Visual Data Compression for Multimedia Applications

TOURADJ EBRAHIMI, MEMBER, IEEE, AND MURAT KUNT, FELLOW, IEEE

Invited Paper

In this paper, the compression of visual information in the framework of multimedia applications is discussed. To this end, major approaches to compress still as well as moving pictures are reviewed. The most important objective in any compression algorithm is that of compression efficiency. High-compression coding of still pictures can be split into three categories: waveform, second-generation, and fractal coding techniques. Each coding approach introduces a different artifact at the target bit rates. The primary objective of most ongoing research in this field is to mask these artifacts as much as possible to the human visual system. Video-compression techniques have to deal with data enriched by one more component, namely, the temporal coordinate. Either compression techniques developed for still images can be generalized for three-dimensional signals (space and time) or a hybrid approach can be defined based on motion compensation. The video-compression techniques can then be classified into the following four classes: waveform, object-based, model-based, and fractal coding techniques. The aim of this paper is to provide the reader with a tutorial on major visual data-compression techniques and a list of references for further information on the details of each method.

Keywords— *H.261, H.263, image compression, interactivity, JPEG, MPEG, video compression.*

I. INTRODUCTION

The importance of visual communications has increased tremendously in the last few decades. The progress in microelectronics and computer technology, together with the creation of networks operating with various channel capacities, is the basis of an infrastructure for a new era of telecommunications. New applications are preparing a revolution in the everyday life of our modern society. Emerging applications such as desktop video conferencing, mobile terminals, and Internet-based audio-visual communications will (if they do not already) have a great impact on modern professional life, education, and enter-

tainment. Visual information is one of the richest but also most bandwidth-consuming modes of communication. To meet the requirements of new applications, powerful data-compression techniques are needed to reduce the global bit rate drastically, even in the presence of growing communications channels offering increased bandwidth.

The issue of quality is of prime importance in most applications requiring compression. In fact, although most applications require high compression ratios, this requirement is in general in conjunction with the desire for high quality in the resulting pictures. In this paper, the terms "high compression" or "very-low-bit-rate coding" are used as synonyms and with always the underlying assumption that the resulting pictures bear the highest possible quality (ideally without visible distortions). It is the belief of the authors that the requirement of high quality in the resulting pictures is the most important factor in the success of an application in the marketplace. Without this condition, the widespread acceptance of such systems in consumer or professional sectors cannot be achieved.

This paper gives an overview of major techniques already used or under investigation for compression of visual information. The outline of the paper is as follows. Section II gives a review of still-image coding techniques. Section III discusses video coding. Pre- and postprocessing of the visual information are discussed in Section IV. Section V overviews the existing standards for the compression of still and moving pictures. Dynamic coding of visual information is presented in Section VI. Last, conclusions are drawn in Section VII.

II. STILL-IMAGE CODING

High-compression still-image coding has triggered strong interest in recent years thanks to the progress of Internet imaging and digital photography. In this type of coding, perceptually invisible distortions of the original image are accepted in order to attain very high compression ratios. In its most general case, the compression of a signal (including visual information) can be seen as described by
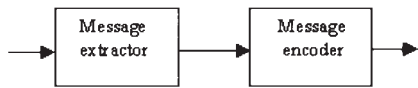
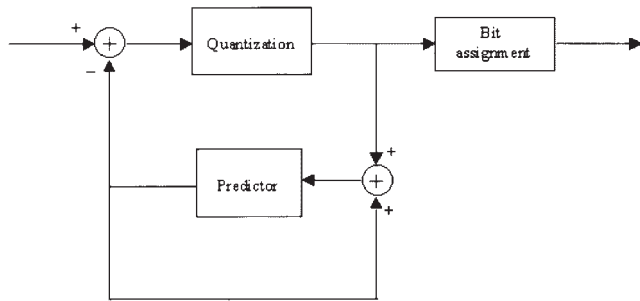**Fig. 1.** General block diagram of a compression scheme.

Fig. 2. General block diagram of predictive coding.

Fig. 3. General block diagram of a transform-based scheme.

the block diagram in Fig. 1. According to this diagram, such a system is composed of a message extractor followed by a message encoder. Message extraction is intimately related to the representation model of the information and/or the transformations to the initial representation and aims at extracting most relevant features from the original data. The message encoder is in general a source coder or an entropy coder assigning bit words to every extracted message, either by a systematic method or by making use of statistical behavior of extracted messages (entropy coding).

Still-image coding techniques can be split into three distinct groups. The first group is called waveform coding and consists of transform and subband coding (including wavelets). The second group, called second-generation techniques, consists of techniques attempting to describe an image in terms of visually meaningful primitives (contour and texture, for example). The third group is based on the fractal theory, in which an image is reconstructed by means of affine transformation of its self-similar regions. Fractal-based coding methods can be seen as a subclass of either waveform or second-generation techniques (depending on their implementations). However, we propose to consider them as a separate class, as they are based on the principle of representing an image by the set of transformations that will create it from any initial starting point rather than from a coded representation of the results of transformations on the initial image.

A. Waveform-Based Techniques

These techniques refer to methods that assume a certain model on the statistics of pixels in an image. The primitives of these techniques are either individual pixels or a block of pixels (or a transformed version from their values) and constitute the message to be encoded. These techniques can be roughly divided into two major classes, namely, predictive coding and transform-based coding. In predictive coding, the spatial correlation between pixels in an image is reduced by predicting the value of every pixel and by encoding the error of this prediction only. Fig. 2 gives the general block diagram of this class of coding techniques.
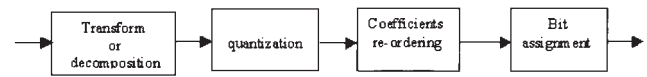
A transform-based coding system can be divided into the following steps: decomposition/transform (also referred to as energy compaction) of the image, quantization of the resulting coefficients, ordering of quantized coefficients, and bit assignment. Fig. 3 depicts these steps in a general block diagram. The first step transforms the image into another representation, where most of the energy of the signal is compacted in a few coefficients. The most popular transforms are linear, which generally speaking belong to the family of pyramidal [23] or subband decomposition schemes. As a general approach, the subband analysis/synthesis system was first introduced for one-dimensional (1-D) data such as audio signals by Croisier *et al.* [24] in 1976. Smith *et al.* [25] proposed solutions having the perfect-reconstruction property. The extension to two-dimensional (2-D) signals was reported by Vetterli [26]. Later, it was applied to image compression [27]. In subband coding, a signal is split into its subbands by using a group of band-pass filters [28] followed by critical subsampling, as depicted in Fig. 4. A special case of subband decomposition is block-based linear transform coding, among which the most popular is discrete cosine transform (DCT) [22] coding. This technique is used in the Joint Photographic Experts Group (JPEG) still-image compression standard. At compression factors of about 30–40, the DCT approach produces strong blocking artifacts (see Fig. 7). All block transform coders suffer from this distortion. Unfortunately, the human eye is very sensitive to such distortions, and therefore, block coders are not appropriate for high-compression image coding. However, they provide images of reasonably good quality at lower compression factors. Subband coding of images has been the subject of intensive research in the last few years [29]–[32]. Fig. 5 depicts an example of subband decomposition of a natural image. A particularly well-investigated approach is that of wavelet decomposition, which is a subset of subband decomposition in which the transformed representation provides an intrinsic multiresolution data structure useful for a number of applications [33], [34]. The main artifact at high-compression factors (around 50) in such approaches is due to the Gibbs phenomenon of linear filters and is called the ringing effect.

Quantization is one of the most important steps in any compression scheme [144]. The aim of such a process is to project the continuous values of resulting transformed coefficients into a finite set of symbols, each representing a good approximation of the coefficients' values. The most popular quantization scheme is that of uniform quantization, which is widely used due to its simplicity and efficiency. For a higher compression efficiency, uniform quantization can be combined with a dead-zone thresholding in order to augment the number of zero-valued quantized coefficients. Other variants of quantization have been studied, such as
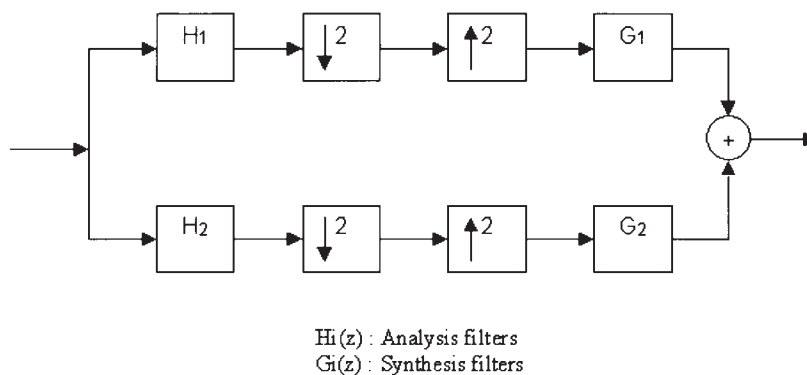
Hi(z) : Analysis filters
Gi(z) : Synthesis filters

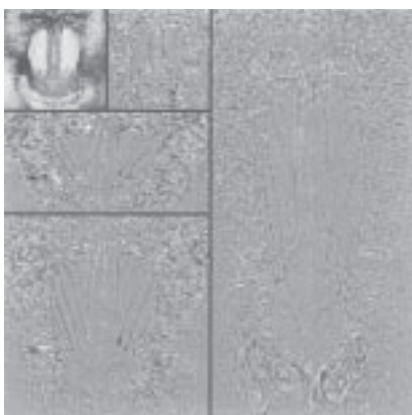**Fig. 4.** Block diagram of a subband decomposition of a 1-D signal.

