*Journal of* ELECTRICAL ENGINEERING, VOL. 56, NO. 7-8, 2005, 189–194

# DEPLOYMENT OF CONSTRAINED DELAUNAY MESH IN VOP SHAPE AND TEXTURE CODING

**Son Minh Tran** * — **Lajos Konyha** * — **Balázs Enyedi** * — **Kálmán Fazekas** * — **Ján Turán** **

The work presents an alternative method for encoding shape-information of video objects, which is a new requirement in the recent revolution of video codec scheme: compressing and transmitting object-based video streams. The proposed method takes the advantage of the latest development in the video encoding — motion compensation on the base of mesh presentation of frame — to complete its target. The so-called constrained Delaunay mesh is deployed to represent video objects, which are detected from a previous video object detector. This special mesh structure in one hand constructs the base for successive motion compensation taking root on mesh-vertices. In another hand, the virtue of its constrained property opens a possibility to encode the shape of video object implicitly, through the bounding polygon of the mesh. Hence, this lossy presentation of the shape-information reduces the amount of transmitted information. We discuss the trade-off between the precision of shape approximation and necessary bandwidth. We integrate the shape encoder into a mesh-based video codec scheme. The simulating results of quality and bitrate are encouraging.

K e y w o r d s: shape coding, Delaunay mesh, VOP, MPEG-4

## 1 INTRODUCTION

The recent MPEG-4 Visual coding technique is the first standard, which allows the transmission of arbitrarily shaped video objects (VOs). In the new scenario of video transmission, scene consists several VOs, which in turn possess their own textures, motions and shape information. The bitstreams of these VOs and the accompanying composition information can be multiplexed such that the decoder is able to decode the VOs separately and arranges them into a video scene in accordance with users' demand. In comparison with the older members of the MPEG family, such as MPEG-1 / 2, the foremost new challenge posed to developers is how to send the transparency and the shape of VOs to decoders efficiently. Within the scope of this work, we restrict ourselves to opaque objects only, where binary shape information (showing whether pixel belongs to associated object or not) is satisfactory. Basically, there are three methods to encode this type of information:

*Chroma-keying*: actually no dedicated transmitting channel is necessary since the shape information is encoded implicitly in the texture information of the VOs. It is possible thanks to the fact that the background object of a simple scene often has a static, homogenous and distinguishable color: the chroma-key (background color). Pixels with a color similar to this predefined value are considered as the external ones. The remains belong to the VO. The 1important advantage of this method is its low computational and arithmetic complexity.

*Pixel-based*: at first, the silhouette of VOs is created by assigning two different intensities (1 and 0) to internal and to external pixels of a given VO. The resulting binary mask then can be treated as an independent "texture" and encoded as usual. Exploiting the bi-values of the mask, special encoding methods are dedicated to pixel-based shape encoder, such as context-based arithmetic encoder (CAE) in [5], modified modified read (MMR) in [6]. The CAE method is well integrated into the current MPEG-4 scheme since CAE is also the block-based approach (conventional coding method for texture in MPEG family). It has the benefit of short processing time because it only has a delay time of a macro-block unit. The block-based conversion in CAE, however, shows visually annoying staircase effects [11].

*Contour-based*: only those pixels, which lie on the boundary of VOs become targets of shape encoder. Numerous techniques are developed to reconstruct the positions of these outlining pixels in either lossy or lossless manner. They are modified vertex-based, chain coding, baseline-based B-splines, polar coordinate based, etc. as in [2] [7]. Contour-based shape coding can be thought of as an alternative to avoid the staircase effect.

Another revolutionary approach of MPEG-4 Visual standard is the mesh-based video presentation. Blocks/ macro blocks are no longer used as elementary units for encoding texture in a motion compensation manner. Instead, video frames/VOs are now constructed by quadrilateral or trilateral mesh with either regular or irregular structure. The topology consisting of vertices' positions and connectivity is sent to the decoder while the texture is reconstructed by compensating the predicted topology and its warped texture with error residual. This new method in video presentation drastically reduces the block effect, a typical artifact of the block-based counterpart in low bitrate applications.

