Content-Based Indexing for Search and Browsing

Storage and archiving of digital video in shared disks and servers in large volumes, browsing of such databases in real-time, and retrieval over switched and packet networks pose many new challenges, one of which is efficient and effective description of content. The simplest method to index content is by means of a thesaurus of keywords, which can be assigned manually or semiautomatically to programs, shots, or visual objects [232]. It is also desirable to supplement these keywords with visual features describing appearance (color, texture, and shape) and action (object and camera motion), as well as sound (audio and speech) and textual (script and close-caption) features [227, 233]. Furthermore, it is of interest to browse and search for content using compressed data since almost all video data will likely be stored in compressed format [234].

Video-indexing systems may employ a frame-based, scene-based, or object-based video representation [235]. The basic components of a video-indexing system are temporal segmentation, analysis of indexing features, and visual summarization. The temporal segmentation step extracts shots, scenes, and/or video objects. The analysis step computes content-based indexing features for the extracted shots, scenes, or objects. Content-based features may be generic or domain-dependent. Commonly used generic indexing features include color histograms, type of camera motion [236], direction and magnitude of dominant object motion, entry and exit instances of objects of interest [237], and shape features for objects. Domain-dependent feature extraction requires a priori knowledge about the video source, such as news programs, particular sitcoms, sportscasts, and particular movies. Content-based browsing can be facilitated by a visual summary of the contents of a program, much like a visual table of contents. Among the proposed visual summarization methods are story boards, visual posters, and mosaic-based summaries. With the upcoming MPEG-7 standardization effort on content-based video description, this subject is sure to remain an active research topic in the near future.

Image and Video Coding

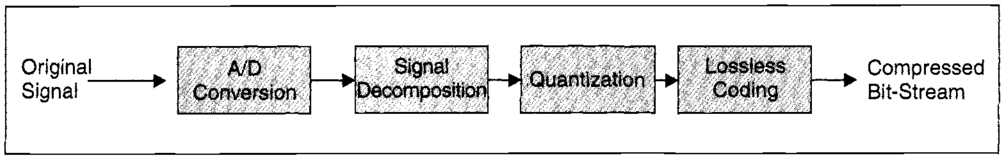
Bernd Girod, University of Erlangen-Nuremberg; Robert Gray, Stanford University; Jelena Kovacevic, Lucent Technologies; and Martin Vetterli, Swiss Federal Institute of Technology and University of California, Berkeley

Due to the vast amount of data associated with images and video, compression is a key technology for their digital transmission and storage. The availability and demand for images and video continue to outpace increases in network capacity. Hence, the importance of compression is not likely to diminish, in spite of the promises of unlimited bandwidth.

Compression takes advantage of the structure of images and video, especially of the statistical redundancy inherent in such signals. It can also exploit the limitations of human visual perception to omit components of the signal that will not be noticed.

A typical compression system has several stages as depicted in Fig. 8. The analog-to-digital converter samples and finely quantizes an image, producing a digital representation. A signal decomposition uses linear transforms or filter banks to break the signal into parallel channels of separate images (bands or groups of coefficients). Such decompositions serve to compact the energy into a few coefficients, to decorrelate separate subbands or groups of coefficients, and to convert the images into a form where perceptual processing is naturally incorporated. Most of the compression occurs in the quantization stage, which operates in the transform domain on individual pixels (scalar quantization) or on groups of pixels (vector quantization). Lossless compression (entropy coding) typically involves run-length coding combined with Huffman or arithmetic codes to further save bits in an invertible fashion. Occasionally specific components will be combined or omitted, but most current image-coding standards possess all of the components in some form.

The enormous commercial potential of image and video coding has stimulated rapid growth of the research efforts to find improved or entirely new techniques. In this short survey we do not attempt a comprehensive overview; rather we selectively summarize some of the most common themes and principles of this exciting field. We start by reviewing source-coding principles, which are not specific to image coding, but are fundamentally important nevertheless. We then discuss subband and transform coding, move on to predictive coding, motion compensation and rate-distortion methods in compression systems, and close with a discussion of image communication systems issues.



▲ *8. A typical compression system. The analog-to-digital converter samples and finely quantizes an image, producing a digital representation. A signal decomposition uses linear transforms or filter banks to break this digital representation into parallel channels of separate images. Most of the compression occurs in the quantization stage. Lossless compression (entropy coding) typically involves run-length coding combined with Huffman or arithmetic codes to further save bits in an invertible fashion.*

Source Coding Principles

In an ideal image coder (or any source coder) one would like high fidelity of the reconstructed image at the receiver in combination with low bit-rate and low complexity of the coder and the decoder. Alas, these requirements are at odds with each other, and good design is a question of finding optimal trade-offs. Source-coding theory focuses on the basic trade-off between average rate, the bit-rate in terms of bits per pixel, and average distortion, measured commonly by mean-squared error. Although often maligned, weighted versions of squared error (especially when applied to transformed signals and allowed to depend on the input) have proved quite useful as indications of perceived quality. Complexity considerations may enter through structural constraints on the specific types of codes considered.

The optimal trade-off between rate and distortion can be precisely defined in several equivalent ways: by minimizing the (average) distortion, D, subject to a constraint on the (average) rate, R, by minimizing the rate subject to a constraint on the distortion, or by an unconstrained minimization of the Lagrangian cost function, $D + \lambda R$, where the Lagrange multiplier, λ, provides a weighting of the relative importance of distortion and bit-rate.

The theory of data compression has two principal branches, both of which are celebrating their 50th birthday: Shannon's rate-distortion theory, a branch of information theory sketched in his 1948 paper [238] and developed in remarkably complete form in his 1959 paper [239], and high-rate or high-resolution quantization theory, an approach involving approximations for low distortion and high bit-rate that began with the classical work on PCM (pulse coded modulation) of Oliver, Pierce, and Shannon [240] and the work on quantization error spectra by Benett [241]. Rate-distortion theory [242- 245] provides unbeatable lower bounds to the obtainable average distortion for a fixed average rate, or vice versa. It also promises that codes exist that approach these bounds when the code dimension and delay become large. High-rate quantization theory [246, 247] provides approximations for distortion and rate that can be optimized for fixed and finite dimension and delay. These results imply many useful comparisons of the gains achievable by transform coding and other structured codes in comparison with the Shannon optima.

