NEW TRENDS IN IMAGE AND VIDEO COMPRESSION

Luis Torres*
Polytechnic University of Catalonia
Barcelona, Spain
luis@gps.tsc.upc.es

Edward J. Delp
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu

ABSTRACT

Image and video compression have been the object of intensive research in the last thirty years. The field is now mature as is proven by the large number of applications that make use of this technology. Digital Video Broadcasting, Digital Versatile Disc, and Internet streaming are only a few of the applications that use compression technology. Image and video standards have played a key role in this deployment. Now is time to ask: are there any new ideas that may advance the current technology? Have we reached a saturation point in image and video compression research? Although the future is very difficult to predict, this paper will try to provide a brief overview to where this exciting area is heading.

1. INTRODUCTION

Image and video coding are one of the most important topics in image processing and digital communications. During the last thirty years we have witnessed a tremendous explosion in research and applications in the visual communications field. There is no doubt that the beginning of the new century revolves around the "information society." Technologically speaking, the information society will be driven by audio and visual applications that allow instant access to multimedia information. This technological success would not be possible without image and video compression. The advent of coding standards, adopted in the past years, has allowed people around world to experience the "digital age." Each standard represents the state of the art in compression at the particular time that it was adopted. It is important then, to summarise the state of the art for each of the standards[1]. Section 2

presents a brief summary of the technology related to present still image and video compression standards. Further developments in the standards will also be presented.

Section 3 presents ideas as to how compression techniques will evolve and where the state of the art will be in the future. We will also described new trends in compression research such joint source/channel coding and scalable compression. Section 4 will introduce preliminary results of face coding in which a knowledge-based approach will be shown as a promising technique for very low bit rate video coding. Section 5 describes media streaming which is a new and exciting area for compression research.

2. STANDARDS AND STATE OF THE ART

2.1 Still Image Coding

For many years the Discrete Cosine Transform (DCT) has represented the state of the art in still image coding. JPEG is the standard that has incorporated this technology [1]. JPEG has been a success and has been deployed in many applications reaching worldwide use. However, for some time it was very clear that a new still image coding standard needed to be introduced to serve the new range of applications which have emerged in the last years. The result is JPEG2000 that will be standardized at the end of 2000. It is currently in the Final Committee Draft stage [2]. The JPEG2000 standard uses the Discrete Wavelet Transform. Tests have indicated that at low data rates JPEG2000 provides about 20% better compression efficiency for the same image quality than JPEG. JPEG2000 also offers a new set of functionalities. These include error resilience, arbitrarily shaped region of interest, random access, lossless and lossy coding as well as a fully scalable bit stream. These functionalities introduce more complexity for the encoder. MPEG-4 has a "still image" mode known as Visual Texture Coding (VTC) which also uses wavelets but supports less functionalities than JPEG2000 [3]. For a comparison between the JPEG2000 standard, JPEG,

---

[1] The reader should be aware that the authors have somewhat differing views on the role and impact of compression standards. LT feels very strongly that standards are important and necessary. EJD is somewhat skeptical of this view. We are still very good friends.

MPEG-4 VTC and other lossless JPEG schemes see [4]. For further discussion on the role of image and video standards see [5].

2.2 Video Coding

During the last ten years, the hybrid scheme combining motion compensated prediction and DCT has represented the state of the art in video coding. This approach is used by the ITU H.261 and H.263 standards as well as for the MPEG-1 and MPEG-2 standards. However in 1993, the need to add new content-based functionalities and to provide the user the possibility to manipulate the audio-visual content was recognized and a new standard effort known as MPEG-4 was launched. In addition to these functionalities, MPEG-4 provides also the possibility of combining natural and synthetic content. MPEG-4 phase 1 became an international standard in 1999 [3]. MPEG-4 is having difficulties finding wide-spread use, mainly due to the protection of intellectual property and to the need to develop automatic and efficient segmentation schemes.

The frame-based part of MPEG-4 which incorporates error resilience tools, is finding its way in the mobile communications and Internet streaming. H.263, and several variants of it [6], are also very much used in mobile communication and streaming and it will be interesting to see how these two standards compete in these applications.

The natural video part of MPEG-4 is also based in motion compensation prediction followed by the DCT, the fundamental difference is that of adding the coding of the object shape. Due to its powerful object-based approach, the use of the most efficient coding techniques, and the large variety of data types that it incorporates, MPEG-4 represents today the state-of-the-art in terms of visual data coding technology [5]. How MPEG-4 is deployed and what applications will make use of its many functionalities is still an open question.