* Budapest University of Technology and Economics, Department of Broadband Infocommunication Systems, Hungary, Budapest 1111, Goldmann tér 3, sony@mht.bme.hu

** Department of Electronics and Multimedia Telecommunication, University of Technology Košice, Slovakia, jturan@ccsun.tuke.sk

ISSN 1335-3632 

We here introduce a new technique to encode the shape of VOs, which can be grouped into the class of contour-based shape encoders. In our scheme, the shape information — the 2-D positions of boundary vertices — are encoded and transmitted as a part of a constrained Delaunay mesh, which is used to present associated VOs. By using an unique constrained Delaunay mesh for both VO shape and texture carrier, we not only implicitly encode the shape information but also combine together the above advantages of contour-based shape coding and mesh-based presentation. We dedicate the Section 2 to introduce in detail the implementation of this double usage of the constrained Delaunay mesh. In Section 3, we integrate the proposed shape encoder into a full scheme of mesh-based VO compression and transmission. We close our discussion with conclusions and our future developments in Section 4.

## 2 PRESENTATION OF VIDEO OBJECT WITH CONSTRAINED DELAUNAY MESH

In our shape coder, the Cartesian coordinates of bounding outline are encoded like other members in the contour-based family. The approach starts with extracting the exact contour of the VO. The resulting collection of pixels is then estimated with a polygon shape for a given tolerance of errors. The constrained Delaunay mesh topology is generated upon the suitably fitted polygon. Finally the coordinates of vertices are encoded in a differential manner for high compression.

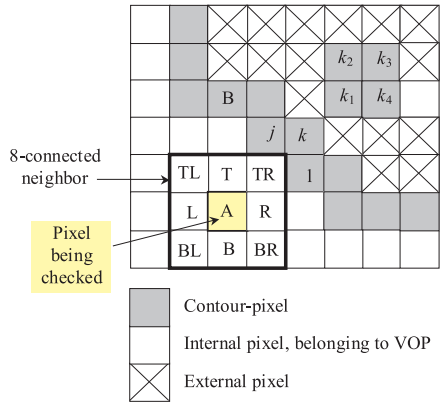


**Fig. 1.** Searching method for contour-pixels.

### 2.1 Pixel-based contour extraction

In general, we can assume that the silhouettes of VO at any time $t$ — the binary mask of the VOP (Video Object Plane) — are given. They are the results of video object extraction, which are the preceding phase of video object encoder. This common input of shape encoder is more suitable for pixel-based encoding method only. Therefore the extraction of contour — the collection of pixels lying on the boundary of video object — is necessary in our case. We define contour-pixels as such internal pixels of VOP, which have at least one external pixel among their 8-connected neighbors — they are pixel Top (T), Bottom (B), Left (L), Right (R), Top Left (TL), Top Right (TR), Bottom Left (BL) and Bottom Right (BR) in Fig. 1.