Fig. 5. Example of subband decomposition of a natural image.

nonuniform quantization and perceptual quantization. In the latter, larger quantization step sizes are used to quantize higher frequency transformed coefficients by taking into account the sensitivity of the human visual system to frequency content of an image. In contrast to scalar quantization discussed above, vector quantization (VQ) is the process of mapping sets of values in the form of vectors into a predefined set of patterns. Although VQ can be seen as a generalization of scalar quantization, it can be shown that from a rate-distortion point of view, it results in an optimum performance even when the data to be quantized is made of independent samples [145]. In principle, the design of a vector quantizer requires a common definition-of-patterns codebook used in both the encoder and decoder. The choice of the codebook depends on the statistical properties of the source data to be quantized and as a consequence plays an essential role in the performance of VQ. Several strategies have been developed in order to design appropriate codebooks in an efficient way requiring minimal computational load. The most well-known approach for the design of a codebook is that of a generalized Lloyd algorithm, also known as LBG algorithm [141]. Other variants of vector quantization may update the initial codebook as the process of quantization is in progress. Such quantizers are known as adaptive vector quantizers [140], [143]. Another class of vector quantizers makes use of no explicit codebook in order to avoid storage requirements. Lattice vector quantization

(LVQ) is one such approach. It is motivated by the fact that early investigations in information theory have shown that the optimal high resolution entropy-constrained scalar quantization in that of a uniform quantization [146]. Based on similar results, for large vector sizes, the optimal vector quantizer for a constrained entropy is also obtained using a codebook with uniform cells [147]. Formation of vectors is another important issue in the design of VQ. In the earlier realizations of vector quantizers, vectors were formed by the values of pixels in an image on square blocks [142]. In more recent techniques, vectors are created from the coefficients of a transformed image [140]. The coefficients can be either formed from those of a given subband around the same spatial position or from the coefficients of different subbands in the same spatial position. Although from the theoretical point of view, it is possible to show that VQ remains superior to scalar quantization, from the complexity point of view, the additional overhead due to implementation of VQ does not seem to favor such approaches for most applications.

Reordering of nonzero coefficients is performed to better exploit the statistical occurrence of zero-valued coefficients in an energy-compacting transform. The most popular coefficient reordering is that of incremental zigzag scanning of DCT coefficients from its zero-frequency (DC) value up to its highest frequency components. This process will reduce the number of code words needed to code nonzero coefficients. For high-compression subband coding [35] (compression ratios higher than 50), it is of major importance to exploit the existing zero correlation across the subbands as proposed in [36]–[38] in order to maintain a good quality. Fig. 6 gives an example of such a coefficient reordering. A scheme combining VQ and the prediction of insignificance across the bands has been proposed in [39] and [40]. Although it is possible to reduce the ringing effect by an appropriate design of the subband filters [41]–[44], it is not possible to find linear subband filters that do not have any ringing effect. To avoid this artifact, morphological subband decompositions have been proposed [45]–[47] that lead to decoded pictures of good quality at compression ratios as high as 70–80. Fig. 7 shows a comparison between a DCT-based image-compression scheme (JPEG) and a wavelet-based method for compression ratios of $40:1$ in both cases.
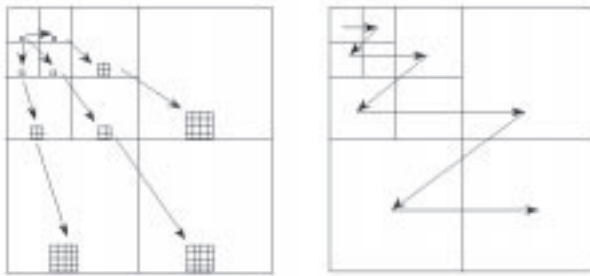
Fig. 6. Parent-children and its corresponding scan path for coding of wavelet coefficients.

(a)

(b)

Fig. 7. Results of compression of a still image using (a) JPEG and (b) wavelet.

B. Second-Generation Techniques

The second group of methods is based on second-generation techniques. They attempt to decompose the data into visual primitives such as contours and textures [48], [49]. One approach is to divide the image into directional primitives as proposed in [50]. Segmentation-based coding techniques [51] extract regions from the image data, which are represented by their shape and textural content. Following similar ideas, sketch-based image coding [52] is based on extracting the contours of an image, namely, their geometric and intensity information, resulting in the so-called sketch picture. The texture is then defined by the difference between the original and the sketch

image and is coded using waveform-coding techniques. An extension of this technique has been proposed by Ran *et al.* [53], [54] and is based on a three-component image model. This technique divides the image into strong-edge, texture, and smooth components. The strong-edge component is encoded separately, whereas the texture and smooth components are encoded using waveform-coding techniques. A solution to find the most important image features has been proposed by Mallat *et al.* [55] using multiscale edges. A double-layer technique based on multiscale edges and textures has been proposed in [56]. In general, second-generation techniques become efficient at higher compression ratios (about 50) when compared to other methods [49], [55], [56].

C. Fractal-Based Techniques

Iterated functions systems (IFS) theory, closely related to fractal geometry, has recently found an interesting application in image compression. Barnsley [57] and Jacquin [58] pioneered the field, followed by numerous contributions [59], [60]. The approach consists of expressing an image as the attractor of a contractive functions system, which can be retrieved simply by iterating the set of functions starting from any initial arbitrary image. The form of redundancy exploited is named piece-wise self-transformability. This term refers to a property that each segment of an image can be properly expressed as a simple transformation of another part of higher resolution. IFS-based still-image compression techniques can pretend to have very good performance at high compression ratios (about 70–80), as shown in [62], [63]. The major problem with fractal-based coding techniques is that of complexity at the encoding stage. However, the complexity of the decoder remains reasonable when compared to the encoding. Fractal-based techniques produce outstanding results in terms of compression in images, retaining a high degree of self-similarity. Another interesting feature of fractal-based techniques is their ability to produce a good-quality rendered image for an arbitrary scaling factor.

III. VIDEO CODING

In addition to still-image coding, the compression of video signals has to process one more dimension: the temporal coordinate. The existing compression techniques for still images can serve as the basis for the development of video-coding techniques. Either compression algorithms for still images can be generalized to three-dimensional (3-D) signals or a hybrid approach based on motion compensation can be defined. The input video data for very-low-bit-rate applications is typically composed of small-size images [approximately 144×176 pixels in luminance and a quarter of this resolution in chrominances, known as quarter-common intermediate format (QCIF)] with a frame rate of about 5–15 frames/s. The target bit rates vary from 4.8 to 64 Kb/s depending on the application. Medium-bit-rate video deals with images of average size [approximately 288×352 pixels in luminance and a quarter of this resolution

in CIF, or interlaced 576 (480) × 720 pixels in luminance and half-horizontal resolution in chrominances, called ITU-R 601] at a frame rate of 25 or 30 Hz. The bit rates delivered by most conventional medium-bit-rate schemes vary between a few hundred Kb/s up to 1.5 Mb/s. In professional and high-end applications, image sizes and frame rates are at least that of ITU-R 601 or larger (high-definition television). The expected compressed bit rates are in the range of 1.5–35 Mb/s. Although at high end, digital television delivers excellent quality, it cannot compete with the quality of existing 35-mm film as far as the spatial resolution is concerned. Recent developments in superhigh-definition (SHD) imaging technology allow a real-time display of images as large as 2000 × 2000 pixels with frame rates as high as 60 Hz. Compression of SHD digital video is a new axis of research. In the following, the video-compression techniques have been classified into four classes: waveform, object based, model based, and fractal coding.

A. Waveform-Based Techniques

Viewing the temporal axis as a third dimension, all the waveform-coding techniques developed for image compression can be generalized to the compression of video signals. Only limited work has been published in the use of 3-D transforms for 3-D visual data compression [64]. The blocking artifacts at low bit rates, however, make it improper to code image sequences. Three-dimensional subband coding of video was first introduced by Karlsson *et al.* [65]. In this work, standard subband filters are used for the spatial directions while a DCT-derived filter bank is applied to the temporal dimension. Variations of this scheme have been reported in literature [66], [67]. The drawback of 3-D subband coding is that the temporal filtering is not performed along the direction of motion. The most popular as well as efficient approach to reduce the temporal correlation between consecutive frames of a video signal is that of motion compensation (MC). MC makes use of the motion field of a video signal at a given instance to predict its following instance. This process is known as motion estimation (ME) and is similar in principle to that of predictive coding, discussed earlier. Several ME techniques have been investigated in the past in order to reduce its computational complexity while keeping a good performance in prediction [69]. Similarly, several methods have been investigated in order to reduce the energy (and therefore the bit rate required to code) of prediction error. The most popular MC techniques are those making use of prediction from the past frames (predictive) or past and future frames (bidirectional) [68] and those using overlapping motion compensation [11], [15]. Other variants, such as a combination of the temporal subband-coding component with MC, have been proposed in [70] and [71]. This scheme has then been extended by the use of LVQ and MC with subpixel accuracy [72]. The problem of coding the resulting prediction error images, also called displaced frame differences (DFD's), has been addressed by using linear transforms such as the DCT [73], [74] and

(a)

(b)

Fig. 8. Result of postprocessing (deblocking) of a picture compressed by DCT (a) without postprocessing and (b) after postprocessing.