2.3 What can be done to improve the standards?

Can something be done to "significantly" improve the performance of compression techniques? How will this affect the standards? We believe that no significant improvements are to be expected in the near future. However, compression techniques that require new types of functionalities driven by applications will be developed. For example, Internet applications may require new types of techniques that support scalability modes tied to the network transport. We may also see proprietary methods developed that use variations on standards, such as the video compression technique used by RealNetworks, for

applications where the content provider wishes the user to obtain both the encoder and decoder from them so that the provider can gain economic advantage.

2.3.1 Still image coding

JPEG2000 represents the state of the art with respect to still image coding standards. This is mainly due to the 20% improvement in coding efficiency with respect to the DCT as well as the new set of functionalities incorporated. Non-linear wavelet decomposition may bring further improvement [7]. Other improvements will include the investigation of color transformations for color images [8] and perceptual models [9].

Although other techniques, such as fractal coding or vector quantization have also being studied, they have not found their way into the standards. Other alternate approaches such as "second generation techniques" [10] raised a lot of interest for the potential of high compression ratios. However, they have not been able to provide very high quality. Second generation techniques and, in particular, segmentation-based image coding schemes, have produced a coding approach more suitable for content access and manipulation than for strictly coding applications. These schemes are the basis of MPEG-4.

There are many schemes that may increase the coding efficiency of JPEG2000. But all these schemes may only improve by a small amount. We believe that the JPEG2000 framework will be widely used for many applications.

2.3.2 Video coding

All the video coding standards based on motion prediction and the DCT produce block artifacts at low data rate. There has been a lot of work using post-processing techniques to reduce blocking artifacts [11, 12, 13]. A great deal of work has been done to investigate the use of wavelets in video coding. This work has taken mainly two directions. The first one is to code the prediction error of the hybrid scheme using the DWT [14]. The second one is to use a full 3-D wavelet decomposition [15, 16]. Although these approaches have reported coding efficiency improvements with respect to the hybrid schemes, most of them are intended to provide further functionalities such as scalability and progressive transmission.

One of the approaches that reports major improvements using the hybrid approach is the one proposed in [17]. Long-term memory prediction extends motion compensation from the previous frame to several past frames with the result of increased

coding efficiency. The approach is combined with affine motion compensation. Data rate savings between 20 and 50% are achieved using the test model of H.263+. The corresponding gains in PSNR are between 0.8 and 3 dB.

It can be said that MPEG-4 and H.263+ represent the state of the art in video coding. H.263+ provides a framework for doing frame-based low to moderate data rate robust compression. MPEG-4 combines frame-based and segmentation-based approaches along with the mixing of natural and synthetic content allowing efficient coding as well as content access and manipulation. There is no doubt that other schemes may improve the coding efficiency established in MPEG-4 and H.263+ but no significant breakthrough has been presented to date. The basic question remains: what is next? The next section will try to provide some clues.

3. NEW TRENDS IN IMAGE AND VIDEO COMPRESSION

Before going any further, the following question has to be raised: if digital storage is becoming so cheap and so wide spread and the available transmission channel bandwidth is increasing due to the deployment of cable, fiber optics and ADSL modems, why is there a need to provide more powerful compression schemes? The answer is, with no doubt, mobile video transmission channels and Internet streaming. For a discussion on the topic see [18, 19].

Coding generation	Approach	Technique
0th Generation	Direct waveform coding	PCM
1st Generation	Redundancy removal	DPCM,DCT DWT,VQ
2nd Generation	Coding by structure	Image segmentation
3rd Generation	Analysis and Synthesis	Model-based coding
4rd Generation	Recognition and reconstruction	Knowledge-based coding
5th Generation	Intelligent coding	Semantic coding

Figure 1. Image and video coding classification

3.1 Image and video coding classification

In order to have a broad perspective, it is important to understand the sequence of image and video coding developments expressed in terms of "generation-

based" coding approaches. Figure 1 shows this classification according to [20].

It can be seen from this classification that the coding community has reached third generation image and video coding techniques. MPEG-4 provides segmentation-based approaches as well as model based video coding in the facial animation part of the standard.

3.2 Coding through recognition and reconstruction

Which techniques fall within the "recognition and reconstruction" fourth generation approaches? The answer is coding through the understanding of the content. In particular, if we know that an image contains a face, a house, and a car we could develop recognition techniques to identify the content as a previous step. Once the content is recognized, content-based coding techniques can be applied to encode each specific object. MPEG-4 provides a partial answer to this approach by using specific techniques to encode faces and to animate them. Some researchers have already addressed this problem. For instance, in [21] a face detection algorithm is presented which helps to locate the face in a videoconference application. Then, bits are assigned in such a way that the face is encoded with more quality than the background.

