A Multiresolution Approach for All-Digital HDTV

Martin Vetterli and Dimitris Anastassiou
Department of Electrical Engineering
and Center for Telecommunications Research
Columbia University, New York, NY 10027-6699

Abstract

In this paper, we investigate the feasibility of a fully digital HDTV production, processing, coding and transmission standard that will be a true alternative to analog television. An all digital HDTV should become economically feasible towards the end of the century. Such a television system is defined as a medium range technological goal, and will have a direct impact on various technologies, including computers, telecommunications and high speed VLSI. We describe research directions as well as current progress, especially with respect to hierarchical representations and coding methods, nonlinear interpolation techniques and joint source-channel coding for digital channels. Multiresolution techniques are central to our approach, because they are useful for compression, joint source/channel coding and for compatibility.

1. Introduction

By now, it is widely acknowledged that quality improvement of television signals is best achieved through digital enhancement signals. The purpose of this research is to investigate the ultimate case when the whole system is actually digital from end to end. While such a system is not economically feasible at the present time, it should so become in the near future. Such an all digital television system will achieve what the compact disc has done for the audio world, that is, a quantum step towards the ideal system.

To achieve the goal of an all digital television system, a number of key signal processing and communications problem have to be solved. They include the following issues:

- Efficient and hierarchical video signal representations: scanning structures in three dimensions and for the various color components have to be investigated, so as to achieve maximum perceptual quality at a given initial pixel rate.
- Medium compression transparent video coding techniques: a compression by a factor of about 10 has to be achieved, in a perceptually lossless fashion. This requires motion based processing, but in a robust manner.
- Novel interpolation methods: in order to recover higher spatial and temporal resolution imagery, new nonlinear interpolation techniques are required, e.g., for deinterlacing, or for purely spatial or temporal interpolation.
- Joint source-channel coding methods: source coding and transmission have to be well matched, so as to achieve graceful degradation in adverse conditions. It will be necessary to allocate "high-quality" channels to perceptually important components of the HDTV signal.
- Efficient digital transmission methods: to remain bandwidth efficient, several bits per hertz will have to be transmitted using sophisticated modulation and error correction techniques.
- Low complexity algorithms: to be economically feasible, all algorithms involved have to be of low complexity, especially in the demodulation and decoding part in the receiver. This constrains the possible schemes to be considered.

As far as transmission of HDTV is concerned, Figure 1 depicts the various alternatives. Analog transmission of advanced television is currently under study and often includes some digital component as well (of the order of a few megabits/sec typically). Transmission of digital HDTV over Broadband ISDN at 140 Mbits/s and 45 Mbits/s is also studied. We will concentrate on digital transmission over other possible channels, like cable. In that context, the quality of the channel should allow transmission of the order of several bits/hertz at reasonable cost. The cable transmission case will thus be taken as an example medium for all digital transmission of an HDTV signal coded at around 45 Mbits/sec but in a perceptually lossless fashion.

In this paper, we will indicate in more detail current progress on the project, especially as far as signal processing issues are concerned. First, multiresolution methods will be reviewed, in light of video applications in particular. New hierarchical schemes for video representation will be indicated, including in particular trees of progressive and interlaced sequences at various resolutions. This representation is useful both for compatibility purposes as well as for coding. A multiresolution approach to motion estimation and interpolation is presented, which permits both efficient motion estimation and hierarchical representation of the video signal. Then, new nonlinear interpolation techniques for both spatial and temporal resolution enhancement will be described. These techniques can be used both to increase the resolution of still frames, as well as to increase the number of frames, for example for deinterlacing or slow motion rendition purposes. Finally, initial work on joint source/channel coding for robustness purposes will be indicated.

2. Multiresolution Techniques: Subband and Pyramid Coding for HDTV

The representation of a signal in a hierarchy of resolutions or as successive approximations has been used in several disciplines, like speech and image compression, as well as computer vision. From a mathematical point of view, the unifying theory of these multiresolution techniques is the the theory of wavelets

CH2829-0/90/0000-0881 $1.00 © 1990 IEEE

[11]. In the digital signal processing literature, the subject has been investigated with multirate filter banks [10, 15, 13], and in computer vision, the methods are known, in a slightly different form, as image pyramids [5].