Unfortunately, the theory does not provide us with an explicit method for constructing a practical optimum coder and decoder. It can nevertheless give very important hints about the properties of an optimum coder/decoder. For example, for wide-sense stationary Gaussian sources with memory and mean-squared error distortion, the mathematical form of the rate-distortion function suggests that an optimum coder splits the original signal into spectral components of infinitesimal bandwidth and encodes these spectral components in an independent manner [248]. This is consistent with the high-rate quantization theory, which demonstrates that the decorrelating KLT is optimal for transform coding of Gaussian sources. The corresponding bit allocation for the separate subband quantizers should make the rate proportional to the logarithm of energy in each subband. For high bit-rate, uniform scalar quantization coupled with high-order or adaptive entropy coding can at best achieve a performance 0.255 bits or 1.5 dB away from the Shannon bound. The gap can be closed further if vector quantization or trellis-coded quantization is used [249, 250]. These theoretical insights have motivated the widespread use of transform and subband coders, even when the rates are not high and the images are certainly not Gaussian.

Transform and Subband Coding

Transform coding and subband coding (SBC) refer to compression systems where the signal decomposition (Fig. 8) is implemented using an analysis filter bank. At the receiver, the signal is reassembled by a synthesis filter bank. By transform coding, we usually mean that the linear transform is block-based (such as a block-wise DCT in JPEG). When transform coding is interpreted as an SBC technique, the impulse responses of the analysis and synthesis filters are at most as long as the subsampling factor employed in the subbands; thus, the image can be subdivided into blocks that are processed in an independent manner. General SBC, on the other hand, allows the impulse responses to overlap and thus includes transform coding as a special case.

As pointed out earlier in this article, one of the most important tasks of the transform is to pack the energy of the signal into as few transform coefficients as possible. The DCT yields nearly optimal energy concentration for images, while being a lot easier to implement than the KLT, which is the theoretically best orthonormal energy-packing transform. As a result, almost all image transform coders today employ the block-wise DCT, usually with a block size of 8 x 8 pixels. The transform is followed by quantization (most often scalar uniform quantization) and entropy coding. Typically, the transform coefficients are run-level encoded; that is, successive zeros along a zig-zag path are grouped with the first nonzero amplitude into a joint symbol which is then Huffman coded. For example, the widely used JPEG standard works in this fashion, and so do the prediction error encoders for MPEG, H.261, and H.263. The LOT could be substituted for the DCT in the above process to avoid some of the typical blocking artifacts that become visible with coarse quantization. Figures 9(b)-(c) show the DCT decomposition of the *Barbara* image in Fig. 9(a) as well as the JPEG coding result at 0.5 bits/pixel.

The full potential of SBC is unleashed when nonuniform bandsplitting is used to build multiresolution representations of an image. Beside excellent compression, multiresolution coders provide the successive approximation feature;

Visual Coding Standards: A New World of Visual Communication

Gary J. Sullivan, PictureTel Corporation

In the past few years, the advent of international standards for digital image and video coding has given the world a vast new array of capabilities for visual communication. Widespread communication is impossible without the use of a common language, and a visual-coding "standard" is a specialized language that defines how to interpret the digital representations of pictures.

The compressed coding of pictures into digital form was primarily the domain of cloistered research laboratories just a few years ago. Today, millions of people worldwide watch television pictures that have been digitally transmitted, view and create coded pictures and video on the Web, view coded video on their personal computers, and even use digital videotelephony for interactive remote conversations. The key to making this transition from the research lab to the casual user has been the creation of standards for visual communication.

These standardization activities have provided great opportunities for researchers to directly influence the world of the future. The standardization groups meet to closely examine the capabilities, performance, and practicality of the various concepts, features, and designs that are brought forth from the research community. They then collaborate to merge the best of these ideas into a coherent and fully defined specification that all can use. Repeatedly, the final written standard has become a better overall technical solution than any of the individual proposals that were brought into the collaborative process.

The standards have themselves become touchstones for new creative research, since they provide a well-known reference for comparison. The creation and promulgation of a standard organizes the collective thoughts of the technical community, creating a breadth of understanding and experience that could not have been achieved by research alone.

The biggest names in the realm of standardization of visual information coding are the ITU-T and the ISO/IEC JTC1 organizations. The ITU-T (formerly called the CCITT) approved the first digital video-coding standard (Rec. H.120) in 1984, and has been updating its methods periodically since then by revising H.120 in 1988, then moving on to increasingly successful standards in 1990 (Rec. H.261) and 1995 (Rec. H.263), and enhancing its latest standard this year (Rec. H.263+). In 1993, the ISO/IEC JTC1 completed the MPEG-1 video-coding standard (IS 11172-2) and joined with the ITU-T to develop the JPEG standard for still pictures in 1994 (IS 10918-1) and the MPEG-2 standard for video in 1996 (IS 13818-2). Each of these standards has led to increasing growth in the use and variety of applications for digital coded visual information, and both groups are working on new efforts for the future (such as the MPEG-4 project in ISO/IEC JTC1 and the H.263++ and H.26L projects in the ITU-T).

Today's standards development process has become very responsive to the progress of research, and the research world has been helped by the progress of standardization. This symbiotic relationship will continue into the future, providing a fertile field for new technology development.

as higher-frequency components are added, higher-resolution, better-quality images are obtained. Moreover, multiresolution techniques fit naturally into joint source-channel coding schemes. Figures 9(d)-(e) show the uniform subband decomposition of the *Barbara* image as well as the SBC coding result at 0.5 bits/pixel, while Figures 9(f)-(g) show the octave-band subband decomposition and the coding result at 0.5 bits/pixel. Subband coders with octave band decomposition such as illustrated in Figures 9(f)-(g) are also often referred to as discrete wavelet transform (DWT) coders, or wavelet coders.