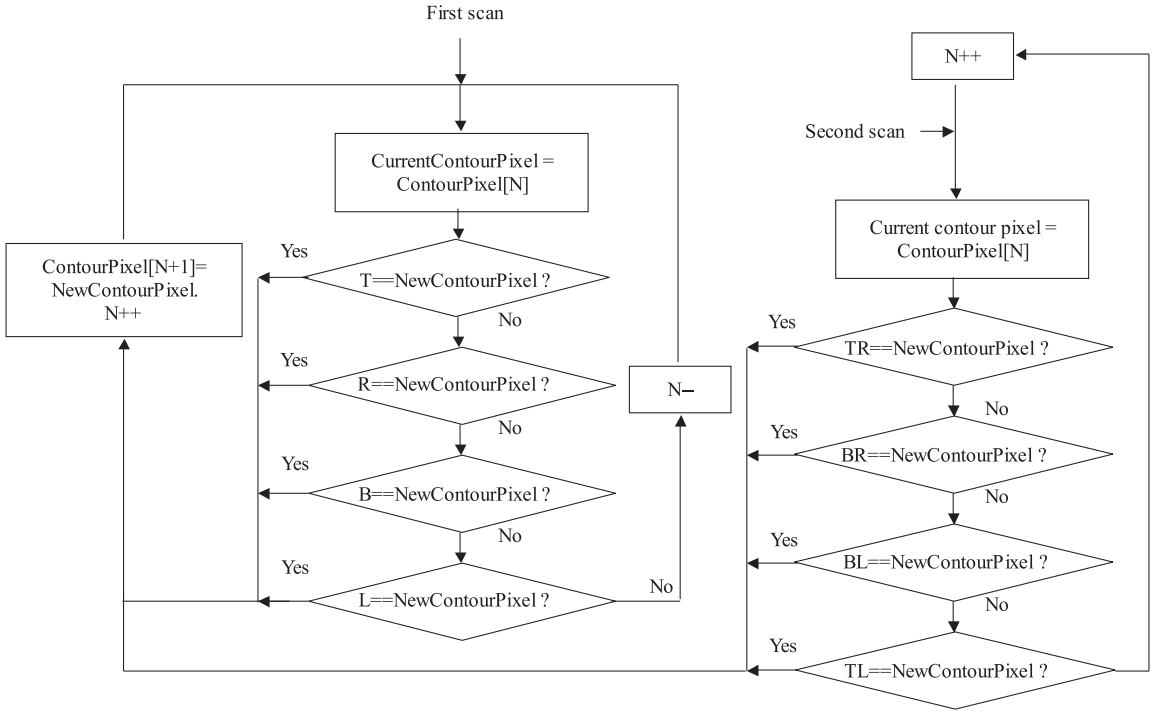


**Fig. 2.** Process of detecting contour pixels.

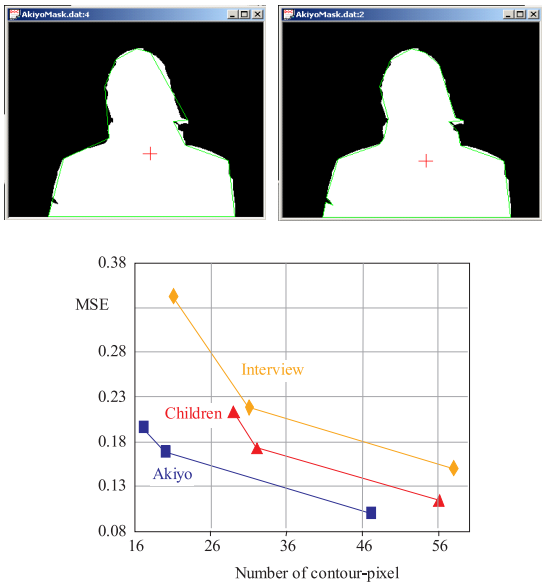


**Fig. 3.** Approximation error versus vertices number of outlining polygon.

The first contour-pixel is detected by selecting an arbitrary internal pixel as a start point (pixel "A" in Fig. 1). Then the process traces upward vertically from this pixel until the topmost contour-pixel is found (the pixel "B" in Fig. 1). Knowing the position of the $i^{\text{th}}$ contour-pixel, the next $(i+1)^{\text{th}}$ one is searched in two successive scans as followings

*First scan*: the process checks whether the contour-pixel classification is matched at the neighbor T, L, B or R. The qualified pixel will be added into the ordered list containing all, previously found contour-pixels of the VOP. The scanning process then repeats again at this new contour-pixel. The iteration will rolls back to the previous contour-pixel (in correspondence with the ordered list) if all the neighbors T, L, B and R are internal.

*Second scan*: it will start when the first scan returns to the first contour-pixel in the ordered list. The process now checks the contour-possibility of the remaining neighbor TR, BR, BL and TL for every contour-pixel as it occurs in the ordered list. If at the $k^{\text{th}}$ position of the list, a new contour-pixel is found (pixel $k_1$ in Fig. 1), the first scan will be applied to this new target; the resulting new segment of contour-pixel (pixel $k_1$, $k_2$, $k_3$ and $k_4$ Fig. 1) will be inserted in to the ordered list immediately right after the $k^{\text{th}}$ position.