In a typical subband coding scheme, the input signal is sliced into frequency bands that can be coded "independently", at least in a first approximation. The various subbands have varying perceptual importance, and this can be used for compression, for joint source/channel coding and for compatibility. Typically, the lowest band has the highest semantic content, while higher bands increase the quality or detail over that of the "rough" approximation given by the "baseband". Such a subband scheme can be derived for 2D or 3D signals as well [14, 8] and one of the main benefits, besides compression [20], is that various bands can be protected differently. For example, in transmission over asynchronous networks, where high loss rates can occur, a 3D subband coding of video was demonstrated [9], where the low band would have high priority or protection, while higher bands would compete for resources and possibly suffer high loss rates at times. Due to the hierarchical nature of the decomposition, the overall scheme is quite robust because delivery of the lowband is guaranteed. Such joint source/channel coding is also essential in a broadcast environment, where various receivers have different receiving conditions, and a certain level of quality has to be guaranteed at all times. Note that transform coding, which can be seen as a subband coding method with particular filters, has therefore also a hierarchical decomposition built in. Recursive schemes like DPCM are not hierarchical, and can only be made hierarchical at the price of loss of performance. However, some high compression DPCM-base coding techniques [3] could be used to encode the low-resolution version of hierarchical coders.

Pyramid methods, introduced by several authors in various contexts [5], derive a low resolution version of the original signal, from which an approximation to the original signal is calculated. The difference between the approximation and the original is computed and has to be sent together with the low resolution version. The scheme is depicted for the simplest one dimensional case in figure 2. When compared to the subband coding scheme, we see that the difference signal (which is typically a high pass version of the original) is not subsampled, whereas the subband scheme is critically sampled. While this redundancy of the pyramid might seem to be a drawback at first sight, there are a number of attractive features. In particular, any function (including nonlinear ones) can be used to derive the low resolution version, as well as the interpolation, and the system can be designed so that there is only a single source of quantization noise, namely at the quantization of the difference signal. Because the quantization of the low resolution signals is included in the prediction of the higher resolution, only the top level quantization affects the reconstruction. This is unlike transform or subband coding, were there are a number of "independent" noise sources which can potentially add up in certain cases. These features of the pyramid coding scheme make it very flexible (virtually no condition on the decimation and interpolation functions) and robust (good control over quantization effects). Figure 3 shows such a generalized pyramid coding scheme, where NLI stands for nonlinear interpolation, E for Encode and D for Decode. The function of this figure will be described in more detail in section 5.

Note that in the linear case and when quantization can be neglected, it can be shown that the difference signal in the pyramid case can be subsampled as well. This is shown in Appendix A. Of course, the same is true for the generalized pyramid coding in the case that the low resolution version of the signal is subsampled without prefiltering, as sometimes is the case with interlacing.

In what follows, we will see that the possibility of using nonlinear operators in a pyramid scheme is quite attractive, in particular for image and video coding, since model based processing can be included. That is, in still image pyramids, edge modeling can be used [7], and in moving image pyramids, motion of solid objects can be included as part of the prediction in the 3D pyramid [19, 12]. Note that such models are hard to include in a subband scheme, where critical sampling makes the system very sensitive to nonlinearities.

The oversampling of the pyramidal schemes becomes negligible as the dimensionality D grows, since the oversampling ratio equals:

$$r_{over} = 1 + \frac{1}{2^D} + \frac{1}{(2^D)^2} + \cdots \tag{1}$$

$$= \frac{1}{1 - 1/2^D} = \frac{2^D}{2^D - 1} \tag{2}$$

that is, 33% for still images and 14% for moving images (assuming that the low resolution version is subsampled by 2 in each dimension, and that the pyramid is iterated indefinitely).

While subband methods are usually more delicate to design than pyramids, their simplicity can still be attractive. We will discuss a scheme that starts with a progressive video sequence and derives two interlaced subsequences, one lowpass and the other highpass, via a two channel perfect reconstruction filter banks using quincunx subsampling [17]. This is similar to the helper signal concept [6], but here we guarantee perfect reconstruction. The system is shown in figure 4. With appropriate filter bank designs [17], the lowpass version is a clean interlaced version of the progressive sequence, thus creating compatibility with existing interlaced standards. When using the low entropy highpass channel as an augmentation channel, the progressive HDTV sequence can be perfectly recovered. The filters can be designed to have only integer coefficients, an attractive feature for VLSI implementation. Note that the same scheme can be used with interlaced input to derive two progressive subsequences, a decomposition that can be useful as a first step in a coding scheme because the progressive lowpass version is well suited for motion based coding, while the highpass version can be coded by other means. While we have investigated other subband based schemes, like 3D subband coding [8], we believe that motion can be better incorporated into pyramids, and we will discuss this further.

