WO/2006/040793   Sign in

Patents   English   French   Find prior art   Discuss this application   ⚙ ▾

Scalable video coding method

EP 1810519 A2 (text from WO2006040793A2)

ABSTRACT

A method of scalable video coding in which the video signal is processed and coded at various layers of spatial resolution where between each pair of spatial layers a prediction mechanism is enabled which allows a comparison at homologous layers of spatial resolution with the aim to obtain a good efficiency.

DESCRIPTION (OCR text may contain errors)

DESCRIPTION "Scalable video coding method"

[0001] The invention refers to the systems that need to exchange information in the form of a numerical video signal (videotelephony (mobile) , videocommunications

(mobile) , terrestrial digital TV, satellite digital TV,

DVD, video streaming, video-on-demand...) and refers in particular to a mechanism of cross-prediction between various spatial resolution representations of a same video signal to be used within a process of coding/decoding of the same video signal in order to ensure a full spatial, temporal and reconstruction quality scalability, starting from a single compressed stream. [0002] Temporal sequences of images acquired and represented in numerical format are commonly called digital video signals. In today's information society, the compression of video signals is a necessary operation to enable and/or to facilitate the recording of such signals or their, distant communication. The compression in fact is applied to reduce the number of recording unit

(commonly bit) for the digital representation of the video signal, which in turns lends to a reduction of the bit-rate that is necessary for the transmission of the same signal on a digital communication channel. [0003] The compression or coding system receives an incoming video signal and returns a bit stream of an inferior size with respect to the size (expressed in bits) of the original signal. It is typically possible "to trade" a smaller compressed bit-stream at the expense of a lower reconstruction quality of the data after the inverse operation, called decompression or better decoding. [0004] The coding is carried out, by who holds the original signal, by means of a system said encoder, while decoding is performed by the receiver by means of a system called decoder. Normally the original video signal has spatial (coordinates x and y) and temporal (rhythm of images per second, or frame rate) characteristics that are not changed by the coding and decoding process that operates at a so-called working point defined by an appropriate quality of the decoded data. The most spread standards for video compression work according to this paradigm, as an example those pertaining to the MPEG or the H.26x families.

[0005] Such a paradigm becomes obsolete in the context of Scalable Video Coding (SVC) . In such a context, the coded data with respect to a certain working point can be decoded according to a not necessarily defined a priori large number of working points so called inferior, in the sense that they can be decoded from

---

Publication number	EP 1810519 A2
Publication type	Application
Application number	EP20050802518
PCT number	PCT/IT2005/000599
Publication date	25 Jul 2007
Filing date	12 Oct 2005
Priority date	15 Oct 2004
Also published as	US8233526, US20090285306, WO2006040793A2, WO2006040793A3
Inventors	Riccardo Leonardi, Nicola Adami, Michele Brescianini, Alberto Signoroni
Applicant	Universita' degli studi di Brescia
Export Citation	BiBTeX, EndNote, RefMan

Non-Patent Citations (1), Classifications (22), Legal Events (7)

External Links: Espacenet, EP Register

CLAIMS (OCR text may contain errors)

