

IMAGE AND VIDEO COMPRESSION TECHNIQUES IN IMAGE PROCESSING AN OVERVIEW

¹S.SUKUMARN & ²T.VELUMANI

¹Associate Professor, Department of Computer Science, Erode Arts & Science College, Erode, Tamil Nadu, India

²Research Scholar, Department of Computer Science, Kongu Arts & Science College, Erode, Tamil Nadu, India

ABSTRACT

At the present time, we stand upon the threshold of a Revolution in the means available to us for the widespread dissemination of information in visual form through the rapidly increasing use of international standards for image and video compression. Yet, such standards, as observed by the casual user, are only the tip of the coding iceberg. Something like half a century of scientific and technological development has contributed to a vast body of knowledge concerning techniques for coding still and moving pictures, and the present article presents a survey of developments which have taken place since the first (predictive) coding algorithms were implemented in the 1950s. Initially, we briefly review the characteristics of the human eye which influence how we approach the design of coding algorithms; then we examine the still picture techniques of major interest—predictive and transform coding, vector quantization, and subband and wavelet multiresolution approaches. Recognizing that other forms of algorithm have also been of interest during this period, we next consider such techniques as quad tree decomposition and segmentation before looking at the problems arising from the presence of motion and its compensation in the coding of video signals. In the next section, various approaches to the coding of image sequences are reviewed, and we concentrate upon the now universally used hybrid motion compensated transform algorithm before examining more advanced techniques such as model and object based coding. Of course, the key to widespread acceptance of any technological development is the establishment of standards, and all major proposals—JPEG, MPEG-I,II,and IV, H.261, and H.263, are considered with emphasis on the way in which the coding algorithm is implemented rather than on protocol and syntax considerations. Finally, comments are offered in respect of the future viability of coding standards, of less well researched algorithms, and the overall position of image and video compression techniques in the rapidly developing field of visual information provision.

KEYWORDS: MPEG, JPEG, Quadtree, Ltiresolution, H.261.

INTRODUCTION

As this century draws to a close, it is virtually impossible to imagine what life must have been like for the average man in the street 100 years ago. No cars meant either cycling to work or going by rail or on foot, no aircraft left international journeys to be undertaken by sea, no domestic appliances meant that almost all jobs around the house had to be done manually, and no communication meant, well, no communication. Telephone systems were in their infancy, there were no broadcast or television services, and this left newspapers as the only widespread information provider (and only one way at that). Person-to-person contact was carried out either face-to-face or by letter. Probably the greatest influence (although there are many contenders) on changes in social attitudes has been our growing

Ability to engage in almost instant broadcast and person-to-person

communication over ever greater distances (in some cases it might even be argued that there now exists more communication provision than we need or is, strictly speaking, good for us). Undoubtedly, this change has been fostered by the widespread supplanting of analogue by digital technology over the past 3 decades or so (although, paradoxically, the final link in the chain, the telephone line, radio link, or whatever, may still well be analogue in nature), for this has allowed us to do three things much more easily than before: (a) carry out signal processing operations very rapidly; (b) build very complex large scale systems; and, most important, (c) store data easily. Where would telecommunications technology be, for example, if it were still as difficult to store information as it was, say, 50 years ago?

So where does image coding fit into all this? Throughout history, pictures have always had a high profile role to play in communication. In the days when the majority of people could not read whatever written words were available to them anyway, images allowed an immediacy of impact and directness of contact achievable in no other way. Later, representation of moving pictures, as provided by the film and, subsequently, television, enhanced this capability to an enormous degree. It is impossible to appreciate the impact of moving color picture presentation (something we take for granted) on anyone who has not grown up with the idea. In the case of television, however, it was quickly realized that communicating a video image was vastly more expensive in terms of necessary channel capacity than was speech or music transmission; and even from early days, methods of reducing this requirement were sought. Given the constraints operating at the time, the development of interlace, for example, has to be seen as an elegant practical solution to the problem (despite the drawbacks it presents today for digital processing of conventional video material). Again, early studies showed that the video signal could advantageously be split into different frequency bands and these sent separately and yet needing, overall, less capacity than the original signal. With the development of digital services, a growing problem soon appeared in that it was no longer a matter of dealing with a few highly specialized application areas, but rather a wide ranging spectrum encompassing person-to-person (videophone) communication at one end through videoconference and conventional television to high definition television at the other. Of course, supply and demand are inextricably linked in this field, as elsewhere. Given the capability of sending

images and video efficiently, more people want to make use of the facility, and at the same time, they generate new application areas, and so on. Widespread use over many different fields of application.