The above scanning processes ensure the lossless presentation of shape information for VOP, which consists of several connected segments.

## 2.2 Shape approximation with polygon

The efficiency of a contour-based shape coding method depends so much on the number of contour-pixels. Therefore the reduction in number of contour-pixels is necessary when the shape detail is uninteresting (when origin shape resulted from an inadequate segmenting preprocess is already noisy, or in low bitrate applications). The art of lossy coding method here lies in selecting appropriate contour-pixels, which construct the acceptable fitting polygon to the origin shape-curve. As in [8], the fitness of the polygon is defined as constraining the approximation error. We consider that the edge $P_kP_{k+n}$ of the polygon approximates the original contour drawn by the $n+1$ pixels located from $k^{\text{th}}$ to $(k+n)^{\text{th}}$ position in the ordered list. With $\mathrm{d}(P_kP_{k+n}, P_x)$ being the Euclidean distance between contour-pixel $P_x$ $(x \in [k, k+n])$ and the edge $P_kP_{k+n}$, the approximation error for this segment of contour is given by $\mathrm{d}_{\max}(k, k+n)$:

The construction the polygon has the following steps:

1. The first pixel in the ordered list of the contour-pixels is chosen as the first vertex of the polygon.
2. Search in the ordered list for the farthermost pixel from the first vertex. It is the second vertex of the polygon. We already create the first edge of the polygon.
3. For each edge of the polygon, it is checked whether the approximation lies within a given tolerance $\mathrm{d}_{\max}(k, k+n) < threshold$. If not, a new vertex for the polygon is inserted at the pixel $P_x$ having the largest approximation error. The polygon is then added with two more edges.
4. Step 3 is repeated until the approximation errors for all edge lie within the allowable range.

Figure 3 demonstrates the results of approximating Akiyo VOP with 20 contour vertices, $MSE = 0.169$ (Upper left image) and 12 contour vertices $MSE = 0.2354$ (Upper right image). The lower graphic shows the trend between used contour vertices and approximation error for three types of VOP: Children, Akiyo and Interview.

## 2.3 Constrained Delaunay mesh generation

Our VOP now has an arbitrary shape of the well-fitted polygon created as above. Its texture will be encoded in the motion compensation manner based on mesh structure. Making use of the constrained Delaunay mesh topology, we have several advantages as the followings:

- The mesh retains the shape of the VOP.
- In order to reconstruct the mesh at the decoder, there is no need to transmit the connectivity information of the mesh. The geometry information vertex position-uniquely defines the topology.
- In the constrained Delaunay structure, most of the triangles possess the property that no vertex in the mesh falls in the interior of the so-called circumcircle (circle that passes through all three vertices) of any triangle of the mesh. Hence the structure warranties the "regular shape" of every piecewise triangle area. It eases the estimation of vertices' motion vectors and piecewise warping process applied to the internal texture.

The constrained Delaunay mesh is generated onto the shape VOP as in [3]. The mesh generation must be controlled so that: 1. Minimum number of vertices are created inside the VOP to reduce the cost of encoding their

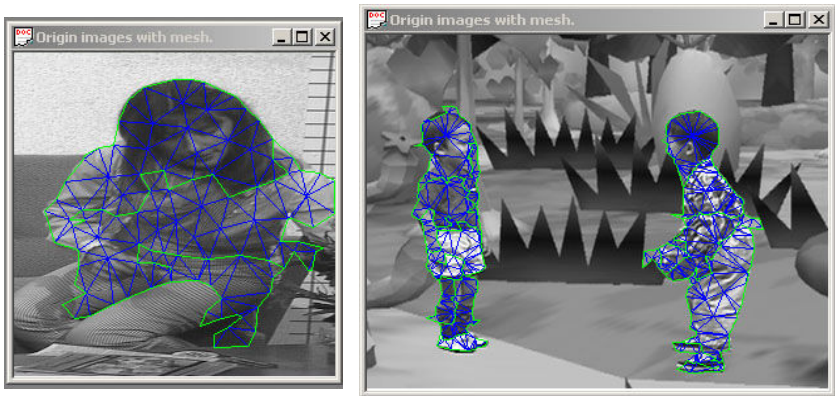


**Fig. 4.** Representation of a VOP from Interview (Left) and Children (Right) sequences with constrained Delaunay meshes.