Claims

1. Method of coding of a video signal comprising the steps of: a) acquiring a first digital signal (SI) representative of a video signal, said digital signal comprising a sequence of numerical images; b) operating a transformation (TS) of that first digital signal (SI) in order to generate at least a second digital video signal (S2) having a different spatial resolution with respect to that of said first digital signal (SI) ; c) carrying out a transformation (TT) of the at least second digital video signal (S2) generating a third digital signal (S3) , such transformation leaving substantially unvarying said spatial resolution of the second digital video signals (S2) ; d) carrying out an operation of compression (CE) of said third digital signal (S3) in order to generate a first digital compressed signal (Sc) ; e) operating a transformation (CE^{-1}) of said first digital compressed signal (Sc) in order to generate a reconstructed signal (Sr) equivalent to that obtainable by means of a decompression of the said first digital compressed signal (Sc) ; f) carrying out a processing of the first digital video signal (SI) comprising a transformation (TT) of the kind of the transformation of phase c) in order to generate a digital signal (St) , the method being characterized in that it further comprises the steps of: g) carrying out a transformation (TS) of at least a portion of said signal (St) of the kind of the transformation of the phase b) in order to obtain a digital signal (S4) comprising a first component (10) and a second component (11) , where said first and second components represent in combination the same informative content of the at least a portion of said digital signal (St) , said transformation (TS) being adapted to generate said first component (10) so that it has substantially the same spatial resolution of the digital signal (S3) ; h) generating a prediction error signal (Sp) representative of the difference between said first component (10) and at least a portion of the reconstructed signal (Sr) corresponding to the portion of the signal (St) ; and i) carrying out an operation of compression (CE) of said prediction error signal (Sp) and of said second component (11) in order to generate a second digital compressed signal (Sc2) . 2. A coding method according to claim 1, wherein in the case of generation at step b) of two or more video signals (s2, s2(I), ..., s2(i), ..., s2 (L-I),) having a spatial resolution differing from that of said first

only a fraction of the originally coded bit-stream. Such working points allow not only to reconstruct (to decode) the video signal with a scaled, reduced quality, spatial dimension (or resolution) and frame rate with respect to the signal that can be decoded from the whole compressed bit-stream. [0006] Figure 1 shows a typical SVC system. The example refers to the coding of a video signal at an original CIF spatial resolution and a rate of 30 fps (images or "frames" per second) . The scalable encoder typically produces a bit-stream where it is possible to determine one or more portions referring to the texture information, typically corresponding to static image coding, and optionally one or more portions referring to motion information (typically represented by a motion vector field) used in coding of motion-compensated temporal prediction operations. In the example, the originally coded stream is generated according to a bit rate equal to 2 megabits per second (2Mbps) , such rate being linked to a maximum layer of quality chosen for the original spatial and temporal resolutions. For a scaled decoding in terms of spatial and/or temporal and/or quality resolution, the decoder only works on a portion of the original coded bit stream according to the indication of the desired working point. Such stream portion is extracted from the originally coded stream by a block called "extractor" which in Figure 1 is arranged between the encoder and the decoder and which in general, according to the application filed, can represent an independent block of the whole chain or it can be an integral part of the encoder or decoder. The extractor receives the information referring to the desired working point (in the example of Figure 1, a lower spatial resolution QCIF, a lower frame rate (15 fps) , and a lower bit rate (quality) (150 kilobit per second) and extracts a decodable bit stream matching or almost matching the specifications of the indicated working point. The difference between an SVC system and a transcoding system is the low complexity of the extraction block that does not require coding/decoding operations and that typically consists in simple "cut and paste" operations. [0007] The application scenarios that can benefit from SVC are numerous, as an example the production and distribution of video over communication channels of diverse capacity, and for receiving terminal devices with different spatial or temporal resolution capability

(television, cellular videophones, palm-desktop...) ; video streaming on heterogeneous IP (Internet Protocol) networks; advanced tele-surveillance systems; videoconferencing applications with non guaranteed bandwidth; video streaming on mobile networks; fast video archives query and retrieval; and others.

[0008] Recently a strong interest has been focused around SVC coding solutions also thanks to important technological advances that enable spatial as well as temporal scalability, the most important one likely being both spatial and temporal wavelet transform, this last one in its motion compensated version. [0009] Within the scientific community and above all as a result of an explorative work made within working groups operating in the standardization organization ISO-MPEG, it has been possible to reach a classification of SVC systems according to the order in which the aforesaid transformations are applied to the original data. [0010] Scope of the present invention is that to propose a method of coding and of decoding of video signals that allows overcoming some limits of state-of-the-art SVC architectures and at the same time to give competitive if not improved coding performances with respect to the current state-of-the-art in video coding.