Also depends upon the development of and agreement on, standards, which is the point at which we stand at the moment. In a broad sense, we can now assume that the first phase of the development of image coding has been completed. Established, well regarded techniques are in place in several standards for still picture, video communication, broadcast use, etc., and there is some uncertainty about where we should be going next. There are also other well researched and efficient techniques which did not make it into the standards, but which nevertheless may still be of use in particular areas of application. On the other hand, it is hard to escape the realization that, notwithstanding past successes, some fresh ideas are now necessary to impart a new direction to coding development and also free us from some of the problems which still bedevil what we have come to call conventional approaches. Again, the enormous bandwidth offered to us by optic fiber may remove the need for image coding algorithms to be used in fixed service applications, leaving mobile systems as the major user of such techniques in an effort to conserve finite, and thus precious, spectrum space. There are thus many factors, not all of a purely technical nature, which will determine the future course of image coding research.

In this article, we initially review the status of well researched techniques; both those now incorporated into standards and others of similar standing, and then consider how the standards work of the last 10 years has employed such algorithms for the processing of a wide variety of image data. Thereafter, we consider the possibilities for new directions in image coding

BASIC CODING TECHNIQUES

In this and the next section, basic algorithms for coding still images (intraframe coding) are described. Typically, the image will be around 512 × 512 in extent, quantized to 8-bit (256 levels) amplitude resolution and either monochrome or color. In the latter case, it is usual to move from a three color plane (RGB) representation to a television style YUV or YIQ formulation. It is found that the additional load on the coding system due to the color information is relatively small—the luminance signal carries all of the detail resolution information and the U and V or I and Q data can be subsampled and coded using the same algorithm as the luminance term, but only needing something like 20% of the rate. Figure 1 shows the first frame of the sequence “CLAIRE”: 256 × 256 in extent with 8-bit amplitude resolution.

A. Predictive Coding. It is usually the case that useful images comprise recognizable objects—cars, buildings, people, etc.—portrayed against similarly recognizable areas of reasonable extent.

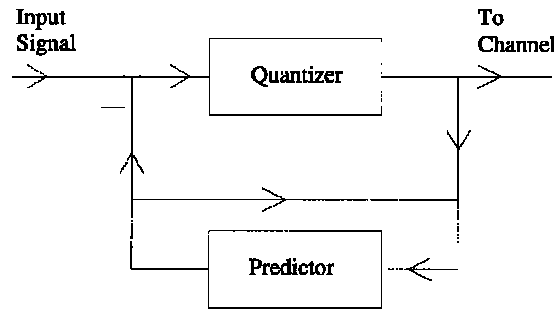


Figure 1. Predictive coding.

Only for regions of high detail or texture are this unlikely to be true and these tend to occur over only a small part of the image or video sequence. This being so, we can make use of the high degree of correlation existing within the picture to reduce the number of bits necessary for reproduction, and the algorithm which does this in the most obvious fashion is predictive coding. Deriving from work on general signal prediction in the 1940s, this algorithm was the first image coding technique to be extensively researched starting in the

1950s, and in its basic form it still finds place (albeit in the time dimension) in the standards algorithms of today. It is very simple and capable of reasonable rate reduction with excellent quality when applied to still pictures (O'Neal, 1996; Musmann, 1979). From the values of previously scanned picture elements at both transmitter and receiver, predictions are made of the next sample which the coder is to process. This prediction is subtracted from the actual value of the next sample and the error signal (difference) quantized and transmitted, to be added to the prediction made at the decoder to form the reconstructed signal. The procedure is shown in Figure 2. For a good prediction the error will be small, but just as important is the fact that its probability distribution is very well behaved compared with that of the original picture, and of Laplacian or even Gamma distribution form—very highly peaked around zero with relatively few large values, which latter will be a consequence of the edge detail in the picture. Coding this signal with maybe three bits per element and using no uniform quantization and variable word length coding allows good quality to be achieved for images containing television type detail (Musmann et al., 1985). Of course many refinements are possible; both predictor and quantized may be made adaptive to account for localized detail (needing transmission of overhead information to the receiver), possibly through the use of a least mean squares updating algorithm (Alexander and Rajala, 1984) or, more recently, through the use of higher order statistics (Tekalp et al., 1990). Alternatively, neural approaches may be used to optimize predictor structure (Dony and Haykin, 1995). Setting the basic element by-element algorithm in a block context in this way and maybe using arithmetic coding for the output signal allows rates down to around one bit per element to be achieved.

