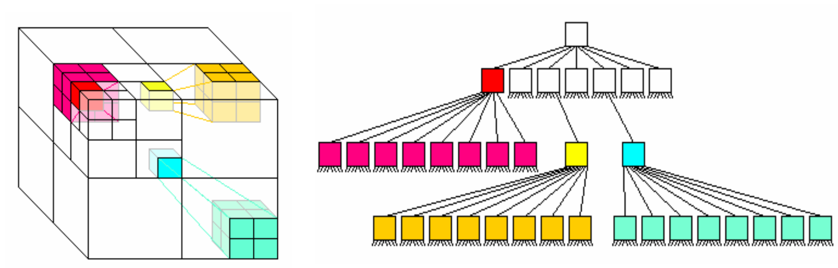

Figure 1 – 3D dyadic wavelet decomposition (a) and the tree structured graph reflecting the relations between coefficients (b)

The distribution of the coefficients of three-dimensional dyadic wavelet transform is shown in figure 1 a **[4]**. During SPIHT algorithm the order of collection of coefficients is determined by spatial orientation tree describing the relations between wavelet coefficients. I have modified this tree in a way shown in the figure 1 b. So SPIHT algorithm has become to be able to process 3D wavelet coefficients.

Results made through dyadic decomposition and SPIHT algorithm can be seen in figure 2. The upper curve corresponds to the results obtained using Akiyo sequence, the bottom curve corresponds to the results obtained using Bream sequence. Bream sequence gives worse quality at the same compression ratio because it contains more dynamic, more complex details and motion. The results show that a video of simple content can be compressed to acceptable quality at 0.021-0.042bpp ratio. Video of more complex structure requires about 0.1-0.2bpp information for acceptable quality.

T1/a: I have modified the SPIHT algorithm in order to be able to process the coefficients of dyadic three dimensional wavelet transform to exploit the temporal redundancy efficiently. *Experimental results obtained by modified structured tree have proven that the algorithms driven by spatial orientation tree are suitable to process three dimensional data. The results obtained by SPIHT algorithm shows that SPIHT outperforms the conventional algorithms. The full processing time is increased because of the huge data amount of three dimensional structure.*

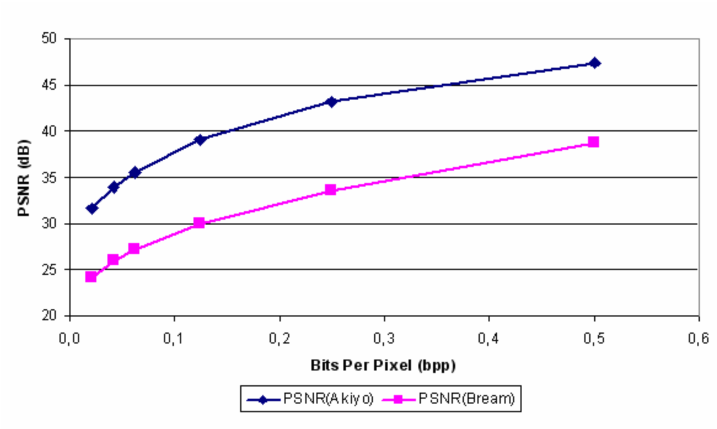

Figure 2 – Results of 3D SPIHT: quality in function of bit rate

A serious disadvantage of the three dimensional data processing is the huge amount of data that requires a lot of computational resources. It is especially true by SPIHT algorithm because the most computational complex part of this algorithm is to find the new significant coefficients, which is equivalent to the analysis of the subtrees of the graph shown in figure 1.b. This means that the algorithm have to walk through the whole graph starting from the root node till it reaches the bottom nodes and all descendants (in some cases all descendants have to be considered while in other cases the direct descendants have to be ignored) have been examined. In the three-dimensional case each node has eight descendants on the lower level (exceptions are the top and the bottom levels), therefore the number of nodes to be examined is increasing exponentially as the algorithm walks through each levels of the graph. This makes the real-time processing impossible in practice.