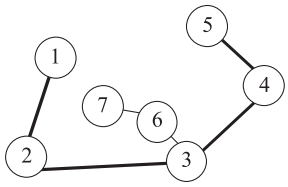


**Fig. 5.** Prim's traversal with forced border nodes.

positions; 2. The areas of emerged triangles must be relatively small for better performance of texture compensation based on piece-wise warping of every area inside the associated triangles.

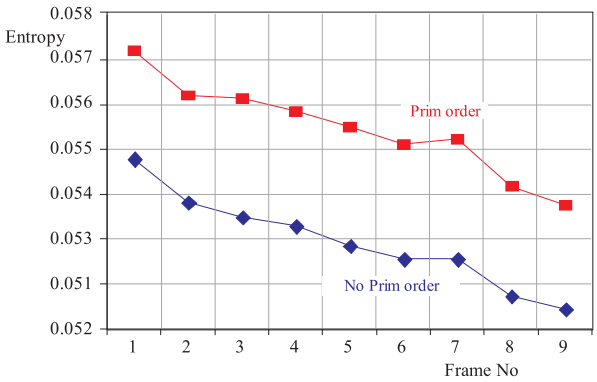


**Fig. 6.** Performance of Prim ordering.

### 2.4 Geometry encoder

We encode both horizontal and vertical coordinates of vertices in two steps: Prim ordering and predictive coding. The first process is necessary to generate an optimal estimator for high compression in the second one. We propose optimal graph traversal to encode each vertex position differentially with regard to the previous one. One way to compute a minimum spanning tree is to grow the tree in a successive stage; in each stage, one vertex is picked as the root, and we add an edge, thus an associated vertex, to the tree. At any point in the algorithm, we can see that we have a set of vertices that have already been included in the tree; the rest of them have not. The algorithm then finds, at each stage, a new vertex to add to the tree by choosing the edge $(u, v)$ such that the its cost $c(u, v)$ is the smallest among all edges, where $u$ is in the tree and $v$ is not. The iteration is going on util all nodes are included in the tree. It is the core implementation of the so-called Prim's algorithm [12]. In our case, the Euclidean distance between adjacent nodes is chosen as the cost function for optimal graph traversal. For the vertices lying on the border of a region, their cost function is forced to be zero. In consequence, the Prim's algorithm chooses these contour vertices in the first positions along the spanning tree path (the bold line in Fig. 5). We thus implicitly aid the decoder to distinguish the boundary nodes from the normal ones, which is essential for reconstruction of the constrained Delaunay mesh (obviously number of boundary node must be transmitted as side-information). After the minimum spanning tree is created, each vertex is traversed and at the same time, its position is encoded differentially using the position of the previously vertex (in that order of the tree) as a predictor. The absolute value of the first node is sent directly. Besides these integer-values (vertices' positions and their differences are integers), some marking symbols can be found in the bit-stream whenever a node with a degree higher than 2 is met (it has some branches), the end of the branch is encountered and the end of the bit-stream is reached. They are specific flag Pu (Push), Po (Pop) or En (End) respectively. For demonstration, the bit-stream for the example shown in Fig. 5 is the following: $X_1$, $Y_1$, $\Delta X_2$, $\Delta Y_2$, $\Delta X_3$, $\Delta Y_3$, $Pu$, $\Delta X_4$, $\Delta Y_4$, $\Delta X_5$, $\Delta Y_5$, $Po$, $\Delta X_6$, $\Delta Y_6$, $\Delta X_7$, $\Delta Y_7$, $(Po)$, $En$. The subscripts, $X$, $Y$, $\Delta X$ and $\Delta Y$ are indexes of vertices, two coordinate value and their differences in that order. As $\Delta X_6$, $\Delta Y_6$ follows the $Po$ flag, the coordinates are encoded differentially with regard to the last point which is encoded just before the farthest $Pu$ flag from the current $F$ flag (Last In First Out LIFO order as in Push and Pop operator with memory stack). The last $Po$ flag is optional.

