Coding Algorithms for 3DTV—A Survey

Aljoscha Smolic, Karsten Mueller, *Senior Member, IEEE*, Nikolce Stefanoski, *Student Member, IEEE*, Joern Ostermann, *Fellow, IEEE*, Atanas Gotchev, *Member, IEEE*, Gozde B. Akar, *Senior Member, IEEE*, Georgios Triantafyllidis, *Member, IEEE*, and Alper Koz, *Member, IEEE*

(Invited Paper)

***Abstract*—Research efforts on 3DTV technology have been strengthened worldwide recently, covering the whole media processing chain from capture to display. Different 3DTV systems rely on different 3-D scene representations that integrate various types of data. Efficient coding of these data is crucial for the success of 3DTV. Compression of pixel-type data including stereo video, multiview video, and associated depth or disparity maps extends available principles of classical video coding. Powerful algorithms and open international standards for multiview video coding and coding of video plus depth data are available and under development, which will provide the basis for introduction of various 3DTV systems and services in the near future. Compression of 3-D mesh models has also reached a high level of maturity. For static geometry, a variety of powerful algorithms are available to efficiently compress vertices and connectivity. Compression of dynamic 3-D geometry is currently a more active field of research. Temporal prediction is an important mechanism to remove redundancy from animated 3-D mesh sequences. Error resilience is important for transmission of data over error prone channels, and multiple description coding (MDC) is a suitable way to protect data. MDC of still images and 2-D video has already been widely studied, whereas multiview video and 3-D meshes have been addressed only recently. Intellectual property protection of 3-D data by watermarking is a pioneering research area as well. The 3-D watermarking methods in the literature are classified into three groups, considering the dimensions of the main components of scene representations and the resulting components after applying the algorithm. In general, 3DTV coding technology is maturating. Systems and services may enter the market in the near future. However, the research area is relatively young compared to coding of other types of media. Therefore, there is still a lot of room for improvement and new development of algorithms.**

***Index Terms*—Depth coding, disparity coding, MPEG, multiview video coding (MVC), multiple description coding (MDC), stereo video coding, watermarking, 3-D mesh compression, 3DTV.**

Manuscript received March 6, 2007; revised June 2, 2007. This paper was recommended by Guest Editor L. Onural.

A. Smolic and K.Müller are with the Fraunhofer Institute for Telecommunications-Heinrich-Hertz-Institut, (HHI), 10587 Berlin, Germany (e-mail: smolic@hhi.de).

N. Stefanoski and J. Ostermann are with the Leibniz Universität Hannover, Institut für Informationsverarbeitung, 30167 Hannover, Germany.

A. Gotchev is with the Institute of Signal Processing, Tampere University of Technology, Korkeankoulu 1, FI-33720 Tampere, Finland.

G. B. Akar and A. Koz are with the Department of Electrical and Electronics Eng., M.E.T.U., 06531 Ankara, Turkey.

G. Triantafyllidis is with ITI-CERTH, Thermi 57001, Thessaloniki, Greece.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.909972

I. Introduction

EXTENDING visual sensation to the third dimension has been investigated over decades. However, significant consumer mass markets haven't developed yet. 3-D video is established in niche markets, including professional applications (e.g., scientific visualization, medicine) and entertainment (IMAX cinemas, 3-D gaming). In recent years, research efforts have been strengthened worldwide to remove technological obstacles that encumber wider success of 3-D video applications [1]–[5]. Significant improvements have been achieved for all components of the processing chain, from acquisition and signal processing, over 3-D scene representation, coding and transmission, to rendering and 3-D display. Very likely various 3-D video systems, components, applications and services will enter the market in the near future.

Overviews of the state-of-the-art of technology for 3-D video are given in different papers of this Special Issue, each focusing on a specific component of the processing chain. This paper is devoted to coding. Various 3-D scene representation formats are used in different 3-D video systems and applications as described in [6]. These formats integrate various types of data, such as multiview video, and geometry data in form of depth or 3-D meshes. In general, 3-D video results in a tremendous amount of data that needs to be transmitted, stored or watermarked. Therefore, efficient compression is a key condition for the success of 3-D video. There is also a strong necessity for developing robust 3-D watermarking techniques, which protect the ownership rights. Further, the availability of open international standards is in general an important enabling factor for the development of markets in the media business. ISO/IEC JTC 1/SC 29/WG 11 (Moving Picture Experts Group—MPEG) is one of the international standardization bodies that play an important role in digital media standardization. Therefore, this paper highlights MPEG standards for the different data as available and under development.

The following section gives an overview of coding of multiview video, depth and associated data, with focus on available and emerging MPEG standards. Section III is devoted to compression of static and dynamic 3-D mesh data as used for 3-D video representations as well as for 3-D computer graphics. Error resilience and multiple description coding (MDC) for 3-D is outlined in Section IV. Then, Section V elaborates on protection of 3-D content using watermarking. Finally, Section VI summarizes and concludes the paper.

II. Coding of Multiview Video, Depth, and Associated Data

Many 3DTV systems are based on scenarios, where a 3-D scene is captured by a number of N cameras (see e.g., [7]–[14]). The simplest case is classical stereo video with two videos. More advanced systems apply 8, 16, and more cameras. Some systems additionally apply per sample depth data that can also be treated as video signals. This section gives an overview of compression algorithms and standards for such data. An early overview of this research area can be found in [15]. Depending on the degree of common content, shared by a subset of the cameras, a coding gain can be achieved in comparison to single-view coding. In multiview coding, correlations between adjacent cameras are exploited in addition to temporal correlations within each sequence. Therefore, multiview coding adds another compression dimension on top of single-view coding: the inter-view direction.

A. Conventional Stereo Video Coding

Stereo view is the most important special case of multiview with $N = 2$ views. Compression of conventional stereo video has been studied for a long time and the corresponding standards are available. A conventional stereo pair consists of two images showing the same scene from two slightly different viewpoints corresponding to the distance of human eyes. The images are in general very similar, which makes them well suited for compression, e.g., with one image predicting the other. For instance one of them can be compressed without reference to the other stereo image. Then, the second image can be predicted from the already encoded one, just like temporally related images can be motion-compensated in video compression.

The samples of both images correspond to each other through the 3-D geometry of the scene and camera properties, including positions and internal camera parameters such as the focal length. The displacement or disparity of each sample in one image with respect to the other is equivalent to a dense motion field in between two consecutive images of a video sequence. Therefore, it is justified to use the same principles of motion estimation and motion compensation for disparity estimation and disparity compensation for image prediction and then to only encode the prediction error or residual further.

Nevertheless, some specific differences between motion compensation and disparity compensation need to be considered. The statistics of disparity vector fields is different from the statistics of motion vector fields. Disparities are biased and relatively large. Zero disparity means a very large depth of the corresponding point in 3D, while 3-D points close to the camera may have very large disparity values. This may require adjustments of entropy coding of the disparity vectors. In general temporally adjacent images of a video sequence tend to be more similar than views of a stereo pair at practical frame rates. Disocclusion effects, i.e., content that is visible in one image is occluded in the other and can therefore not be predicted, are on average more evident in a stereo pair than in between two temporally adjacent video images. Further, specific differences in a stereo pair may come from incorrect white and colour balance but also due to scene lighting and surface reflectance effects.

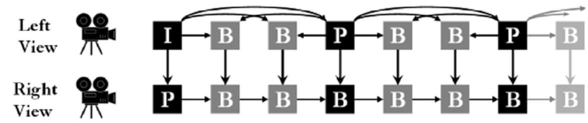

Fig. 1. Illustration of prediction in H.262/MPEG-2 Video multiview profile.

The combination of inter-view and temporal prediction is the basic principle for efficient compression of conventional stereo video. A corresponding standard specification has already been defined in ITU-T Rec. H.262/ISO/IEC 13818-2 MPEG-2 Video, the Multiview Profile [16], [17], as illustrated in Fig. 1. The left eye view is encoded without reference to the right eye view, using standard MPEG-2. This ensures backward compatibility with Main Profile of H.262/MPEG-2 Video, since it is possible to decode the left eye view bit stream and to display 2-D video. For the right eye view, inter-view prediction is allowed in addition to temporal prediction.

However, the gain in compression efficiency compared to independent encoding of both video streams is rather limited. This is mainly due to the fact that temporal prediction already provides very good performance. Typically, if temporal prediction is efficient for a certain image (e.g., B-pictures for right view in Fig. 1) then additional inter-view prediction does not increase the coding performance significantly. Temporal neighbouring images are typically on average more similar than spatially neighbouring images.

For images that are coded as I-pictures, i.e., without reference to other temporally adjacent images in the video sequence, a significant gain can be achieved by inter-view prediction. Typically every 0.5–1 s such I-pictures are inserted into a video stream to enable random access and error robustness. In Fig. 1, the left-hand side picture of the left view is encoded as I-picture. The corresponding left-hand side picture of the right view would also be encoded as I-picture for random access when independently encoding both video streams. However, in H.262/MPEG-2 Video multiview coding, inter-view prediction can be applied, resulting in a significant increase of compression efficiency compared to coding this picture as I-picture.

Research on compression of conventional stereo video has continued into several directions, including for instance optimum joined bit allocation for both channels, or abandoning backward compatibility to design more efficient inter-view prediction structures. Algorithms have been based on more up-to-date video codecs such as H.263 [18], MPEG-4 Visual [19], or H.264/AVC [20]–[22]. However, none of the developments including the original Multiview profile have reached commercial relevance so far, since the application of stereo video did not develop into a relevant mass market yet.

B. Compression of Video Plus Depth Data

An alternative to classical stereo video as described in the previous section is to transmit a video signal and a per sample depth map. From the video and depth information, a stereo pair can be rendered at the decoder [23], [24]. This extends the functionality since it enables head motion parallax viewing if the user's head motion is tracked. Additionally, this format is interesting from compression efficiency point of view. Per sample depth data can be regarded as a monochromatic, luminance-only video signal. The depth is restricted to a range between two extremes $Z_{\rm near}$ and $Z_{\rm far}$ indicating the minimum and maximum distance of the corresponding 3-D point from the camera respectively. The depth range is linearly quantized with 8 bit, i.e., the closest point is associated with the value 255 and the most distant point is associated with the value 0. With that, the depth map is specified, resulting in a grey scale image. These grey scale images can be fed into the luminance channel of a video signal and the chrominance can be set to a constant value. The resulting standard video signal can then be processed by any state-of-the-art video codec.

Results from the European ATTEST project [23] have shown that depth data can be very efficiently compressed this way. Several state-of-the-art video codecs have been tested (MPEG-2, MPEG-4, H.264/AVC). A course estimate indicates that 10%–20% of the bit rate which is necessary to encode the colour video is sufficient to encode the depth at good quality. This is due to the specific statistics of depth data, being on average smoother and less structured that colour data.

Based on these observations, a new backward compatible (with respect to classical DVB) approach for 3DTV was developed in the ATTEST project. It uses a layered bit stream syntax. The base layer is a conventional 2-D colour video encoded using MPEG-2. This base layer can be processed by any existing MPEG-2 decoder providing backward compatibility. Additionally the bit stream contains an advanced layer carrying the encoded depth information. Advanced systems may access this layer to decode the depth stream and then generate a stereo pair to be displayed stereoscopically by view interpolation.