by using the wavelet transform [76]. Most video-coding standards, such as Moving Picture Experts Group (MPEG)-1, MPEG-2, MPEG-4 [5], [7], [11], and recommendations H.261, H.263, and H.263+ [14]–[16], suggest the use of DCT-based algorithms for coding the DFD's. A preprocessing of the DFD using a morphological segmentation has been proposed by Li *et al.* [77], [78]. Applications of waveform-coding algorithms to very-low-bit-rate video coding have been proposed based on 3-D subband coding [79], [80], motion-compensated subband coding, and motion-compensated transform coding [75], [81]–[86].
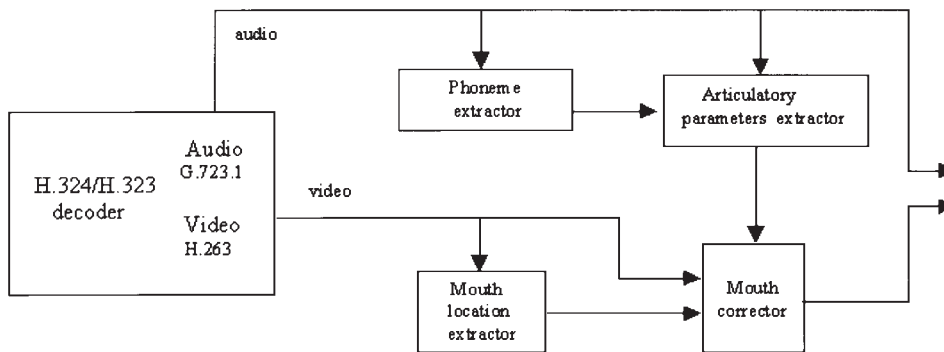
Fig. 9. Block diagram of audio-assisted postprocessing of head-and-shoulder video.
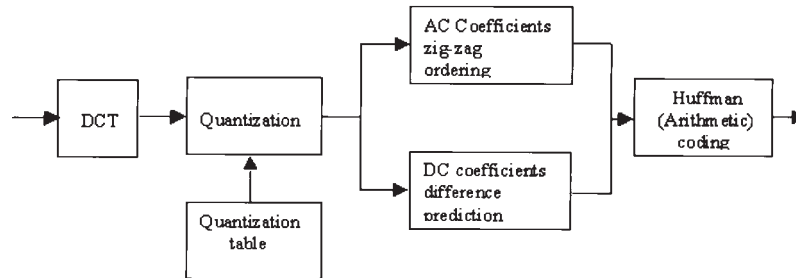

Fig. 10. Block diagram of JPEG algorithm (lossy-mode encoder).

B. Object-Based Techniques

The promising results obtained with second-generation techniques for still images motivated its extension to image sequence compression. A straightforward solution is to extend the 2-D techniques used previously in a 3-D context. One approach is to perform a 3-D segmentation of a sequence viewed as a 3-D volume. On this topic, Willemin *et al.* [87], [88] proposed an octree split-and-merge segmentation as a generalization of the quad-tree segmentation previously used in still-image coding [48]. Similarly, another technique has been introduced by Salembier *et al.* [89], [90] based on mathematical morphology allowing coding of arbitrary region shapes.

Along the same lines, an object-oriented scheme in which objects are defined as regions with three associated parameters—shape, textural content, and motion—was proposed in [92]. The parameters are obtained by image analysis based on source models of either moving 2-D or 3-D objects [93].

All these approaches require the transmission of the objects created at the encoder side. The textural content of the objects can be coded efficiently using transform-based techniques similar to those used in block-based methods [110]–[113]. Typically, the shape information is represented by bitmap coding [11], [16], [91], chain coding of the contour information [95], quad-tree shape representation [96], [97], or the medial axis transform [94], [98], [99]. Simulation results show that shape coding requires an important portion of the global bit rate. One solution to reduce this cost is to use more efficient techniques for shape representation, such as the geodesic morphological skeleton as proposed by Brigger *et al.* [99], and/or to perform a simplification of the contours by appropriate preprocessing operations prior to their encoding [100]. Another solution

is to avoid frequent transmission of contour information by object tracking [101]–[105]. A third solution is to define objects with simple shapes that need less bits to be transmitted for their shape representation [106]–[109], [114], [115].

Two imminent standards use the principle behind object-based coding in their design, namely, MPEG-4 and H.263+. The major reason for the use of object-based coding in both standards is to allow some sort of interactivity with objects rather than to achieve high compression ratios. The methods used for motion estimation as well as texture coding in both techniques are extensions of their block-based counterparts. The main difference between the two techniques resides in the method used for the coding of the shape information. MPEG-4 shape coding relies on a bitmap-based context arithmetic coding with motion compensation [11], whereas H.263+ makes use of a chroma-keying method and DCT-based coding [16], [116].

C. Model-Based Techniques

It is obvious that all techniques developed for compression rely on a certain model. The term model-based coding, however, refers to an approach seeking to represent the projected 2-D image of a 3-D scene by a 2-D or 3-D predefined model. The goal consists of finding an appropriate model with its corresponding parameters in order to make it as similar as possible to the desired objects in the scene to be coded.

This technique can be divided into two main steps, namely, analysis and synthesis. The analysis block is the most difficult task due to the complexity of the natural scenes. So far, the main effort has been concentrated on simple scenes such as head-and-shoulder sequences [117]–[119]. The synthesis block, however, is easier be-

Table 1 Summary of Present and Emerging Standards for Coding and Representation of Visual Information

Standard	Standardization body	Main Target bitrate	Main compression technologies	Main target applications
JPEG	ISO/IEC	Compression ratios 2-30	- DCT - Perceptual quantization - Zig-zag reordering - Huffman coding - Arithmetic coding	- Internet imaging - Digital photography - Image and video Editing
JPEG2000	ISO/IEC	Compression ratios 2-50	To be defined	- Internet imaging - Digital photography - Image and video editing - Printing - Medical imaging - Mobile applications - Color fax - Satellite imaging
MPEG-1	ISO/IEC	Bitrates up to about 1.5 Mb/s	- DCT - Perceptual quantization - Adaptive quantization - Zig-zag reordering - Predictive motion compensation - Bi-directional motion compensation - Half-sample accuracy motion estimation - Huffman coding - Arithmetic coding	- Storage on CD-ROM - Consumer video
MPEG-2	ISO/IEC	Bitrates 1.5 Mb/s to about 35 Mb/s	- DCT - Perceptual quantization - Adaptive quantization - Zig-zag reordering - Predictive motion compensation - Bi-directional motion compensation - Frame/field based motion compensation - Half-sample accuracy motion estimation - Spatial scalability - Temporal scalability - Quality scalability - Huffman coding - Arithmetic coding - Error resilient coding	- Digital TV - Digital HDTV - High quality video - Satellite TV - Cable TV - Terrestrial broadcast - Video editing - Video storage
MPEG-4	ISO/IEC	Bitrates 8 Kb/s to about 35 Mb/s	- DCT - Wavelet - Perceptual quantization - Adaptive quantization - Zig-zag reordering - Zero-tree reordering - Predictive motion	- Internet - Interactive video - Visual editing - Content manipultion - Consumer video - Professional video - 2D/3D computer

cause the techniques developed for image synthesis in the field of computer graphics have already addressed this problem. We will not enter into any further details of model-based techniques in this paper. An excellent tutorial exists on this subject by Pearson [120]. It is important to mention that the MPEG-4 standard will include technology for face (version 1) and body (version 2) animation, developed in the context of the MPEG-4 Synthetic Natural Hybrid Coding subgroup. This technology is described in "MPEG-4 Part 2: Visual," devoted to the coding of visual data.

D. Fractal-Based Techniques

The promising performance provided by fractal-based still-image compression techniques led to its applica-

			compensation - Bi-directional motion compensation - Frame/field based motion compensation - Half-sample accuracy motion estimation - Advanced motion estimation - Overlapping motion compensation - Spatial scalability - Temporal scalability - Quality scalability - View dependent scalability - bitmap shape coding - Sprite coding - Face animation - Dynamic mesh coding - Huffman coding - Arithmetic coding - Error resilient coding	graphics - Mobile
MPEG-7	ISO/IEC	To be defined	To be defined	- Visual content search
H.261	ITU-T	Bitrates p x 64 Kb/s (p:1-31)	- DCT - Adaptive quantization - Zig-zag reordering - Predictive motion compensation - Integer-sample accuracy motion estimation - Huffman coding - Error resilient coding	- ISDN video-conferencing
H.263	ITU-T	Bitrates 8 Kb/s up to 1.5 Mb/s	- DCT - Adaptive quantization - Zig-zag reordering - Predictive motion compensation - Bi-directional motion compensation - Half-sample accuracy motion estimation - Advanced motion estimation - Overlapping motion compensation - Huffman coding - Arithmetic coding - Error resilient coding	- POTS video-telephony - Desktop video telephony - Mobile video telephony

tion to video compression. Different approaches have been proposed. Beaumont suggested a straightforward extension of the 2-D approach to 3-D data volumes [122]. To reduce the computational burden, Li *et al.* proposed a 3-D approach without domain block search but increasing contractive transformation complexity [123]. Reusens worked on a scheme where sequence volume is adaptively segmented along an octree structure and 3-D blocks coded either by contractive transformation or 3-D temporal block matching [114]. Independently, Lazar *et al.* followed the same approach but allowing only contractive transformations [125]. Hürtgen *et al.* introduced a 2-D approach where regions classified as foreground are coded by intraframe fractal approximation [124].