[0011] The idea at the base of the present invention regards a new SVC system architecture in which a video signal is processed and coded at various layers of spatial resolution where between pairs of spatial layers a prediction mechanism is enabled which allows a comparison between homologous spatial resolution layers to obtain a good coding efficiency.

[0012] The basic idea will be further detailed and analysed in a series of preferred embodiments to investigate the basis idea itself and/or exemplify possible embodiments thereof.

[0013] The possible SVC systems resulting from the idea at the basis of the present invention intrinsically exhibit spatial scalability features. [0014] A further aspect of the invention relates to a coding mechanism aimed at enabling temporal

digital signal (si) , the reconstructed signal (Sr) used in step h) is obtained from the sequence of the following steps, repeated for each signal (s2 (i)) at an intermediate spatial resolution (i) : h(i)) carrying out a transformation (TT) of the video digital signal at intermediate spatial resolution (S2(i)) generating a third digital signal (S3 (i)) , said transformation leaving said spatial resolution of the video digital signal at intermediate spatial resolution (s2 (i)) substantially unchanged; i(i)) carrying out a transformation (TS) of at least one portion of said signal at intermediate spatial layer (S3 (i)) of the kind of the transformation of the phase b) in order to obtain a digital signal (S4(i)) comprising a first component (10 (i)) and a second component (11 (i)) , where said first and second components represent in combination the same informative content of the at least one portion of said digital signal (S3(i)) , said transformation (TS) being adapted to generate said first component (10 (i)) so that it has substantially the same spatial resolution as the digital signal (s3(i-l), where s3 (0) =s3) ; j (i)) generating a prediction signal (Sp (i)) representative of the difference between said first component (10 (i)) and at least one portion of the reconstructed signal (Sr(i-D), where Sr(O)=Sr) corresponding to the portion of signal (S3 (i)) ; k(i)) carrying out an operation of compression (CE) of said prediction error signal (Sp(i)) and of said second component (11 (i)) in order to generate a digital compressed signal at intermediate layer (Sc (i)) ; l(i)) operating a transformation (CE^{-1}) of said digital compressed signal at intermediate layer (Sc (i)) in order to generate a residual reconstructed signal

(Srr(i)) equivalent to that obtainable by means of a decompression of said digital compressed signal at intermediate layer (Sc (i)) ; m(i)) operating a combination (C) of said reconstructed residual signal Srr(i) and of the reconstructed signal at a previous spatial layer Sr(i-l) which preserves the layers of spatial resolution of the above signals for generating a combined signal Sb(i) and n(i)) operating a transformation (TS^{-1}) of said combined signal Sb(i) , said transformation (TS^{-1}) being adapted for generating a reconstructed signal at intermediate layer Sr(i) so that it substantially has the same spatial resolution as the digital signal (s3(i)) . 3. Method of coding according to claim 1 or 2, in which the processing of the first digital video signal (SI) at step f) comprises in addition a transformation (TS) of said first digital video signal (SI) in order to generate at least a digital video signal (S2) having different spatial resolution from that one of the said first digital signal (SI) .

4. Method of coding according to any one of the previous claims, in which the transformation (TS) generates an output digital video signal having a spatial resolution lower than that of the input signal.

5. Method of coding according to any one of the previous claims, in which the transform (TS) at step b) is a spatial transform. 6. Method of coding according to the claim 5, in which said spatial transformation is a spatial wavelet transform.

7. Method of coding according to any one of the previous claims, in which the transform (TT) at step c) is a temporal transform.

8. Method of coding according to claim 7, in which said temporal transform it is a temporal wavelet transform.

9. Method of coding according to claim 8, in which said temporal wavelet transform is a motion-compensated temporal wavelet transform. 10. Method of coding according to any one of the previous claims, in which the steps from c) to i) are repeated for ulterior second signals (S2, S2', ...) .