This concept is highly interesting due to the backward compatibility, compression efficiency and extended functionality compared to conventional stereo video. Moreover, it does not introduce any specific coding algorithms. It is only necessary to specify high-level syntax that allows a decoder to interpret 2 incoming video streams correctly as colour and depth. Additionally information about depth range ($Z_{\rm near}$ and $Z_{\rm far}$) needs to be transmitted. Therefore, MPEG specified a corresponding container format "ISO/IEC 23002-3 Representation of Auxiliary Video and Supplemental Information," also known as MPEG-3 Part 3, for video plus depth data [25], [26]. Moreover, H.264/AVC contains an option to convey the depth images through its auxiliary picture syntax. Here, the video codec for the colour video signal and associated depth video signal are both H.264/AVC. This approach is backwards compatible with any existing deployment of H.264/AVC.

A general problem of the video plus depth format is content creation, i.e., the generation of depth information. Cameras that automatically capture per pixel depth with the video are available and are being further enhanced, but the quality of the captured depth fields is currently still limited. Algorithms for depth estimation have been studied extensively in computer vision literature and powerful solutions are available. However, it always remains an estimation that can only be solved up to a residual error probability. Estimation errors influence the quality of rendered views. A fully automatic, accurate and reliable depth capturing system is still to be developed. User-assisted content generation is an option for specific applications. Even having perfect depth available, artifacts may occur in rendered views due to dis-occlusion. This effect increases with the distance of the virtual view from the original camera position. Additional occlusion layers (layered depth video as extension of layered depth images [27]) or extension to multiview video plus depth [28], [29] help to minimize these problems at the cost of increased data rate and complexity.

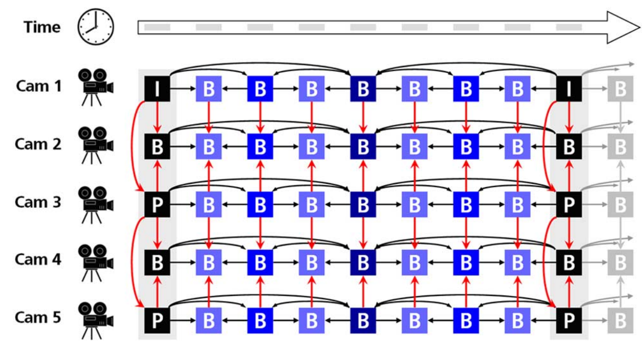

Fig. 2. Temporal/inter-view prediction structure for MVC.

C. Multiview Video Coding

A common element of many 3DTV systems is the use of multiple views of the same scene that have to be transmitted to the user. The straight-forward solution for this would be to encode all video signals independently using a state-of-the-art video codec such as H.264/AVC. However, multiview video contains a large amount of inter-view statistical dependencies, since all cameras capture the same scene from different viewpoints. These can be exploited for combined temporal/inter-view prediction, as illustrated in Fig. 2. Images are not only predicted from temporally neighbouring images but also from corresponding images in adjacent views. Statistical evaluations have shown that significant gain can be expected from such combined temporal/inter-view prediction [30], [31]. Pioneering work on multiview image coding is reported in [32], [33].

Several research groups addressed multiview video coding (MVC) and developed dedicated inter-view/temporal prediction structures to efficiently exploit all statistical dependencies within the multiview video data sets (see e.g., [34]–[41]). Among those, algorithms that are based on hierarchical B-pictures [42] as supported by H.264/AVC syntax in temporal and inter-view dimension (Fig. 2) proved best performance in exhaustive experiments conducted in the context of MPEG standardization [43]–[46]. In these experiments, it has been shown by objective and subjective measurements that dedicated MVC outperforms independent encoding of the multiple video streams significantly. However, the achievable gain strongly depends on the content and its properties such as camera distance, frame rate and complexity of the content (motion, texture). For some data sets the peak signal-to-noise ratio (PSNR)gain was reported as 0.5 dB and below. Maximum reported gains were up to 3 dB.

A drawback of combined temporal/inter-view prediction as illustrated in Fig. 2 is the complexity. This includes computational complexity, memory requirements and delay. In [46], [47] it has been shown that the complexity can be significantly decreased without sacrificing much coding efficiency. Inter-view prediction is restricted here to the key pictures that would be treated as I-pictures in independent encoding of the views (e.g., time t_0 and t_8 in Fig. 2). Most of the coding gain of MVC comes from inter-view prediction of these pictures that do not use temporal prediction for temporal random access reasons. Omitting inter-view prediction for pictures that have a temporal reference does not cost much coding efficiency whereas it decreases complexity significantly.

In addition to combined temporal/inter-view prediction, specific MVC algorithms have been proposed. The basic idea of depth/disparity-based view interpolation prediction [48]–[50] is to estimate depth or disparity either at the encoder (this requires overhead for sending the depth/disparity) or the decoder, and to perform view interpolation or 3-D warping for prediction. However, the gains reported so far are marginal. Only for very few test data sets with very close camera settings such view interpolation prediction provides a gain of up to 5% bit rate saving at the same visual quality. Further investigations are needed to optimize the performance.

Further, illumination and color inconsistencies affect the exploitation of inter-view statistical dependencies. Usually such effects should be minimized by proper setting of the conditions, however, an MVC algorithm should also be able to cope with this as well, since proper white and color balancing of the input can not be guaranteed. Also, the illumination (spotlights, shadows, etc.) varies largely over the multiview images due to the lighting conditions. These problems might be handled by proper illumination and color compensation as proposed in [51] and [52]. The basic idea is to modify the motion compensation on macroblock level. Before subtracting the pixel values of the block to be encoded and the reference block, the mean of each is compensated from the corresponding pixel values. This assumes locally constant illumination and color variations, which is an appropriate model trading-off accuracy and complexity. Gains of up to 0.7 dB have been reported for some test data using illumination compensation. However, this strongly depends on the test data and in some cases the gain is negligible or there is none at all. On average over all test data sets a bit rate reduction of 5% was reported for the same visual quality.

An alternative to illumination compensation on macroblock level integrated into the encoding process, also an appropriate preprocessing can be applied, prior to encoding. Algorithms for illumination correction are well-known in image and video processing. Then the corrected data can be passed to a standard encoder. The big advantage of such an approach is that no changes are necessary to the encoder, decoder and bit stream syntax. A preliminary investigation in this direction is presented in [53], however, results are not complete and performance in comparison to integrated illumination compensation is not clear.

Another research direction is improving disparity estimation, compensation, and coding [54]. Mostly disparity is treated equally to motion, however, it is known that the statistical properties of disparity vectors can be quite different compared to those of motion vectors. Disparity estimation has been studied extensively in the computer vision literature. Usually, basic geometric properties and constraints are taken into account. For instance, the search can be done along epipolar lines. This may lead to better estimates and reduce the complexity. Further, specific disparity coding may improve the efficiency of inter-view prediction. Specific coding modes for MVC such as the inter-view direct mode [55] are under investigation.

Combination of scalability with MVC is investigated for instance in [56]–[60]. So far the feature of scalability only comes with decreased compression efficiency. Distributed MVC is investigated for instance in [61]. Also first work on efficient transport and delivery taking user interactivity into account has been presented [62], [63]. Finally, efficient mechanisms for optimum random access, parallel processing and memory management for MVC have been investigated.

Since the results clearly indicate that MVC outperforms independent encoding of the multiple video signals, and there is a clear demand from industry for a corresponding standard, ISO/MPEG and ITU/VCEG decided developing a dedicated MVC specification [64], [65]. It will be an extension of H.264/AVC (Amendment 4), which is scheduled to be finalized in 2008. MVC is the main focus of this Special Issue. We therefore refer the reader to the dedicated articles in this volume that contain all details of state-of-the-art MVC technology.

D. Conclusions and Future Research Directions

Research on coding of stereo video, multiview video and associated depth or disparity data has reached a good level of maturity. Related international standards are available enabling a variety of 3DTV systems and applications. However, compared to coding of other types of media data the scientific field is relatively young, and therefore there is still a lot of room for improvement of algorithms. This includes for instance optimization of MVC and development of new specific coding MVC algorithms. Depth or disparity coding may be improved further by development of dedicated algorithms. Further, there are more complex types of data representations for 3DTV such as layered depth video and multiview video plus depth that provide extended functionality. Efficient coding algorithms for such data are still under investigation. Initial results are reported in [66] and [67]. Other types of multiview representations are also under investigation often related to specific types of 3-D displays. An example is presented in [68].

III. 3-D Mesh Compression

3-D meshes are used in 3-D video representations to represent the shape of static or dynamic 3-D objects. 3-D triangle meshes are a common form to represent an object's shape. In fact, a triangle mesh represents only a piecewise-linear approximation of an object's surface. Consequently, the approximation error can be high unless the number of triangles is sufficiently large. On the other hand a large number of triangles makes these *static 3-D meshes* expensive to store or to transmit. Static 3-D meshes consist of two types of data: *connectivity*, which describes the triangulation of mesh vertices, and *geometry*, which assigns 3-D locations to vertices. *Dynamic 3-D meshes* are in their generic form represented as a sequence of static meshes with a common connectivity called frames. They require even several times more storage than a single static mesh, unless frames are compressed. Static as well as dynamic meshes show topological as well as geometrical dependencies in spatial and spatio–temporal domain, respectively, which are exploited for compression. Subsequently, we give a survey on compression techniques for static and dynamic 3-D meshes, as they are used in 3-D video representations and 3-D computer graphics.

A. Compression of Static 3-D Meshes

During the last decade there has been a lot of research in the area of static mesh compression. In 1996, Taubin and Rossignac [69] proposed the method *Topological Surgery* (TS), which was the first method for lossless compression of mesh connectivity and compression of locations of mesh vertices with controllable loss. They decomposed connectivity in so called vertex and triangle spanning trees, which were encoded. Linear predictive coding was employed for compression of vertex locations. They were predicted in an order guided by connectivity using already encoded locations. Later, Taubin *et al.* [70] extended the TS approach in order to obtain a Levels of Detail representation of a compressed static mesh, providing progressive decoding from low to high resolution. They introduced the *Progressive Forest Split* (PFS) scheme using a forest, i.e., a set of trees, in order to describe connectivity in different resolution levels. Both, TS and PFS, are building the basis of the *MPEG-4 3-D Mesh Coding* standard (3DMC tools) for Single Resolution and Progressive Levels of Detail static mesh compression. Many improvements and generalizations upon the TS approach were presented later. Here, we want to point out the *Edgebreaker* technique [71] of Rossignac and the *Cut-Border Machine* [72] of Gumhold and Straßer. Both methods are very similar and use a finite state machine to compactly describe mesh connectivity. They apply face based vertex traversal emitting one out of five symbols each time a new vertex is visited, with each symbol describing the configuration of this vertex relative to the traversed region. Subsequently this stream of only five different symbols is entropy encoded. It can be shown that Edgebreaker and the Cut-Border-Machine encode planar mesh connectivity with no more than 4 and 5 bpv (bpv = bits/vertex), respectively. Vertex location prediction is improved by employing the so-called *Parallelogram Prediction Rule* [73]. A compression technique similar to Edgebreaker called *Face Fixer* [74] was developed hereon. It encodes polygonal meshes, i.e., it is not restricted to encode only triangle meshes like previous approaches.

The Parallelogram Predictor was first presented by Touma and Gotsman [73]. It is the most commonly applied predictor in the literature for Single Resolution Compression of vertex locations. A multiway extension of this predictor and a higher order predictor are presented in [77] and [78]. Touma and Gotsman proposed a valence-based technique (TG) for mesh connectivity compression. The TG technique produces a stream of symbols that mainly consists of valences of processed vertices. Due to the Euler characteristic of triangle meshes, vertex valences show low valence dispersion around the average valence 6. Consequently, this stream of few different valences can be well compressed. Recently, based on the TG coder, Alliez and Desbrun proposed improved valence based approaches for single resolution [75] and progressive levels of detail [76] static mesh compression. They state that valence based compression is optimal in the sense that it reaches asymptotically the upper entropy bound for planar mesh connectivity compression of about 3.25 bpv.