3. Multiresolution motion estimation and interpolation

It turns out that a multiresolution view is appropriate both for the motion estimation problem, where hierarchical motion estimation is efficient and robust, and motion interpolation, which is used to build video pyramids.

For motion estimation, a computationally effective way to to catch large displacements is to perform the estimation first on lowpass subsampled frames, and then to use the resulting motion field as an initial guess for the next resolution estimation [4]. Besides the computational advantage (due to the reduced search space at each resolution), there is a built-in robustness in the sense that large displacement are usually related to large moving objects. Of course, we do not claim that there is a hierarchical optimal solution to the global optimization problem posed by motion estimation. However, the hierarchical approach is very likely to be correct, and the few possible failures (like very small objects having a very large displacement between two frames) will be caught in the difference signal since we are in a pyramidal scheme.

While hierarchical motion estimation has been used in a 2D fashion [4], we actually extend it by subsampling the time dimension as well [12]. Indeed, similar displacements are achieved over larger time periods in subsampled frames. This has however one catch: the linear motion model, which works well on a frame by frame basis, is more likely to be violated over several frames. Thus, care has to be taken not to rely on linear motion over too many frames. Results obtained with this 3D hierarchical motion estimation, including some motion segmentation in critical areas and provisions for occlusions and uncovered areas, are quite promising [12]. Furthermore, the hierarchical motion estimation can be used directly to build 3D video pyramids, where difference signals are now interpolation errors between original frames and interpolated frames based on adjacent frames and motion vectors. This is done at various levels in the pyramid, which also includes increase in spatial resolution with associated difference signals. The initial results show that actually, interpolation errors are not needed in the final level (degradations are unnoticeable), unless post-processing is intended (like slow motion). This indicates how to build hierarchies of signals, where only contribution quality video would use all levels of the hierarchy (including all "error signals") and distribution quality would use only certain subsignals.

4. Resolution Enhancement using nonlinear interpolation

If the interpolation process is nonlinear, as in figure 3, the frequencies that are higher than the Nyquist frequency of the subsampled signal can be estimated by exploiting knowledge of an assumed source model, and the energy of the difference (high frequency) signals becomes lower, resulting in higher compression ratios. If the sampling rate of a signal is increased using linear interpolation, the signal will remain blurred, because the frequencies above the Nyquist limit of the original sampling rate cannot be estimated, unless there is additional knowledge about the signal. In that, latter, case, nonlinear interpolation, based on a source model, can provide a more accurate representation of the signal at a higher sampling rate, thus "defeating" the sampling theorem.

Nonlinear interpolation of video signals can be done temporally and spatially, in either a separable fashion, or jointly. There are various ways of achieving this. A reasonable assumption is that video scenes typically contain a number of moving objects with well defined edges. This assumption is used to make a more accurate prediction of the intensity values between the sampling points. For example, the edges of objects can be predicted from a low spatial resolution image, and the pictures at a higher spatial resolution can be shown sharper than they would if traditional interpolation techniques were used [7]. Similarly, temporal resolution is enhanced by motion compensated temporal interpolation. Further, if the sequence is subsampled from a noninterlaced to an interlaced format, a progressive sequence will also be nonlinearly interpolated based on the same assumptions [19].

Another way of looking at this process is to observe that the two signals (the nonlinearly interpolated signal and the difference from the original signal) become "more independent" of each other, than they would be in linear subband decomposition. It must be noted that even if a signal is decomposed into two bands that are disjoint in frequency domain, this does not guarantee that the two bands are independent of each other. For example, the edges of an object will show as spikes in the high-frequency band, and as blurred (but still relatively high slope) edges in the low frequency band. It is suboptimal to independently encode two components that are not independent of each other. On the other hand, independent lossy "two-component" compression of two independent processes lead to near-optimum coding under quite general conditions [2], thus justifying this approach.