The multiresolution image representation in Figures 9(f)-(g) is a critically sampled subband pyramid. However, overcomplete representations, first introduced as the Laplacian pyramid by Burt and Adelson [251], are also very powerful. An input image is fed into a lowpass filter followed by a downsampler to produce a coarse approximation that is then used to interpolate the original (by upsampling and filtering) and calculate the difference as the interpolation error. This process can be recursively applied to the coarse version. Thus, instead of compressing the original image one compresses the coarse version and the interpolation errors at various resolutions. The interpolation can be based on lower-resolution images with or without quantization error (referred to as open-loop and closed-loop pyramid coders). The overcomplete pyramid provides energy concentration and possesses the successive approximation property, since one can start with the coarsest version and then add detail (interpolation errors) to reconstruct higher-resolution versions. Moreover, the pyramid coding scheme allows for nonlinear operations for producing the coarse version and the details. Its only disadvantage is that it produces a redundant representation.

Today, many state-of-the-art multiresolution image coders draw on the ideas introduced by Shapiro in his embedded zero-tree wavelet algorithm (EZW) [252]. The algorithm employs a data structure called zero-tree, where one assumes that if a coefficient at a low frequency is zero, it is highly likely that all the coefficients at the same spatial location at all higher frequencies will also be zero; thus, when encountering a zero-tree root, one can discard the whole tree of coefficients in higher-frequency bands. Moreover, the algorithm uses successive approximation quantization, which allows termination of encoding or decoding at any point. These initial ideas have produced a new class of algorithms aimed at exploiting both frequency and spatial phenomena [253].

While research has shown that wavelet coders can produce superior results, transform coders employing a

Visual Coding Standards: A Research Community's Midlife Crisis?

Michael Orchard, Princeton University

Yes, standards have turned JPEG and MPEG into household terms and brought digital images and video into millions of homes worldwide. But what have they done for us visual compression researchers, the community responsible for developing the algorithms? Now that our best ideas have been perfected, packaged, and polished for public dissemination, what remains for researchers in this field to do? Should research continue on algorithms whose chances of ever becoming a standard might be questionable?

The visual-compression research community has wrestled with these kinds of questions over the past five years, and the answers that have been offered have reshaped the field. Widespread acceptance of visual coding standards have forced us to reassess directions and priorities, but overall, it is clear that standards open more doors than they close, and standards cannot alter the nature or diminish the importance of truly fundamental research advances in the field.

By accelerating the development of visual applications, standards have helped uncover challenging new problems offering exciting opportunities for the research community. Robust transmission of images and video over packet networks and video transmission over wireless channels have become hot research topics. Digital video libraries, content-based retrieval, and digital watermarking are examples of active new research areas spawned by the widespread application of coding standards and involving problems of visual representation that are closely related to the coding problem.

Balanced against their positive effects, standards have also had the unfortunate effect of diverting attention from important fundamental questions in image and video compression. The success of standards has suggested that they are based on sound technical approaches to the coding problem and has focused the community's attention on the refinement of those approaches for improved performance. In fact, today's standards are built on ad-hoc frameworks that reflect our very limited understanding of the fundamental structure of image and video sources. There is very little reason to believe either that today's standards come close to the ideal performance possible for these sources (that is, it is unlikely that they are near the fundamental entropy of these sources), or that there cannot exist simple, practical coding algorithms performing much better than today's standards. In particular, the standard hybrid framework for motion-compensated video coding is based on a naive understanding of the relationship between motion and intensity uncertainty models for video sequences.

The gaps in our understanding are wide, and progress in bridging those gaps requires continued strong research efforts by the community. Unfortunately, the fundamental advances that are needed are not likely to produce immediate practical algorithms to challenge today's standards, and this has discouraged research in these directions. It is particularly important that young researchers entering the field be encouraged to apply their creativity and healthy skepticism toward challenging accepted frameworks, engaging basic issues, and proposing sound alternative approaches, no matter how far-fetched they may appear. In the long term, these efforts promise progress on important fundamental questions, a more vibrant research community AND superior standards.

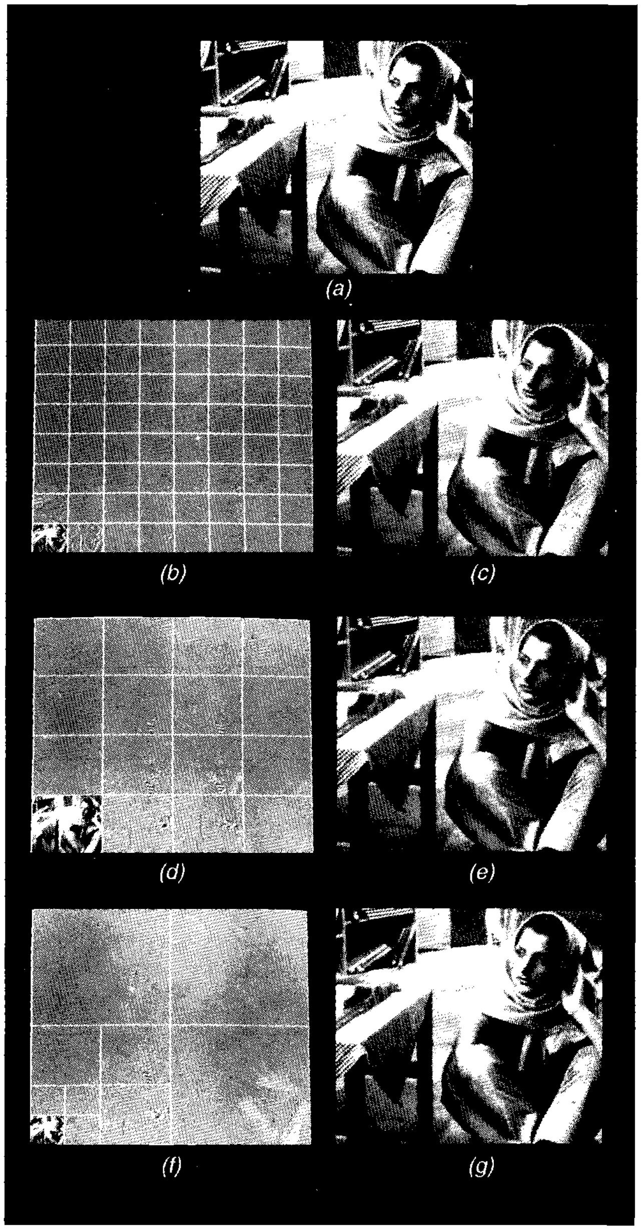
block-wise DCT are still dominant today. After years of use, DCT coders are very well understood and many improvements have been made, for example in the area of fast algorithms or by imposing perceptual criteria. The next still-image coding standard, JPEG 2000, as well as the next in the line of MPEG standards, MPEG-4, might very well include wavelet coding, in addition to or in place of the DCT.