TABLE I
COMPRESSION RATIOS FOR CONNECTIVITY COMPRESSION

Author	Name	Connectivity (bpv)	Remarks
Taubin & Rossignac 98, [66]	Topological Surgery	~ 4	part of MPEG-4 standard
Gumhold & Straßer 98, [72]	Cut Border Machine	~ 4.36	5 bpv guaranteed
Rossignac 99, [71]	EdgeBreaker	~ 3	4 bpv guaranteed (later improved to 3.55)
Isenburg & Snoeyink 00, [74]	FaceFixer	~ 2.5	supports arbitrary polygon meshes
Touma & Gotsman 98, [73]	The TG coder	~ 2	~ 0 when regular! (i.e. all valences are equal to 6)
Alliez & Desbrun 01, [75]	Adaptive Valence-based	~ 1.85	~ 0 when regular! (i.e. all valences are equal to 6)

While in all approaches so far mesh compression is manly guided by mesh connectivity, in [79], a geometry-guided progressive compression approach is presented. Here, quantized 3-D vertex locations are partitioned in an octree-structure and subsequently, by traversing the tree from the root to the leaves, all geometry and connectivity changes are encoded. As other geometry-guided compression techniques we want to mention the wavelet (WL)-based approach [80] of Khodakovsky *et al.*, and the approach [81] of Karni and Gotsman based on spectral decomposition.

In order to increase the compression efficiency for the geometry part of a mesh, image-based compression techniques are developed. Geometry Images [82] are a lossy coding technique where a manifold triangular mesh is remeshed on a regular grid, i.e., an image. The geometry is represented by the RGB channels of the image while the connectivity is given implicitly by the regular structure of the grid, i.e., original mesh connectivity is not encoded. The image is compressed using a standard image encoder. Improvements and variants of this technique can be found in [83]–[85].

Overall, the area of static mesh compression has already reached a high level of maturity. In practice, connectivity is encoded with about 3 bpv, while the bit rate for vertex locations is about 16–23 bpv. Connectivity compression ratios are summarized in Table I. Altogether, we have bit rates of about 22 bpv for static meshes. For more extensive surveys on static mesh compression we want to refer the interested reader to [86] and [87].

B. Compression of Dynamic 3-D Meshes

Compression of dynamic 3-D meshes is a young research area. First compression approaches were presented 1999, and in the following years, only a few papers dealing with this topic appeared. However, since 2004, along with emerging research efforts in 3DTV, a high level of research activity can be noticed in this prospective area, which is reflected in many publications. Dynamic 3-D meshes, i.e., sequences of static meshes with nonchanging connectivity, show not only geometrical dependencies in spatial direction but also dependencies in temporal direction. Compression approaches exploiting these dependencies can be classified in two types: *transform dominated* and *prediction dominated*.

In 1999, Layngyel [89] first presented a transform dominated clustering-based approach for compression of dynamic 3-D meshes. Vertex trajectories, i.e., paths of single vertices in time, are grouped in clusters and for each cluster a transformation is selected which best approximates the global behavior of vertices. Selected transformations and resulting approximation errors are encoded subsequently. One year later Alexa and Müller [90] proposed a compression approach based on principal component analysis (PCA). They treat a dynamic mesh as a large *3VxF* matrix, where V and F denote the number of vertices and frames of the dynamic mesh respectively. Compression is achieved by decomposing this matrix using PCA and omitting basis vectors. Several authors combined and improved these two compression approaches. Karni and Gotsman [91] employed PCA and linear predictors (PCA+LP) in order to further exploit dependencies between interpolation coefficients for compression. Sattler *et al.* [92] applied the PCA approach to clusters of vertex trajectories (CPCA) and obtain even higher compression gains. Váša *et al.*, [93] suggest using trajectory based PCA in combination with EdgeBreaker mesh traversal. PCA coefficients are predicted using the Parallelogram Predictor, which yields lower entropy of residuals. This method has been shown to work very well for high precision meshes, where the number of vertices is much greater than the number of frames. Recently, Amjoun and Strasser [94] proposed an compression approach which applies the PCA approach to clusters of vertex trajectories, where trajectories were previously represented in a local coordinate frame showing significant gains compared to CPCA.

Guskov and Khodakovksy [95] employed WLs to encode dynamic meshes. They first determine a consistent parameterization for all frames and subsequently, based on this parameterization, they decorrelate each frame spatially using WLs. Simple differential encoding is then applied in order to exploit remaining temporal dependencies between WL coefficients for compression. Briceno *et al.* introduced Geometry Videos [88] that extend the concepts of geometry images to dynamic meshes. A geometry video provides a technique to handle an animated mesh like a video sequence. Recently, Mamou *et al.*, [96] introduced a novel technique for compression of mesh animations based on a skinning animation technique. They employ vertex clustering and use weighted affine transforms in order to exploit inter and intra cluster dependencies for compression.

Ibarria and Rossignac [97] presented 2003 a prediction dominated vertex traversal based compression algorithm. They applied a linear predictor which is based on the Parallelogram Prediction Rule, to exploit local inter and intra frame coherence between vertex locations. Jang *et al.*, [98] proposed a compression technique (AFX-IC) based on removing frames in a dynamic mesh and replacing them by interpolated ones, which are obtained by using remaining frames. Remaining frames are encoded using a classical DPCM and entropy coding structure. This coder is already part of the MPEG-4 standard. Müller *et al.*, [99] introduced a predictive octree-based motion-vector clustering approach (D3DMC) for compression of dynamic meshes. Octree-clustering is applied in order to group motion vectors, which are calculated between subsequent frames, assuming that motion vectors located in one cube or cluster of the octree describe similar motion. Recently, a rate-distortion optimized version of this approach (D3DMC-RD) was presented in [100]. Different coding modes for compression of single motion vectors and motion-vector clusters are introduced. The procedure of octree-clustering is improved by employing rate-distortion criteria for determining the optimal mode and granularity of the octree, in order to compress all motion vectors in the most efficient way. The authors show that D3DMC-RD leads to significant gains compared to AFX-IC. Stefanoski and Ostermann [101] presented a nonlinear predictor for vertex traversal based compression improving the prediction accuracy at high bit rates compared to linear predictors. Recently, Stefanoski *et al.*, [102], [103] introduced a layered predictive approach supporting and exploiting spatial (SSLPC) and spatio–temporal (STSLPC) scalability. Layers are defined by employing patch-based mesh simplification techniques, and a layer-wise predictor selection is applied in order to exploit linear and nonlinear spatio–temporal dependencies between layers and frames. They show that a layered configuration of vertices improves the exploitation of spatio–temporal dependencies for compression. Operational rate-distortion curves are presented in Fig. 3 comparing these coders with PCA and WL-based coders (progressive coders). Resulting curves demonstrate the superiority of layered approaches SSLPC and STSLPC. Currently, the layered approach of Stefanoski *et al.*, shows the best compression performance at quality levels relevant for applications when compared to other state-of-the-art approaches. Furthermore, it provides the feature of spatial and temporal scalability for free.

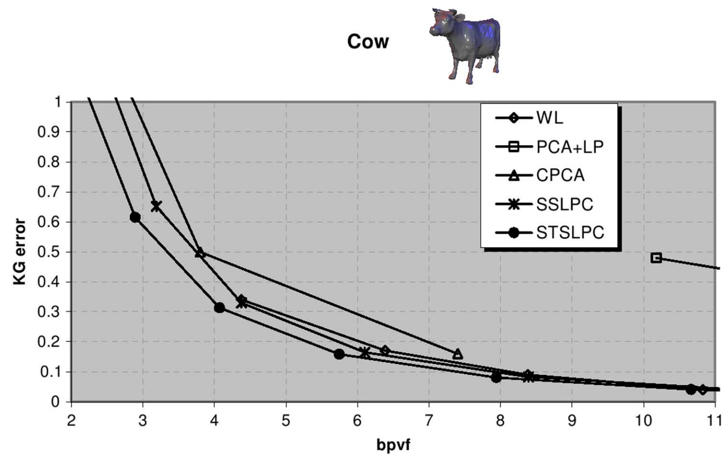

Fig. 3. Distortion over bit rate for selected coders. A KG error below 0.15 can be regarded as visually lossless.

C. Conclusions and Future Research Directions

The area of compression of static 3-D meshes has already reached a high level of maturity, which is reflected in many publications and an already existing MPEG standard. Compression of dynamic 3-D meshes is a young but nevertheless active and prospective research area, which receives many impulses from static mesh compression. Recently, MPEG started a new activity on the specification of an improved standard for compression of dynamic 3-D meshes.

Textured dynamic meshes, i.e., dynamic meshes with varying texture information at each time-instance, are used for 3-D video object representations. They allow a free-viewpoint representation, as it is required for interactive 3DTV applications. Only little attention has been paid so far to compression of dynamic meshes with time-varying texture information, pointing out promising directions for future research.

IV. Multiple Description Coding for 3DTV

In a typical scenario, compressed 3-D data content such as multiview video or 3-D dynamic meshes has to be transmitted over error-prone channels. Therefore, error resilience should be utilized in order to ensure robust transmission. One of the most promising approaches, specifying the error resilience as a joint source-channel coding problem, is the MDC. MDC relies on splitting the single data source into two or more independently decodable bit streams, called descriptions, which are mutually refining. At the decoder, the initial data source can be restored with acceptable quality by any of these subsets while by receiving more descriptions higher fidelity up to perfect reconstruction, is achieved. The cost of this operation is the insertion of a certain amount of redundancy in the descriptions. Due to its flexibility, MDC is considered a very robust and reliable tool for information transmission and a variety of practical MDC algorithms have been proposed for compression of images and video sequences (cf. the overview papers [104] and [121] and the references therein). In this section, we investigate its applicability to 3-D visual data, such as multiview/stereo video and 3-D mesh representations, through surveying relevant papers.

A. MDC of Single-View Visual Data

In many 3-D systems one of the views is completely coded and transmitted while the other(s) is/are just predicted and differential information, such as disparity, is coded and transmitted as a side channel. Therefore, it is of crucial importance to ensure the best possible error-proof coding of the leading view. Current attempts in standardizing multiview video compression have been based on single-view prediction techniques, such as h.264, modifying the prediction loop to include frames from other views [see Section II–A3)].

1) MDC of Still Images: MDC is essentially about adding controllable redundancy to the compressed image bit stream, thus compromising between compression efficiency and error protection. The redundancy can be added during any of the stages of a classical compression scheme consisting of (decorrelating) transform, transform coefficients scanning and quantization and entropy coding.

Early approaches suggest introducing redundancy at the stage of quantization, that is, the stage where the loss of information happens for the sake of higher compression. In the case of scalar quantization, two or more course scalar quantizers with overlapping cells process the data to create multiple coarsely quantized descriptions. Each of them is sufficient for reconstructing the source with acceptable quality. If received together, they form a finer quantization grid yielding a higher quality reconstruction. The fixed-rate quantization algorithm has been described in [105], and then extended to entropy-constrained quantization in [106]. Practical schemes working in DCT domain [111] or in WL domain [112] have been developed. Scalar quantization has been improved later by changing the geometry of the quantization cells [107] and developing multiple description lattice vector quantizers [109], [110].