The nonlinear effects of quantization and other coding distortions in the various channels are also taken into consideration: for each channel, the encoder has access to the reconstructed values of the signals, and can use those values before interpolating and subtracting from the signal of the next higer hierarchical level. Therefore, in the absence of transmission errors, the encoder will accurately duplicate the reconstruction stages of the decoder, and the distortion of the finally reconstructed signal will be due to the coding distortion of the last channel only.

In figure 3, the "upsampling plus linear filtering" interpolation blocks are substituted by one block performing some form of temporal or spatial nonlinear interpolation. Encoding of the bands is done inside the loop, so that the reconstructed values (after decoding) are added to the result of the nonlinear interpolation of the previous level of hierarchy, at the encoder site. Note that the resulting codec is asymmetric, because the extra complexity of considering the coding distortion of the previous levels of hierarchy affects only the encoder and not the decoder.

5. Conclusion and directions

We have described some on-going work on multiresolution video processing, which fits in an all digital HDTV scheme that is based on a hierarchy of video sequences. This is necessary for three main reasons: compression (pyramidal and subband schemes are well suited for high quality compression), joint source/channel coding (various signals in the hierarchy will have different protection) and compatibility (certain subchannels will be compatible with existing standards, while the current one might be a subchannel of a future standard).

Acknowledgements

This research was supported in part by NSF under Grants CDR-84-21402, ECS-84-51499 and MIP-88-08277 and by Bell Communications Research. The authors would also like to thank K. Jensen, J.Kovacevic, K.M.Uz and F.M. Wang for collaborative work described in part in this paper.

Appendix A

We show that the difference signal in a pyramid scheme (see figure 2) can be subsampled when linear filters are used and that quantization can be neglected. For this demonstration, we will assume that lossless filters are used, but the result holds in a more general case as well [18]. The operation of lowpass filtering and subsampling by 2 can be written as an operator H_0 equal to:

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & h_0(L-1) & \cdots & h_0(0) & 0 & 0 \\ 0 & 0 & 0 & h_0(L-1) & \cdots & h_0(0) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{3}$$

that is, it is like a convolution matrix but with stepping by 2 due to the subsampling. Similarly, the operation of upsampling and interpolation is given by an operator G_0^* ($*$ stands for hermitian conjugation) which is actually equal to H_0^* because the interpolation filter (G_0) is equal to $z^{-N+1}H_0(z^{-1}$, that is, it is the time reversed version of the decimation filter. Therefore, the difference signal is equal to:

$$d = (I - H_0^* H_0)x \tag{4}$$

where x and d are the input and difference signal, respectively. But, because we have perfect reconstruction filters, the following relation holds [11]:

$$H_0^* H_0 + H_1^* H_1 = I \tag{5}$$

that is, (4) becomes:

$$d = H_1^* H_1 x \tag{6}$$

where H_1 corresponds to the highpass version in a subband decomposition having H_0 as a lowpass filter (that is, $H_1(z) = z^{-N+1}H_0(-z^{-1}$. Now, because the impulse responses of the filters $H_0(z)$ and $H_1(z)$ form an orthonormal set [16], we also have:

$$H_0 H_0^* = H_1 H_1^* = I \tag{7}$$

where (7) corresponds to the subsampled domain. Thus, (6) is a projection onto the "half" space spanned by H_1, and one can subsample (6) by simply reapplying H_1:

$$H_1 d = H_1 H_1^* H_1 x = H_1 x \tag{8}$$

The reconstruction from the lowpass version and the subsampled difference signal is done by applying H_0^* and H_1^*, respectively, and perfect reconstruction follows from (5).

References

[1] D. Anastassiou, "Generalized three-dimensional pyramidal coding for HDTV using nonlinear interpolation", 1990 Picture Coding Symposium, Cambridge, Massachusetts, March 1990.

[2] D. Anastassiou and D.J. Sakrison, "New bounds to R(D) for additive sources and applications to image encoding", IEEE Transactions of Information Theory, vol. IT-25, No. 2, March 1979, pp. 145-155.

[3] D. Anastassiou, W.B. Pennebaker and J.L. Mitchell, "Gray scale image coding for freeze-frame videoconferencing", IEEE Transactions on Communications, Vol. COM-34, NO. 4, April 1986, pp. 382-395.