IV. PRE- AND POSTPROCESSING

In a majority of multimedia applications, the materials used for the capture of the data (such as the camera) should be cheap, making it affordable for a large number of users. In addition, compact solutions (miniaturized terminals) are desired. However, the quality of such equipment drops when compared to their more expensive and professional counterparts. It is mandatory to use a preprocessing stage prior to coding in order to enhance the quality of the final pictures and to remove the various noise that will affect the performance of compression algorithms. Solutions have been proposed in the field of image processing to enhance the quality of images for various applications [126]–[128]. A more appropriate approach would be to take into account
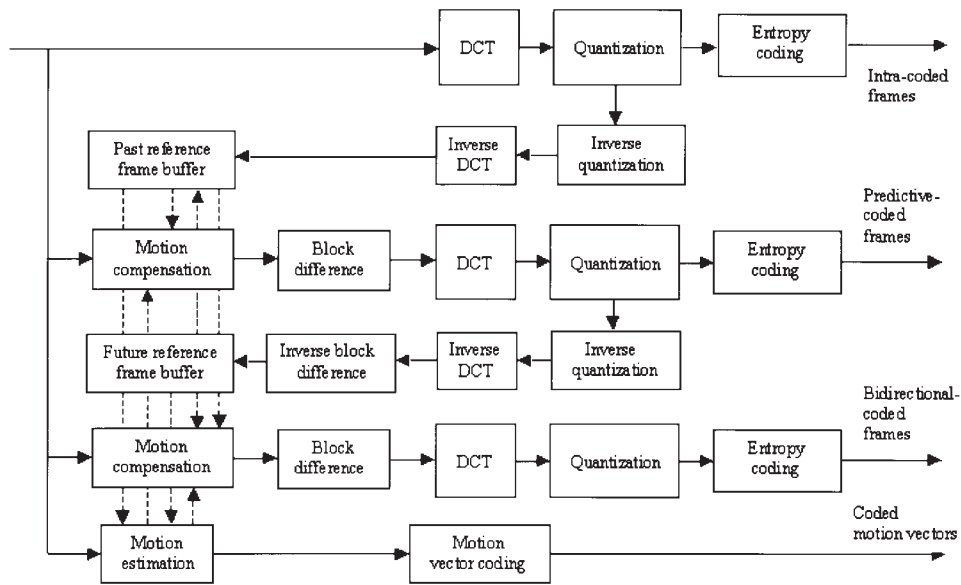
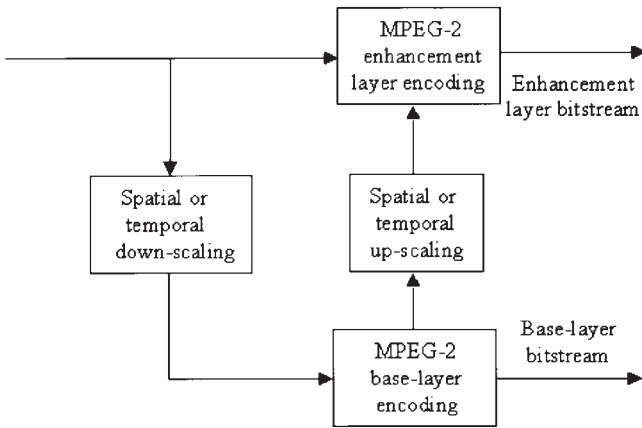**Fig. 11.** Block diagram of MPEG-1 algorithm (encoder).

Fig. 12. Block diagram of scalable compression in MPEG-2 (encoder).

the characteristics of the coding scheme when designing such operators.

In addition to the above, pre- and postprocessing operators are extensively used in order to render the input or output images in a more appropriate format for the purpose of coding or display.

Mobile communications is an important class of applications in multimedia. Terminals in such applications are usually subject to different motions such as tilting and jitter, translating into a global motion in the scene due to the motion of the camera. This component of the motion can be extracted by appropriate methods' detecting the global motion in the scene and can be seen as a preprocessing stage. Results reported in the literature show an important improvement of the coding performance when a global motion estimation is used [128]–[131].

It is normal to expect a certain degree of distortion of the decoded images for very-low-bit-rate applications. An appropriate coding scheme, however, will introduce these distortions in areas that are less annoying for the users. An additional stage could be added to further reduce the

distortions due to compression as postprocessing operators. Shortly after the introduction of the first coding schemes based on block transforms, solutions were proposed in order to reduce the blocking artifacts appearing at high compression ratios [132]–[137]. Fig. 8 gives an example of blocking-effect removal after decoding a compressed bitstream at very high compression ratios. The same types of approaches have been used in order to improve the quality of decoded signals in other coding schemes, reducing different kinds of artifacts such as ringing, blurring, mosquito noise, and so on [138], [139]. More recently, advanced postprocessing mechanisms have been studied to improve lip synchronization of head-and-shoulder video coding at a very low bit rate by using the knowledge of decoded audio in order to correct the positions of the lips of the speaker [148]. Fig. 9 gives an example of the block diagram of such a postprocessing operation.

V. STANDARDS

A number of standards have been defined for the compression of visual information. Table 1 summarizes the major ones. The JPEG still-image compressor was proposed by the Joint Photographic Experts Group [1], [2] and is a general-purpose still-image compression tool [3]. JPEG allows both lossless and lossy compression of pictures. In its lossless mode, the values of pixels in the image are coded by means of a predictive scheme. The prediction is based on a linear combination of causal neighbors of already coded pixels. The JPEG algorithm in its lossless mode provides modest compression ratios of about $2:1$. The block diagram of the baseline algorithm used in the lossy mode of JPEG is given in Fig. 10. It mainly consists of a block DCT, followed by scalar quantization and Huffman or arithmetic coding of zigzag scanned coefficients. The JPEG algorithm in lossy mode provides good-quality results for a large class of natural images at compression ratios of 5–20. Its main drawbacks are the strong distortions (mainly blockiness
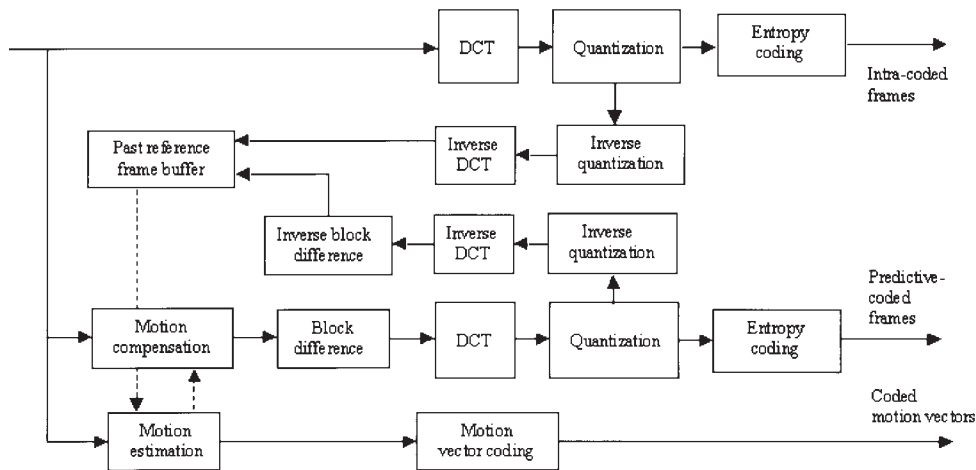
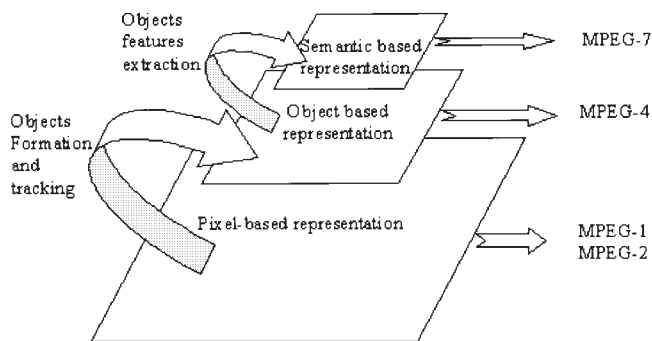**Fig. 13.** Block diagram of H.261 compression algorithm (encoder).

Fig. 14. Visual data-representation pyramid.

and mosquito noise) at higher compression factors and its lack of bit-rate control, as well as weak error-resilience capabilities. The JPEG algorithm also provides a number of extended modes such as progressive mode and hierarchical mode, where it is possible to partially decode a compressed picture at lower quality (only a subset of DCT coefficients are decoded) or lower spatial resolution (pyramidal coding) [3].

The MPEG standards address the compression of video signals. MPEG-1 operates at bit rates of up to about 1.5 Mb/s and targets storage on media such as CD-ROM, as well as transmission over narrow communication channels such as the integrated services digital network (ISDN) or local- and wide-area networks [6]. The basic principle of its compression algorithm is depicted in Fig. 11. Macroblocks of size 16 × 16 can be coded using up to three different modes for a given frame in the image sequence. All macroblocks in a so-called intraframe are coded using a scheme similar to that of JPEG, in which special care has been devoted to allow rate-control and error-resilience mechanisms. The macroblocks in consecutive frames are coded by motion compensation from a past frame (predictive mode) or past and future frames (bidirectional mode). All macroblocks in a frame are regularly forced to be coded in intramode (intracoded frames) to prevent error propagation and to provide rapid access to any decoded frame in a sequence of pictures. The MPEG community proposed yet another class of coding algorithms denoted

as MPEG-2 for generic compression of high-quality video of various types and bit rates. The basic principle behind MPEG-2 algorithms is similar to that of MPEG-1 to which special features have been added to allow an intrinsic coding of frames as well as fields in interlaced sequences. It also allows a scalable coding of video signals by which it is possible to decode a signal with lower temporal or spatial resolutions or qualities from a same compressed bitstream. Fig. 12 gives the block diagram of scalable coding of the MPEG-2 compression algorithm. MPEG-2 mainly operates at bit rates around 1.5–35 Mb/s and provides higher quality video signals at the expense of more complex processing when compared to MPEG-1. MPEG-2 defines several profiles and levels allowing its efficient use in various applications from consumer up to professional categories. Standards such as the Digital Audio Visual Council [8], digital video disc [9], and digital video broadcast [10] make use of MPEG-2 video algorithms in their respective applications. Currently, MPEG is finalizing the first version of a new standard known as MPEG-4 [11], scheduled to become an international standard by 1998 [12]. The latter aims at providing an integrated solution for a multitude of multimedia applications, ranging from mobile video telephony up to professional video editing as well as Internet-like interactive communications.