Another group of algorithms operate in transform domain rearranging the transform coefficients and utilizing their spatial and scale positions in a way appropriate for creating redundant descriptions. In a setting, where the image is decomposed in a transform domain, and the transform coefficients are ordered by their variances, a subsequent pairwise correlating transform (PCT) has been suggested with a purpose to couple and correlate them into pairs [113]. Thus, two descriptions are created and the amount of correlation inserted allows for predicting the missing coefficients, if one description is lost. Proper PCTs, both orthogonal and nonorthogonal have been designed in [113] and [114]. In another algorithm, WL coefficients have been subsampled into polyphase components to create multiple descriptions out of them [115]. In the case of block transforms, such as lapped orthogonal transforms (LOT), multiple descriptions have been generated by interleaving blocks of transform coefficients [116]. Algorithms, utilizing the coefficient structure of the SPIHT and JPEG 2000 image coders have been designed as well [117], [118]. Pyramidal-like algorithms, realizing the so so-called whitening transforms have been proved quite effective, while working in the low redundancy rate-distortion region [119], [120].

2) MDC of Video: Recent video coding standards employ motion compensated prediction to efficiently remove the temporal correlation in video sequences. Consequently, a typical MDC for video has to address both the coding of prediction errors and motion vectors. In addition, the synchronization between the coder and decoder has to be kept intact [121].

The motion compensated prediction combined with multiple descriptions gives rise to a problem of mismatch between the states of the encoder and decoder as loss of one description may cause loss of synchronization. This problem can be solved either by modeling all possible mismatches at the encoder side and generating the corresponding extra prediction loops [122], or by embedding the prediction in each description separately [123], [124].

MDC techniques for motion vectors are based on an overlapped block motion compensation, which generates a smoother motion field than conventional block-based motion compensators [125]. Blocks forming the so generated motion fields are then partitioned into two or more coarser fields using quincunx subsampling and are split between the two or more descriptions [126]. Motion information can be treated also by applying temporal subsampling, in an approach known as

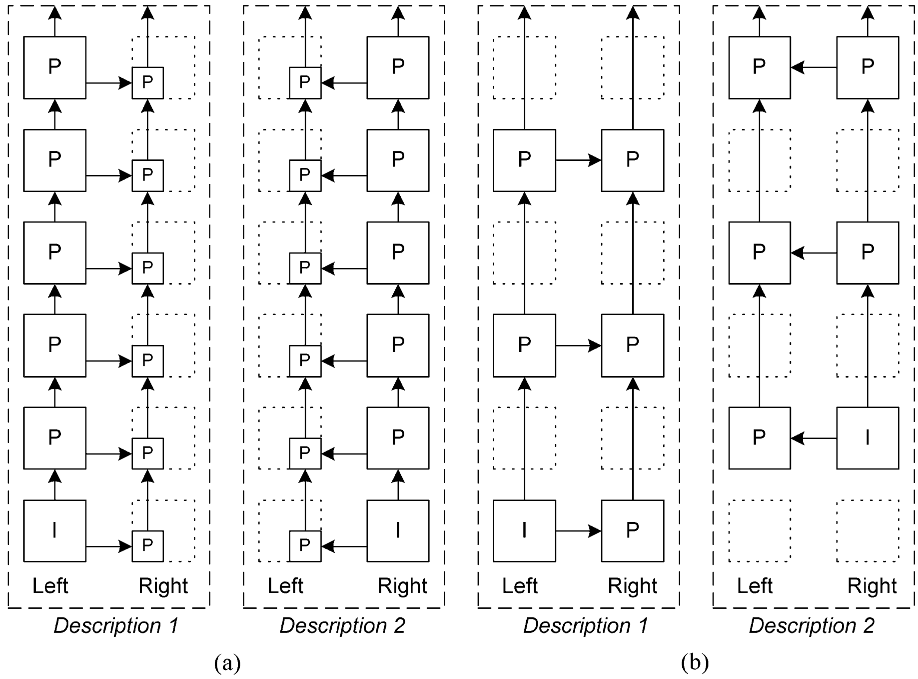

Fig. 4. MDC of stereoscopic video. (a) Spatial scaling. (b) Multistate.

video redundancy coding [127]. Temporally subsampled frame sequences (*threads*) are processed in parallel thus becoming independently decodable. To handle the synchronization issue, in a fixed time the threads converge to a *synch frame*. Variants of the video redundancy coding approach utilize second-order motion predictors, taking both even and odd preceding frames [128]. Yet another alternative in handling the temporal correlation is to completely drop the motion estimation/motion compensation stage. This is a reasoned approach in low-complexity applications, where the motion estimation is replaced by a decorrelating transform along the temporal axis and the encoder is then based on a 3-D block transform. Such an encoder is advantageous in error resilience as it enjoys low redundancy and no error propagation in the subsequent frames [129].

B. MDC of Multiview Video

MDC of multiview video employs ideas borrowed from monoscopic video coding and combined with proper handling of 3-D scene information, such as disparity. We review mainly some recent methods developed for MDC of stereoscopic video. Similar ideas can be easily extended to multiview video.

There are two theories studying the effects of unequal bit allocation between left and right video sequences. The fusion theory states that the total bit budget should be equally distributed between the two views, while the suppression theory argues that the overall perception in a stereo-pair is determined by the higher quality image [135], [136]. Experiments with a stereoscopic video encoder have demonstrated that spatial and spatiotemporal scaling provide acceptable perception performance with a reduced bit rate [137]. These findings have grounded the development of two MDC techniques for stereoscopic video [138], both producing balanced descriptions and ensuring stereoscopic video reconstruction in case of one channel failure for the price of moderate coding redundancy.

In the first approach, two descriptions are formed by combining a *leading view* (view with spatial resolution unchanged) and a *spatially scaled* view [138]. Frames from the leading view are predicted by only frames of the same view while frames from the downscaled view are predicted by frames from the two views, as illustrated in Fig. 4(a). In *Description 1*, the left view is the leading view and the right view is the scaled one, while in *Description 2* it is vice versa. When both descriptions are received, left and right view sequences are reconstructed in full resolution. If one description is lost due to channel failures, the reconstructed stereoscopic video pair gets one of the views with lower spatial resolution, but it is still perceptually acceptable. The other stereoscopic video MDC technique is based on temporal subsampling (Fig. 4(b)) [138]. Odd frames of both left and right sequences go to *Description 1*, while even frames of both sequences go to *Description 2*. Motion compensated prediction is performed separately in each description where left frames are predicted from preceding left frames, and right frames are predicted from preceding right frames or from the left frames corresponding to the same time instant. If the decoder receives both descriptions, the stereoscopic video sequence is reconstructed with the original frame rate. If one description is lost, the sequence is reconstructed with half the frame rate. While this scheme does not allow adjusting the coding redundancy, it reaches bit rates lower than simulcast coding bit rates. It also introduces no mismatch between the states of the encoder and decoder in case of description loss.

C. MDC of 3-D Meshes

A mesh description consists of geometry (vertex coordinates) and connectivity (vertex adjacency), as described in Section III. In an MDC setting, the connectivity should be coded in a lossless manner and included in each description. Lossy compression of geometry should be done is such a way as to generate independently decodable bit streams for each description so that the precision of 3-D coordinate values is increased by receiving any of the descriptions.

Depending on the principal idea employed, the MDC methods for meshes can be summarized in the following categories: *partitioning vertex geometry* [130], *Multiple Description Scalar quantization* (MDSQ) [131], and *partitioning wavelet zero-trees* [133]. In [130], multiple descriptions are generated by splitting the mesh geometry into submeshes and including the connectivity information, properly coded, e.g., by topological surgery, in each description. In [131], multiple description scalar quantization (MDSQ) is applied to WL coefficients of a multiresolution compression scheme. The obtained two sets of coefficients are then independently compressed by the SPIHT coder [117]. In the above two MDC schemes, descriptions are created by heuristic methods and no optimum solutions are proposed for varying network conditions. In [133], zero-trees obtained by *progressive geometry compression* (PGC) algorithm [132] are split into several sets and each set is independently coded (e.g., by SPIHT). The coded sets are then partitioned into multiple descriptions in such a way that each description contains a tree set coded with higher bit rate and several redundant tree sets coded with lower bit rates. Optimized procedures for grouping the WL trees into sets, their partitioning among descriptions, and bit allocation per set depending on the packet los rate have been developed in [133]. In [134], an MDC scheme for packet-loss resilient coding of 3-D geometry combines an unequal protection of embedded coded bit stream with forward error correction (FEC) codes. The embedded coded bit stream is obtained by compressing the 3-D mesh by PGC making use of WL transform and SPIHT-based zero-tree coding.

D. Conclusions and Future Research Directions

Multiple description coding is a promising approach for error resilient 3-D video coding. As summarized in this section several approaches are found in the literature for both mesh and multiview coding. Future research may include studying the performance of these algorithms on real communications channels, such as wireless channels or best-effort networks. Another important issue requiring further research is the issue of video quality assessment. Trivial objective quality measures, such as PSNR, are not quite adequate to the human visual perception. Therefore, new coding methods have been assessed through extensive subjective tests. However, such tests are costly and cannot be done in real time, when a perceptually-driven objective quality metric is expected to return a feedback to the encoder. The design of such metrics is much more complicated for the multiview case, where 3-D specific artifacts, such as *loss of disparity*, *keystone distortions*, and *inter-channel crosstalk*, are superimposed over the "traditional" 2-D artifacts, such as blockiness, and blurriness. An early work, addressing this problem, has shown quite promising results [139]. As far as 3-D geometry is concerned, optimization of the coarse mesh bit allocation and partitioning coarsest mesh vertices into different sets in order to further decrease the amount of redundancy is a research area. More advanced channel model and packetization strategy which better emulate the behavior of a real network and tree grouping strategies should also be investigated.

V. Watermarking of 3-D Data

The utilization of 3-D information for the modeling and representation of a real world scene brings copyright problems by itself for the content providers [140]. 3-D watermarking is proposed as a solution to the copyright problems of 3-D information by means of embedding a secret imperceptible signal, called *watermark*, into the main components of a 3-D representation and extracting the watermark from the resulting components of a scene after any applications.

Considering the dimensions of the main components of scene representations and the resulting components after the application of the algorithm, the watermarking methods in the literature can be classified into three groups, which we denote as 3D/3D, 3D/2D, and 2D/2D. The first pair of symbols identifies whether the watermark is embedded in the 3-D model or a 2-D rendering of it, and similarly the second pair of symbols identifies whether the watermark is detected in the 3-D model or a 2-D rendering. The definitions and requirements of the corresponding watermarking problems are different in each of the aforementioned groups.

A. 3D/3D Watermarking

The first group, 3D/3-D watermarking, mostly focuses on protection of the intellectual property rights of the 3-D geometrical structure, which is the most significant part of any scene [141]–[149]. The watermark is embedded into the 3-D geometric structure of an object used in a scene and tried to be extracted from the 3-D geometry after any attacks on the geometry (Fig. 5). These attacks might include rotation, translation, uniform scaling, polygon simplification, randomization of points, mesh compression, remeshing, mesh smoothing operations, cut operations, local transformations, global transformations and other operations on the geometry that changes the structure but preserving the visual quality in a desired level [141]. Actually, such variety of attacks makes 3-D geometry watermarking more challenging, compared to image and video watermarking.