In order to accelerate the procedure I have introduced new variables containing the maximum of absolute values of descendants at the nodes of the graph (figure 3). So instead of the walk through the whole subtree it is sufficient to check

Conference lecture issued in a publication in Hungarian language:

[S26] Enyedi B., *Konyha L.*, Fazekas K.: **Valósidejű mozgóképtömörítés wavelet-transzformációval**, *TV2005, Televízió- és hangtechnikai konferencia és kiállítás*, 2004. június 1-2., Budapest, Magyarország

[S27] Enyedi B., *Konyha L.*, Fazekas K.: **Megfigyelő kamerák vezérlése wavelet transzformációval, hatékony adattárolás megvalósítása**, *TV 2002 konferencia*, pp.149-159, 2002 május, Budapest

[S28] Enyedi B., *Konyha L.*, Szombathy Cs., Tran Min Son, Gschwindt A., Szokolay M., Fazekas K.: **MPEG-4 rendszerek kutatása és beintegrálása a DVB-T átviteltechnikába**, *TV2005, Televízió- és hangtechnikai konferencia és kiállítás*, 2005. június 1-2., Budapest, Magyarország

[S29] Enyedi B., *Konyha L.*, Szombathy Cs., Tran Min Son, Gschwindt A., Szokolay M., Fazekas K.: **MPEG4 alapú műsorszórás a DVB-T technikában**, *HTE-BME 2005 diákkonferencia*, 2005. május 19., Budapest, Magyarország

Edited periodical article, published abroad in foreign language:

[S30] S. M. Tran, *L. Konyha*, B. Enyedi, K. Fazekas, J. Turan Jr.: **Deployment of Constrained Delaunay Mesh in VOP Shape and Texture Coding,** *Journal of Electrical Engineering*, vol.56, 07-08 (2005) pp.189-194, Slovakia, 2005

[S31] *L. Konyha*, B. Enyedi, K. Fazekas: **Real Time Number Plate Localization Algorithms**, *Journal of Electrical Engineering,* vol.57, 2 (2006) pp.69-77, Slovakia, 2006

[S32] *L. Konyha*, B. Enyedi, K. Fazekas: **Motion Compensated Video Compression with 3D Wavelet Transform and SPIHT**, *Radioengineering, Proceedings of Czech and Slovak Technical Universities and URSI Committees*, vol.15, n.1, pp.42-46, Slovakia, April 2006

Edited periodical article, published Hungary in Hungarian language:

[S33] Enyedi B., *Konyha L.*, Szombathy Cs., Tran Min Son, Gschwindt A., Szokolay M., Fazekas K.: **MPEG-4 alapú átvitel megvalósítása a DVB-T technikában**, *Híradástechnika (LIX. 2004/7)*, pp.23-27, 2004. július, Budapest, Magyarország

[S34] Enyedi B., *Konyha L.*, Szombathy Cs., Tran Min Son, Gschwindt A., Fazekas K.: **Új lehetőségek a digitális televíziózásban**, *Híradástechnika (LX. 2005/9)*, pp.53-57, 2005. szeptember, Budapest, Magyarország

Other:

[S35] Enyedi B., *Konyha L.*, Szombathy Cs., Tran Min Son, Gschwindt A., Szokolay M., Fazekas K.: **MPEG-4 alapú átvitel megvalósítása a DVB-T technikában**, *HTE előadás, Postamúzeum előadóterme*, 2004. szeptember 14., Budapest, Magyarország

[S13] B. Enyedi, *L. Konyha*, S.M. Tran, Cs. Szombathy, K. Fazekas: **A Survey on the Interactivity Feature of MPEG-4**, *46th International Symposium ELMAR-2004 focused on Navigation, Multimedia and Marine*, pp.30-38, Zadar, Croatia, June 16-18, 2004

[S14] B. Enyedi, *L. Konyha*, Cs. Szombathy, K. Fazekas: **Real-time Video Compression with 3D Wavelet Transform and SPIHT**, *4th International Conference Intelligent Systems Design and Applications (ISDA 2004)*, Budapest, Hungary, August 26-28, 2004

[S15] S. M. Tran, *L. Konyha*, **B. Enyedi**, Cs. Szombathy, K. Fazekas: **Experiments on Transmitting MPEG-4 Content over MPEG-2 Transport Stream**, *5th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2004*, Lisboa, Portugal, April, 2004

[S16] B. Enyedi, *L. Konyha*, S.M. Tran, Cs. Szombathy, K. Fazekas: **A Survey on the Interactivity Feature of MPEG-4**, *11th International Workshop on Systems, Signal and Image Processing IWSSIP'04*, pp.87-90, Poznan, Poland, September 13-15, 2004

[S17] B. Enyedi**,** *L. Konyha*, K. Fazekas: **Fast Video Compression Based on 3D Wavelet Transform and SPIHT**, *7th COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pp.121-124, Ankara, Turkey, November 4-5, 2004

[S18] B. Enyedi, *L. Konyha*, Cs. Szombathy, S.M.Tran, A. Gschwindt, K. Fazekas, M. Szokolay: **Implementation of MPEG4-Based Transmission in DVB-T Technology**, *47th International Symposium ELMAR-2005 focused on Multimedia Systems and Applications*, Zadar, Croatia, June 08-10, 2005