3.3 Coding through metadata

If it is clear that understanding the visual content helps provide advanced image and video coding techniques then the efforts of MPEG-7 may also help in this context. MPEG-7 strives at specifying a standard way of describing various types of audio-visual information. Figure 2 gives a very simplified picture of the elements that define the standard. The elements that specify the description of the audio-visual content are known as *metadata*.
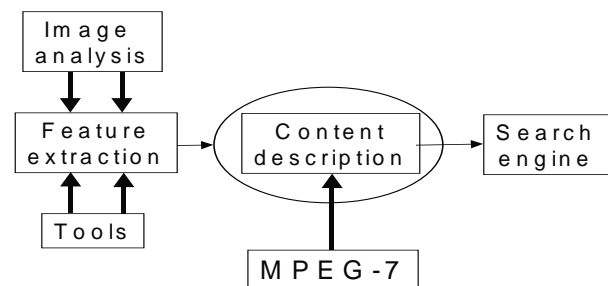

Figure 2. MPEG-7 standard

Once the audio-visual content is described in terms of the metadata, the image is ready to be coded. Notice that what is coded is not the image itself but the

description of the image (the metadata). An example will provide further insight.

Let us assume that automatic tools to detect a face in a video sequence are available. Let us further simplify the visual content by assuming that we are interested in high quality coding of a videoconference session. Prior to coding, the face is detected and represented using metadata. In the case of faces, some core experiments in MPEG-7 show that a face can be well represented by a few coefficients, for instance by using the projection of the face on an eigenspace previously defined. The image face can be well reconstructed, up to a certain quality, by coding only a very few coefficients. In the next section, we will provide some very preliminary results using this approach.

Once the face has been detected and coded, the background remains to be coded. This can be done in many different ways. The simplest case is when the background is roughly coded using conventional schemes (1st generation coding). If the background is not important, then it can not even be transmitted and the decoder adds some previously stored background to the transmitted image face.

For more complicated video sequences, we need to recognize and to describe the visual content. If this is available, then coding is "only" a matter of assigning bits to the description of each visual object.

MPEG-7 will provide mechanisms to fully describe a video sequence (in this section, a still image is considered a particular case of video sequence). This means that knowledge of color and texture of objects, shot boundaries, shot dissolves, shot fading and even scene understanding of the video sequence will be known prior to encoding. All this information will be very useful to the encoding process. Hybrid schemes could be made much more efficient, in the motion compensation stage, if all this information is known in advance. This approach to video coding is quite new. For further information see [18, 22].

It is also clear that these advances in video coding will be possible only if sophisticated image analysis tools (not part of the MPEG-7 standard) are developed. The deployment of new and very advanced image analysis tools are one of the new trends in video coding. The final stage will be intelligent coding implemented through semantic coding. Once a complete understanding of the scene is achieved, we will be able to say (and simultaneously encode): this is a scene that contains a car, a man, a road, and children playing in the background. However we have to accept that we are still very far from this 5th generation schemes.

3.4 Coding through merging of natural and synthetic content

In addition to the use of metadata, future video coding schemes will merge natural and synthetic content. This will allow an explosion of new applications combining these two types of contents. MPEG-4 has provided a first step towards this combination by providing efficient ways of face encoding and animation. However, more complex structures are needed to model, code, and animate any kind of object. The needs arisen in [23] are still valid today. No major step has been made concerning the modeling of any arbitrary-shaped object. For some related work see [24].

Video coding will become multi-modal and cross-modal. Speech and audio will come to the rescue of video (or viceversa) by combining both fields in an intelligent way. To the best of our knowledge, the combination of speech and video for video coding purposes has not yet been reported. Some work has been done with respect to video indexing [25].

3.5 Other Trends in Video Compression: Streaming and Mobile Environments

The two most important applications in the future will be wireless or mobile multimedia systems and streaming content over the Internet. While both MPEG-4 and H.263+ have been proposed for these applications, more work needs to be done.

In both mobile and Internet streaming one major problem that needs to be addressed is: how does one handle errors due to packet loss and should the compression scheme adapt to these types of errors? H.263+ [26] and MPEG-4 [27] both have excellent error resilience and error concealment functionalities.

The issue of how the compression scheme should adapt is one of both scalability and network transport design. At a panel on the "Future of Video Compression" at the Picture Coding Symposium held in April 1999, it was agreed that rate scalability and temporal scalability were important for media strea-ming applications. It also appears that one may want to design a compression scheme that is tuned to the channel over which the video will be transmitted. We are now seeing work done in this area with techniques such as multiple description coding [28, 29].