3-D geometry model watermarking has some specific problems due to its different type of representation, compared to image and video signals. These problems can be briefly summarized as problems in *representation*, *synchronization*, *handling and editing*, *robustness and imperceptibility* [141], [142]. The problems are summarized in Table II.

The methods on 3-D geometry watermarking can be divided into two main groups based on the embedding domain of the watermark: spatial-domain and transform-domain methods. Spatial domain methods [141]–[145] embed the watermark directly into the values of the geometric primitives. In general, most of the methods in this group first extract perceptually significant geometric primitives from the model and then embed a watermark into those primitives by using a specially proposed method. The geometric primitives of a 3-D mesh used for watermarking might be coordinates of a point, length of a line, area of polygon, volume of a polyhedron, ratios of the areas of two polygons, ratios of volumes of two polyhedrons, etc. [142].

The other group contains transform domain methods [146]–[149], where a 3-D object is decomposed into subsignals by applying a 3-D geometry based transformation, such

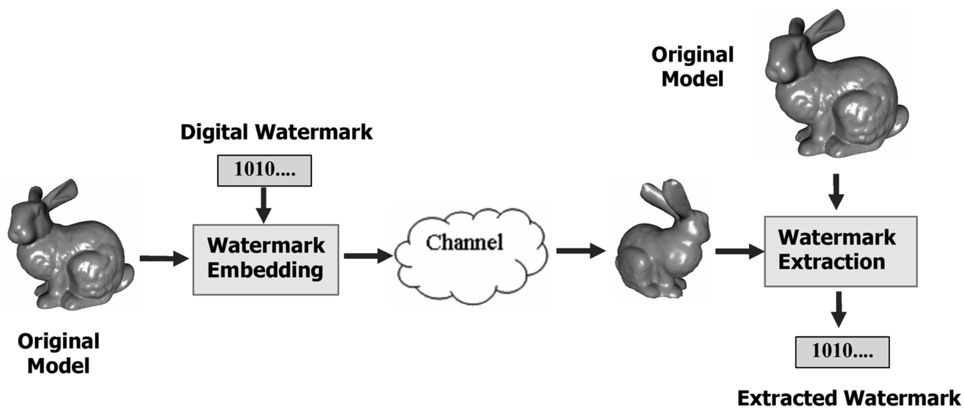

Fig. 5. General scheme for 3-D geometry watermarking (model taken from Stanford 3-D Scanning Repository [154]).

TABLE II
PROBLEMS IN 3-D GEOMETRY WATERMARKING COMPARED TO THE IMAGE AND VIDEO WATERMARKING

	Image and Video	**3D meshes**
Representation	Two and three dimensional functions on a manifold grid	Lack of unique representation, different mesh can represent the same surface
Synchronization	Scan line ordering	Requires a fixed orientation and a position of data in space
Handling and Editing	Well-defined standards for compression (JPEG, JPEG 2000, MPEG, H.26X,), Well-analyzed transformations used in transmission, compression, filtering. (DCT, FFT, wavelet etc.)	No well-accepted standard for compression, Multi-resolution representation techniques are comparatively new.
Robustness	Common signal processing during compression and transmission, synchronization attacks, cropping, etc.	A high number of the diverse attacks, including geometry transformations (translation, scaling, rotation, affine etc.), local modifications, topological changes.
Imperceptibility	Experimentally well studied perceptual analysis inherited from vision research to watermarking research.	No experimental study on the perceptual limits of the watermark for the geometry

as WL transform or mesh spectral analysis. In this group of methods, after applying the transformation to the mesh model, watermark is embedded to the resulting transform coefficients. Compared to the spatial domain methods, transform domain methods generally satisfy the trade off between imperceptibility and robustness in a more reliable manner, since the applied transformation mostly separates the perceptually significant and insignificant part of a 3-D model. A brief comparison between the two categories of geometry watermarking methods is presented in Table III.

TABLE III
COMPARISONS BETWEEN TWO DIFFERENT CATEGORIES IN 3-D GEOMETRY WATERMARKING

	Pros	**Cons**
Spatial Domain Methods	Lower complexity, Robustness against Cropping	Difficulty in finding perceptually significant regions, Weakness against local deformations
Transform Domain Methods	Robustness against compression and noise addition, Well integration with visual perception	Higher Computational Cost, Weakness against Cropping

B. 3D/2D Watermarking

The second group, 3D/2D watermarking, aims to extract the watermark that was originally hidden in the 3-D object, from the resulting images or videos (obtained after projection of 3-D object into 2-D image planes), thus protecting any visual representation of the object. The watermark can be both embedded

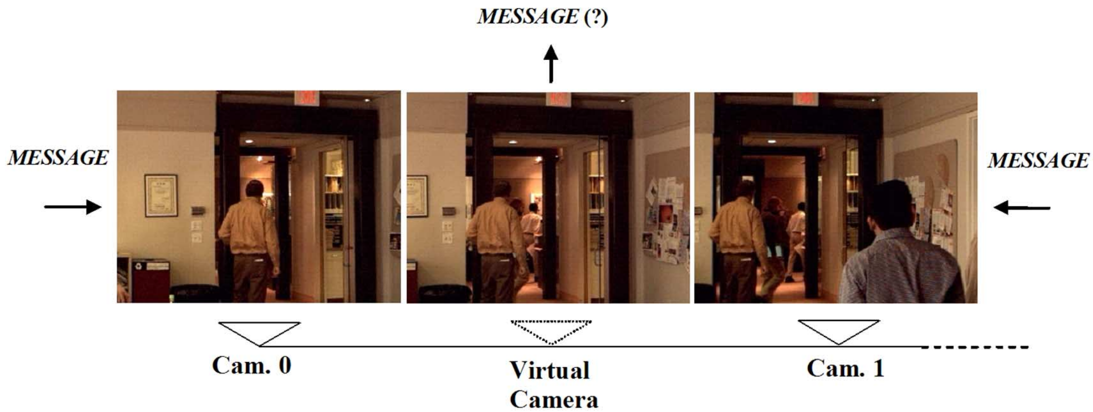

Fig. 6. Watermarking problem for Image-based Rendering.

to the geometry or the texture of the object [150], [151]. The attacks on the texture, not only include the operations on the texture of the object, such as subsampling, JPEG compression or malicious attacks, but also involve modifications in texture mapping or distortions in the geometrical descriptions of the object [150]. Therefore, the problem is much more complicated compared to 3D/3D watermarking.

In the literature, there are only a few methods for 3D/2D watermarking. The method in [150] embeds the watermark into the texture information of a 3-D object. Watermark detection consists of extracting the watermark from the recovered texture by using the 2-D projections (images) of the object. The main problem in 3D/2D watermarking is the estimation of the perspective projection between the original 3-D object and a given 2-D projection in order to recover the original texture from a given 2-D projection. The proposed method in [150] assumes that the original 3-D object is available in the watermark detector and proposes an estimation method in order to find the perspective projection.

C. 2D/2D Watermarking

The final group [152], 2D/2D watermarking, try to protect the image-based representation of a 3-D scene. While the first two groups try to protect the intellectual property rights for the two important components of a traditional representation of a 3-D scene, geometry and texture, the third group approaches to this problem, by watermarking sequences of images, which represent the 2-D projections of the same 3-D scene, and extracting the watermark from any 2-D rendered image, generated for an arbitrary angle of the scene via these sequences.

An application of this category should be expected in the copyright protection of multiview video content in emerging new technologies such as free view televisions (FTV) where the TV-viewers freely select the viewing position and angle for the transmitted multiview video. Noting that a TV-viewer might also record a personal video from the arbitrarily selected views and misuse this content, it is apparent that copyright problems also exist and should be solved for multiview video. However, the problem is more complicated compared to conventional single-view video watermarking.

First of all, concerning with the robustness requirement, the watermark should not only be resistant to common video processing and multiview video processing operations, it should also be extracted from a rendered video frame of an arbitrary view (see Fig. 6). In order to extract the watermark from such a rendered view, the watermark detection scheme should involve an estimation procedure for the camera position and orientation of the underlying recorded 3-D content. In addition, the watermark should also withstand image-based rendering operations, such as frame interpolation between neighboring cameras and pixel interpolation in camera pictures.

The only representative of 2D/2D watermarking is proposed in [152] for scenes where the rendered view is generated by using light field rendering [153]. The method embeds the same watermark into each view of an object and extracts the watermark from any virtual view of the object. However, method [152] assumes that original watermarked views are available during watermark extraction, in order to estimate the position and rotation of a virtual camera by considering the light field rendering operations during the generation of a virtual view.

D. Conclusion and Future Research Directions

There is a strong necessity for developing techniques, which protect the ownership rights of the original 3-D data, as well as prevent unauthorized duplication or tampering. Most of the techniques in the literature belong to 3D/3D watermarking group. On the other hand, 3D/2D and 2D/2D watermarking receives less attention. However, once a 3-D model is used in synthetic video within applications, such as computer generated animations, virtual/augmented reality or any kind of simulators, the projection of the 3-D model has more importance compared to the 3-D model itself. In addition, considering the increased research and progress in the virtual view synthesis and its applications, such as free-view TV and 3-D TV in the recent years, the aforementioned scenarios of 3D/2D and 2D/2D watermarking should become more significant for the copy right protection of 3-D representations.

VI. General Summary and Conclusion

3DTV systems and applications have gained significant attention in research and development recently. Technology for the whole media processing chain from capture to display is maturating, enabling the development of significant markets in the near future. This paper gives an overview of the state-of-the-art in coding and protection of 3DTV content. In general, efficient algorithms for compression of different types of 3-D content are available, as well as related international standards enabling interoperability of systems and services. However, since 3DTV is a relatively young research area compared to other types of media, there is still a lot of room for improvement. This includes optimization of algorithms, efficient implementation, integration into complete 3DTV systems, as well as development of completely new components and technology. Some data types related to 3DTV are not or not sufficiently supported yet.

Compression of pixel-type data such as stereo video, multiview video, and associated depth and disparity has reached a good level of maturity, but there is still a lot of room for improvement. Compression of static meshes is already a well-established research area. On the other hand, compression of dynamic 3-D meshes is a young and prospective area of research showing a high level of activity in the last few years. Error resilience for and MDC of 3DTV data is still a pioneering research area, as well as intellectual property protection for instance by watermarking.

In conclusion, after decades 3DTV is ready to enter everyone's daily life, as far as coding is concerned. However, this research area is still in its infancy and there is a lot of work to be done during the next years.

Acknowledgment

The authors would like to thank R. Sondershaus from WSI/GRIS University of Tübingen, Germany, L. Vasa from the Department of Computer Science and Engineering, University of West Bohemia, Germany, P. Klie from the Institut für Informationsverarbeitung, Leibniz Universitaät Hannover, Germany, A. Norkin from Institute of Signal Processing, Tampere University of Technology, Finland, C. Bilen, A. Aksay, M. Oğuz Bici, A. Aydın Alatan, and C. Cigla from Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, C. Fehn, P. Kauff, P. Merkle, and T. Wiegand from Fraunhofer Institute for Telecommunications—Heinrich-Hertz-Institut, Germany, for their help and support. Further, the authors would like to thank M. Müller (ETH Zürich, Switzerland) for providing the mesh sequence "Cow" for evaluation experiments in mesh compression (Fig. 3) and the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, for providing the "Exit" multiview video data (Fig. 6).

References

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3-D video and free viewpoint video-technologies, applications and MPEG standards," in *Proc. IEEE ICME*, Toronto, ON, Canada, Jul. 2006, pp. 2161–2164.