[S19] B. Enyedi, *L. Konyha*, K. Fazekas: **Threshold Procedures and Image Segmentation**, *47th International Symposium ELMAR-2005 focused on Multimedia Systems and Applications*, Zadar, Croatia, June 08-10, 2005

[S20] B. Enyedi, *L. Konyha*, K. Fazekas: **Fast License Plate Localization**, *6th International Scientific Conference DSP-MCOM'05*, pp.10-14, Kosice, Slovakia, September 13-14, 2005

[S21] B. Enyedi, *L. Konyha*, S.M. Tran, K. Fazekas: **Toward a High Abstaraction Level of MPEG-4 BIFS Composition**, *6th International Scientific Conference DSP-MCOM'05*, pp.50-54, Kosice, Slovakia, September 13-14, 2005

[S22] B. Enyedi, *L. Konyha*, K. Fazekas: **Character Localization**, *13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006)*, pp.317-326, Budapest, Hungary, 21-23 September 2006

[S23] B. Enyedi, *L. Konyha*, K. Fazekas: **Segmentation – Region Growing Based on Adaptive Threshold Algorithms**, *13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006)*, pp.327-332, Budapest, Hungary, 21-23 September 2006

[S24] B. Enyedi, *L. Konyha*, K. Fazekas, J.Turán: **License Plate Localization and Storage Method**, *14th International Conference on Systems, Signals and Image Processing (IWSSIP 2007)*, pp.441-445, Maribor, Slovenia, 27-30 June 2007

[S25] B. Enyedi, *L. Konyha*, K. Fazekas, J.Turán: **Small Tricks to Enhance the Accuracy of License Plate Character Recognition**, *2nd International Conference on Signals Processing and Multimedia Applications (SIGMAP 2007)*, pp.125-128, Barcelona, Spain, 28-31 July 2007

Lajos Konyha – ***Wavelet based efficient video compression methods***

the value of the new variable. This means we need to execute only one operation instead of walking through the whole subtree.

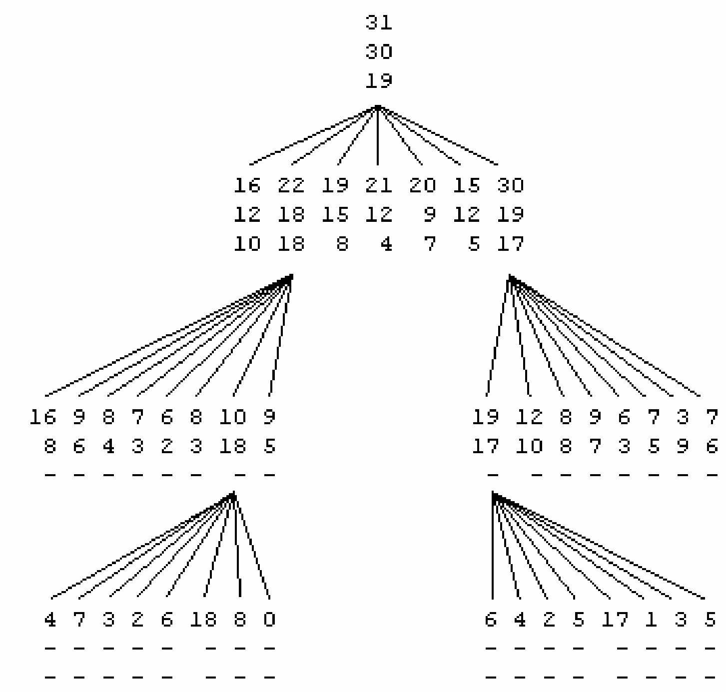

Figure 3 – The new variables in the tree of SPIHT

The coefficients on the bottom level have no descendants (7 of 8 coefficients are on the bottom level) so no new data belongs to them. One level higher the coefficients have no far descendants, so there is no need to store their maximum value either. On the whole only every eighth coefficient needs to store the maximum of its descendants, and every 64^{th} needs to store the maximum of its far descendants, which requires 14% extra memory.

With the new variables the time consumption of each step of coding and decoding according to the compression ratio can be seen on figure 4. The sequences used during simulations were 352x288 pixels in size, and contained 128 frames. The time limit of real-time coding assuming one frame per 30 seconds is a little more than 4.26 seconds. The whole coding requires 4 seconds in worst case, decoding requires 2.4 seconds, so both of the processes can be executed in real time.