[4] M.Bierling, "Displacement estimation by hierarchical blockmatching," SPIE Conf. on Visual Communications and Image Processing, Boston, Nov 1988, pp.942-951.

[5] P.J.Burt and E.H.Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. on Com., Vol. 31, No.4, April 1983, pp.532-540.

[6] M.A.Isnardi, J.S. Fuhrer, T.R. Smith, J.L. Koslov, B.J. Roeder and W.F. Wedam, "Encoding for compatibility and recoverability in the ACTV system," IEEE Trans. on Broadcasting, Vol. BC-33, No.4, Dec. 1987, pp.116-123.

[7] K. Jensen and D. Anastassiou, "Spatial image resolution enhancement using nonlinear interpolation", Proceedings, IEEE 1990 International Conference on Acoustics, Speech and Signal Processing.

[8] G.Karlsson and M.Vetterli, "Three dimensional sub-band coding of video," Proc. of IEEE Int. Conf. on ASSP, April 1988, pp.1100-1103.

[9] G.Karlsson and M.Vetterli, "Packet video and its integration into the network architecture," IEEE Journal on Selected Areas in Communications, Special Issue on Packet Speech and Video, Vol. 7, No.5, June 1989, pp.739-751.

[10] M.J.T.Smith, T.P.Barnwell, "A New Filter Bank Theory for Time-Frequency Representation," IEEE Trans. on Acoust., Speech and Signal Processing, Vol. ASSP-35, No.3, March 1987, pp. 314-327.

[11] G.Strang, "Wavelets and dilation equations: a brief introduction," SIAM Review, to appear.

[12] K.M.Uz, M.Vetterli, and D.LeGall, "A multiresolution approach to motion estimation and interpolation with application to coding of digital HDTV," submitted to ISCAS-90.

[13] P.P.Vaidyanathan, "Quadrature Mirror Filter Banks, M-band Extensions and Perfect-Reconstruction Technique," IEEE ASSP Magazine, Vol. 4, No. 3, pp.4-20, July 1987.

[14] M.Vetterli, "Multi-Dimensional Sub-Band Coding: Some Theory and Algorithms," Signal Processing, Vol. 6, No.2, pp. 97-112, Feb. 1984.

[15] M.Vetterli, "A Theory of Multirate Filter Banks," IEEE Trans. on ASSP, Vol. 35, No. 3, pp. 356-372, March 1987.

[16] M.Vetterli and D. Le Gall, "Perfect reconstruction FIR filter banks: some properties and factorizations," IEEE Trans. on ASSP, Vol. 37, No. 7, July 1989, pp.1057-1071.

[17] M.Vetterli, J.Kovacevic and D.LeGall, "Perfect reconstruction filter banks for HDTV representation and coding," Proc. Third Intern. Workshop on HDTV, Torino, Italy, Aug. 1989.

[18] M.Vetterli and C.Herley, "Wavelets and Filter Banks: Relationships and New Results," to appear, ICASSP-90, Albuquerque, April 1990.

[19] F.M. Wang, D. Anastassiou and A.N. Netravali, "Motion compensated deinterlacing of video sequences", Proceedings, Third International Workshop on HDTV, Torino, Italy, August 1989.

[20] J. W. Woods and S. D. O'Neil, "Sub-Band Coding of Images," IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-34, No. 5, May 1986, pp. 1278-1288.

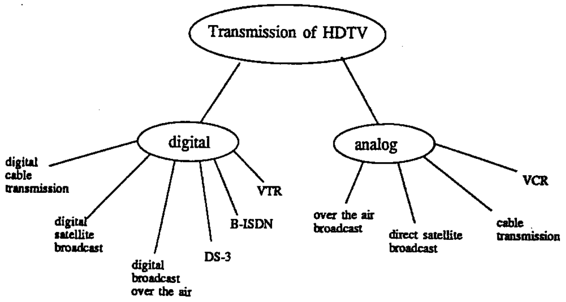

Figure 1: Transmission of HDTV.