[2] L. Onural, A. Smolic, and T. Sikora, "An overview of a new European consortium: Integrated three-dimensional television—Capture, transmission and display (3DTV)," presented at the EWIMT, London, U.K., Nov. 25–26, 2004.

[3] A. Smolic and D. McCutchen, "3DAV exploration of video-based rendering technology in MPEG," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 348–356, Mar. 2004.

[4] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," *Proc. IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 98–110, Jan. 2005.

[5] L. Onural, T. Sikora, J. Ostermann, A. Smolic, R. Civanlar, and J. Watson, "An assessment of 3DTV technologies," in *Proc. NAB 2006*, Las Vegas, NV, Apr. 22–27, 2006, pp. 456–467.

[6] A. A. Alatan *et al.*, "Scene representation technologies for 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1587–1605, Nov. 2007.

[7] M. Tanimoto, "Free viewpoint television—FTV," in *Proc. PCS*, San Francisco, CA, Dec. 2004.

[8] T. Fujii and M. Tanimoto, "Free-viewpoint TV system based on ray-space representation," *ITCom*, vol. 4864-22, pp. 175–189, 2002.

[9] T. Matsuyama, X. Wu, T. Takai, and T. Wada, "Real-time dynamic 3-D object shape reconstruction and high-fidelity texture mapping for 3-D Video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 357–369, Mar. 2004.

[10] T. Hamaguchi, T. Fujii, Y. Kajiki, and T. Honda, "Real-time view interpolation system for multiview 3-D display," *Electron. Imag.*, vol. 4297A, pp. 212–221, Jan. 2001.

[11] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *Trans. Graphics*, vol. 22, no. 3, pp. 569–577, Jul. 2003.

[12] S. Würmlin, E. Lamboray, O. Staadt, and M. Gross, "3-D video recorder: A system for recording, processing and playing three-dimensional video," in *Computer Graphics Forum*. Oxford, U.K.: Blackwell, 2003, vol. 22, no. 2, pp. 181–193.

[13] S. Würmlin, E. Lamboray, and M. Gross, "3-D video fragments: Dynamic point samples for real-time free-viewpoint video," *Comput. Graph.*, vol. 28, no. 1, pp. 3–14, 2004.

[14] W. Matusik and H. Pfister, "3-D TV: A scalable system for real-time acquistion, transmission and autostereoscopic display of dynamic scenes," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 814–824, Aug. 2004.

[15] H.-Y. Shum, S. B. Kang, and S.-C. Chan, "Survey of image-based representations and compression techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 11, pp. 1020–1037, Nov. 2003.

[16] B. Haskell, A. Puri, and A. Netrevali, *Digital Video: An Introduction to MPEG-2*, 1997.

[17] "Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video," ITU-T Rec. H.222.0|ISO/IEC 13818-1 (MPEG 2 Systems), ITU-T and ISO/IEC JTC 1, Nov. 1994.

[18] "Video Coding for Low Bit Rate Communication," ITU-T Rec. H.263; version 1, Nov. 1995; version 2 Jan. 1998; version 3, Nov. 2000.

[19] "Coding of audio-visual objects-Part 2: Visual," ISO/IEC 14496-2 (MPEG-4 visual version 1), Apr. 1999; ISO/IEC JTC 1, Amd. 1 (ver. 2), Feb. 2000; Amd. 2, 2001; Amd. 3, 2001; Amd. 4 (streaming video profile), 2001; Amd 1 to 2nd ed. (studio profile), 2001; Amd. 2 to 2nd ed. 2003.

[20] S. Sun and S. Lei, "Stereo-view video coding using H.264 tools," in *Proc. SPIE Image and Video Commun. and Process. 2005*, Mar. 2005, vol. 5685, pp. 177–184.

[21] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[22] Advanced Video Coding for Generic Audiovisual Services 2005, ITU T and ISO/IEC JTC 1, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, 2003, most recent Version.

[23] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An evolutionary and optimised approach on 3D-TV," in *Proc. Int. Broadcast Conf.*, Amsterdam, The Netherlands, Sep. 2002, pp. 357–365.

[24] C. Fehn, "3D-TV using depth-image-based rendering (DIBR)," in *Proc. Picture Coding Symp.*, San Francisco, CA, USA, Dec. 2004.

[25] *Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information*, ISO/IEC JTC1/SC29/WG11, Jan. 2007, Doc. N8768, Marrakesh, Morocco.

[26] *Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data*, ISO/IEC JTC1/SC29/WG11, Jan. 2007, Doc. N8799, Marrakech, Morocco.

[27] J. Shade, S. Gortler, L. W. He, and R. Szeliski, "Layered depth images," in *Proc. SIGGRAPH'98*, Orlando, FL, Jul. 1998.

[28] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *SIGGRAPH04*, Los Angeles, CA, Aug. 2004.

[29] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Process: Image Commun., Special Issue on 3DTV*, Feb. 2007.

[30] P. Merkle, K. Müller, A. Smolić, and T. Wiegand, "Statistical evaluation of spatiotemporal prediction for multiview video coding," in *Proc. ICOB 2005*, Berlin, Germany, Oct. 27–28, 2005.

[31] A. Kaup and U. Fecker, "Analysis of multireference block matching for multiview video coding," in *Proc. 7th Workshop Digital Broadcasting*, Erlangen, Germany, Sep. 2006, pp. 33–39.

[32] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 338–343, Apr. 2000.

[33] M. Magnor, P. Ramanathan, and B. Girod, "Multi-view coding for image-based rendering using 3-D scene geometry," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 11, pp. 1092–1106, Nov. 2003.

[34] K.-J. Oh and Y.-S. Ho, "Multi-view video coding based on the lattice-like pyramid GOP structure," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, Apr. 2007.

[35] X. Cheng, L. Sun, and S. Yang, "A multiview video coding scheme using shared key frames for high interactive application," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, Apr. 2006.

[36] Y. Yang, G. Jiang, M. Yu, F. Li, and Y. Kim, "Hyper-space based multiview video coding scheme for free viewpoint television," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, Apr. 2006.

[37] A. Vetro, W. Matusik, H. P. Pfister, and J. Xin, "Coding approaches for end-to-end 3-D TV systems," in *Proc. Picture Coding Symp. (PCS)*, San Francisco, CA, Dec. 2004.

[38] H. Kalva, L. Christodoulou, L. M. Mayron, O. Marques, and B. Furht, "Design and evaluation of a 3-D video system based on H.264 view coding," in *Int. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 2006)*, Newport, RI, May 22–23, 2006.

[39] D. Socek, D. Culibrk, H. Kalva, O. Marques, and B. Furht, "Permutation-based low-complexity alternate coding in multiview H.264/AVC," in *IEEE Int. Conf. on Multimedia & Expo (ICME) 2006*, Toronto, Canada, Jul. 9–12, 2006.

[40] H. Kalva, L. Christodoulou, L. Mayron, O. Marques, and B. Furht, "Challenges and opportunities in video coding for 3-D TV," in *Special Session on 3-D TV: Primed for Success?, IEEE Int. Conf. on Multimedia & Expo (ICME) 2006*, Toronto, Canada, Jul. 9–12, 2006.

[41] C. Bilen, A. Aksay, and G. Bozdagi Akar, "A Multi-view video codec based On H.264," in *IEEE ICIP 2006*, Atlanta, GA, Oct. 2006.

[42] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B-pictures and MCTF," in *ICME 2006, IEEE Int. Conf. on Multimedia and Expo*, Toronto, Ontario, Canada, Jul. 2006.

[43] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for video camera arrays," in *Proc. Picture Coding Symp.*, Beijing, China, Apr. 2006.

[44] P. Merkle, K. Müller, A. Smolic, and T. Wiegand, "Efficient compression of multiview video exploiting inter-view dependencies based on H.264/MPEG4-AVC," in *Proc. ICME 2006 IEEE Int. Conf. on Multimedia and Expo*, Toronto, Ontario, Canada, Jul. 2006.

[45] K. Müller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, and T. Wiegand, "Multi-view video coding based on H.264/AVC using hierarchical b-frames," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, Apr. 2006.

[46] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[47] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Coding Efficiency Versus Complexity of MVC Prediction Structures," in *Proc. EUSIPCO 2007*, Poznan, Poland, Sep. 3, 2007.

[48] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, Apr. 2006.

[49] M. Kitahara, H. Kimata, S. Shimizu, K. Kamikura, Y. Yashimata, K. Yamamoto, T. Yendo, T. Fujii, and M. Tanimoto, "Multi-view video coding using view interpolation and reference picture selection," in *ICME 2006, IEEE Int. Conf. Multimedia and Exposition*, Toronto, ON, Canada, Jul. 2006.

[50] E. Martinian, A. Behrens, J. Xin, A. Vetro, and H. Sun, "Extensions of H.264/AVC for multiview video compression," in *IEEE Int. Conf. on Image Processing (ICIP)*, Atlanta, GA, Oct. 2006.

[51] J.-H. Kim, P.-L. Lai, A. Ortega, Y. Su, P. Yin, and C. Gomila, "Results of CE2 on multiview video coding," in *Joint Video Team*, Klagenfurt, Austria, Jul. 2006, Doc. JVT-T117.

[52] Y.-L. Lee, J.-H. Hur, Y.-K. Lee, S.-H. Cho, H. J. Kwon, N. H. Hur, and J. W. Kim, "Results of CE2 on multiview video coding," in *Joint Video Team*, Klagenfurt, Austria, Jul. 2007, Doc. JVT-T110.

[53] U. Fecker, M. Barkowsky, and A. Kaup, "Improving the prediction efficiency for multiview video coding using histogram matching," in *Proc. Picture Coding Symp. (PCS 2006)*, Beijing, China, Apr. 2006.

[54] J. Lu, H. Cai, J.-G. Lou, and J. Li, "An effective epipolar geometry assisted motion-estimation technique for multiview image coding," in *IEEE Int. Conf. on Image Process.*, Atlanta, GA, Oct. 8–11, 2007.

[55] X. Guo, Y. Lu, F. Wu, and W. Gao, "Inter-view direct mode for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1527–1532, Dec. 2006.

[56] J. Garbas, U. Fecker, T. Tröger, and A. Kaup, "4D scalable multiview video coding using disparity compensated view filtering and motion compensated temporal filtering," in *Int. Workshop on Multimedia Signal Processing (MMSP)*, Victoria, Canada, Oct. 2006.

[57] M. Droese, C. Clemens, and T. Sikora, "Extending single-view scalable video coding to multiview based on H.264/AVC," in *Proc. ICIP 2006*, Atlanta, GA, Oct. 2006.

[58] W. Yang, Y. Lu, F. Wu, J. Cai, K. N. Ngan, and S. Li, "4D wavelet-based multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 11, pp. 1385–1396, Nov. 2006.

[59] W. Yang, F. Wu, Y. Lu, J. Cai, K. N. Ngan, and S. Li, "Scalable multiview video coding using wavelet," in *IEEE Int. Symp. on Circuits Syst.(ISCAS)*, 2005, pp. 6078–6081.

[60] N. Ozbek and A. M. Tekalp, "Scalable multiview video coding for interactive 3DTV," in *ICME 2006 IEEE Int. Conf. on Multimedia and Expo.*, Toronto, ON, Canada, Jul. 2006.

[61] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multiview video coding," *Vis. Commun. and Imag. Process.*, vol. 6077, 2006.

[62] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Interactive multiview video delivery with view-point tracking and fast stream switching," in *MRCS 2006*, Istanbul, Turkey, Sep. 2006, pp. 1558–1565.