VECTOR QUANTIZATION.

Transform coding makes use of the interelement correlation within a picture to concentrate

coding capacity on the dominant low frequency terms produced in the corresponding low 'frequency' block representation. Another technique which makes use of similarities within the data, albeit at a higher level, is vector quantization (Gray, 1984). In scalar quantization, data amplitudes are reconstructed as the nearest predetermined value to that actually occurring within any given (one-dimensional) space between two adjacent decision levels. Given the usual level of interelement correlation within pictures, it is evident that we could jointly quantize pairs of neighboring picture elements as points in a two-dimensional decision space and, apart from quantization in accuracies, gain efficiency by so doing; pairs of similar values are much more likely to occur than those which are widely different (these latter representing edges). Vector quantization extends this idea to a larger region (a usual approach is to take a $4 \times 3 \times 4$ picture block, considered as a vector of length 16) to produce results which rival transform coding in terms of the quality of reproduction at rates around 0.5 bit/element, particularly if adaptive schemes are used [Panchanathan and Goldberg, 1991]. In vector quantization, use is made of the fact that many picture blocks will be very similar (background, interiors of large objects, etc.) in terms of luminance, color, and so on, or maybe contain strong detail of the same orientation. Such blocks will all be displayed as the same representative block chosen from a codebook of typical blocks (vectors) via some appropriate distance measure. The system has the advantage for certain applications that all the processing power is required at the transmitter, the receiver/decoder being trivially simple— one transmitted index word per block is all that is needed to access the codebook (look-up table) entry for display, maybe with some simple scaling operation. As a simple example, if we have 1024 representative entries, a 10-bit index is needed. If this identifies a $4 \times 3 \times 4$ block, then the data rate is about $2/3$ bit/element. In practice, some further sophistication is needed in the algorithm to cope with the artifacts which would be produced by such a basic scheme. There are many methods of generating the codebook, one tried and tested example of which was reported in 1980 (Linde et al., 1980). Given a first try reproduction codebook, all vectors from a suitable training sequence are allocated the closest entry according to some distance measure (mean square, or mean absolute, energy of the error vector, for example). Optimization proceeds by determining the new best code word for each of the partitions of training vectors so produced and then iterating the process. An initial codebook may be produced by first finding the one optimum code word for the whole of the training sequence, splitting it into two close but different vectors, and then proceeding as above. This codebook generation process is intensive both in time and computation, as is the other basic operation needed for coding an input vector: full search of the codebook for the nearest reproduction vector to determine the appropriate index to be transmitted. Most research on vector quantization since its introduction for image coding in the early 1980s has concentrated on these two problems, and a multiplicity of methods is now available for their (partial) solution: applying the splitting process preferentially to those nodes giving greatest decrease in distortion for the smallest increase overall in the number of codebook entries, and maybe using multiple splits as well. Separating out block mean value and standard deviation (corresponding to activity level) for separate transmission can also be helpful (Murakami et al., 1982) (Figure 4), as can classification of codebooks according to the presence of strong directional detail (Gersho and Ramamurthi,

1982). Neural optimization techniques can also be employed (Dony and Haykin, 1995; Lee and Petersen, 1990). Likewise, fast search methods have been intensively researched, with all manner of partial, tree, and approximate searches contributing to the speed up of the process. It is also possible to use a regular lattice structure for the codebook (Chen, 1984). This has the advantage that no actual codebook need be stored and processing is very rapid— especially beneficial in video applications (see later). Over the years, vector quantization has evolved into an efficient coding technique which may either be used on its own or as a postprocessing step to code the output of a previous algorithm—arrays of transform coefficients, for example.

Sub Band and Wavelet Coding

One of the earliest approaches suggested for the reduction in bandwidth or channel capacity for the transmission of image data was frequency division, in which the total bandwidth is split, at the simplest level, into low and high frequency segments. The low frequency band has reduced resolution and so needs far fewer bits for transmission; the upper frequency band will generally have few significant components and can likewise be easily coded. Over the past 10 years or so, this basic

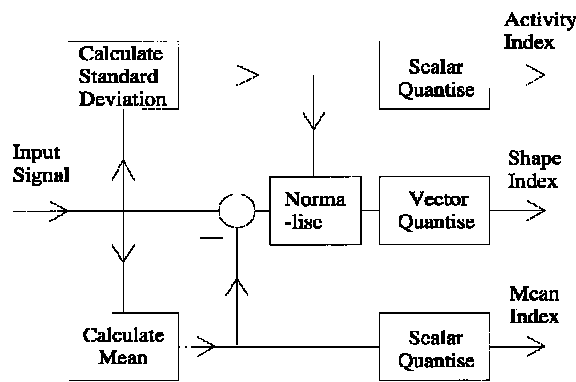


Figure2. Normalized vector quantization.