MPEG-4 is proposing a new "fine grain scalability" mode and H.263+ is also examining how multiple description approaches can be integrated into the standards. We are also seeing more work in how the compression techniques should be "matched" to the network transport [30, 31, 32].

3.6 Protection of Intellectual Property Rights

While the protection of intellectual property rights is not a compression problem, it will have impact on the standards. We are seeing content providers demanding that methods exist for both conditional access and copy protection. MPEG-4 is studying watermarking and other techniques. The newly announced MPEG-21 [33] will address this in more detail.

4. FACE CODING USING RECOGNITION AND RECONSTRUCTION

This section presents very preliminary results on face coding using recognition and reconstruction of visual data. Although the main objective of this research work has been for video face recognition [34] it can be easily extended to face coding. Related work but in a different context, has been presented in [35]. Our application assumes that the video sequence to be coded contains image faces whose identity is known previously. A set of training images for each face contained in the video sequence is previously known. Figure 3 shows five views of the image *Ana* and Figure 4 five views of the image *José Mari*. These images come from the test sequences accepted in MPEG-7.

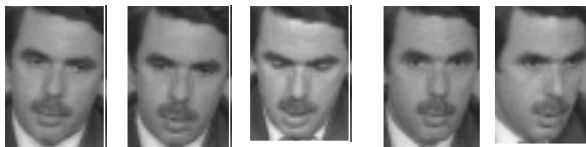Figure 3. Five training views of the image *Ana*

Figure 4. Five training views of the image *José Mari*

Once these training images have been found (usually coming from an image data base), a Principal Component Analysis (PCA) is performed for each individual using the corresponding training set of each person. This means that we obtain a PCA decomposition for every face image to be coded. The PCA is done previously to the encoding process. The first stage of the encoding process is automatic face segmentation and extraction of the video sequence. To that end we have used the face detection algorithm proposed in [36]. Once detected, all faces have been projected and reconstructed using each set of different eigenvectors (called eigenfaces) obtained in the PCA stage. If the reconstruction error using a specific set of eigenfaces is less than a threshold, then the face is said to match the training image which generated this set of eigenfaces. In this case we code the recognized face using only the five coefficients used in the reconstruction. It is clear that the corresponding eigenfaces of each person have to be transmitted previously to the decoder. However this can be done using conventional still image coding techniques such as JPEG and no significant increment in bit rate is generated.

Figure 5 provides some results. Figure 5 shows the original image Ana, the reconstruction of the detected face image Ana using the eigenvectors and corresponding projected coefficients of the PCA using the training images of Ana and the error done. Figure 6 shows the equivalent result for José Mari. Only 5 real numbers have been used to decode the shown images which means a very high compression ratio.

Our scheme is at a very early stage of development and we have not yet designed any bit assignment scheme to encode the face and the background. Our purpose here is to show that face coding using recognition and reconstruction is a promising approach and to indicate that much more work needs to be done in order to have good results. Although the presented results are not yet of very high quality, we believe that image coding using recognition and reconstruction may be the next step forward in video coding. Good object models will be needed, though, to encode any kind of object following this approach.

Figure 5. Decoded (reconstructed) image *Ana*. Left: original image. Center: reconstructed image. Right: Error image.

Figure 6. Decoded (reconstructed) image *José Mari*. Left: original image. Center: reconstructed image. Right: Error image.

Although not directly related to source video coding, let us mention that many efforts are being dedicated to provide robust video transmission through a variety of channels, Internet and mobile being the most significant. For a good review of the topic see [37].

5. CONCLUSIONS

We feel that any advances in compression techniques will be driven by applications such as databases, wireless and Internet streaming. New semantic-based techniques so far have promised much but have delivered little new results. Much work needs to be in the area of segmentation of video.

REFERENCES

[1] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compresion Standard*, Van Nostrand Reihold, New York, 1992.

[2] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information Technology – JPEG 2000 image coding system: Core coding system, WG 1 N 1646, March 2000. http://www.jpeg.org/FCD15444-1.htm

[3] ISO/IEC ISO/IEC 14496-2: 1999: Information technology – Coding of audio visual objects – Part 2: Visual, December 1999.

[4] D. Santa Cruz and T. Ebrahimi, "A study of JPEG2000 still image coding versus other standards," *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 5-8, 2000.

[5] F. Pereira, "Visual data representation: recent achievements and future developments," *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 5-8, 2000.

[6] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 8, no. 7, November 1998.

[7] D. Wajcer, D. Stanhill, and Y. Zeevi, "Representation and coding of images with nonseparable two-dimensional wavelet," *Proceedings of the IEEE International Conference on Image Processing*, Chicago, USA, October 1998.