[63] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Interactive transport of multiview videos for 3DTV applications," in *Packet Video Workshop 2006*, Hangzhou, China.

[64] A. Vetro, Y. Su, H. Kimata, and A. Smolic, "Joint draft 1.0 on multiview video coding," in *Joint Video Team*, Hangzhou, China, Oct. 2006, Doc. JVT-U209.

[65] A. Vetro, Y. Su, H. Kimata, and A. Smolic, "Joint multiview video model," in *Joint Video Team*, Hangzhou, China, Oct. 2006, Doc. JVT-U207.

[66] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient compression of multiview depth data based on MVC," in *Proc. 3DTV-CON 2007*, Kos Island, Greece, May 2007.

[67] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE Int. Conf. on Image Process. (ICIP'07)*, San Antonio, TX, Sep. 2007.

[68] T. Saishu, S. Numazaki, K. Taira, R. Fukushima, A. Morishita, and Y. Hirayama, "Flatbed-type autostereoscopic display system and its image format," in *Proc. SPIE*, 2006, vol. 6055, pp. 261–268.

[69] G. Taubin and J. Rossignac, "Geometry compression through topological surgery," *ACM Trans. Graphics*, pp. 84–115, 1998.

[70] G. Taubin, A. Gueziec, W. Horn, and F. Lazarus, "Progressive forest split compression," in *SIGGRAPH'98*, 1998, pp. 123–132.

[71] J. Rossignac, "Edgebreaker: Connectivity compression for triangle meshes," *IEEE Trans. Vis. Comp. Graph.*, vol. 5, no. 1, pp. 47–61, 1999.

[72] S. Gumhold and W. Straßer, "Real time compression of triangle mesh connectivity," in *SIGGRAPH'98*, 1998, pp. 133–140.

[73] C. Touma and C. Gotsman, "Triangle mesh compression," in *Proc. Graphics Interface'98*, 1998, pp. 26–34.

[74] M. Isenburg and J. Snoeyink, "Face fixer: Compressing polygon meshes with properties," in *Proc. SIGGRAPH 2000*, 2000, pp. 263–270.

[75] P. Alliez and M. Desbrun, "Valence-driven connectivity encoding for 3-D meshes," *Comput. Graph. Forum*, vol. 20, no. 3, 2001.

[76] P. Alliez and M. Desbrun, "Progressive compression for lossless transmission of triangle meshes," in *SIGGRAPH'01*, 2001, pp. 195–202.

[77] D. Cohen-Or, R. Cohen, and R. Irony, Multi-way geometry encoding Sch. Comput. Sci., Tel Aviv Univ., Haifa, Israel, Tech. Rep. 2002, 2002 [Online]. Available: www.cs.tau.ac.il/~dcor/online_papers/papers/multiway.pdf

[78] S. Gumhold and R. Amjoun, "Higher order prediction for geometry compression," *Shape Modeling Int.*, p. 286, 2003.

[79] J. Peng and J. Kuo, "Geometry-guided progressive lossless 3-D mesh coding with octree (OT) decomposition," in *ACM SIGGRAPH*, 2005, pp. 609–616.

[80] A. Khodakovsky, P. Schröder, and W. Sweldens, "Progressive geometry compression," presented at the SIGGRAPH, 2000 [Online]. Available: http://citeseer.ist.psu.edu/khodakovsky00progressive.html , ACM Press/Addison-Wesley Publishing

[81] Z. Karni and C. Gotsman, "Spectral compression of mesh geometry," in *Proc. 27th Annu. Conf. Comput. Graphics and Interactive Techn. Int. Conf. Comput. Graphics Interactive Techn.*, 2000, pp. 279–285.

[82] X. Gu, S. Gortler, and H. Hoppe, "Geometry images," in *ACM SIGGRAPH 2002*, 2002, pp. 355–361.

[83] P. Alliez and C. Gotsman, "Shape compression using spherical geometry images," presented at the Symp. Multiresolution in Geometric Modeling, Cambridge, U.K., 2003.

[84] E. Praun and H. Hoppe, "Spherical parametrization and remeshing," in *ACM SIGGRAPH 2003*, 2003, pp. 340–349.

[85] K. Kose, A. E. Cetin, U. Gudukbay, and L. Onural, "Connectivity-guided adaptive lifting transform for Image like compression of meshes," in *3DTV-CON07*, Kos Island, Greece, 2007.

[86] P. Alliez and C. Gotsman, "Recent advances in compression of 3-D meshes," presented at the Symp. Multiresolution in Geometric Modeling, Cambridge, U.K., Sep. 2003.

[87] J. Peng, "Technologies for 3-D mesh compression: A survey," *J. Vis. Commun. Image Rep.*, vol. 16, pp. 688–733, 2005.

[88] H. Briceno, P. Sander, L. McMillan, S. Gortler, and H. Hoppe, "Geometry videos: A new representation for 3-D animations," in *ACM Symp. on Computer Animation 2003*, pp. 136–146.

[89] J. E. Lengyel, "Compression of time-dependant geometry," in *Proc. 1999 Symp. on Interactive 3-D Graph.*, 1999, pp. 89–95.

[90] M. Alexa and W. Müller, "Representing animations by principal components," *Comp. Graph. Forum*, vol. 19, no. 3, pp. 411–418, 2000.

[91] Z. Karni and C. Gotsman, "Compression of soft-body animation sequences," *Comput. Graph.*, vol. 28, pp. 25–34, 2004.

[92] M. Sattler, R. Sarlette, and R. Klein, "Simple and efficient compression of animation sequences," in *Proc. ACM/EG Symp. Comp. Anim*, 2005, p. 209.

[93] L. Vasa and V. Skala, "CoDDyaC: Connectivity driven dynamic mesh compression," presented at the Proc. 3DTV-CON, 2007.

[94] R. Amjoun and W. Straßer, "Efficient compression of 3-D dynamic mesh sequences," *J. WSCG*, 2007 [Online]. Available: http://wscg.zcu.cz/wscg2007/Papers_2007/journal/C31-full.pdf

[95] I. Guskov and A. Khodakovsky, "Wavelet compression of parametrically coherent mesh sequences," in *Proc. ACM/EG Symp. Comput. Anim.*, 2004.

[96] K. Mamou, T. Zaharia, and F. Preteux, "A skinning approach for dynamic 3d mesh compression," *Comput. Animat. Virtual Worlds*, vol. 17, no. 3–4, pp. 337–346, 2007.

[97] L. Ibarria and J. Rossignac, "Dynapack: Space-time compression of the 3-D animations of triangle meshes with fixed connectivity," in *Proc. EG Symp. on Comput. Anim.*, 2003.

[98] E. S. Jang, J. D. K. Kim, S. Y. Jung, M. J. Han, S. O. Woo, and S. J. Lee, "Interpolator data compression for MPEG-4 animation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 7, pp. 989–1008, Nov. 2004.

[99] N. Stefanoski and J. Ostermann, "Connectivity-guided predictive compression of dynamic 3-D meshes," presented at the IEEE Conf. Image Process. ICIP'06, Atlanta, GA, 2007.

[100] N. Stefanoski, X. Liu, P. Klie, and J. Ostermann, "Scalable linear predictive coding of time-consistent 3-D mesh sequences," presented at the 3DTV-CON, The True Vision-Capture, Transmission and Display of 3-D Video, Kos Island, Greece, May 2007.

[101] N. Stefanoski, P. Klie, X. Liu, and J. Ostermann, "Layered coding of time-consistent dynamic 3-D meshes using a nonlinear predictor," presented at the IEEE ICIP, Thessaloniki, Greece, 2007.

[102] K. Müller, A. Smolic, M. Kautzner, P. Eisert, and T. Wiegand, "Rate-distortion-optimazed predictive compression of dynamic 3-D mesh sequences," *Signal Process: Image Commun.*, vol. 21, no. 9, pp. 812–828, 2007.

[103] K. Müller, A. Smolic, M. Kautzner, P. Eisert, and T. Wiegand, "Predictive compression of dynamic 3-D meshes," in *Proc. IEEE ICIP'06*, 2005, pp. 621–624.

[104] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, pp. 74–93, 2001.

[105] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 821–834, May 1993.

[106] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple description scalar quantizer," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 245–250, Jan. 1994.

[107] V. A. Vaishampayan and J. C. Batllo, "Asymptotic analysis of multiple description qantizers," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 278–284, Jan. 1998.

[108] L. Ozarow, "Ona source-channel coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, 1980.

[109] V. Vaishampayan, N. Sloane, and S. Servetto, "Multiple description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1718–1734, May 2001.

[110] V. Goyal, J. Kelnr, and J. Kovacevic, "Multiple description vector quantization with a coarse lattice," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 781–788, Mar. 2002.

[111] V. Vaishampayan, "Application of multiple description codes to image and video transmission over lossy networks," in *Proc. 7th Int. Workshop Packet Video*, Brisbane, Australia, Mar. 1996, pp. 55–60.

[112] S. D. Servetto, K. Ramcadran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 813–826, May 2000.

[113] Y. Wang, M. Orchard, V. Vaishampayan, and A. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, pp. 351–366, Oct. 2001.

[114] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2199–2224, Sep. 2001.

[115] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. Image Process.*, vol. 12, pp. 1211–1225, Oct. 2003.

[116] D. Chung and Y. Wang, "Lapped orthogonal transforms designed for error resilient image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 752–764, Sep. 2002.

[117] A. Miguel, A. Mohr, and E. Riskin, "SPIHT for generalized multiple description coding," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1999, vol. 3, pp. 842–846.

[118] T. Tillo and G. Olmo, "A novel multiple description coding scheme compatible with the JPEG 2000 decoder," *IEEE Signal Process. Lett.*, vol. 11, pp. 908–911, 2004.

[119] K.-P. Choi and K.-Y. Lee, "An efficient multiple description coding using whitening transform," *IEICE Trans. Fundam. E86-A*, vol. 6, pp. 1382–1389, 2003.

[120] A. Norkin, A. Gotchev, K. Egiazarian, and J. Astola, "Two-stage multiple description coders: Analysis and comparative study," *Signal Process.: Image Commun.*, vol. 21, no. 8, pp. 609–625, Aug. 2006.

[121] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proc. IEEE*, vol. 93, pp. 57–70, 2005.

[122] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion-compensated temporal prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 193–204, 2002.

[123] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "A space domain approach for multiple description video coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, vol. 3, pp. 253–256.

[124] N. Boulgouris, K. Zachariadis, A. Leontaris, and M. Strintzis, "Drift-free multiple description coding of video," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, 2001, vol. 1, pp. 105–110.

[125] M. Orchard and G. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Trans. Image Proc.*, vol. 3, no. 5, pp. 693–699, 1994.

[126] C.-S. Kim and S.-U. Lee, "Multiple description coding of motion fields for robust video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 999–1010, May 2001.

[127] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. Circ. and Syst.*, vol. 8, no. 7, pp. 867–877, 1998, for Video Tech..

[128] Y. Wang and S. Lin, "Error-resilient coding using multiple description motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 438–452, 2002.

[129] A. Norkin, A. Gotchev, K. Egiazarian, and J. Astola, "Low-complexity multiple description coding of video based on 3-D block transforms," *EURASIP J. Embedded Syst.*, 2007, to be published.

[130] P. Jaromersky, X. Wu, Y. Chiang, and N. Memon, "Multiple description geometry compression for networked interactive 3-D graphics," in *Proc. ICIG'2004*, Dec. 2004, pp. 468–471.