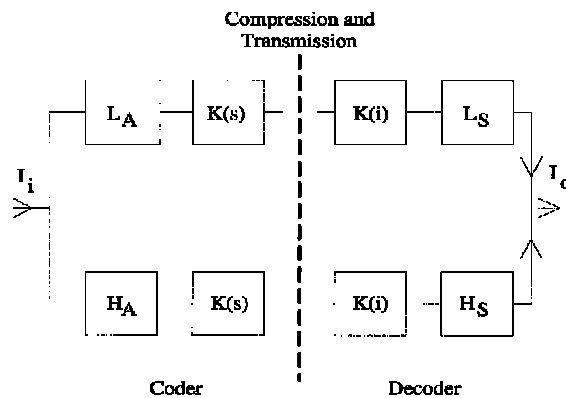


Figure 3. Basic filter structure for subband coding. I_i is input signal; I_o is output signal; L_A and H_A are low and high pass analysis filters; L_S and H_S are low and high pass synthesis filters; $K(s)$ is sub sampling at the coder; $K(i)$ is interpolation at the decoder.

Idea has developed into a powerful and flexible image coding technique: sub band coding (Woods, 1991). Here, the image spectrum is split basically into two components as above, but now both horizontally and vertically, which are then sub sampled by the same factor to give four two dimensional frequency bands containing combinations such as low horizontal and low vertical frequencies, low horizontal and high vertical frequencies, and so on, the total number of samples being unchanged. The basic one dimensional scheme is shown in Figure 5. As with predictive and transform coding, no data reduction is achieved by the first part of the algorithm, and it is the efficient coding of the various sub bands using predictive coding, or vector quantization, for example, which is responsible for this process. Such latter algorithms can be much more efficiently matched to the individual sub band characteristics than to the image as a whole, and this allows good results in the 0.5 bit/element region to be achieved (Lookabaugh and Perkins, 1990). Usually, the band split will be more extensive than the simple 2×2 split described above. The reapplication of this step to the outputs of the initial split will result in a 16-band (4×4) structure. Alternatively, it may be better to split lower frequency bands more finely still and leave the highest horizontal and vertical frequency band unsplit—it will rarely contain any detail of real significance. It might be noted here that there are close connections between sub band and transform coding. Sequential multilevel filtering using a simple high/low split of the kind mentioned above can be shown to produce the same result as a transform operator; indeed, transform coding may be considered to be a form of sub band coding in which each sub band contains only one coefficient.

Lately there has been much interest in the analysis of signals whose properties change with time. For this purpose, the Fourier transform is not suitable, since its baseline (in theory at least) is infinite, and the necessary window function added for practical implementation always introduces undesirable artifacts into the analysis. Out of this interest has come intensive research into wavelets (Chui, 1992). Such functions have restricted support and can allow flexible and efficient time (or space)/frequency processing, with good resolution at low frequencies via long windows, and good spatial resolution at high frequencies obtained by using short, wide bandwidth ones. As far as image coding is concerned, wavelet processing is not unlike sub band coding in basic outline, save that it is usually carried out in a multiresolution context. Here, an initial

Split and subsample operation is carried out as previously described, following which only the lowest frequency sub band is resplit. At each level, this results in one low resolution image and three so called detail images. We can thus generate a hierarchy of images at various levels of resolution which will allow image reconstruction for a wide variety of applications from the one data stream. The smallest and lowest-resolution image might be used as the initial output of a database search, for example. If higher resolution is needed, this signal is up sampled (interpolated) and the three detail signals at the next level added, and so on, until the final full resolution version is obtained. As usual, predictive coding or vector quantization can be used to code the various individual bands involved, and the scheme may be significantly improved by the inclusion of tree structured

algorithms for tracking significant coefficients through the various wavelet levels (Schapiro, 1993; Said and Pearlman, 1996).