11. Method of decoding of a video signal coded according to any one of the previous claims, comprising the steps of: a) acquiring a digital signal comprising at least the first and the second digital compressed signals (Sc, Sc2; SC , Sc2') ; b) carrying out a decompression operation (CE^{-1}) of the said

scalability features. The idea at the basis of such mechanism is to jointly code information portions resulting from the video signal and belonging to a homogeneous temporal resolution. Even though it could be taken into consideration in a generic SVC system, the application of such mechanism is preferred in the architectures subordinate to the basic idea of the present invention. In accordance with a preferred embodiment, an adaptation of a morphological coding technique known in the literature according to the idea at the basis of the above mechanism shall be illustrated.

[0015] An ulterior aspect of the invention refers to a coding stream partitioning representation system suitable for being submitted to the decoding method of the present invention.

[0016] The above system comprises indexing mechanisms that permit a coding stream partitioning with the aim of actually allowing to determine the stream that must be associated to a certain working point in terms of spatial and temporal resolution, and that possibly allow to satisfy some given maximum coding-decoding delay requirements. [0017] Even though it could be taken into consideration in a generic SVC system, the application of such coding stream partitioning representation system is preferred in the architectures subordinate to the basic idea of the present invention. In accordance with a preferred embodiment of said partitioning representation system, an embodiment shall be illustrated which is based on the morphological coding algorithm described above and used in an SVC architecture complying to the basis of the present invention. Such embodiment and all those arising from the mentioned ideas, allows an effective extraction of a decodable stream relative to a given working point and also enables and makes effective the possibility of carrying out a multiplicity of further concatenated extractions according to admissible paths among dependent working points. [0018] The characteristics and the advantages of the coding and decoding method according to the invention will result more apparent from the indicative and non limiting description of preferred examples of realization, with reference to the attached deawings, in which: [0019] - figure 1 represents a block diagram of a generic scalable video coding system SVC according to the prior art, including the coding, extractor and decoding blocks; [0020] - figure 2 represents a block diagram that illustrates the fundamentals of the encoder according to the present invention;

[0021] - figure 3 represents a block diagram that illustrates the fundamentals of the decoder associated to the encoder; [0022] - figure 4 schematically illustrates the extractor block only;

[0023] - figure 5 represents a block diagram of an embodiment of a two layer coding and decoding method according to the invention; [0024] - figures 5a and 5b show block diagrams of the coding and decoding method of figure 5, but with three layers;

[0025] - figure 6 shows an advantageous variation of the example of figure 5; [0026] - figure 7 shows a further advantageous embodiment of the coding method according to the present invention; [0027] - figure 8 represents a diagram of an example of embodiment of a temporal transform;

[0028] - figure 9 is a graphical representation of direct and inverse motion vectors for the description of the movement of an object between two frames of a sequence of images (or video sequence) ;

[0029] - figure 10 is a graphical representation of the

MCTF concept (motion compensated temporal filtering) , in which the trajectory of filtering is performed along time so as to follow the direction of the motion fields;

[0030] - figure 11 represents a block diagram of an additional example of a coding and decoding method according to the invention;

[0031] - figure 12 shows a detailed scheme for the video coding and predictive

first compressed digital signal (Sc; SC) in order to obtain the reconstructed signal (Sr; Sr') ; c) carrying out a transformation ($TT^{"1}$) of said signal (Sr; Sr') in order to generate a decoded signal (Sd) representative of the second digital video signal (S2) ; d) carrying out an operation of decompression ($CE^{"1}$) of said second digital compressed signal (Sc2; Sc2') in order to reconstruct a signal (Sp') representative of the prediction error signal (Sp) and a signal (H') representative of the second component (11) of the digital signal (S4) ; e) carrying out a sum between said prediction error signal (Sp') and at least one portion of said reconstructed signal (Sr; Sr') in order to reconstruct a signal (10') representative of the first component (10) of the digital signal (S4) ; f) carrying out a transformation ($TS^{"1}$) of said signals (10', 11') in order to reconstruct at least one portion of a signal (St') representative of the digital signal (St) ; and g) carrying out a transformation ($TT^{"1}$) of said digital signal (St') in order to generate a decoded signal (SdI) representative of the first video digital signal (SI) .