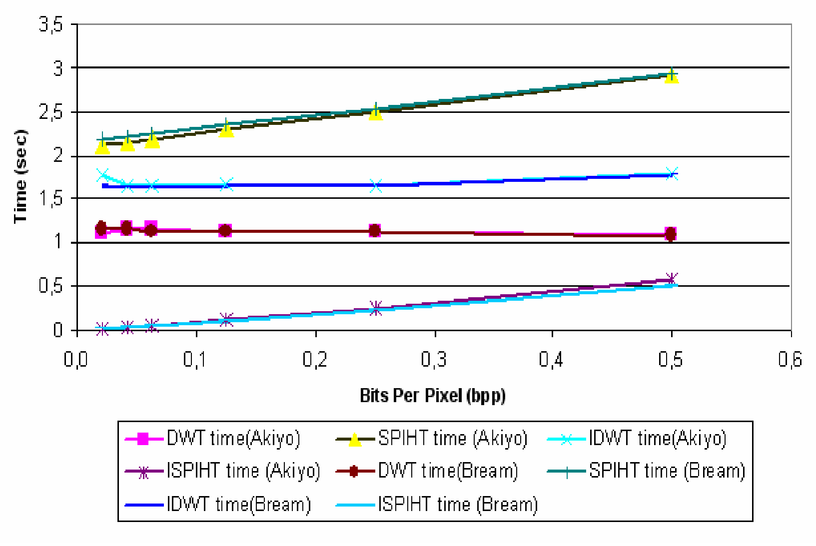

Figure 4 – Execution time of accelerated 3D SPIHT algorithm

T1/b: I have introduced new data in order to accelerate the 3D SPIHT algorithm. With these new data the 3D SPIHT procedure is able to process the wavelet coefficients in real-time. *With the introduction of maximum values of descendants and far descendants the node searching step (the most time-consuming step) dropped to a fraction (one operation instead of walking through the whole subtree). The storage of new data requires only 14% extra memory.*

Publications

Edited lecture in foreign language issued in the publication of an internationally attended conference:

[S1] *L. Konyha*, B. Enyedi, K. Fazekas: **Multimedia Distance Learning – Orthogonal Transformations**, *EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services,* pp.196-198, Budapest, Hungary, September 11-13, 2001

[S2] *L. Konyha*, S. M. Tran, B. Enyedi, K. Fazekas, A. Gschwindt: **Multiresolution Image Decomposition Using Wavelet Transformation**, *Conference on Using Technology in Open and Distance Learning*, pp.49-56, Maribor, Slovenia, September 13-14, 2001

[S3] *L. Konyha*, S. M. Tran, K. Fazekas, A Gschwindt, **Deployment of wavelet transform to VOP codec scheme**, pp. 233-237, Video/Image Processing and Multimedia Communications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom, Zadar, Croatia, June 2002

[S4] B. Enyedi, *L. Konyha*, K. Fazekas: **New Method for Security Systems: On-line Camera Control, Digital Data Processing**, *3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pp.154-160, Budapest, Hungary, October 11-12, 2002

[S5] S. M. Tran, *L. Konyha***,** B. Enyedi, K. Fazekas, A. Gschwindt, J. Turan Jr.: **Deployment of Constrained Delaunay Mesh in VOP Shape Coding**, *3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pp.161-165, Budapest, Hungary, October 11-12, 2002

[S6] S. M. Tran, *L. Konyha*, B. Enyedi, K. Fazekas, A. Gschwindt, J. Turan Jr.: **Deployment of Constrained Delaunay Mesh in VOP Shape Coding**, *9th International Workshop on Systems, Signals and Image Processing (IWSSIP'02)*, pp.205-211, Manchester, United Kingdom, November 7-8, 2002

[S7] *L. Konyha*, B. Enyedi, S. M. Tran, K. Fazekas: **Multimedia Education – Distant Learning**, *4th European Workshop on Image Analysis for Multimedia Interactive Services*, pp.483-486, Queen Mary University, London, United Kingdom, April 9-11, 2003

[S8] *L. Konyha*, B**.** Enyedi, S. M. Tran, K. Fazekas: **Content-Based Mesh Generation Algorithm**, *4th EURASIP Conference focused on Video / Image Processing and Multimedia Communications*, pp.175-180, Zagreb, Croatia, July 2-5, 2003

[S9] *L. Konyha*, B. Enyedi, S. M. Tran, K. Fazekas: **e-Learning – Multimedia Education**, *4th cost 276 workshop, Information and Knowledge Management for Integrated Media Communication*, pp.95-97, Bordeaux, France, March 31- April 1, 2003

[S10] S. M. Tran, *L. Konyha*, B. Enyedi, K. Fazekas: **Toward a High Abstraction Level of MPEG-4 BIFS Composition**, *10th International Workshop on Systems, Signals and Image Processing IWSSIP'03*, Czech Technical University in Prague, Prague, Czech Republic, September 10-11, 2003