Wavelet coding represents a sophisticated version of multi resolution decomposition whereby one algorithm may have a variety of image qualities at the output. It is appropriate to mention here the original impetus for the idea, which has been with us for some 20 years now. In this realization, lower resolution images were produced from higher level ones by Gaussian filtering and sub sampling and then expanded (interpolated) again to be used as a prediction for the upper level image. The prediction error then corresponds to the detail signal and, since it has a Laplacian probability distribution, and, moreover, appears as a set of image signal levels, one above the other, the idea of the Laplacian pyramid emerged (Burt and Adelson,

1983). Multi resolution decomposition is of great significance, given the variety and scope of present day digital image services, and the wavelet approach is more elegant and flexible than many other methods (using the DCT, for example) which have been proposed for this purpose.

MOTION

If there is one thing above all that the prevalence of television throughout the world as a provider of information and entertainment demonstrates, it is the overwhelming preference of the human observer for moving images. As in the case of still pictures, there has long been a concomitant interest in ways of processing these at as low a rate, given quality of reproduction constraints, as possible. It is only with the comparatively recent development of large scale, on chip storage, however, combined with the ready availability of ultra high speed hardware, which it has become practicable to implement such schemes. One way, of course, is simply to process image sequences on a frame byframe basis, i.e., without regard for any interrelation between them. Just as it is logical, though, to consider objects rather than arbitrary square blocks in the case of still images, so too, these objects are not only present, but also move within image sequences, and so the estimation of motion and its compensation have assumed increasing importance in image coding over the past 20 years or so.

Early work on motion estimation is represented by algorithms involving both the space and the frequency domain (Limb and Murphy, 1975; Haskell, 1974). In the former, the ratio of frame-to- frame differences, over the moving area, to the sum of element to element differences in the present frame was used to give an object speed measure, whereas in the latter the Fourier shift theorem can give an indication of motion via its phase shift term. Work started in earnest, however, in the late 1970s with the development of a recursive steepest descent algorithm which minimized the inter frame difference signal, in an algorithm which could also be modified to account for problems with changing illumination (Netravali and Robbins, 1979; Stuller et al., 1980). Intensive development of this algorithm by various workers continued for the next decade or so, but problems with reliable determination of changing areas and choice of a suitable initial estimate for the recursion meant that alternative schemes came into prominence and, as will be seen, were eventually incorporated into standard algorithms. The technique most widely used at present is based upon the taking of a block of elements in the present

frame and simply searching for a similar block in the previous frame which minimizes some function of the frame-to-frame difference over the area—mean square or mean absolute error (Jain and Jain, 1981). Prior to search, it can be advantageous to test the initial error against a threshold; if it is small enough motion compensation is not needed, anyway. The relative locations of blocks in present and previous frames are then characterized as a motion vector which must be transmitted to the decoder. This simple correlation like technique is computationally intensive (search over a 67 element displacement in both x and y directions requires the error to be calculated at over 200 locations); and although it is now possible at real time rates and is indeed the preferred approach, as if fully searched it guarantees to find a true minimum in the error function, the literature contains a long history of reduced search approaches— using a search route covering only a selection of locations, minimizing x and y error terms sequentially, and so on (Kappagantula and Rao, 1985). An added advantage is that, having all possible error values, displacement can be determined via interpolation to accuracy better than a single element. A refinement which can aid fast determination of the location corresponding to the minimum frame-to-frame error is hierarchical block matching (Wang and Clarke, 1990). Here, an image pyramid consisting of a sequence of planes, each formed by averaging over a small region of the previous one, is used top to bottom: A rapid search at the lowest resolution level forms the initial estimate for the next, and so on. This approach is also useful in dealing with large frame-to-frame differences.

Motion estimation and compensation form part of the standard present day approach to low rate video transmission, but their applicability is wider than simply optimizing frame-to-frame prediction. They can also be used for interpolation, when maybe every other frame in a sequence is dropped to achieve a minimum bit rate (Thoma and Bierling, 1989). In this situation, simple static interpolation produces unacceptable motion artifacts when used to reconstruct the missing frames, and motion compensation enables the movement of the object(s) to be accounted for in this operation (note that in this case, vectors representing true motion are necessary, not simply those which indicate a minimum in the frame difference signal). Areas covered up and uncovered by the moving object are accounted for by forward and backward extrapolation, respectively. Another application of motion compensation is in noise filtering, where, if the motion estimate is good, strong low pass filtering may be applied along the object path to reduce noise (Dubois, 1992). Naturally, the assumption that square block translation represents true object motion is only approximate, and experiments have been done in an attempt to allow more refined tracking of object rotation, scale change, and general motion throughout a video sequence (Wu and Kittler, 1990). There is also continuing work on effective frequency domain algorithms for motion compensation (Young and Kingsbury, 1993). The area in general is one where we are yet some way from the goal of being able to track object motion reliably through a sequence of frames and employ the knowledge so gained in reducing yet further the data rate for video transmission.