To allow interactivity with visual objects, the MPEG-4 algorithm is based on an object-based representation similar to those used in second-generation coding techniques (see Fig. 14). This is the most basic and major difference between MPEG-4 and all other existing standards. In an object-based representation, pixels of an object are considered as inseparable entities the same way molecules of objects in the real world are kept together by a force field. The function of the force field in an MPEG-4 object-based representation is fulfilled by object shape coding. Fig. 15 depicts the general block diagram of the MPEG-4 coding algorithm. In MPEG-4, video signals are composed of distinct objects, each with its respective shape, motion, and texture information. These entities are coded per object independently in order to allow direct access and manipulation (cut, paste, deformation, etc.). Fig. 16 gives an example
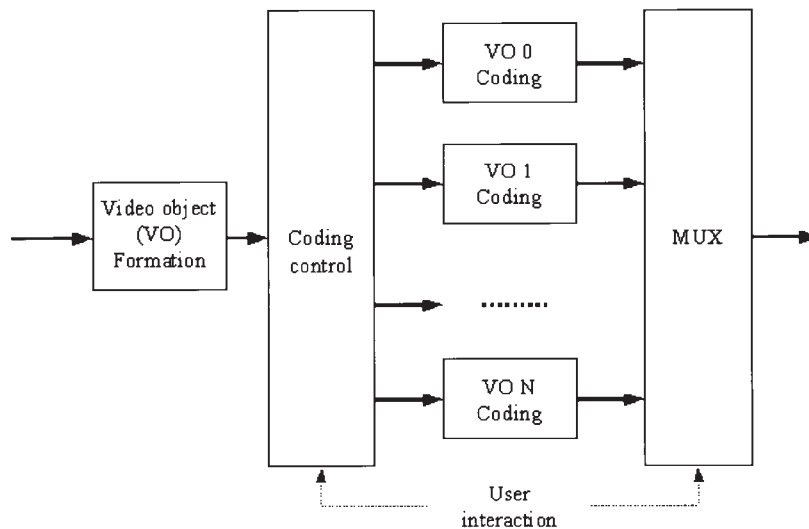
Fig. 15. General block diagram of MPEG-4 algorithm (encoder).

Fig. 16. Examples of manipulation of objects in an MPEG-4 object-oriented bitstream.

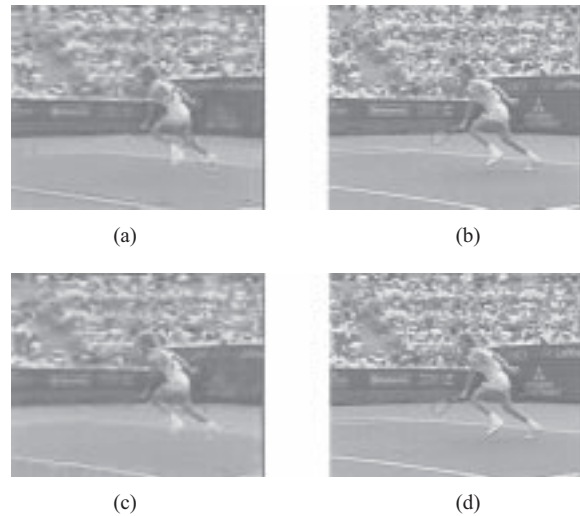of possible manipulations of objects in a scene that have been compressed using the MPEG-4 algorithm.

Fig. 17. Results of compression of a video sequence using (a) H.263 at 100 Kb/s, (b) H.263 at 1 Mb/s, (c) MPEG-4 at 100 Kb/s, and (d) MPEG-4 at 1 Mb/s.

Recommendation H.261 (also known as p × 64) was proposed by the International Telegraph and Telephone Consultative Committee (CCITT, now the ITU-T) [14]. Based on this standard, video conferencing at bit rates down to 64 Kb/s has become feasible. This requires the capacity of one channel of the ISDN. The basic principle of the H.261 algorithm is similar to that of MPEG-1, in which spatial redundancies are reduced by DCT and temporal correlation by motion-compensated predictive coding. Due to computational complexity and delay issues, no bidirectional coding mode is used in H.261, and the motion estimation is performed with an integer pixel accuracy only. Fig. 13 gives the block diagram of the H.261 compression algorithm.

In the near future, modern visual communication applications will be possible for the general public. For that objective, the transmission media must switch to widely available public switched telephone networks or mobile channels. The transmission of the video sequences at bit rates as low as 9.6 Kb/s will be strongly needed. A
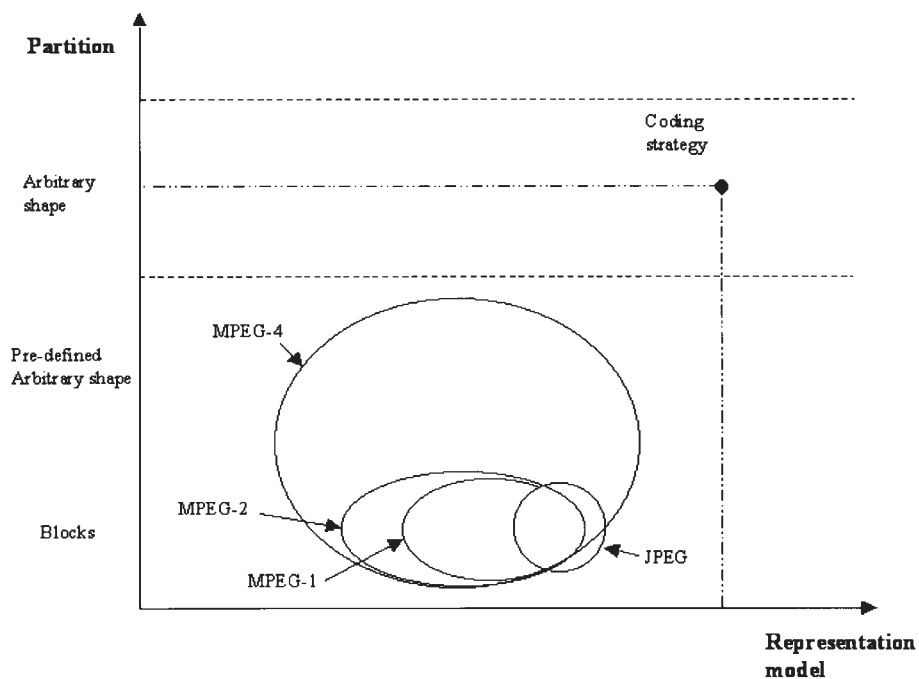
Fig. 18. Dynamic coding syntax.

series of standards are being developed for this purpose. Recommendation H.263 is the first one for very-low-bit-rate video compression, allowing transmission of audio-visual information on narrow channels down to 9.6 Kb/s. The compression scheme used in H.263 is similar to that of H.261, to which a few enhancements have been added. The major enhancements are:

- a new type of coded frame called PB, in which both predicted and bidirectional macroblocks can reside;
- an advanced motion estimation with a motion field resolution down to 8×8 pixels;
- half-sample motion estimation;
- overlapping motion compensation.

H.263 will be followed by H.263+, [16] which is currently under development by Study Group XVI of the ITU-T. As mentioned earlier, MPEG-4, as an integrated standard covering a large number of applications, will also offer a solution for this class of applications in one or several of its profiles [17]. Fig. 17 shows some results in terms of compression efficiency using MPEG-4 and H.263 standards for coding of a video signal at various bit rates.

An uncompressed video sequence for very-low-bit-rate applications typically requires a bit rate of up to 10 Mb/s. To achieve very low data rates, compression ratios on the order $1000:1$ are required to meet the needs of a large public, which has access only to very narrow communication channels for the few coming years. Intensive research has been carried out in the last decade to attain this objective [20], [21]. As mentioned above, variations of H.261 and MPEG for very-low-bit-rate applications have been defined as simulation models and have given birth to new techniques such as H.263, H.263+, and MPEG-4. For these simulation models, severe blocking artifacts may still occur at very low data rates. Much ongoing research is devoted

to developing methods based on philosophies differing drastically from the existing standards for higher bit rates. Work in this direction has already begun, as one can witness an increasing effort in bringing to standardization activities techniques such as sprite coding and matching pursuits in MPEG-4 [150]–[152] and wavelet-based compression in MPEG-4 and JPEG2000 [18], [19].

The problem of search for visual information is of prime importance in a number of future multimedia applications. The MPEG community has started to seek a standardized solution to this problem in a new work item called MPEG-7. At the time of this writing, only preliminary work has been achieved to clarify issues related to the problem of search, such as identification of its context, objectives, requirements, and applications. Very likely, the solution would reside in another fundamental change of the representation to a so-called semantic representation (see Fig. 14), which will be built on top of those used in other standards (pixel based and object based) [149].

VI. DYNAMIC CODING OF VISUAL INFORMATION

It is well known that visual information is of a highly nonstationary nature. In multimedia applications, all sort of visual data could be transmitted between terminals. Among all the techniques already investigated in literature, some perform better in particular regions of an image than others. Typically, subband/wavelet schemes are known to perform well in areas with texture, whereas techniques based on object representation or morphological operators perform well in areas with sharp edges and contours. Similarly, methods using linear transforms produce poor results in areas with text or graphics. Dynamic coding is a solution to solve the drawbacks existing in a given scheme while still maintaining its strong performance where appropriate. The

Fig. 19. Dynamic coding of a still picture at a compression factor of 50 : 1. (a) Compressed by a linear subband decomposition. (b) Compressed by a nonlinear subband decomposition. (c) Compressed by dynamic coding putting in competition the previous two methods.

basic idea behind dynamic coding is simple yet powerful [153]. The visual information is divided into several regions with similar compression characteristics. Every region is encoded using a multitude of compression techniques. Among all these techniques, the one that is the most efficient is chosen, and the compressed bitstream of the region using the best coding technique is sent to the decoder along with information specifying which technique

was chosen for its coding. As an example, in areas with texture, a subband/wavelet technique would be used, while areas with strong edges and contours will be coded using morphological-based or other more appropriate techniques. Similarly, text areas will use an encoding technique more appropriate for an efficient compression of such data.