[131] M. O. Bici and G. Bozdagi Akar, "Multiple description scalar quantization based 3-D mesh coding," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, Oct. 2006.

[132] A. Khodakovsky, P. Schroder, and W. Sweldens, "Progressive geometry compression, Siggraph 2000," in *Computer Graphics Proc. 2000*, pp. 271–278.

[133] A. Norkin, M. O. Bici, G. Bozdagi Akar, A. Gotchev, and J. Astola, "Wavelet-based multiple description coding of 3-D geometry," in *VCIP2007*, San-Jose, U.S., Jan. 2007.

[134] A. Norkin, M. O. Bici, G. Bozdagi Akar, A. Gotchev, and J. Astola, "Multiple description coding of 3-D geometry with forward error correction codes," in *Proc. 3DTV Con 2007*, Greece.

[135] I. Dinstein, M. G. Kim, A. Henik, and J. Tzelgov, "Compression of stereo images using subsampling and transform coding," *Optical Engineering*, vol. 30, no. 9, pp. 1359–1364, 1991.
[136] B. Julesz*, Foundations of Cyclopeon Perception*. Chicago, IL: Univ. of Chicago Press, 1971.
[137] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. B. Akar, M. R. Civanlar, and A. M. Tekalp, "Temporal and spatial scaling for stereoscopic video compression," in *EUSIPCO 2006*, Florence, Italy, Sep. 2007.
[138] A. Norkin, A. Aksay, C. Bilen, G. B. Akar, A. Gotchev, and J. Astola, "Schemes for multiple description coding of stereoscopic video," presented at the MRCS 2006, Istanbul, Turkey, Sep. 2007.
[139] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. Akar, "Towards compound stereo-video quality metric: A specific encoder-based framework," in *Proc. Southwest Symp. Image Analysis and Interpretation (SSIAI 2006)*, pp. 218–222.
[140] I. J. Cox, M. L. Miller, and J. A. Bloom*, Digital Watermarking*. New York: Academic, 2002.
[141] O. Benedens, "Geometry-based watermarking of 3-D models," *IEEE CGA*, pp. 46–55, Jan. 1999.
[142] R. Ohbuchi, H. Masuda, and M. Aono, "Watermarking three dimensional polygonal models through geometric and topological modifications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 551–560, 1998.
[143] Z. Yu, H. S. Horace Ip, and L. F. Kwok, "A robust watermarking scheme for 3-D triangular mesh models," *Pattern Recognit.*, vol. 36, no. 11, pp. 2603–2614, Nov. 2003.
[144] S. Zafeiriou, A. Tefas, and I. Pitas, "Blind watermarking schemes for copyright protection of 3-D mesh objects," *IEEE TVCG*, vol. 11, no. 5, Sep./Oct. 2005.
[145] A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 687–701, Mar. 2006.
[146] E. Praun, H. Hoppe, and A. Finkelstein, "Robust mesh watermarking," in *SIGGRAPH 99*, 1999, pp. 69–76.
[147] R. Ohbuchi, S. Takahashi, T. Miyazawa, and A. Mukaiyama, "Watermarking 3-D polygonal meshes in the mesh spectral domain," in *Proc. Graph. Interface 2001*, 2001, pp. 9–17.
[148] S. Kanai, H. Date, and T. Kishinami, "Digital watermarking for 3-D polygons using multiresolution wavelet decomposition," in *6h IFIP WG 5.2 GEO-6*, Tokyo, Japan, Dec. 1998, pp. 296–307.
[149] K. Yin, Z. Pan, J. Shi, and D. Zhang, "Robust mesh watermarking based on multiresolution processing," *Comput. Graph.*, vol. 25, pp. 409–420, 2001.
[150] E. Garcia and J.-L. Dugelay, "Texture-based watermarking of 3-D video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 853–866, Aug. 2003.
[151] J. Bennour and J.-L. Dugelay, "Protection of 3-D object through silhoutte watermarking," in *IEEE ICASSP*, Toulouse, France, May 14–19, 2007.
[152] A. Koz, C. Cigla, and A. Alatan, "Free view watermarking for free view television," in *Proc. IEEE ICIP*, Oct. 2007, pp. 1405–1408.
[153] M. Levoy and P. Hanrahan, "Light field rendering," in *Comput. Graphics (SIGGRAPH'96)*, New Orleans, LA, Aug. 1996, pp. 31–42.
[154] *"The Stanford 3-D Scanning Repository"* Stanford Univ., Stanford, CA, 2007 [Online]. Available: http://graphics.stanford.edu/data/3Dscanrep

Aljoscha Smolic received the Dipl.-Ing. degree in electrical engineering from the Technical University of Berlin, Berlin, Germany in 1996, and the Dr.-Ing. degree in electrical and information engineering from Aachen University of Technology (RWTH), Aachen, Germany, in 2001.

He joined the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut (HHI), Berlin, Germany, in 1994 and has been Scientific Project Manager since 2001. He has been involved in several national and international research projects, where he conducted research in various fields of video processing, video coding, computer vision and computer graphics and published more than 80 referred papers in these fields. In this context, he has been involved in ISO standardization activities where he contributed to the development of the multimedia standards MPEG-4 and MPEG-7. In current projects, he is responsible for research in free viewpoint and 3-D video processing and coding, augmented reality, 3-D reconstruction and video-based rendering. Since 2003, he has been an Adjunct Professor at the Technical University of Berlin, Berlin, Germany, and teaches multimedia communications and statistical communications theory. He chaired the MPEG ad hoc group on 3DAV pioneering standards for 3-D video. Currently , he is the Editor of the Multiview Video Coding (MVC) standard and co-chairs a working group of the JVT for MVC.

Dr. Smolic received the "Rudolf-Urtlel-Award" of the German Society for Technology in TV and Cinema (FKTG) for his dissertation in 2002. He is Area Editor for *Signal Processing: Image Communication* and Guest Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE *Signal Processing Magazine.* He is a Committee Member of several conferences, including ICIP, ICME, and EUSIPCO.

Karsten Müller (M'98–SM'07) received the Dipl.-Ing. and Dr.-Ing. degree in electrical engineering from the Technical University of Berlin, Berlin, Germany, in 1997 and 2006, respectively.

He has been with the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Berlin, Germany, since 1997, where he is currently Project Manager. His research interests are mainly in the field of representation, coding, and reconstruction of 3-D scenes in free viewpoint video scenarios and coding, multi-view applications and combined 2D/3D similarity analysis. He has been involved in MPEG activities, where he contributed to visual MPEG-7 descriptors and MPEG-4.

Nikolce Stefanoski (S'07) received the Dipl.-Math. degree from the University of Hannover, Hannover, Germany, in 2003, where he studied mathematics, computer science, and telecommunications. Since 2004 , he has been working toward the Ph.D. degree at the Institut für Informationsverarbeitung (TNT), University of Hannover.

His research interests are 3-D graphics compression, signal processing and stochastics. He is an active contributer to the Motion Picture Experts Group (MPEG) of ISO/IEC SC29.

Mr. Stefanoski received a prize for best diploma thesis in the field of computer science from the Fachbereich für Informatik of the University of Hannover.

Jörn Ostermann (SS'86–M'88–SM'00–F'05) studied electrical engineering and communications engineering at the University of Hannover, Hannover, Germany, and Imperial College London, London, U.K. He received the Dipl.-Ing. and Dr.-Ing. degrees from the University of Hannover in 1988 and 1994, respectively.

From 1988 till 1994, he worked as a Research Assistant at the Institut für Theoretische Nachrichtentechnik conducting research in low bit rate and object-based analysis-synthesis video coding. In 1994 and 1995, he worked in the Visual Communications Research Department, AT&T Bell Laboratories on video coding. He was a member of Image Processing and Technology Research within AT&T Labs—Research from 1996 to 2003. Since 2003, he has been a Full Professor and Head of the Institut für Informationsverarbeitung , University of Hannover. From 1993 to 1994, he chaired the European COST 211 sim group coordinating research in low bit rate video coding. Within MPEG-4, he organized the evaluation of video tools to start defining the standard. He chaired the Adhoc Group on Coding of Arbitrarily-shaped Objects in MPEG-4 Video. Jörn was a scholar of the German National Foundation. He published more than 50 research papers and book chapters. He is coauthor of a graduate level text book on video communications. He holds 22 patents. His current research interests are video coding and streaming, 3-D modeling, face animation, and computer-human interfaces.

Dr. Ostermann received the AT&T Standards Recognition Award and the ISO award in 1998. He is a member of the IEEE Technical Committee on Multimedia Signal Processing and past chair of the IEEE CAS Visual Signal Processing and Communications (VSPC) Technical Committee. Joern served as a Distinguished Lecturer of the IEEE Circuits and Sytems Society.

Atanas Gotchev (M'93) received the M.Sc. degrees in communications engineering and in applied mathematics from Technical University of Sofia, Sofia, Bulgaria, in 1990 and 1992, respectively, the Ph.D. degree in communications engineering from Bulgarian Academy of Sciences, Sofia, Bulgaria, in 1996, and the Dr.Tech. degree from Tampere University of Technology, Tampere, Finland, in 2003.

Currently, he is Senior Researcher in the Institute of Signal Processing, Tampere University of Technology. His research interests are in transform methods for signal, image, and video processing.

Gozde B. Akar (S'86–M'94–SM'98) received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1988 and the M.S. and Ph.D. degrees from Bilkent University, Bilkent, Turkey, in 1990 and 1994, respectively, all in electrical and electronics engineering.

She was with the University of Rochester, Rochester, NY, Center of Electronic Imaging Systems as a Visiting Research Associate from 1994 to 1996. From 1996 to 1998, she worked as a Member of Research and Technical Staff at Xerox Corporation—Digital Imaging Technology Center, Rochester. From 1998 to 1999, she was with Baskent University, Department of Electrical and Electronics Engineering. During the summer of 1999, she worked as a Visiting Researcher at the Multimedia Laboratories, New Jersey Institute of Technology. Currently, she is an Associate Professor with the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey. Her research interests are in face recognition, 2-D and 3-D video compression, and multimedia streaming.

Dr. Akar is an Editor of the *EURASIP Journal on Signal Processing: Image Communincation*.

Georgios A. Triantafyllidis was born in Thessaloniki, Greece, in 1975. He received the Diploma degree in electrical engineering and the Ph.D. degree in image/video processing from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1997 and 2002, respectively.

He is currently a Senior Researcher in the Informatics and Telematics Institute of CERTH, Thessaloniki, Greece. His main research interests include image compression and analysis, 3-D data processing, and medical image communication. His involvement with those research areas has led to the co-authoring of ten journal publications, five book chapters, and more than 35 conference papers. He is involved in several research European and National projects and he serves as a regular reviewer for a number of international journals and conferences.

Dr. Triantafyllidis is a member of the Technical Chamber of Greece.

Alper Koz (S'05–M'06) was born in Kahramanmaras, Turkey, in 1978. He received the B.S. and M.S. degrees in electrical and electronics engineering, from Middle East Technical University, Ankara, Turkey, in 2000 and 2002, respectively, where he is currently working toward the Ph.D. degree

He is currently a Research Assistant with Middle East Technical University, Ankara, Turkey. He has participated in the projects in the area of image and video watermarking supported by The Scientific and Technological Research Council of Turkey from 2000 to 2004. Since 2004, he is also a researcher in 3DTV, Network of Excellence, funded by the European Commission 6th Framework Programme. His research interests include watermarking techniques for image, video, Free View, and 3-D Television.