[S11] *L. Konyha*, B. Enyedi, S. M. Tran, K. Fazekas: **Paradigm of Multimedia Distance Teaching**, *10th International Workshop on Systems, Signals and Image Processing IWSSIP'03*, Czech Technical University in Prague, Prague, Czech Republic, September 10-11, 2003

[S12] B. Enyedi, *L. Konyha*, Cs. Szombathy, K. Fazekas: **Strategies for Fast Licence Plate Number Localization**, *46th International Symposium ELMAR-2004 focused on Navigation, Multimedia and Marine*, pp.579-584, Zadar, Croatia, June 16-18, 2004

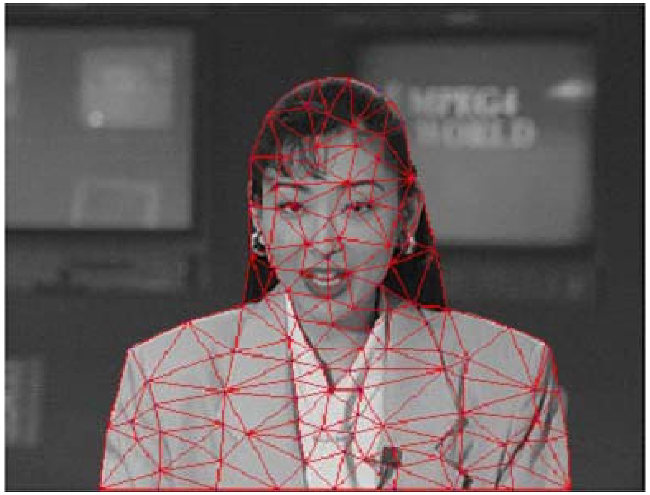

Figure 14 – Mesh structure in case of object based video processing

Lajos Konyha – ***Wavelet based efficient video compression methods***

2nd Thesis: Coding of objects of arbitrary shapes

It is the MPEG-4 standard spreading due to its effectiveness and versatility that makes possible the coding of object of arbitrary shape for the first time. The significant effectiveness of wavelet transform and SPIHT algorithm makes it reasonable to modify the procedures so that it will be able to process not only objects of a rectangle but objects of any shape. It is well-known from the literature that wavelet transform can be used to code object of any shapes **[6]**. The transform acting upon shapes has the important attribute, that only coefficients which's spatial position is covered by the original shape are being created. Figure 5 shows the mask of a video object and the position of the coefficients arose.

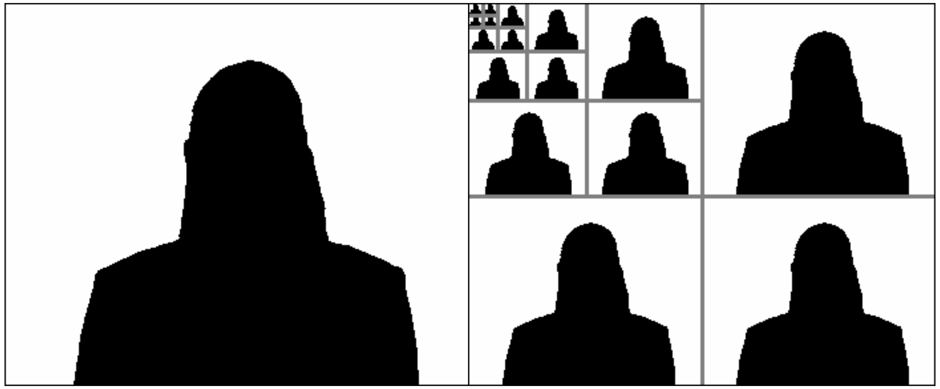

Figure 5 – Arbitrary shaped video object and the shape adaptive wavelet coefficients

SPIHT algorithm must be made able to process the result of such modified wavelet transform. Observing the relation between the coefficients it can be seen that three coefficients of one level lower belong to the top level coefficient. And on lower levels each coefficient may have 0, 1, 2, 3 or 4 coefficients of one level lower (except coefficient on the bottom level).

So I have modified the graph controlling the SPIHT algorithm so that it takes into account only the nodes where wavelet coefficients have been generated (figure 6). The markings and remainder error of objects can be coded more efficiently through this modified algorithm than through the current cosine transform acting upon shape and EZW procedures (figure 7).

The speed of the procedure is determined by two parameters of the modification: first the complex searching rules make the working of algorithm slower. Second if lesser amount of data has to be processed, the procedure will finish sooner.