Predictive Coding

Except when used with subband or transform coding, predictive coders do not decompose the image into independent components. Instead, both the coder and the decoder calculate a prediction value for the current signal sample. Then, the prediction error, rather than the signal sample itself, is transmitted. This principle can be used for both lossy and lossless image coding. Most commonly, the predictors calculate linear combinations of previous image samples, since general nonlinear predictors, addressed by combinations of, say, 8-bit pixels, would often require enormous look-up tables for the same performance.

For lossy predictive coding, differential code pulse modulation (DPCM), invented by Cutler in 1952, has been used since the early days of image coding. Intraframe DPCM exploits dependencies within a frame of a video sequence. Typically, pixels are encoded in line-scan order and the previous sample in the current line and samples from the previous line are combined for prediction. Today, this simple scheme has been displaced by vastly superior transform SBC schemes, without doubt a result of the unnatural causal half-plane constraint for the region of support of the predictor. In fact, lossy predictive intraframe coding is alive and well in the form of predictive closed-loop pyramid coders that feed back the quantization error before encoding the next higher-resolution layer (see the section on Image Transforms). It has been shown recently that closed-loop pyramid coders even outperform the equivalent open-loop overcomplete pyramid representations when combined with scalar quantizers [254].

For interframe coding where statistical dependencies between successive frames of a video sequence are exploited, DPCM is the dominating scheme today and for the foreseeable future. Other than, for example, spatio-temporal SBC, interframe DPCM avoids the undesirable delay due to buffering of one or several frames. Moreover,



▲ 9. *Subband image coding results: (a)* Barbara *image of size 512 x 512 pixels with 8 bits/pixel; (b) 8 x 8 DCT transform of* Barbara *used in JPEG; (c) JPEG-coded* Barbara *with 0.5 bits/pixel and 28.26 dB SNR; (d) Uniform subband decomposition of* Barbara; *(e) SBC coded* Barbara *using uniform subband decomposition at 0.5 bits/pixel with 30.38 dB SNR; (f) Octave-band subband/wavelet decomposition of* Barbara; *(g) SBC coded* Barbara *using octave-band subband/wavelet decomposition at 0.5 bits/pixel with 29.21 dB SNR.*

it is straightforward to incorporate motion adaptation and motion compensation into a temporal prediction loop and combine motion-compensated prediction with other schemes for encoding of the prediction error.

Motion-Compensated Video Coding

All modern video-compression coders such as those standardized in the ITU-T Rec. H.261 [255] and H.263 [256], or in the ISO MPEG standards [257], are motion-compensated hybrid coders. Motion-compensated hybrid coders estimate the displacement from frame to frame and transmit the motion vector field as side information in addition to the motion-compensated prediction error image. The prediction error image is encoded with an intraframe source encoder that exploits statistical dependencies between adjacent samples. This intraframe encoder is an 8 x 8 DCT coder in all current video-coding standards [255-257], but other schemes, such as subband coders or vector quantizers, can be used as well.

Motion-compensated hybrid coding can theoretically outperform an optimum intraframe coder by at most 0.8 bits/pixel in moving areas of an image, if motion compensation is performed with only integer-pixel accuracy [258]. For half-pixel accuracy, this gain can be up to 1.3 bits/pixel. In addition, in nonmoving areas (or other parts of the image that can be predicted perfectly) no prediction error signal has to be transmitted and these areas can simply be repeated from a frame store, a technique often referred to as conditional replenishment.

Motion compensation works well for low spatial frequency components in the video signal; for high spatial frequency components even a small inaccuracy of the motion compensation will render the prediction ineffective. Hence, it is important to spatially lowpass filter the prediction signal by a loop filter. This loop filter is explicitly needed for integer-pixel accurate motion compensation. For subpixel accurate motion compensation, it can be incorporated into the interpolation kernel required to calculate signal samples between the original sampling positions. The loop filter also improves prediction by acting as a noise-reduction filter. Prediction can be further improved by combining multiple independently motion-compensated prediction signals. Examples are the bidirectionally predicted B-frames in MPEG [257] or overlapped block motion compensation [259] that has also been incorporated in the ITU-T Rec. H.263 [256].

Especially at low bit-rates, motion compensation is severely constrained by the limited bit-rate available to transmit the motion vector field as side information. Rate-constrained estimation [260] and a rate-efficient representation of the motion vector field are therefore very important. For simplicity, most practical video-coding schemes today still employ block-wise constant motion compensation. More advanced schemes interpolate between motion vectors, employ arbitrarily shaped regions, or use triangular meshes for representing a smooth motion vector field. Ultimately, we might expect 3-D models to be incorporated into motion compensation—one day, we hope with such success that transmission of the prediction error is no longer required. This is a goal of ongoing research into model-based video coding, although the success of such schemes for general types of video material is still uncertain.