## 3 INTEGRATION OF SHAPE CODING IN A FULL VOP CODEC SCHEME

Figure 7 outlines the block diagram of our full VO codec scheme. The shape encoder discussed in section 2 plays a center role in this closed chain of codec. The scheme is a quasi- automatic process. The set of segments resulted from VOP detector unit (in a simple case, it is a segmentation process for each frame) for an intra-frame

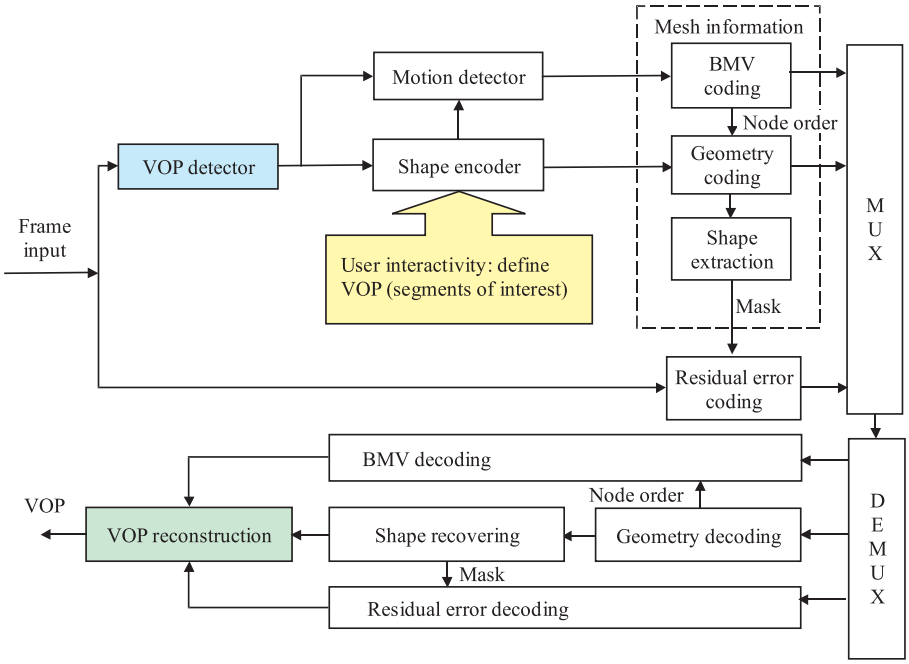


**Fig. 7.** Mesh-based VO codec scheme.

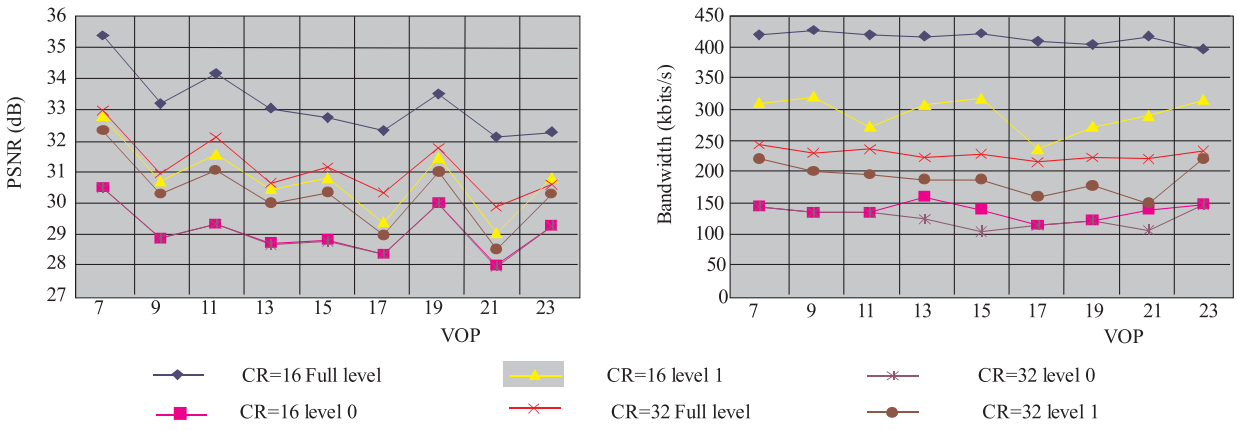


**Fig. 8.** Overall performance of the full codec scheme at several bitrate.

(where the VO of interest first appears) is offered to end user (author, composer) so that he/she can link them together according to their logical meaning, that is, to declare several VOPs possessing one or more regions. It is the only one phase, in which human interaction is involved (therefore it is quasi automatic). Then the shape coder constructs the constrained Delaunay mesh onto the segments of interest. In the successive inter-VOP, motion vector of every node of the mesh is predicted. In the current work, we deal with mesh consisting of a constant number of nodes along the time. As a result, possessing of nodal vectors in an inter-VOP (they are backward motion vectors BMVs) together with the adjustment of geometry applied to the previous VOP (they are forward motion vectors FMVs), decoder can reconstruct the texture of VOP compensated with residual error transmitted in a separate channel (Fig. 7). We refer readers to [9] for detail discussion of how we encode nodal motion vectors, nodal positions and residual errors. There the VOP detector is processed in a sophisticated way, deploying temporal-spatial segmentation, which produces constrained Delaunay mesh structure and their related nodal motion vectors at the same time. The goal of the scheme proposed in this work is to simplify the VO extraction phase. As a result, the shape and VOP presentation can be deployed independently with a larger scale of VO detecting techniques.

Figure 8 presents the overall performance of the scheme applied to Interview sequence with bounding rectangle of size $256 \times 256$ pixel, 25, fps. The MSE for shape approximation is kept under 0.2 for every segments of VOP. FMVs and BMVs are encoded as in [9]. Residual errors are transformed into 7 subbands (two levels) with biorthogonal 9/7 Daubechies wavelet - proposed filter set in MPEG-4 standard. The obtained coefficients are then fed to a dynamic allocator with a settable compression rate (CR): CR = (No of Bit/1 Coeff)/(8 Bit). The coefficients of subbands can be decoded partly, depending