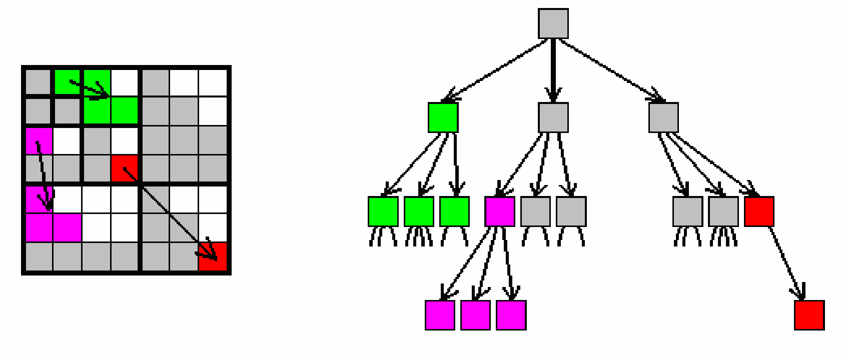

Figure 6 – Spatial orientation tree of shape adaptive SPIHT algorithm

In practice this algorithm is slower than the original SPIHT procedure because of the complex rules, but with the new variables introduced in T1/a the working will be accelerated the same way. This time 25% extra memory required for storing the maximum values of descendants (every 4th coefficient is not on the bottom level), and storing the far descendants needs 6.25% extra memory (Every 16th coefficient is not on the two lowest level). This makes 31.25% extra memory requirement. This way the whole SPIHT coding acting upon shapes can be executed real-time.

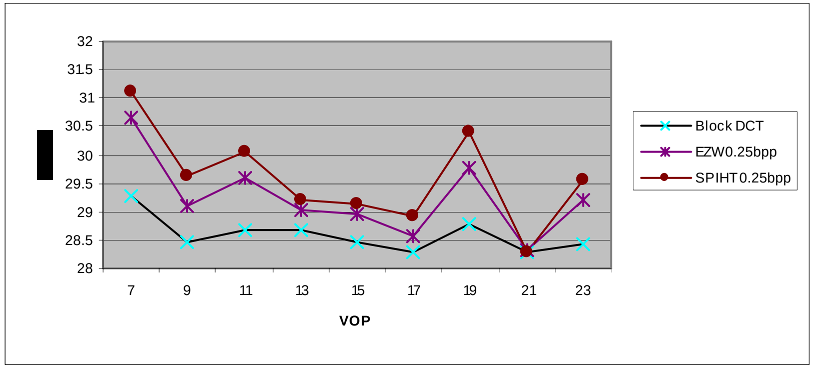

Figure 7 – Comparison of SA SPIHT with EZW and DCT

T2: I have modified the SPIHT procedure so that it will ignore the wavelet coefficients that are not covered by the object. It won't waste data for these areas, and the effectiveness of the compression will be higher. *By using the maximum values of descendants and far descendants this procedure is also much faster, its price is the 31.25% extra memory requirement.*

Lajos Konyha – ***Wavelet based efficient video compression methods***

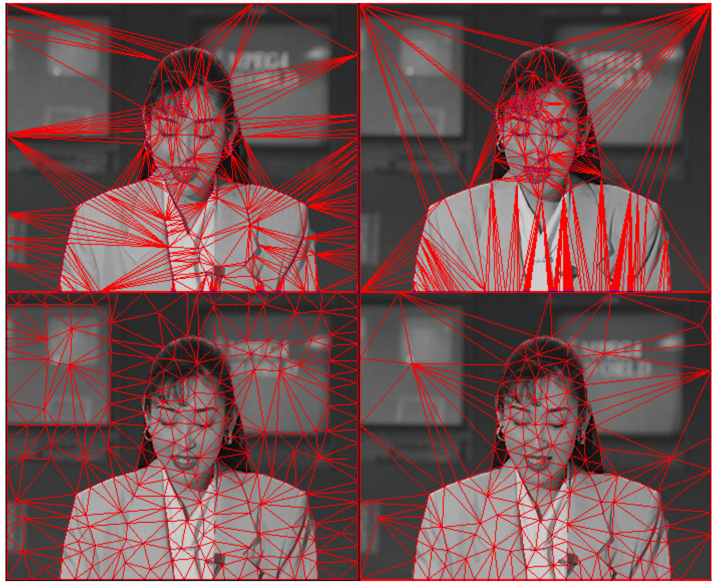

Figure 13 – The mesh in cases of different settings (frame based video processing)

Figure 13 shows some results of frame-based video processing. The weight of the time constant is 2 in the pictures on the left side, while it is 100 in the pictures on the right side. The minimal distance between mesh nodes is 1 in the upper pictures and 10 in the bottom pictures. In the upper examples the nodes are placed very close to each other therefore any inaccuracy of motion estimation can lead to unexpected results and the consistency will be violated. The lower results are free of these problems, and the mesh edges are likely to be at the object contours therefore the mesh is useful for motion compensation. The 16×16 pixels block based motion compensation used in MPEG standards requires 396 motion vectors in the case of CIF image (352×288 pixels). The mesh contains 280 nodes (this is equal to the number motion vectors) in the bottom left picture, and 172 nodes in the bottom right picture where the most of the nodes are placed to temporal dynamic regions.