Rate-Distortion Methods in Compression Systems

As outlined earlier, the formal framework for compression methods consists of rate distortion theory and high-rate quantization theory. How can this theory be applied to practical compression schemes? Recent work has made progress in this direction by bridging at least in part the gap between theory and practice in source coding. The idea is to use standard optimization procedures such as Lagrangian methods to find local optimal operating points in a rate-distortion sense, under some assumptions about the source. Such techniques were first introduced in the context of PCM by Lloyd in the 1950s, and were later used for vector quantization designs (for example, [261]) and other problems involving transforms and quantization. As an example, consider the problem of finding best orthonormal bases for compression from a large collection of possible transforms. This can be posed as a Lagrangian optimization problem: each set of transform coefficient generates an operational rate-distortion curve, and optimal allocation of bit-rate between transform coefficients is standard. Among all possible transforms, Lagrange optimization allows one to choose the winning transform, and this can be done in an efficient tree-pruning manner if the transforms have some structure [262].

Similar ideas can be used for many of the other problems appearing in practical compression schemes. As examples, we can cite rate control for video coders using dynamic programming [263], allocation of rate between competing units (for example, motion and residual), and optimization of quantization schemes. The important point is that, under certain assumptions such as independence, an optimal or locally optimal solution is sought, as opposed to the somewhat ad-hoc methods that are often used in practical compression schemes.

Image Communication System Issues

Image and video compression is usually not done in isolation, but integrated in to a larger system, typically a communication system. This poses some interesting challenges to the designer of the compression system.

In his groundbreaking 1948 paper [238] that laid the foundations of information theory, Shannon showed that for point-to-point communication over a well-defined channel, a separate optimization of source coding and channel coding can lead to a performance arbitrarily close to the information-theoretic bound for the entire system. Therefore, traditionally, the coding problem has been split into two independent subproblems: source compression and channel coding [264]. This has resulted in largely independent research in the two areas. However, many practical situations do not satisfy the assumption required for the separation principle to hold. For example, most communication is done with a finite-delay constraint, which leads to finite block sizes. Under such delay constraints, error probabilities can be non-negligible, and thus the separation principle has to be revisited.

In such cases, practical schemes using unequal error protection of different parts of the coded stream are used. For example, in video coding for noisy channels, the motion vectors are highly protected, since their loss would be catastrophic. Motion residuals (errors after motion compensation) are either not protected, or much less protected than the motion vectors. In multiresolution source coders, it is natural to protect the coarse resolutions (which are absolutely necessary) more than the fine details; this is another instance of unequal error protection. Note that multiresolution video coding requires multiscale motion compensation [265]. An interesting application of multiresolution coding is the progressive transmission of images, in particular in browsing applications. Instead of coding at a fixed rate, one has to accommodate many rates, depending on the resolution desired by the particular user accessing the image.

Given a source and a channel, how do we allocate resources to source and channel coding (still under the finite-delay constraint)? The answer turns out to be more complicated than expected and is only partly known [266].The question becomes even more intricate when protocol issues are included (as in channels with feedback [267]).

All methods that allow some interaction of the source and the channel coders go under the name of a joint source/channel coding system. They do not fit the classic separation principle; rather, they solve the practical problem of robust communication when errors do occur. An instance where the separation principle cannot be used is in the case of multiple channels, as in broadcast or multicast scenarios. Then, users with very different channels have to be accommodated, which leads to schemes that adapt to the particular channels they face. For example, embedded modulation together with multiresolution source coding leads to a robust scheme with graceful degradation when the channel degrades [268]. Similar ideas can be adapted to multicast over packet networks [269]. Finally, recent work on multiple description coding addresses the question of transmitting several source descriptions over multiple channels. For example, two descriptions of a source are sent to three receivers, where the first two receive either description, while the third receives both. This interesting theoretical question is relevant to transmission over lossy or delay-constrained packet networks, where random drops may occur. Recently, Vaishampayan derived quantization schemes for this problem [270].

Finally, let us stress the importance of protocol issues. If an error occurs, it can have catastrophic consequences (for example, loss of synchronization in a variable-length lossless code). Therefore, there exists a need for a powerful mechanism to recover from errors, using feedback channels (in point-to-point communications) or resynchronization points (in multicast situations). In a practical image/video communication system, one can do

Exploiting Self-Similarities: Fractal Image Compression

Geoff Davis, Dartmouth College

The prototypical fractal coder, first proposed by Barnsley and Jacquin [271], encodes images via a process resembling vector quantization. The twist is that fractal coders employ a vector codebook derived from the image being coded rather than using a prespecified codebook. The result is a coded image representation very different from that for transform coders: the stored data defines a contraction map of which the original image is an approximate fixed point. Images are decoded by iterating this stored map to convergence.

Transform coders take advantage of spatial redundancy in images. Fractal coders take a somewhat different approach: they exploit redundancy in *scale*. Common image features such as edges and linear gradients are self-similar, in the sense that they are invariant under contractions up to an affine transform. This self-similarity of key image features motivates the particular codebook used by fractal coders—a set of affine transforms of contracted image blocks. One intriguing property of fractal coders is that this self-similarity property can be used to synthesize finer detail, particularly at edges, than was present in the original image.

A major drawback of fractal coding is the high complexity of the encoding process, and considerable effort has been devoted to finding efficient encoding algorithms. Additional important research areas have included bounding reconstruction errors and determining conditions under which the iterative decoding algorithm converges.

Although the mechanics of fractal coding are quite different from transform coders, fractal coders have recently been shown to be closely related to wavelet coders [272]. The link is a natural one, since wavelet bases possess a dyadic self-similar structure similar to that found in fractal coders. This wavelet/fractal synergy provides important insights into the workings of fractal coders. The new understanding has improved the performance of fractal coders considerably, and it has revealed some basic limitations of current coders that will require further research to overcome.