on the bandwidth the decoder can process. The level 0 in Fig. 8 means that only the coefficients of all-pass subband are taken into account.

## 4 CONCLUSION AND PERSPECTIVES

The boom of the Internet together with a rapid development in digital storage capacity makes multimedia applications more and more popular and indispensable, thanks to their possible integration of multi information-source through a uniform network and media. In this new era, the efficiency of image/video compression, the object-oriented compression and manipulation for audio/video become real challenges to researchers. Taking the advantages of some new key-techniques defined in MPEG-4 standard, our scheme tends to reach closer to these targets. VOs are extracted in a quasi-automatic manner, constrained Delaunay mesh is deployed to characterize the concerned VOs. Transmission of their positions not only saves the transmission of mesh connectivity, but also creates a vertex-base for processing the revolutionary mesh-based motion compensation for texture of VOs.

The proposed scheme still has several points for optimization. We are working on the presentation of shape in a scalable way: bounding polygon can be refined in a progressive manner. Nodal motion vectors in the current work is simply block-based searching. Topology of the mesh can also be involved as a hint for the motion detector to reduce the uncertainty of decision.

## References

[1] ISO/IEC JTC1/SC29/WG11 N3908 MPEG-4 Video Verification Model version 18, January 2001/Pisa.

[2] OSTERMANN, J.—JANG, E. S.—SHIN, J. S.—CHEN, T.: Coding of Arbitrarily Shaped Video Objects in MPEG-4, Proceedings of ICIP '97, Santa Barbata, 1997.

[3] SHEVCHUK, J. R.: Triangle Program, http://www.cs.cmu.edu/~quake/triangle.html.

[4] CHEN, T.—SWAIN, C. T.—HASKELL, B. G.: Coding of Subregions for Content-Based Scalable Video, IEEE Trans. On Circuits and Systems for Video Technology **7** (1997), 256–260.

[5] BRADY, N.—BOSSEN, F.: Shape Compression of Mving Objects Using Context-Based Arithmetic Encoding, Signal Processing: Image Communication **15** (2000), 601–617.