In case of object based video processing (figure 14) the poly-line contains several vertices at complex part of the contour (head), and a few nodes at simple parts (bottom edge). The inner nodes are placed according to the texture of the object similar to the previous example. The mesh covering the object requires only 122 nodes.

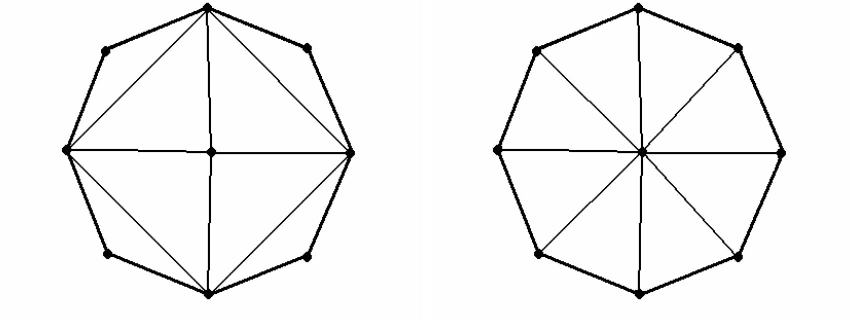

Figure 11 – Mesh structure before (a) and after (b) optimization

In figure 11 it can be seen that by optimization of structure, always two neighboring triangles must be changed. The common edge must be eliminated, and the not common nodes must be connected. Goal of optimization is elimination of distorted triangles, so by changing of edges the new triangles must be checked whether they are more distorted than the old ones.

Based upon this the procedure works the following way: it seeks the most distorted triangle and tries to change its edges with all of its neighbors. Change with the lowest shape factor of the most distorted triangle will be kept. If the change improved the situation the change will be executed. Then the procedure seeks the next triangle until it finds none.

In some cases edges between common nodes and not common nodes cannot be exchanged. In case of triangles in figure 12 one of the new triangles contains the other after the exchange. So consistency of the mesh is damaged. In this case the changed edges do not intersect each other. So the algorithm must be modified that it will check distortion only if the edges intersect each other.

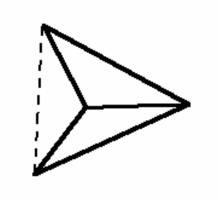

Figure 12 – Triangles that cannot be optimized

T3/c: I have developed a method for optimization of mesh connectivity. This method increases the resilience of motion estimation error. *The goal of mesh optimization is minimization of impact of motion estimation error. This will be executed by rearranging the edges of triangles with sharp angles.*

3rd Thesis group: Mesh based motion compensation

The universal block based motion compensation has a disadvantage: it can handle only the simple movements. By more complex motion (e.g. camera zooming, spinning, distortion motion) there will always be great residual error after compensation. Another disadvantage that it has been developed for processing rectangular frames. Because of this it cannot be used for motion compensation of objects of arbitrary shape. Mesh based motion compensation procedures are able to handle more complex shapes and motions. These procedures bend a mesh made of polygons (usually triangles) onto the frames or objects of video, and then during motion estimation the motion vector of nodes will be determined.

During compensation motion vectors of pixels inside the polygons are being determined through an interpolation **[7].** One common mesh creating procedure is Delaunay method that covers the object with a mesh appropriate for the geometric constrains. It has the disadvantage that only given geometric compulsions are taken into consideration. Because of this the created mesh is not always the best from the point of view of motion compensation.

My new method creates an optimal mesh for motion compensation, and it takes into consideration the geometric data of objects as well as their texture. The method consists of three algorithms. The three tasks are determination of nodes of the mesh, creating edges between nodes, and optimization of structure by rearranging the edges.