Model-Based Video Coding

Don Pearson, University of Essex, UK

Mention *model-based coding* (MBC) to someone and they tend to think of animated texture-mapped wire-frame heads, some of them looking distinctly zombie-like! This is indeed the way MBC began back in the early 1980s; it is also where many young people start their research today. But it is not necessarily the way it will end; coding methods invented for one purpose sometimes find their application elsewhere. An example is run-length coding, which was first investigated for gray-scale picture compression before finding its home in facsimile.

What will determine MBC's ultimate fate is its coding efficiency. This depends on the picture material, as it does with all image-coding methods; no method works well for all types of objects and all types of movement. We may have many different options within a coder, each suited to a particular type of visual material, and that we (or rather the coder) will choose or switch between. Experiments alone will show where MBC fits in. Those conducted so far tend to indicate that the method works best for large, relatively rigid moving areas in translational or rotational movement. This is not surprising when we think about it, since shape information has to be added to that for motion and texture. The additional overhead must save texture bits to be worth sending.

With increasing levels of sophistication in facial analysis and modeling, it is quite likely that the traditional approach to MBC will eventually yield highly efficient and believable talking heads in low bit-rate applications. But it is also possible that MBC will be found to be useful on a selective basis for coding large nonfacial moving objects in higher-resolution, higher bit-rate video. MBC is theoretically efficient for such objects, and they are the very objects that cause difficulties in the current generation of MPEG-2 coders [273, 274].

better by jointly designing source coder, channel coder, and transmission protocols. Research that addresses these issues is still in its infancy.

Better Compression Forever?

Students of image coding often ask how many bits at least are required to represent an image or a motion video with reasonable quality. They ask this question not only out of scientific curiosity, but they also want to find out whether research in the field has a future, or whether all the interesting problems have already been solved.

As illustrated by the success of image and video coding standards (see the "Visual Coding Standards: A New World of Communication" and "Visual Coding Standards: A Research Community's Midlife Crisis" sidebars), image and video coding is a mature discipline today. It rests solidly on the foundations of source coding theory. Often, practical schemes perform close to their information-theoretic bounds. Note, however, that most of these bounds are calculated on the basis of crude models about the structure of images. As image models become more refined, compression ratios can improve further. Moreover, many interesting open problems are yet to be solved on how to gracefully integrate image and video codecs into communication systems, where previously neglected requirements, such as robustness, delay, or random access, have to be taken into account. We believe that image and video coding will remain a quick-paced, exciting field well into the next millennium.

Image-Processing Software and Hardware

Ed J. Delp, Purdue University

The hardware and software tools available to acquire and process digital images have changed a great deal in 50

210. K. Aizawa and T.S. Huang, "Model-based image coding," *Proc. IEEE*, 83, pp. 259-271, 1995.

211. N. Jojic, T.S. Huang, "On analysis of cloth drape range data," *Proc. Asian Conf. Computer Vision*, 2, pp. 463-470, 1998.

212. C. Shu and R. Jain, "Vector field analysis for oriented patterns," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 16, pp. 946-950, 1994.

213. J.Y. Aloimonos, Z. Duric, C. Fermuller, L. Huang, E. Rivlin, and R. Sharma, "Behavioral Visual Motion Analysis," in *Proc. DARPA Image Understanding Workshop*, San Diego, CA, pp. 521-533, 1992.

214. M. Tistarelli and G. Sandini, "Direct Estimation of Time-to-impact from Optical Flow," in *Proc. IEEE Workshop on Visual Motion Princeton*, NJ, pp. 226-233, 1991.

215. R. Cipolla and A. Blake, "Surface Orientation and Time to Contact from Image Divergence and Deformation," in *Proc. European Conf. on Computer Vision*, Genoa, Italy, pp. 189-202, 1992.

216. F. Meyer and P. Bouthemy, "Estimation of time-to-collision maps from first order motion model and normal flows, " in *Proc. Intl. Conf. Patt. Recog.*, The Hague, Netherlands, pp. 78-82, 1992.

217. P. Burlina and R. Chellappa, "Time-to-x: analysis of motion through temporal parameters," *Proc. IEEE Computer Society Conf. Computer Vision and Patt. Recn.*, Seattle, WA, pp. 461-468, 1994.

218. A.M. Tekalp, *Digital Video Processing*, Prentice Hall, Englewood Cliffs, NJ, 1995.

219. B.G. Haskell, A. Puri, and A. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman and Hall, 1997.

220. A.J. Patti, M.I. Sezan, and A.M. Tekalp, "Robust methods for high-quality stills from video in the presence of dominant motion," *IEEE Trans. Circ. Syst. Video Tech.*, 7, pp. 328-342, 1997.

221. J.S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1990.

222. M. Irani and P. Anandan, "Video indexing based on mosaic representations," to appear in special issue of *Proc. IEEE*, 1998.

223. M. Irani and S. Peleg, "Improving resolution by image registration," *Graph. Models and Image Proc.*, 53, pp. 231-239, 1991.

224. R.R. Schultz and R.L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, 5, pp. 996-1011, 1996.

225. A.J. Patti, M.I. Sezan, and A.M. Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and non-zero aperture time," *IEEE Trans. Image Processing*, 6, pp. 1064-1076, 1997.

226. H.J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Systems*, 1, pp. 10-28, 1993.

227. S.-F. Chang, A. Eleftheriadis, and R. McClintock, "Next-generation content representation, creation and searching for new media applications in education," to appear in special issue of *Proc. IEEE*, 1998.

228. B. Günsel, A. Müfit Ferman, and A. Murat Tekalp, "Temporal video segmentation using unsupervised clustering and semantic object tracking," to appear in the special issue of *J. Electronic Imaging*, 1998.