12. A coding method according to the previous claim, wherein if at step a) , besides the first signal (Sc;

Sc'), two or more compressed digital signals are acquired (Sc2, Sc2(I), ..., Sc2(i); Sc2' , Sc2(I)', ..., Sc2 (i) ') relating to different spatial resolutions and according to a predetermined order of such spatial resolutions, steps from d) to g) are repeated for each signal (Sc2(i); Sc2 (i) ') at intermediate spatial resolution (i) and wherein the reconstructed signal used in step e) is obtained through a composition of the decompressed signals (Sp', 11') belonging to all the lower spatial resolutions and of the reconstructed signal (Sr) at the base layer.

13. A decoding method according to claim 11 or 12, wherein said first and second compressed digital signals

(Sc', Sc2') come from an extractor block (E) having the compressed digital signals (Sc, Sc2) in input. 14. Method of decoding according to any one of claims 11 to 13, wherein said reconstructed signal (Sr') is a signal equivalent to the reconstructed signal used in the coding step. 15. A method of decoding of digital video signals which is able to decode a fraction of the coding stream in order to reconstruct a transformed domain representation of the digital video signal corresponding to a certain reference layer (in terms of quality, spatial resolution and temporal resolution) , which representation is used in order to predict the same type of transformed domain representation of the digital video signal at a different layer of spatial resolution. 16. Method of coding of a video signal, comprising in succession the steps of: a) generating representations of the video signal at various spatial resolutions; b) temporally transforming said representations; c) executing an operation of prediction between such temporal transform by putting in correspondence information inherent to homologous spatial resolutions; and d) entropy coding the resulting information in a progressive way, beginning from a first reference layer, said base-layer. 17. Method of coding according to claim 14, wherein at step c) , a temporal transform is predicted from a temporal transform at lower spatial resolution, and wherein said prediction is carried out through a matching of information regarding such lower spatial resolution.

18. Method according to claim 17, wherein, for each temporal transform, the conservation of the entire information regarding the related spatial resolution is provided. 19. Method of coding according to any one of claims 16 to 18, wherein at step b) , different temporal transforms are used (for example that differ in the number of temporal layers used) for different layers of spatial resolution, and wherein at step c) , a prediction operation is carried out between such temporal transforms through a matching of information regarding homologous spatial and temporal resolutions.

20. Method according to claim 19, wherein, in order to equalize the temporal

video coding blocks of the scheme of figure 11;

[0032] - figure 13 represents a block diagram of a method of synchronous evaluation of the motion field;

[0033] - figure 14 is a block diagram of a progressive encoder;

[0034] - figure 15 shows an example of organization of a coded bit-stream;

[0035] - figure 16 represents a scheme of extraction of a bit-stream organized as indicated in figure 15, and [0036] - figure 17 shows a schematic view of a possible structure of the truncation instructions of a progressive coding method.

[0037] The general principles of operation that govern the process of coding are described from the schemes reported in Figures 2 and 3.

[0038] The coding method comprises in succession the phases of:

[0039] a) generating representations of the video signal at various spatial resolutions; [0040] b) temporally transforming said representations;

[0041] c) executing an operation of prediction between such temporal transform by putting in correspondence the information inherent to homologous spatial resolutions; and [0042] d) entropy coding according to appropriate scalability criteria, the resulting information beginning from a first reference layer, called base-layer.

[0043] With reference to the scheme of Figure 2, the original video signal, indicated with si, is spatially filtered (blocks Ts), first column on the left of the scheme, generating the signals s2, s2 '.. that represent versions of the original images at gradually decreasing spatial resolutions.