The concept of dynamic coding defines implicitly a general coding syntax. Visual data are segmented into regions, each represented by its respective representation model. The syntax therefore relies on two degrees of freedom, namely, the partition and the associated representation models.

As depicted in Fig. 18, the resulting syntax is both open and flexible. Indeed, different classes of partitioning can be considered, ranging from the whole image as a single object to arbitrary shape segmentation through predefined and adaptive partitioning. Additionally, each region resulting from a particular segmentation can be coded with respect to a model chosen from a multitude of representation methods. Fig. 19 gives an example of dynamic coding of a still image by putting in competition a linear and a nonlinear subband decomposition scheme.

VII. CONCLUSIONS

The goal of this paper has been to provide interested readers with a road map for navigating through coding techniques targeting visual data compression. The compression of still pictures as well as the compression of video sequences were overviewed. It is delicate to point out the most promising approach, since this would strongly depend on the application requirements. Although the compression efficiency is the most important feature in any compression scheme, depending on the situation, other functionalities and parameters, such as scalability, error resilience, complexity, and delay, to mention a few, should be taken into account when choosing a best compression scheme. The hope of the authors is that the discussions in this paper can provide enough insights and information so as to allow readers to be able to make the appropriate choice after identifying the needs of their applications.

REFERENCES

[1] "Digital compression and coding of continuous-tone still images," ISO/IEC JTC1 Committee Draft, JPEG 8-R8, 1990.
[2] Joint Photographic Experts Group. JPEG home page. [Online]. Available WWW: http://www.disc.org.uk/public/jpeghomepage.htm
[3] W. B. Pennebacker and J. L. Mitchell, *JPEG Still Image Compression Standard.* New York: Van Nostrand, 1993.
[4] L. Chiariglione, "MPEG and multimedia communications," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 5–18, Feb. 1997.
[5] "Information technology—Coding of moving pictures and associated audio for digital storage media up to about 1.5

Mbit/s—Part 2: Coding of moving pictures information," ISO/IEC JTC1 CD 11172, MPEG, 1991.

[6] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM,* vol. 34, no. 4, pp. 46–58, Apr. 1991.

[7] "Information technology—Generic coding of moving pictures and associated audio information—Part 2: Video," ISO/IEC DIS 13818-2, MPEG-2, 1994.

[8] Digital Audio Visual Council. What is DAVIC? [Online]. Available WWW: http://www.davic.org/WHATIS.htm.

[9] DVD frequently asked questions. [Online]. Available WWW: http://www.videodiscovery.com/vdyweb/dvd/dvdfaq.html.

[10] DVB—Television for the third millennium. [Online]. Available WWW: http://www.dvb.org/.

[11] "Information technology—Coding of audio-visual object: Visual," ISO/IEC JTC1 CD 14496-2 (MPEG-4), Oct. 1997.

[12] R. Koenen, F. Pereira, and L. Chiariglione, "MPEG-4: Context and objectives," *Signal Process.,* vol. 9, no. 4, pp. 295–304, May 1997.

[13] M. Liou, "Overview of the p × 64 Kbit/s video coding standard," *Commun. ACM,* vol. 34, no. 4, pp. 60–63, Apr. 1991.

[14] "Recommendation H.261—Video codec for audiovisual services at p × 64 Kbit/s," CCITT SG XV, COM XV-R37-E, 1990.

[15] "Video coding for low bitrate communication," ITU-T SG XV, DRAFT H.263, 1996.

[16] "Video coding for low bitrate communication," ITU-T SG XVI, DRAFT 13 H.263+ Q15-A-60 rev. 0, 1997.

[17] Moving Picture Experts Group. MPEG home page. [Online]. Available: http://drogo.cselt.stet.it/mpeg/.

[18] "New work item: JPEG 2000 image coding system," ISO/IEC JTC1/SC29WG1 N390R, JPEG2000, 1997.

[19] "Call for contributions for JPEG 2000 image coding system," ISO/IEC JTC1/SC29WG1 N505, JPEG2000, 1997.

[20] D. E. Pearson and J. A. Robinson, "Visual communications at very low data rates," *Proc. IEEE,* vol. 73, pp. 795–812, Apr. 1985.

[21] T. Ebrahimi, E. Reusens, and W. Li, "New trends in very low bitrate video coding," *Proc. IEEE,* vol. 83, pp. 877–891, June 1995.

[22] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.,* vol. C-23, pp. 90–93, Jan. 1974.

[23] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.,* vol. COM-31, pp. 532–540, Apr. 1983.

[24] A. Croisier, D. Esteban, and C. Galand, "Perfect channel splitting by use of interpolation, decimation, tree decomposition techniques," in *Proc. Int. Conf. Information Sciences/Systems,* Aug. 1976. pp. 443–446.

[25] M. J. T. Smith and T. P. Barnwell, "Exact reconstruction techniques for tree structured subband coders," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 434–441, June 1986.

[26] M. Vetterli, "Multi-dimensional subband coding: Some theory and algorithms," *Signal Process.,* vol. 6, pp. 97–112, Feb. 1984.

[27] J. Woods and S. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 1278–1288, Oct. 1986.

[28] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Apr. 1980, pp. 291–294.

[29] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Oct. 1987, vol. 845, pp. 50–58.

[30] H. Gharavi and A. Tabatabai, "Subband coding of monochrome and color images," *IEEE Trans. Circuits Syst.,* vol. 35, pp. 207–214, Feb. 1988.

[31] J. W. Woods, *Subband Image Coding.* Boston, MA: Kluwer, 1991.

[32] A. Nicoulin, M. Mattavelli, W. Li, and M. Kunt, "Subband image coding using jointly localized filter banks and entropy coding based on vector quantization," *Opt. Eng.,* vol. 32, no. 7, pp. 1438–1450, July 1993.

[33] C. K. Chui, *An Introduction to Wavelets.* New York: Academic, 1992.

[34] G. Strang and T. Nguyen, *Wavelets and Filter Banks.* Cambridge, MA: Wellesley-Cambridge Press, 1996.

[35] K. H. Goh, J. Soraghan, and T. S. Durrani, "Multi-resolution based algorithms for low bit-rate image coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Austin, TX, Nov. 1994, vol. II, pp. 285–289.

[36] J. M. Shapiro, "An embedded wavelet hierarchical image coder," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* San Francisco, CA, Mar. 1992, vol. IV, pp. 657–660.

[37] ——, "Application of the embedded wavelet hierarchical image coder to very low bit rate image coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Minneapolis, MN, Apr. 1993, vol. V, pp. 558–561.

[38] ——, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing,* vol. 41, pp. 3445–3462, Dec. 1993.

[39] I. Moccagatta and M. Kunt, "A pyramidal vector quantization approach to transform domain," in *Proc. European Signal Processing Conf. (EUSIPCO),* Brussels, Belgium, Aug. 1992, vol. III, pp. 1365–1368.

[40] ——, "VQ and cross-band prediction for color image coding," in *Proc. Picture Coding Symp. (PCS),* Sacramento, CA, Sept. 1994, pp. 383–386.

[41] O. Egger and W. Li, "Subband coding of images using asymmetrical filter banks," *IEEE Trans. Image Processing,* vol. 4, pp. 478–485, Apr. 1995.

[42] O. Egger, A. Nicoulin, and W. Li, "Embedded zerotree based image coding using linear and morphological filter banks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Detroit, MI, May 1995.

[43] H. Caglar, Y. Liu, and A. N. Akansu, "Optimal PR-QMF design for subband image coding," *J. Visual Commun. Image Representation,* vol. 4, no. 4, pp. 242–253, Sept. 1993.

[44] A. N. Akansu, R. A. Haddad, and H. Caglar, "The binomial QMF—Wavelet transform for multiresolution signal decomposition," *IEEE Trans. Signal Processing,* vol. 41, pp. 13–19, Jan. 1993.

[45] O. Egger and W. Li, "Very low bit rate image coding using morphological operators and adaptive decompositions," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Austin, TX, Nov. 1994, vol. II, pp. 326–330.

[46] O. Egger, W. Li, and M. Kunt, "High compression image coding using an adaptive morphological subband decomposition," *Proc. IEEE,* vol. 83, pp. 272–287, Feb. 1995.

[47] D. A. F. Florencio and R. W. Schafer, "A nonexpansive pyramidal morphological image coder," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Austin, TX, Nov. 1994, vol. II, pp. 331–335.

[48] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second-generation image coding techniques," *Proc. IEEE,* vol. 73, pp. 549–574, Apr. 1985.

[49] M. Kunt, M. Benard, and R. Leonardi, "Recent results in high-compression image coding," *IEEE Trans. Circuits Syst.,* vol. CAS-34, pp. 1306–1336, Nov. 1987.

[50] A. Ikonomopoulos and M. Kunt, "High compression image coding via directional filtering," *Signal Process.,* vol. 8, no. 2, pp. 179–203, Apr. 1985.

[51] M. Kocher and M. Kunt, "Image data compression by contour texture modeling," in *Proc. SPIE Int. Conf. Applications of Digital Image Processing,* Geneva, Switzerland, Apr. 1983, pp. 131–139.

[52] S. Carlsson, "Sketch-based coding of grey level images," *Signal Process.,* vol. 15, no. 1, pp. 57–83, July 1988.