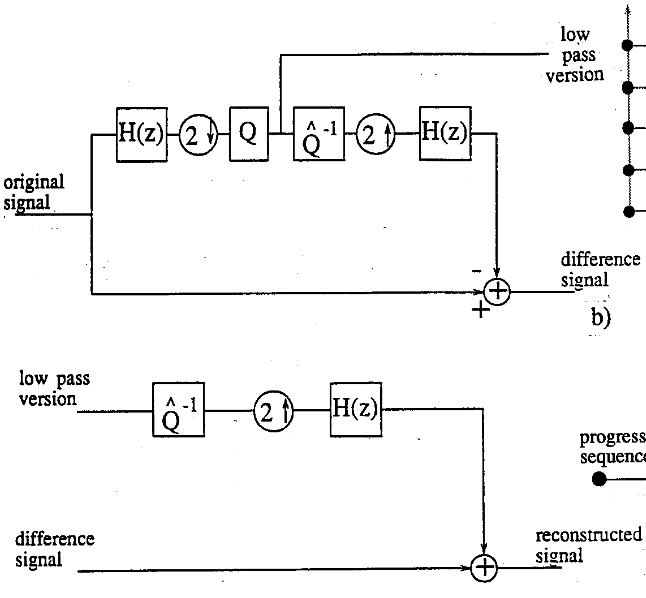

Figure 2: Pyramid coding.

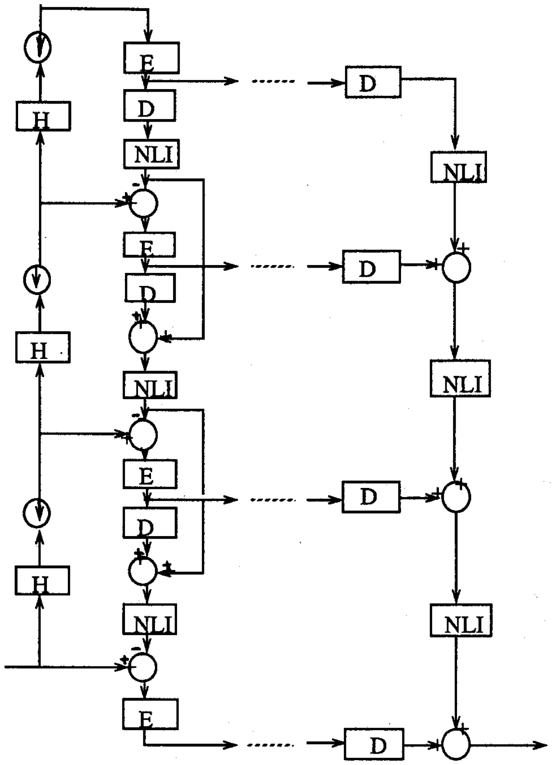

Figure 3: Generalized pyramid coding using nonlinear interpolation.

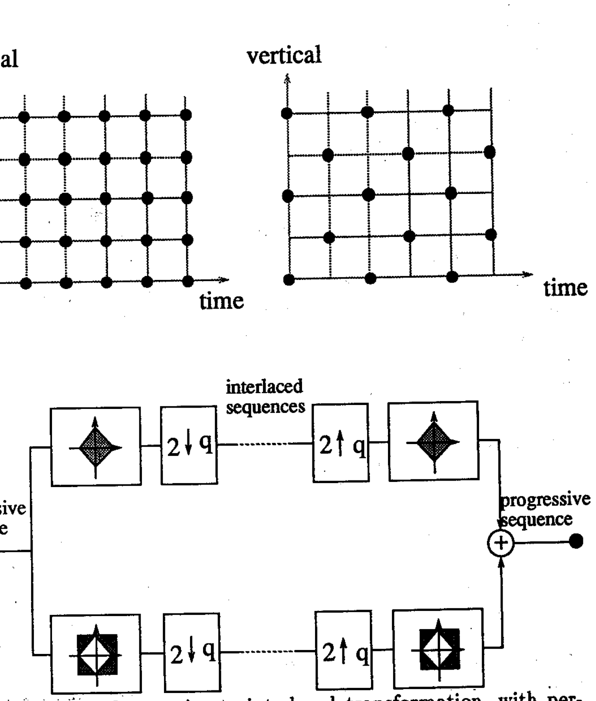

Figure 4: Progressive to interlaced transformation, with perfect inversion. (a) progressive and interlaced scanning. (b) 2 channel filter bank with quincunx subsampling.