[0044] The signals thus obtained are filtered carrying out a temporal transformation (Tt), which generates signals St, S3, S3', ..., followed by a further spatial transformation $(Ts)_1$ suitable for generating a digital signal comprising a first component 10 and a second component 11, where said components 10, 11 in combination represent the same information contents as signals St, S3, S3', ... The transformation Ts, moreover, generates the first component 10 so that it substantially has the same spatial resolution as the digital signal S3, S3', ... [0045] Considering now the output signal from block Ts of the generic layer n, where n coincides with the number of spatial transformations applied in the beginning phase of the process, the signal s3, s3', ... coming from the output of the temporal module (Tt) at the layer n+1 is used in order to predict the portion of information 10 that is compatible to it and generate a prediction error signal Sp. Such operation is represented in Figure 2 by the difference node between the first component 10 and signal S3. The signals thus obtained undergo a successive spatial transformation (Ts) and finally an entropic coding (EC) is applied which generates the final coded bit-stream. The entropy coding block provides a lossless or lossy compressed version of the incoming data. [0046] With respect to the coding procedure, the decoding method works in a dual way (figure 3). The coded bit streams undergo an entropy decoding and then an inverse spatial transformation $(Ts^{"1})$ that brings to the generation of a signal Sp' representative of the prediction error signal Sp and a signal 11' representative of the second component 11 of the digital signal. By signal representative of a reference signal it is meant a signal exhibiting spatial, temporal and quality features not differing from those of the reference signal by such an amount as to impair a useful usage thereof or a usage according to the system operating principles.

[0047] Considering now the signals pertaining to the generic layer n, where n coincides with the number of spatial transformations executed in the beginning phase of the process, the signal Sr coming from the output of the inverse spatial module relative to the layer n+1, is used in order to reconstruct the complete information related to the signal of the layer n. Such an operation is represented in Figure 2 by the addition node that allows obtaining the signal (10', 11') starting

layers to carry out the prediction(s), direct or inverse temporal transformations of different and/or partial nature (as compared to the temporal transforms at step b)) are carried out on the different layers of spatial resolution.

21. Method of coding according to any one of the previous claims, where the type of video representation in a transformed domain is represented by a temporal wavelet transform (also realized with motion compensation, i.e. MCTF) of the video signal for every desired layer of spatial resolution. 22. Method of coding according to any one of the previous claims, where the information at a certain layer of spatial resolution in the transformed domain is obtained through a spatial processing of the transformed information at a different layer of spatial resolution, homologous to the processing used for the generation of the video at the same different layer of spatial resolution.

23. Method of coding according to any one of the previous claims, where the motion information that is used in order to determine the transformed domain of the video representation is obtained in a synchronous way and by a refinement between at least two layers of the considered spatial resolutions.

24. Method of coding according to any one of the previous claims, where the entropic coding of the symbols obtained from the approximation of the transformed frames is carried out on groups of frames relating to the same layers of temporal resolution, so as to guarantee an adequate temporal and quality scalability of the coding stream. 25. Method of coding according to the previous claim, where said entropic coding is obtained through a progressive coding mechanism based on bit-planes (or similar iterative thresholding method from high to low threshold values) of transformed frames relating to same temporal resolution layers on which an eventual further spatial transformation is applied (generation of spatiotemporal coefficients).

26. Method of coding according to the previous claim, where said entropic coding is based on a hypothesis of agglomeration of spatiotemporal transformed coefficients at spatial discontinuities and/or at residual not motion- compensable temporal discontinuities and/or at temporal discontinuities located in areas where it is not possible to obtain an accurate motion estimate.

27. Method of coding according to the previous claim, where said entropic coding provides for an exploration of said agglomeration of spatiotemporal coefficients through morphological operators that exploit the agglomeration hypothesis and the structural information present in the spatiotemporal multi-resolution domain useful for predicting further coefficients higher than a set threshold.

28. A system of representation of the coding stream partitioning adapted to being submitted to the decoding method according to any one of the previous claims, comprising indexing mechanisms that permit a coding stream partitioning in order to actually determine the stream that must be associated to a certain layer of quality, for a given working point in terms of spatial and temporal resolution, and that allow to satisfy certain maximum coding-decoding delay requirements.

29. A system of representation of the coding stream partitioning according to the previous claim, wherein said indexing mechanisms comprise an insertion, in the header of the temporal layers, of the partial lengths of the coded flow for each coding step and for each bit- plane.