[53] X. Ran and N. Farvardin, "Adaptive DCT image coding on a three-component image model," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP),* San Francisco, CA, Mar. 1992, vol. III, pp. 201–204.

[54] ——, "Low bit-rate image coding using a three-component image model," University of Maryland, College Park, Tech. Rep. TR 92-75, 1992.

[55] S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 14, pp. 710–732, July 1992.

[56] J. Froment and S. G. Mallat, "Second generation compact image coding with wavelets," in *Wavelets—A Tutorial in Theory and Applications,* C. Chui, Ed. New York: Academic, Jan. 1992.

[57] M. F. Barnsley, *Fractals Everywhere.* San Diego, CA: Academic, 1988.

[58] A. E. Jacquin, "Image coding based on a fractal theory of

iterated contractive image transformation," *IEEE Trans. Image Processing,* vol. 1, pp. 18–30, Jan. 1992.

[59] Y. Fischer, "A discussion of fractal image compression," in *Chaos and Fractals,* H. O. Peitgen, H. Jurgens, and D. Saupe, Eds. New York: Springer-Verlag, 1992.

[60] G. E. Oien, "L2-optimal attractor image coding with fast decoder convergence," Ph.D. dissertation, Institutt for Teleteknikk, Universitetet I Trondheim, Trondheim, Norway, 1993.

[61] S. Ono, N. Ohta, and T. Aoyama, *Super-High-Definition Images: Beyond HDTV.* Norwood, MA: Artech House, 1995.

[62] K. Barthel, T. Voy'e, and P. Noll, "Improved fractal image coding," in *Proc. Picture Coding Symp. (PCS),* Lausanne, Switzerland, Mar. 1993, no. 1.5.

[63] E. W. Jacobs, Y. Fisher, and R. D. Boss, "Image compression: A study of iterated transform method," *Signal Process.,* vol. 29, pp. 251–263, Dec. 1992.

[64] A. Baskurt and R. Goutte, "3-dimensional image compression by discrete cosine transform," in *Proc. European Signal Processing Conf. (EUSIPCO),* 1988, vol. IV, pp. 79–82.

[65] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Apr. 1988, vol. II, pp. 1100–1103.

[66] T. Ebrahimi and M. Kunt, "Image sequence coding using a three dimensional wavelet packet and adaptive selection," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Boston, MA, Nov. 1992, vol. 1818, pp. 222–232.

[67] E. Chang and A. Zakhor, "Scalable video coding using 3-D subband velocity coding and multirate quantization," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Apr. 1993, vol. V, pp. 574–577.

[68] H. G. Musmann, P. Pirsh, and H. Grallert, "Advances in picture coding," *Proc. IEEE,* vol. 73, pp. 523–548, Apr. 1985.

[69] F. Dufaux and F. Moscheni, "New perspectives in motion estimation techniques for digital TV," *Proc. IEEE,* vol. 83, pp. 858–876, June 1995.

[70] J. R. Ohm, "Temporal domain subband video coding with motion compensation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* San Francisco, CA, Mar. 1992, vol. III, pp. 229–232.

[71] ——, "Three dimensional SBC-VQ with motion compensation," in *Proc. Picture Coding Symposium (PCS),* Lausanne, Switzerland, Mar. 1993, no. 11.5.

[72] ——, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing,* vol. 3, pp. 559–571, Sept. 1994.

[73] M. Gilge, "A high quality videophone coder using hierarchical motion estimation and structure coding of the prediction error," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1988, vol. 1001, pp. 864–874.

[74] C. Stiller and D. Lappe, "Laplacian pyramid coding of prediction error images," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1991, vol. 1605, pp. 47–57.

[75] W. E. Lynch and A. Reibman, "The lapped orthogonal transform (LOT) for motion-compensated video compression," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1991, vol. 1605, pp. 285–296.

[76] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1991, vol. 1605, pp. 301–316.

[77] W. Li and F. Mateo, "Segmentation based coding of motion compensated prediction error images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Minneapolis, MN, Apr. 1993, vol. V, pp. 357–360.

[78] W. Li and M. Kunt, "Morphological segmentation applied to displaced frame difference coding," *Signal Process.,* vol. 38, no. 1, pp. 45–56, July 1994.

[79] K. N. Ngan and W. L. Chooi, "Very low bit rate video coding using 3D subband approach," *IEEE Trans. Circuits Syst.,* vol. 4, pp. 309–316, June 1994.

[80] C. Podilchuk, "Low bit-rate subband video coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Austin, TX, Nov. 1994, vol. III, pp. 280–284.

[81] P. H. Westerink, J. Biemond, and F. Muller, "Subband coding of image sequences at low bit rates," *Signal Process.,* vol. 2, no. 4, pp. 441–448, Dec. 1990.

[82] T. Ebrahimi and M. Kunt, "A video codec based on perceptually derived and localized wavelet transform for mobile applica-

tions," in *Proc. European Signal Processing Conf. (EUSIPCO),* 1992, pp. 1361–1364.

[83] D. Thoreau and J.-F. Vial, "Very low bit rate subband coding," in *Proc. Picture Coding Symposium (PCS),* Sacramento, CA, Sept. 1994, vol. 11-10, pp. 304–307.

[84] D. Qian, "Block-based motion compensated subband coding at very low bit rates using a psychovisual model," in *Proc. Very Low Bitrate Video Workshop (VLBV),* Essex, U.K., 1994, no. 6.8.

[85] K. Dachiku *et al.,* "Motion compensation subband extra/interpolative prediction coding at very low bit rate," in *Proc. Very Low Bitrate Video Workshop (VLBV),* Essex, U.K., 1994, no. 6.2.

[86] J. Katto, J. Ohki, S. Nogaki, and M. Ohta, "A wavelet codec with overlapped motion compensation for very low bit-rate environment," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, pp. 328–338, June 1994.

[87] P. Willemin, T. R. Reed, and M. Kunt, "Image sequence coding at very low bit rates with a 3-D split and merge algorithm," in *Proc. 2nd Int. Workshop 64 Kb/s coding of Moving Video,* Hannover, Germany, Sept. 1989.

[88] ——, "Image sequence coding by split and merge," *IEEE Trans. Commun.,* vol. 39, pp. 1845–1855, Dec. 1991.

[89] P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Processing,* vol. 3, pp. 639–651, Sept. 1994.

[90] P. Salembier, L. Torres, F. Meyer, and C. Gu, "Region-based coding using mathematical morphology," *Proc. IEEE,* vol. 83, pp. 843–857, June 1995.

[91] F. Bossen and T. Ebrahimi, "A simple and efficient binary shape coding technique based on bitmap representation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP).*

[92] M. Hotter, "Object-oriented analysis-synthesis coding based on moving two-dimensional objects," *Signal Process.,* vol. 2, no. 4, pp. 409–428, Dec. 1990.

[93] H. Jozawa and H. Watanabe, "Video coding using segment-based affine motion compensation," in *Proc. Picture Coding Symposium (PCS),* Sacramento, CA, Sept. 1994, pp. 238–241.

[94] A. Rosenfeld and A. C. Kak, *Digital Picture Processing,* vol. 2. New York: Academic, 1982.

[95] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comput.,* vol. EC-10, pp. 260–268, June 1961.

[96] C. R. Dyer, A. Rosenfeld, and H. Samet, "Region representation: Boundary codes from quadtrees," *Commun. ACM,* vol. 23, no. 3, pp. 171–179, Mar. 1980.

[97] H. Samet, "Region representation: Quadtrees from boundary codes," *Commun. ACM,* vol. 23, no. 3, pp. 163–170, Mar. 1980.

[98] H. Blum and R. N. Nagel, "Shape description using weighted symmetric axis features," *Pattern Recognition,* vol. 3, pp. 167–180, 1978.

[99] P. Brigger, S. Ayer, and M. Kunt, "Morphological shape representation of segmented images based on temporally modeled motion vectors," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Nov. 1994, vol. III, pp. 756–760.

[100] C. Gu and M. Kunt, "Contour simplifications and motion compensation for very low bit-rate video coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Nov. 1994, vol. II, pp. 423–427.

[101] Y. Yokoyama, Y. Miyamoto, and M. Ohta, "Very low bit-rate video coding with object-based motion compensation and othogonal transform," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1993, vol. 2049, pp. 12–23.

[102] C. Gu and M. Kunt, "Contour image sequence coding by motion compensation and morphological filters," in *Proc. Very Low Bitrate Video Workshop (VLBV),* Essex, U.K., 1994, no. 7.1.

[103] J. R. Casas and L. Torres, "Coding of details in very low bit-rate video systems," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, pp. 317–327, June 1994.

[104] P. Cicconi and H. Nicolas, "Efficient region-based motion estimation and symmetry oriented segmentation for image sequence coding," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, pp. 357–364, June 1994.

[105] V. Garcia-Garduno and C. Labit, "On the tracking of regions over time for very low bat rate image sequence coding," in *Proc. Picture Coding Symposium (PCS),* Sacramento, CA, Sept. 1994, pp. 257–260.

[106] P. Strobach, "Tree-structured scene adaptive coder," *IEEE*

Trans. Commun., vol. 38, pp. 477–486, Apr. 1990.

[107] L. Lu and W. A. Pearlman, "Multi-rate image sequence coding with quadtree segmentation and backward motion compensation," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1992, vol. 1818, pp. 606–614.

[108] K. Asai, Y. Yamada, and T. Murakami, "Video sequence coding based on segment-model and priority control," in *Proc. Picture Coding Symposium (PCS),* Sacramento, CA, Sept. 1994, pp. 325–328.