CONCLUSIONS

No one needs reminding nowadays that as we approach the millennium, we are living in a time of phenomenal change in our ability to access information. Undoubtedly, this has been due to the development of the computer and large scale high speed integrated circuits, and these have contributed to major advances in communication via speech and the written word. Overwhelmingly, however, it has been the image that has been in the forefront of this revolution as even a cursory examination of our image processing capabilities half a century ago and now will reveal. Oddly enough, we may be in the position we are now just because pictures, especially in moving form, contain so much information that they present an interesting and relevant challenge to the technology existing at any particular point in time as far as data reduction developments are concerned.

From the early days of predictive coding, via extremely powerful transform coding algorithms and the development of variable word length coding and motion compensation, all of which have given us the standards we have today, necessary transmission rates have fallen from megabytes to kilobytes per second; in addition, a plethora of alternative techniques has grown up, all of which, even if not standardized or, in some cases, very successful, have taught us more about the relation between the human observer and the visual representation of scenes. And priorities? In 1985, it was suggested that conventional algorithms had reached their limit and that we should be coding something other than square image blocks. Therefore, here we are 13 years later with all of our standards still based on just that approach, but with quite a bit more compression and flexibility to show, whereas in parallel, object and model based approaches creep painfully toward more generic and robust application. Shall we now say that the old methods have really reached the end of the road? We can always argue, of course, that more research is needed, and so it is, especially in bringing the HVS into the system how do we really perceive images, still or moving? How can we perform segmentation even with a tiny fraction of the ease with which the eye does it? Oddly enough, what was once considered to be the major stumbling block, processing speed (or the lack of it) seems to have disappeared from the equation. No longer do we have to accept compromises in algorithm design because the hardware cannot cope, and the days of an overnight job at the computer center simply to transform code a single image frame now seem like just a bad dream.

So where do we go from here? For fixed services, there seem to be two distinct opinions. One says that given the enormous bandwidth of optical fiber, the provision of new image services can be handled with only moderate levels of compression, or even none at all (this view ignores the fact of life that, given a new source of transmission capacity, human ingenuity will soon find ways not only of filling it up, but of finding reasons for requiring still more). The alternative argues that even so, there will always be a place for satellite and conventional broadcast distribution, in which case compression algorithms will play a vital part in providing choice within an ever increasing volume of program material. Whichever of these holds sway eventually, we can be safe in the knowledge that it is difficult to attach a fiber optic link to a moving vehicle, and finite radio channel space together with an ever

increasing demand for video communication in this context (public service applications, for example) is a guarantee that image and video compression techniques will become and remain common in this area. Again, the standardization activity of the past 10 years has to be seen by any criterion as a monumental achievement in drawing together scientific, technological, and economic and commercial considerations. Yet, in one sense, all arguments about standards and rationalization may not matter at all—the diversity of applications and techniques for image and video coding may mean that I as a service provider can arrange for you as a consumer to download my decoding software prior to transmission of the data, and we can all use whatever algorithm is dictated as economic by the actual application. Therefore, it may well be that there is room for everyone after all.

The Impact of enhanced algorithms can be justified with high speed derivate hardware algorithms. Thus Hardware architectures are the essential base to build the high efficient algorithm.