[8] M. Saenz, P. Salama, K. Shen and E. J. Delp, "An evaluation of color embedded wavelet image compression techniques," *Proceedings of the SPIE/IS&T Conference on Visual Communications and Image Processing (VCIP)*, January 23-29, 1999, San Jose, California, pp. 282-293.

[9] N. S. Jayant, J. D. Johnston and R. J. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, October 1993, pp. 1385-1422.

[10] L. Torres and M. Kunt, Editors, *Video coding: the second generation approach*, Kluwer Academic Publishers, Boston, USA, January 1996.

[11] K. K. Pong and T. K. Kan, "Optimum loop filter in hybrid coders," *IEEE Transactions on Circuits and Systems in Video Technology*, vol. 4, no.2, pp. 158-167, 1997.

[12] T. O'Rourke and R. L. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 5, no. 6, pp. 490-499, December 1995.

[13] R. Llados-Bernaus, M. A. Robertson and R. L. Stevenson, "A stochastic technique for the removal of artifacts in compressed images and video," in *Recovery Techniques for Image and Video Compression and Transmission*, Kluwer, 1998.

[14] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, February 1999, pp. 109-122.

[15] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 125-139, February 1995.

[16] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Transactions on Image Processing,* vol. 3, no. 5, pp. 572-588, September 1994.

[17] T. Wiegand, E. Steinbach, and B. Girod, "Long term memory prediction using affine compensation," *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[18] R. Schäfer, G. Heising, and A. Smolic, "Improving image compression - is it worth the effort?" *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 5-8, 2000.

[19] M. Reha Civanlar and A. Murat Teklap, "Real-time Video over the Internet," *Signal Processing: Image Communication,* vol. 15, no.

1-2, pp. 1-5, September 1999 (Special issue on streaming).

[20] H. Harashima, K.Aizawa, and T. Saito, "Model-based analysis synthesis coding of video-telephone images – conception and basic study of intelligent image coding," *Transactions IEICE*, vol. E72, no. 5, pp. 452-458, 1989.

[21] J. Karlekar and U. B. Desai, "Content-based very low bit-rate video coding using wavelet transform," *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[22] P. Salembier and O. Avaro, "MPEG-7: Multimedia content description interface," *Workshop on MPEG-21*, Noordwijkerhout, the Netherlands, March 20-21, 2000. http://www.cselt.it/mpeg/events/mpeg21/

[23] D. Pearson, "Developments in model-based coding," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 892-906, June 1995.

[24] V. Vaerman, G. Menegaz, and J. P. Thiran, "A parametric hybrid model used for multidimensional object representation," *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[25] T. Huang, "From video indexing to multimedia understanding," *Proceedings of the 1999 International Workshop on Very Low Bitrate Video Coding*, Kyoto, Japan, October 1999. (Keynote speech.)

[26] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 867-877, November 1998.

[27] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 Standard," *IEEE Communications Magazine*, vol. 2, no. 6, pp. 112-119, June 1999.

[28] S. D. Servetto, K. Ramchandran, V.A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Transactions on Image Processing,* vol. 9, no. 5, pp. 813-826, May 2000.

[29] S. D. Servetto and K. Nahrstedt, "Video streaming over the public Internet: Multiple description codes and adaptive transport protocols," *Proceedings of the 1999 International Conference on Image Processing*, Kobe, Japan, October 1999.

[30] W. Tan and A. Zakhor, "Real-time Internet video using error resilient scalable compression and TCP-friendly transport protocol," *IEEE Transactions on Multimedia*, vol. 1, no. 2, pp. 172-186, June 1999.

[31] H. Radha, Y. Chen, K. Parthasarathy and R. Cohen, "Scalable Internet video using MPEG-4," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 95-126, September 1999.

[32] U. Horn, K. Stuhlmüller, M. Link and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 77-94, September 1999

[33] ISO/IECJTC1/SC29/WG11/N3300, MPEG-21 Multimedia Framework, Noordwijkerhout – March 2000.

[34] L. Torres, L. Lorente and J. Vilà, "Face recognition using self-eigenfaces," *Proceedings of the International Syposium on Image/Video Communications Over Fixed and Mobile Networks*, Rabat, Morocco, pp. 44-47, April 2000.

[35] B. Moghaddam, A. Pentland, "Probabilistic visual learning for object representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, July 1997.

[36] F. Marqués, V. Vilaplana, and A. Buxes, "Human face segmentation and tracking using connected operators and partition projection," *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[37] Special session on Robust Video, *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.