[109] T. Ebrahimi, "A new technique for motion field segmentation and coding for very low bitrate video coding applications," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Nov. 1994, vol. II, pp. 433–437.

[110] F. Lavagetto, F. Cocurullo, and S. Curinga, "Texture approximation through discrete legendre polynomials," in *Proc. Very Low Bitrate Video Workshop (VLBV),* Essex, U.K., 1994, no. 6.5.

[111] H. H. Chen, M. R. Civanlar, and B. G. Haskell, "A block transform coder for arbitrarily shaped image segments," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Nov. 1994, vol. I, pp. 85–89.

[112] M. Gilge, T. Engelhardt, and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform," *Signal Process.,* vol. 1, no. 2, pp. 153–180, Oct. 1989.

[113] M. Kocher and R. Leonardi, "Adaptive region growing technique using polynomial functions for image approximations," *Signal Process.,* vol. 11, no. 1, pp. 47–60, July 1986.

[114] E. Reusens, "Sequence coding based on the fractal theory of iterated transformation systems," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1993, vol. 1, pp. 132–140.

[115] M. P. Queluz and B. Macq, "A split-and-merge motion estimation/compensation technique for very-low bitrate image coding," in *Proc. Very Low Bitrate Video Workshop (VLBV),* Essex, U.K., 1994, no. 5.2.

[116] T. Chen, C. T. Swain, and B. G. Haskell, "An approach to region coding for content-based scalable video," in *Proc. IEEE Int. Conf. Image Processing (ICIP),* Lausanne, Switzerland, Sept. 1996, vol. III, pp. 399–402.

[117] K. Aizawa, H. Harashima, and T. Saito, "Model-based analysis synthesis image coding (MBASIC) system for a person's face," *Signal Process.,* vol. 1, no. 2, pp. 139–152, Oct. 1989.

[118] T. S. Huang, S. C. Reddy, and K. Aizawa, "Human facial motion modeling, analysis, and synthesis for video compression," in *Proc. SPIE Conf. Visual Communications and Image Processing,* Nov. 1991, vol. 1605, pp. 234–241.

[119] H. Li and R. Forchheimer, "Two-view facial movement estimation," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, June 1994.

[120] D. Pearson, "Developments in model-based video coding," *Proc. IEEE,* vol. 83, pp. 892–906, June 1995.

[121] P. K. Doenges, T. K. Capin, F. Lavagetto, J. Ostermann, I. S. Pandzic, and E. D. Petajan, "MPEG-4: Audio/video and synthetic graphics/audio for mixed media," *Signal Process.,* vol. 9, no. 4, pp. 433–464, May 1997.

[122] J. M. Beaumont, "Image data compression using fractal techniques," *Br. Telecommun. Technol. J.,* vol. 9, no. 4, pp. 93–109, Oct. 1991.

[123] H. Li, M. Novak, and R. Forchheimer, "Fractal-based image sequence compression scheme," *Opt. Eng.,* vol. 32, no. 7, pp. 1588–1595, July 1993.

[124] B. Hurtgen and P. Buttgen, "Fractal approach to low rate video coding," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* Nov. 1993, vol. 1, pp. 120–131.

[125] M. S. Lazar and L. T. Bruton, "Fractal-based image sequence compression scheme," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 4, pp. 297–308, June 1994.

[126] M. Mattavelli and A. Nicoulin, "Pre and post processing for very low bit-rate video coding," in *Proc. Int. Workshop HDTV,* Torino, Italy, Oct. 1994.

[127] M. I. Sezan, M. K. Ozkan, and S. V. Fogel, "Temporally adaptive filtering of noisy image sequences using a robust motion estimation algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* May 1991, vol. IV, pp. 2429–2432.

[128] Y. T. Tse and R. L. Baker, "Global zoom/pan estimation and compensation for video compression," in *Proc. IEEE Int. Conf.*

Acoustics, Speech, and Signal Processing (ICASSP), May 1991, vol. IV, pp. 2725–2728.

[129] M. Hoetter, "Differential estimation of the global motion parameters zoom and pan," *Signal Process.,* vol. 16, pp. 249–265, 1989.

[130] S. F. Wu and J. Kittler, "A differential method for simultaneous estimation of rotation, change of scale, and translation," *Signal Process.,* vol. 2, pp. 69–80, 1990.

[131] C. Horne, "Improving block based motion estimation by the use of global motion," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP),* 1993, vol. 2094, pp. 576–587.

[132] H. C. Reeve and J. S. Lim, "Reduction of blocking effects in image coding," *Opt. Eng.,* vol. 23, no. 1, pp. 34–37, 1984.

[133] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol. 34, pp. 1258–1268, Oct. 1986.

[134] R. L. Stevenson, "Reduction of coding artifacts in transform image coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Apr. 1993, vol. V, pp. 401–404.

[135] L. Yan, "A nonlinear algorithm for enhancing low bit-rate coded motion video sequence," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Nov. 1994, vol. II, pp. 923–927.

[136] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Iterative projection algorithms for removing the blocking artifacts of block-DCT compressed images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Apr. 1993, vol. V, pp. 401–408.

[137] B. Macq, M. Mattavelli, O. Van Calster, E. van der Plancke, S. Comes, and W. Li, "Image visual quality restoration by cancellation of the unmasked noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* 1994, vol. V, pp. 53–56.

[138] T. Chen, "Elimination of subband-coding artifacts using the dithering technique," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Nov. 1994, vol. II, pp. 874–877.

[139] W. Li, O. Egger, and M. Kunt, "Efficient quantization noise reduction device for subband image coding schemes," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP),* Detroit, MI, May 1995, vol. IV, pp. 2209–2212.

[140] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Norwell, MA: Kluwer, 1992.

[141] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, pp. 84–95, Jan. 1980.

[142] R. M. Gray, "Vector quantization," *IEEE Acoust., Speech, Signal Processing Mag.,* vol. 1, pp. 4–29, Apr. 1984.

[143] H. Abut, *Vector Quantization.* New York: IEEE Press, 1990.

[144] A. N. Netravali and B. G. Haskell, *Digital Pictures Representation, Compression, and Standards,* 2nd ed. New York: Plenum, 1995.

[145] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol. 27, pp. 379–423, 1948.

[146] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Inform. Theory,* vol. IT-18, pp. 759–765, Nov. 1972.

[147] I. Moccagatta and M. Kunt, "Lattice vector quantization approach to image coding," in *Proc. SPIE Visual Communications and Image Processing '92,* Boston, MA, 1992.

[148] T. Chen, H. P. Graf, H. Chen, W. Chou, B. G. Haskell, E. Petajan, and Y. Wang, "Lip synchronization in talking head video utilizing speech information," in *Proc. SPIE Conf. Visual Communication and Image Processing,* Taipei, Taiwan, May 1995.

[149] R. Koenen, "MPEG-7: Context and objectives version 5.0," ISO/IEC JTC1 SC29/WG11 N1920, Fribourg, Switzerland, Oct. 1997.

[150] M. R. Banham and J. C. Brailean, "A selective update approach to matching pursuits video coding," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 119–129, Feb. 1997.

[151] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 158–171, Feb. 1997.

[152] M.-C. Lee, W. Chen, C. B. Lin, C. Gu, T. Markoc, S. I. Zabrinsky, and R. Szelinski, "A layered video object coding system using sprite and affine motion model," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 130–145, Feb. 1997.

[153] E. Reusens, T. Ebrahimi, C. Le Buhan, R. Castagno, V.

Vaerman, L. Piron, C. de Sola, S. Bhattarjee, F. Bossen, and M. Kunt, "Dynamic approach to visual data compression," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 197–211, Feb. 1997.

Touradj Ebrahimi (Member, IEEE) was born on July 30, 1965. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 1989 and 1992, respectively.

From 1989 to 1992, he was a Research Assistant with the Signal Processing Laboratory of the Swiss Federal Institute of Technology (EPFL). During summer 1990, he was a Visiting Researcher with the Signal and Image Processing Institute of the University of Southern California, Los Angeles. In 1993, he was a Research Engineer with the Corporate Research Laboratories of Sony Corporation, Tokyo, Japan, where he conducted research on advanced video-compression techniques for storage applications. In 1994, he was a Research Consultant with AT&T Bell Laboratories working on very-low-bit-rate video coding. He currently is with the Signal Processing Laboratory of EPFL, where he is involved with various aspects of digital video and multimedia applications and is in charge of the Digital TV group. His research interests are multidimensional signal processing, image processing, information theory, and coding. He is the author or coauthor of more than 70 research publications and six patents.

Dr. Ebrahimi is a member of the European Association for Signal Processing. He received the IEEE and Swiss National ASE Award in 1989.

Murat Kunt (Fellow, IEEE) was born in Ankara, Turkey, on January 16, 1945. He received the M.S. degree in physics and the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1969 and 1974, respectively.

From 1974 to 1976, he was a Visiting Scientist with the Research Laboratory of Electronics of the Massachusetts Institute of Technology, Cambridge, where he developed compression techniques for X-ray images and electronic image files. In 1976, he returned to the EPFL, where presently he is a Professor of electrical engineering and Director of the Signal Processing Laboratory, one of the largest laboratories at EPFL. He conducts teaching and research in digital signal and image processing with applications to modeling, coding, pattern recognition, scene analysis, industrial developments, and biomedical engineering. He is the author or coauthor of more than 200 research papers and 14 books. He has received three patents. He is the Editor-in-Chief of *Signal Processing*. He was President of the Swiss Association for Pattern Recognition from its creation until 1997. He consults for governmental offices, including the French General Assembly.

Dr. Kunt is a founding member of the European Association for Signal Processing (EURASIP). He received the gold medal from EURASIP for meritorious services in 1983 and the IEEE ASSP Technical Achievement Award in 1997.