REFERENCES

1. N. Ahmed, T. Natarajan, and K.R. Rao, Discrete cosine transform, *IEEE Trans Compute C-23*, (1974), 90–93.
2. T. Akiyama, T. Takahashi, and K. Takahashi, Adaptive three-dimensional transform coding for moving pictures, *Proc Picture Coding Symp*, Cambridge, MA, 26–28 March, 1990, Paper 8.2.
3. S.T. Alexander and S.A. Rajala, Optimal gain derivation for the LMS algorithm using a visual fidelity criterion, *IEEE Trans Acoustic Speech Signal Process ASSP-32* (1984), 434–437.
4. A.D. Barnsley and A.D. Sloan, A better way to compress images, *BYTE*, January (1988) 215–223.
5. P.J. Burt and E.H. Adelson, The Laplacian pyramid as a compact image code, *IEEE Trans Commun COM-31* (1983), 532–540.
6. C.F. Chen and K.K. Pang, Hybrid coders with motion compensation, *Multidimen Syst Signal Process 3* (1992) 241–266.
7. T.C. Chen, A lattice vector quantization using a geometric decomposition, *IEEE Trans Commun COM-38* (1984), 704–714.
8. W.H. Chen and W.K. Pratt, Scene adaptive coder, *IEEE Trans Commun*
9. *COM-32* (1984), 225–232.
10. W.H. Chen and C.H. Smith, Adaptive coding of monochrome and color images, *IEEE Trans Commun COM-25* (1977), 1285–1292.
11. C.K. Chui, *Wavelets: A tutorial in theory and applications*, Academic Press, San Diego, 1992.
12. R.J. Clarke, *Transform coding of images*, Academic Press, San Diego, 1985. R.J. Clarke, *Digital*

coding of still images and video, Academic Press, San

13. Diego, 1995.
14. R.D. Dony and S. Haykin, Neural network approaches to image compression, *IEEE Proc* 83 (1995), 288–303.
15. E. Dubois, Motion-compensated filtering of time-varying images, *Multidimensional System Signal Process* 3 (1992), 211–239.
16. S. Ericsson, Fixed and adaptive predictors for hybrid predictive/transform coding, *IEEE Trans Comm COM-33* (1985), 1291–1302.
17. R. Forchheimer and T. Kronander, Image coding—from waveforms to animation, *IEEE Trans Acoust Speech Signal Process ASSP-37* (1989),
18. 2008–2023.
 - A. Gersho and B. Ramamurthy, Image coding using vector quantization, *ICASSP Proc*, 1982, pp. 428–431.
 - B. Girod, Psycho visual aspects of image communication, *Signal Process* 28 (1992), 239–251.
 - B. Girod, N. Faerber, and E. Steinbach, Standards based video communications at very low bit rates, *EUSIPCO Proc*, 1996, pp. 427–430.
19. R.M. Gray, Vector quantization, *IEEE ASSP Mag*, April (1984), pp. 4–29. A. Habibi, Hybrid coding of pictorial data, *IEEE Trans Comm COM-22*
20. (1974), 614–624.
21. B.G. Haskell, Frame-to-frame coding of television pictures using two-dimensional Fourier transforms, *IEEE Trans Inf Theory IT-20* (1974), 119–120.
22. A.E. Jacquin, Image coding based on a fractal theory of iterated contractive image transformation, *IEEE Trans Image Proc* 1 (1992), 11–30.
23. J.R. Jain and A.K. Jain, Displacement measurement and its application to interframe image coding, *IEEE Trans Commun COM-29* (1981), 1799–
24. 1808.
25. S. Kappagantula and K.R. Rao, Motion compensated interframe image prediction, *IEEE Trans Commun COM-33* (1985) 1011–1015.
26. M. Kunt, A. Economopoulos, and M. Kocher, Second generation image coding techniques, *IEEE Proc* 73 (1985), 549–574.
27. T.C. Lee and A.M. Peterson, Adaptive vector quantization using a self-development neural network, *IEEE J Select Areas Commun* 8 (1990),

28. 1458–1471.
29. H. Li, A. Lundmark, and R. Forchheimer, Image sequence coding at very low bitrates: A review, *IEEE Trans Image Proc* IP3 (1994), 589–609.
30. J.O. Limb and J.A. Murphy. Measuring the speed of moving objects from television signals, *IEEE Trans Commun COM-23* (1975), 474–478.
31. Y. Linde, A. Buzo, and R.M. Gray, An algorithm for vector quantizer design, *IEEE Trans Commun COM-28* (1980), 84–95.
32. T.D. Lookabaugh and M.G. Perkins, Application of the Princen-Bradley filter bank to speech and image compression, *IEEE Trans Acoust Speech Signal Process ASSP-38* (1990), 128–136.
33. H.S. Malvar, *Signal processing with lapped transforms*, Artech House, London, 1992.
34. T. Murakami, K. Asai, and E. Yamazaki, Vector quantizer of video signals, *Electron Lett* 18 (1982) 1005–1006.
35. H.G. Musmann, “Predictive image coding,” *Image transmission techniques (Advances in electronics and electron physics, Suppl 12)*, W.K. Pratt (Editor), Academic Press, New York, 1979, pp. 73–112.
36. H.G. Musmann, M. Hotter, and J. Ostermann. Object-oriented analysis- synthesis of moving images, *Signal Process Image Commun* 1 (1989),
37. 117–138.
38. H.G. Mussman, P. Pirsch, and H.-J. Grallert, *Advances in picture coding*, *IEEE Proc* 73 (1985), 523–548.
39. A.N. Netravali and J.D. Robbins, Motion-compensated television coding: Part 1, *Bell System Techn* 58 (1979), 631–670.
40. J.B. O’Neal, Jr., Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals, *Bell Syst Techn* 45 (1966),
41. 689–721.
42. S. Panchanathan and M. Goldberg, Min-max algorithm for image adaptive vector quantization, *IEE Proc Commun Speech Vision* 138 (1991), 53–60.
43. F. Pereira and R. Koenen. Very low bit rate audio-visual applications, *Signal*
44. *Process Image Commun* 9, 1996, 55–77.
45. K.R. Rao and J.J. Hwang, *Techniques and standards for image, video and audio coding*, Prentice Hall, Upper Saddle River, New Jersey, 1996.