[6] YAMAGUCHI, N.—IDA, T.—WATANABE, T.: A Binary Shape Coding Method Using Modified MMR, Proceeds of ICIP 97, Santa Barbata, 1997.

[7] LEE, S. H.—CHO, D. S.—CHO, Y. S.—SON, S. H.—JANG, E. S.—SHIN, J. S.—SEO, Y. S.: Binary Shape Coding Using Baseline-based Method, IEEE Trans. on Circuits and Systems for Video Technology **9** (1999), 4458.

[8] KIM, J. I.—BOVIK, A. C.—EVANS, B. L.: Generalized Predictive Binary Shape Coding Using Polygon Approximation, Signal Processing: Image Communication **15** (2000), 643–663.

[9] TRAN MINH, S.—FAZEKAS, K.—BENOIS-PINEAU, J.—GSCHWINDT, A.: Full Scheme of MPEG4-Like Codec Based on Wavelet Transform, Proceedings of ICASSP 2002.

[10] MA, C. C.—CHEN, M. J.: Vertex-Based Shape Coding in Polar Coordinates for MPEG-4, Proceedings of the Taiwan Area Network Conference, November 1998.

[11] Shape Coding Ad-Hoc Group, Core Experiments on MPEG-4 Video Shape Coding, ISO/IEC JTC1/SC29/WG11 N1326, Chicago, IL, October 1996.

[12] WEISS, M. A.: Data Structures and Algorithm Analysis, the Benjamin/Cummings Publishing Company, Inc, 1994.

[13] TURAN, J.: Fast Translation Invariant Transforms and Their Applications, elfa, Košice, 1999.

Received 20 April 2005

**Son Minh Tran**, born in 1973, received the MSc degree in electronic engineering in 1998 from the Budapest University of Technology and Economics (Budapest, Hungary). He then continued his research at the same university as a PhD student until 2004. His dissertation was about enhanced methods for representing multimedia. Currently, he is a Post Doctoral Fellow within the ARTEMIS Project Unit at Institut National des Télécommunications. Besides the main specialization as a digital video broadcasting engineer, his research interests include video compression algorithms and enhanced features for digital video broadcasting.

**Lajos Konyha** was born in Budapest, Hungary, in 1978. He received the MSc degree in electrical engineering from the Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Department of Microwave Telecommunications, in 2001. Since 2001 he has been a PhD student at the Budapest University of Technology and Economics, Department of Broadband Infocommunications and Electromagnetic Theory, Media Technology Laboratory and Rohde&Schwarz Reference Laboratory. As a PhD student, he deals with one and more dimensional signal processing, transform image coding, video and image compression.

**Balázs Enyedi** was born in Budapest, Hungary, in 1978. He received the MSc degree in electrical engineering from the Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Department of Electric Power Engineering, in 2001. He worked for Matáv Rt. (Hungarian Telecommunications Company Limited), IT Directorate, Customer Care Systems Department as IT Project Manager from 2001 to 2003. He dealt with CRM and OSZTR (Nationwide Computerized Information Bureau System) systems. Since 2001 he has been a PhD student at the Budapest University of Technology and Economics, Department of Broadband Infocommunications and Electromagnetic Theory, Media Technology Laboratory and Rohde&Schwarz Reference Laboratory. As a PhD student, he deals with image and signal processing methods, segmentation, video and still image compression.

**Kálmán Fazekas** (Prof, PhD, Dr) studied electronics at TU Budapest, now is working as professor at TU Budapest. His research interests include multimedia signal processing, teleeducation and image coding. Research activities: digital image processing/coding: multiresolution decomposition methods, VLBR coding, object-based coding/processing (MPEG-4, H.264), motion estimation; image communication, DVB, multimedia technology, distance education.

**Ján Turán** (Prof, Ing, RNDr, DrSc, PhD) studied physical electronics at ČVUT and Charles University Prague, PhD and DrSc degree received from TU Košice in radioelectronics, now is working as full professor at TU Košice. His research interests include multimedia signal processing, fiber optics communication and sensing.