229. H.G. Musmann, M.Hötter, and J. Ostermann. "Object-oriented analysis-synthesis coding of moving images," *Signal Processing: Image Communication*, 1, pp. 117-138, 1989.

230. J.Y.A. Wang and E.H. Adelson, "Representing moving images by layers," *IEEE Trans. on Image Processing*, 3, pp. 625-638, 1994.

231. C. Toklu, A. T. Erdem, M.I. Sezan, and A.M. Tekalp, "Tracking motion and intensity-variations using hierarchical 2-D mesh modeling for synthetic object transfiguration," *Graph. Mod. & Image Proc.*, 58, pp. 553-573, 1996.

232. G. Ahanger and T.D.C. Little, "A survey of technologies for parsing and indexing digital video," *J. Vis. Comm. and Image Rep.*, 7, pp. 28-43, 1996.

233. B. Furht, S.W. Smoliar, and H. Zhang, Video and Image Processing in Multimedia Systems, Kluwer Academic Publishers, Boston, MA, 1995.

234. B.-L. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Trans. on Circuits and Systems for Video Technology*, 5 pp. 533-544, 1995.

235. B. Günsel, A.M. Tekalp, and P.J.L Van Beek, "Object-based video indexing for virtual studio productions," *Proc. IEEE Int. Conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 769-774, 1997.

236. A. Akutsu and Y. Tonomura, "Video tomography: An efficient method for camerawork extraction and motion analysis," *ACM Multimedia 94-10/94*, pp. 349-356, 1994.

237. J.D. Courtney. Automatic video indexing via object motion analysis. *Pattern Recognition*, 30, pp. 607-626, 1997.

238. C.E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, 27, pp. 379-423, 623-656, 1948.

239. C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record, Part 4*, pp. 142-163, 1959.

240. B.M. Oliver, J. Pierce, and C.E. Shannon, "The philosophy of PCM," *Proceedings of the IRE*, 36, pp. 1324-1331, 1948.

241. W.R. Bennett, "Spectra of quantized signals," *Bell Systems Technical Journal*, 27, pp. 446-447, 1948.

242. R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York NY, 1968.

243. T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

244. R.M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, Boston, MA, 1990.

245. R.M. Gray, *Entropy and Information Theory*, Springer-Verlag, New York, NY, 1990.

246. A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, 25, pp. 373-380, 1979.

247. S. Na and D. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. on Inform. Theory*, 41, pp. 886-900, 1995.

248. S. Nanda and W. A. Pearlman, "Tree coding of image subbands," *IEEE Trans. on Image Processing*, 1, 2, pp. 133-147, 1992.

249. A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, Boston, 1992.

250. M.W. Marcellin, T.R. Fischer, and J.D. Gibson, "Predictive trellis coded quantization of speech," *IEEE Trans. Acoust., Speech, and Signal Proc.,* 38, pp. 46-55, 1990.

251. P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, 31, pp. 532-540, 1983.

252. J.M. Shapiro. "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Proc.*, 41, pp. 3445-3462, 1993.

253. A. Said and W. Pearlman, "Image compression using the spatial-orientation tree," *Proc. IEEE Int. Symp. Circ. and Syst.*, Chicago, IL, pp. 279-282, 1993.

254. U. Horn, T. Wiegand, B. Girod, "Bit allocation methods for closed-loop coding of oversampled pyramid decompositions," *Proc. IEEE Int. Conf. on Image Proc.*, Santa Barbara, CA, pp. 17-20, 1997.

255. "Video codec for audiovisual services at px64 kbit/s," *ITU-T Rec. H.261*, Geneva, 1990.

256. "Video coding for narrow telecommunications channels at <64 kbit/s," *ITU-T Rec. H.263*, Geneva, 1996.

257. ISO/IEC 13818, "Generic coding of moving pictures and associated audio information, Part 2: Video," *ITU-T Rec. H.262*, 1996.

258. B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. on Commun.*, 41, pp. 604-612, 1993.

259. M.T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Trans. Image Processing*, 3, pp. 693-699, 1994.

260. B. Girod, "Rate-constrained motion estimation," *in Visual Communication and Image Proc. VCIP'94*, A.K. Katsaggelos (ed.), *Proc. SPIE 2308*, pp. 1026-1034, 1994.

261. P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy constrained vector quantization," *IEEE Trans. Acoust., Speech, and Signal Proc.*, 37, pp. 31-42, 1989.

262. K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, 2, pp. 160-175, 1993.

263. A. Ortega, K. Ramchandran and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. on Image Processing*, 3, pp. 26-40, 1994

264. T.M. Cover and J.A. Thomas, *Elements of Information Theory*. Wiley Interscience, New York, NY, 1991.

265. U. Horn and B. Girod, "Perfomance analysis of multiscale motion compensation techniques in pyramid coders," *Proc. IEEE Int. Conf. on Image Proc.*, Lausanne, Switzerland, pp. 255-258, 1996.

266. B. Hochwald and K. Zeger, "Trade-off between source and channel coding," *IEEE Trans. on Inform. Theory*, 43, pp. 1412-1424, 1997.

267. E. Steinbach, N. Färber and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. on Circ. and Syst. for Video Technology*, 7, pp. 872-881, 1997.

268. K. Ramchandran, A. Ortega, K.M. Uz and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source-channel coding," *IEEE JSAC*, Special issue on High Definition Television and Digital Video Commun., 11, pp. 6-23, 1993.

269. S. McCanne, "Scalable Compression and Transmission of Internet Multicast Video," PhD Diss., Dept. of EECS, UC Berkeley, 1996.

270. V.A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Inform. Theory*, 39, pp. 821-824, 1993.

271. M.F. Barnsley and A. Jacquin, "Application of recurrent iterated function systems to images," *Proc. SPIE*, 1001, pp. 122-131, 1988.