By the determination of mesh nodes I have exploited that the motion estimation is more reliable at the locations with high variability in intensity than at solid surfaces. Besides this it was also taken into account that temporal static parts are irrelevant to motion compensation, therefore the temporal variability was also considered during the determination of mesh nodes. Following the mentioned rules the proposed method calculates the variability of each pixel's location which is equal to the sum of the absolute values of differences between the intensities of current pixel and the temporal and spatial adjacent pixels. This quantity reflects how likely the current location is a mesh node. The temporal variability is extremely important during motion compensation therefore the temporal difference may be calculated by higher weight in summation (the weight is an adjustable parameter). The average

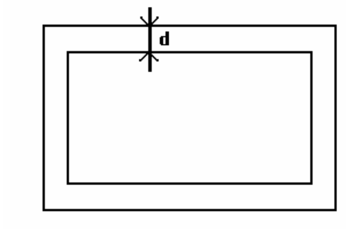

Figure 8 – Band near to the edges with no inner nodes to avoid distorted triangles

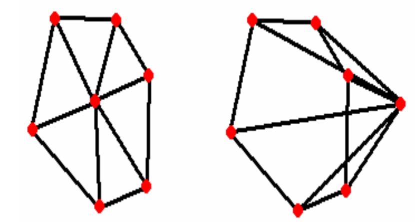

Figure 9 – Violated mesh effected by inaccurate motion estimation

variability of each node can be obtained by dividing the sum of variabilities in all locations by the number of mesh nodes. In the second step the algorithm finds the location with the higher variability value and picks it as a mesh node. Starting from the new node the algorithm zeroes the variability value in the adjacent locations until the sum of the canceled values reaches the average variability of the mesh nodes. Then the algorithm continues from second step by finding the next location with the highest variability value until all variabilities are cancelled (all nodes are found). Using these rules it is guaranteed that no mesh nodes get into solid parts, few nodes get into simple and almost static parts, and the most of the nodes get into the complex and temporal dynamic parts.

The nodes near the edge of the image or object would result always distorted triangles. Because of this the procedure does not seek inner nodes in a given distance of edges (figure 8). In case of object-based video processing the nodes of poly-line approximation of the contour are the outer nodes of the mesh. In case of frame based video processing the variability values near the edges are summed into the nearest edge (figure 8). At first the corners are marked as outer nodes then summed values are used to find the outer nodes.

Too small triangles can also lead to inadequate motion compensation because the vertices of the small triangles are very close to each other and even a small error of motion estimation can violate the consistency of the mesh (figure 9). To avoid this the minimal distance of the mesh nodes is also a parameter of the algorithm, and all values will be zeroed inside this distance after picking up a new node, regardless of the sum of the cancelled values.

T3/a: I have developed an algorithm to find the nodes of mesh optimally for motion compensation. *The picked nodes can be tracked dependably during motion estimation. Beside this the complex, highly-detailed areas get many nodes, the simple areas get only a few ones.*

The next phase is creating a consistent net by connecting the nodes. The structure of the mesh has the following restrictions: edges cannot cross each other, each edge is between two nodes, and no node can be on an edge (violation of these restrictions would result a structure like in figure 9).

To keep these restrictions we have to make connections the following way: two nearest unconnected nodes must be found, and the edge between them would not cross any already existing edge. So edges won't intersect each other. The second restriction will be kept because there cannot be node between two nearest nodes (otherwise the middle node would be closer to end node than the other end node).

Based upon this it can happen that in the case of concave object a connection will be made outside the object. However all rues are complied the dashed lines show erroneous connections in figure 10. As all of these connections are between two outer nodes, the procedure should be modified: at first the outer nodes have to be connected in sequence, and then the new connections will be made between an inner

node and any other node. This ensures that the mesh connections form always a consistent structure.

T3/b: I have developed a procedure which automatically connects the vertices based on the locations to create a triangular mesh. *An advantage of the method is that it does not require any other parameter excluding the location of the nodes, therefore it can be used both in encoder and decoder side (transmitting the connectivity is not required).*

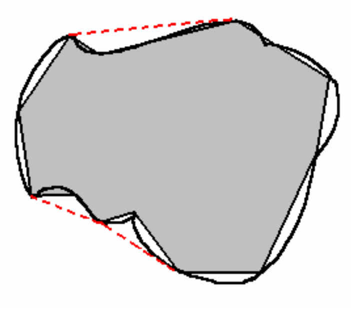

Figure 10 – Erroneous connections in case of concave shape

Meshes generated with the previous procedure contain many triangles that are not optimal for motion compensation (figure 11.a). Rearranging the edges we get more beneficial triangles with less sharp angles, and these triangles are better for motion compensation. In order to mark the distortion of triangles I have introduced the ratio of the longest side and the sum of the other two sides (Shape Factor).

If the triangle has very sharp angle(s) the long side is almost as long as the sum of the other two sides, so shape factor is almost 1. If the triangle is even sided then the longest side is half as long as the sum of the other two, so shape factor is 0.5. With this simple factor triangles can be qualified according to their distortion, so distorted triangles can be detected.