- A. Said and W.A. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans Circuits Syst Video Technol*
46. 6 (1996), 243–250.
47. H. Samet, The quadtree and related hierarchical structures, *Comput surveys*
48. 16 (1984), 187–260.
49. D.G. Sampson and M. Ghanbari, Interframe coding of images using lattice vector quantization, *Proc 4th IEE Int Conf Image Process, Maastricht, Netherlands, 7–9 April, 1992*, pp. 131–134.
50. R.A. Schaphorst, Status of H.324 —the videoconference standard for the public switched telephone network and mobile radio, *Opt Eng* 35 (1996),
51. 109–112.
52. J.M. Schapiro, Embedded image coding using zerotrees of wavelet coefficients, *IEEE Trans Signal Process SP-41* (1993), 3445–3462.
53. T. Sikora, The MPEG-4 video standard verification model, *IEEE Trans Ckts*
54. *Syst Video Technol* 7 (1997), 19–31.
55. M. Soryani and R.J. Clarke, Segmented coding of digital image sequences, *Proc IEE I Commun Speech Vision* 139 (1992) 212–218.
56. P. Strobach. Quadtree-structured recursive plane decomposition coding of images, *IEEE Trans Signal Process SP-39* (1991), 1380–1397.
57. J.A. Stuller, A.N. Netravali, and J.D. Robbins, Interframe television coding using gain and displacement compensation, *Bell Syst Techn J* 58 (1980),
58. 1227–1240.
59. A.M. Tekalp, M.K. Ozkan, and A.T. Erdem, Image modeling using higher-order statistics with application to predictive image coding, *ICASSP Proc*,
60. 1990, pp. 1893–1896.
61. R. Thoma and M. Bierling, Motion compensation interpolation considering covered and uncovered background, *Signal Process Image Commun* 1 (1989),191–212. Q. Wang and R.J. Clarke, Motion compensated sequence coding using image pyramids, *Electron Lett* 26 (1990), 575–576.
62. H. Watanabe and Y. Suzuki 64 kbit/s video coding algorithm using adaptive gain/shape vector quantization, *Signal Process Image Commun* (1989),87–102.
63. P.H. Westerink, J. Biemond, and G. Muller, Subband coding of image sequences at low bit rates,

- Signal Process Image Commun 2 (1990), 441–448.
64. T. Wiegand, M. Lightstone, D. Mukherjee, T.G. Campbell, and S.K. Mitra, Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard, *IEEE Trans Ckts Syst Video Technol* 6 (1996), 182–190.
 65. J.W. Woods (Editor), *Subband image coding*, Kluwer, Dordrecht, 1991.
 66. S. Wu and J. Kittler, A differential method of simultaneous estimation of rotation, change of scale and translation, *Signal Process Image Commun* 2 (1990), 69–80.
 67. S. Yao and R.J. Clarke, Motion-compensated wavelet coding of colour images using adaptive vector quantization, *Proc Conf Image Process, Theory and Applications*, San Remo, Italy, 14–16 June, 1993, pp. 99–102.
 68. R.W. Young and N.G. Kingsbury, Frequency domain motion estimation using a complex lapped transform, *IEEE Trans Image Proc* 2 (1993), 2–17.
 69. L.H. Zetterberg, S. Ericsson, and H. Brusewitz, Interframe DPCM with adaptive quantization and entropy coding, *IEEE Trans Commun COM-30* (1982), 1888–189