272. G.M. Davis, "A wavelet-based analysis of fractal image compression," *IEEE Trans. on Image Processing*, 1998.

273. D. Pearson, "Developments in Model-based Video Coding", *Proc. IEEE*, 83, pp. 892-906, 1995.

274. D. Pearson, "Variability of Performance in Video Coding," *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, pp. 5-8, Munich, Germany, 1997.

275. G. Nagy, "Optical scanning digitizers," *IEEE Computer*, 16, pp. 13-24, 1983.

276. The Chromatic media processor: http://www.chromatic.com/

277. A. Peleg, U. Weiser, "MMX technology extension to the Intel architecture," *IEEE Micro*, 16, pp. 42-50, 1996,

278. A. Peleg, S. Wilkie, U. Weiser, "Intel MMX for multimedia PCs," *Communications of the ACM*, 40, pp. 25-38, 1997

279. T.S. Huang, W.F. Schreiber, and O.J. Tretiak, "Image processing," *Proceedings of the IEEE*, 59, pp. 1335-1346, 1971.

280. H.C. Andrews, A.G. Tescher, and R.P. Kruger, "Image processing by digital computer," *IEEE Spectrum*, 9, pp. 20-32, 1972.

281. M.D. McFarlane, "Digital pictures fifty years ago," *Proc. IEEE*, 60, pp. 768-770, 1972.

282. P. Ruetz and R. Brodersen, "Architectures and design techniques for real-time image processing IC's," *IEEE Journal of Solid-State Circuits*, 22, pp. 233-250, 1987.

283. K. Hwang (editor), *Computer Architectures for Image Processing, Special Issue of IEEE Computer*, 16, 1983.

284. A. Choudhary and S. Ranka (editors), *Parallel Processing for Computer Vision and Image Understanding, Special Issue of IEEE Computer*, 25, 1992.

285. K. Shen, G.W. Cook, L.H. Jamieson, E.J. Delp, "An overview of parallel processing approaches to image and video compression," *Proceedings of the SPIE Conference on Image and Video Compression*, 2186, San Jose, CA, pp. 197-208, 1994.

286. J. Allen, "Conmputer Architectures for Signal Processing," *Proc. of the IEEE*, 63, 5, pp. 624-633, 1975.

287. P. Lapsley, J. Bier. A. Shoham, E.A. Lee, *DSP Processor Fundamentals*, IEEE Press, 1997.

288. Texas Instruments TMS320C67: http://www.ti.com/sc/docs/dsps/products/c67x/index.htm

289. Analog Devices SHARC processor: http://www.analog.com/products/sheets/ADSP21060.html

290. P. Lapsley, "NSP Shows promise on the pentium and PowerPC," *Microprocessor Reports*, 8, pp. 11-15, 1995.

291. "The visual instruction set: on chip support for new-media processing." *Sun Microsystems Whitepaper 95-022*: http://www.sun.com/sparc/vis/

292. R.M. Haralick, "Interactive image processing software," *Proceedings of the NATO Advanced Study Institute on Digital Image Processing and Analysis*, Bonas, France, June 1976.

293. M.S. Landy, Y. Cohen, and G. Sperling, "HIPS: A Unix-based image processing system," *Computer Vision, Graphics and Image Processing*, 25, pp. 331-47, 1984.

294. L H. Jamieson, S.E. Hambrusch, A.A. Khokhar, and E.J. Delp, "The role of models, software tools, and applications in high performance computing," *in Developing a Computer Science Agenda for High Performance Computing*, edited by Uzi Vishkin. ACM Press, pp. 90-97, 1994.

295. Rasure and Kubica, "The Khoros application development environment," *Experimental Environments for Computer Vision and Image Processing*, editor H.I Christensen and J.L Crowley, World Scientific, 1994.

296. Rasure and Young, "An open environment for image processing software development," *1992 SPIE/IS&T Symposium on Electronic Imaging*, SPIE Proceedings 1659, 14, 1992.

297. MPEG-4 Overview, ISO/IEC JTC1/SC29/WG11 N1730, Rob Koenen, ed., 1997

298. G. Bouton, B. Bouton, and G. Kubicek, *Inside Adobe Photoshop 4*, New Riders Publishing, 1997.

299. MATLAB Image Processing Toolbox: http://www.mathworks.com/products/image/

300. S. Wolfram, *The Mathematica Book*, Third Edition Wolfram Media, Inc. and Cambridge University Press, 1996.

301. National Instruments LabView: http://www.natinst.com/

302. R.A. Lerner, "The image understanding environment: progress since IUW'96," *Proceedings of the 1997 DARPA Image Understanding Workshop*, New Orleans, LA, 1997. (also see: http://www.aai.com/AAI/IUE/IUE.html).

303. NIH Image: http://rsb.info.nih.gov/nih-image/

304. R.A. Crowther, D.J. DeRosier, and A. Klug, "The reconstruction of three-dimensional structure from projections and its applications to electron microscopy," *Proc. Roy. Soc. London A*, 317, pp. 319-340,1970.

305. A. Macovski, *Medical Imaging Systems*, Prentice Hall, Englewood Cliffs, NJ, 1983.

306. D. A. Ausherman, A. Kozma, J.L. Walker, H.M. Jones, and E.C. Poggio, "Developments in radar imaging," *IEEE Trans. Aerospace and Electronic Systems*, 20, pp. 363-400, 1984.

307. R.N. Bracewell, "Strip integration in radio astronomy," *Aust. J. Phys.*, 9, pp. 198-217, 1965.

308. A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*, John Wiley and Sons, New York, NY, 1986.

309. R.E. Sabbagha, *Diagnostic Ultrasound Applied to Obstetrics and Gynecology, 3rd ed.*, J.B. Lippincott, Philadelphia, PA, 1994.

310. B. Cornuelle et al., "Tomographic maps of the ocean meoscale. Part 1: Pure acoustics," *J. Phy. Ocean*, 15, pp. 133-152, 1985.