30. A system of representation of the coding stream partitioning according to the previous claim, wherein said headers relating to the partial lengths of the coded stream are updated and stored in case of multiple extractions, or if it is necessary to make further extractions on an already extracted stream. 31. Encoder for coding a video signal, comprising: a) means for generating representations of the video signal at various spatial resolutions; b) means

from signals (Sp', 11') and Sr. Therefore, the signals resulting from this operation undergo an optional inverse spatial transformation (Ts^{-1}) that generates signal St' representing signal St, and subsequently an inverse temporal transformation (Tt^{-1}). In such a way the video signals SvO, SvI... Svn are obtained at several spatial resolutions.

[0048] It should be noted that at the basic layer (bit stream LN), the second inverse spatial transformation (Ts^{-1}) is not present. [0049] The prediction can be defined as "crossed" because the video signals at various spatial resolutions undergo a transformation before being predicted from a layer of spatial resolution to the other. In other words, and according to a preferred implementation, to the signal obtained from a temporal transformation of the video at a certain layer of spatial resolution (said reference layer) a spatial transformation is applied so that it allows to identify (some) data at the same spatial resolution of the signal obtained by temporally transforming the video at a different layer of spatial resolution. A prediction can therefore be operated between the so obtained temporal representations, thus reducing the amount of information to be coded. Such prediction between signals with the same spatial resolution (such as s3 and 10), represented by the difference node in Figure 2, is to be understood in a general meaning and comprising the prediction methods commonly used for coding. Examples thereof are the open loop prediction that is based on the direct comparison of signals S3 and 10 and the closed loop prediction based on the comparison between a lossy version of S3 (typically compressed and decompressed at a certain quality layer) and signal 10, so as to allow the decoder to faithfully compensate the prediction operated. These choices do not interfere with the nature of the "crossed" prediction idea. In the encoder, the residual information generated at different layers of spatial resolution is entropy coded using a procedure of prioritization of the coding stream that guarantees a complete scalability in terms of quality. The described principle of crossed prediction can be similarly applied to all the remaining layers of spatial resolution of the video signal to be coded, either by starting from the original data layer, or from a different reference data layer. As far as the principle of crossed prediction is concerned, the decoder operates in a dual way with respect to the encoder. As described below, the decoder can also decode fractions of the coded stream in case it is desired to reconstruct a single video stream to a wished quality level, and to a particular temporal and spatial resolution.

[0050] Figure 4 shows the possible action of the "extractor" block E which, as already said, in an SVC system, is arranged between the encoder and the decoder and which in general, according to the applications, can be constructed as an independent block or as an integral part of the encoder or of the decoder. Thus, the incoming bit streams in Figure 3 can be interpreted as processed by the extractor block E and thus the corresponding decoded signals can be regarded as lossy versions of the coded ones but to be used in an "equivalent" manner.

[0051] By decoded equivalent signal, or signal to be used in an equivalent manner as compared to the corresponding signal in the encoder (or reference signal) it is meant a signal having a quality, resulting from a choice on the number of decoded bits, which approaches the quality of the reference signal so as to be used in replacement thereof. Typically, the equivalent signal exhibits a smaller number of bits than the number of bits of the reference signal, the number of discarded bits could otherwise be used to ensure a proper operation or improve the performance of the scalable coding/decoding system. [0052] The presence or not of an extractor does not interfere with the nature of the basic idea of this invention. However, the architectures of the resulting SVC systems allow detecting innovative solutions that relate to the actions and structure of the extractor, which shall be illustrated as further aspects of the present invention. [0053] Figures 5-5b show examples of a coding and decoding method according to the present invention. For higher clarity of description, the following description of the diagram of figure 5 does not comprise the action of the extractor block, which shall be better described hereinafter. [0054] The coding method comprises the following

for temporally transforming said representations; c) means for executing an operation of prediction between such temporal transform by putting in correspondence the information inherent to homologous spatial resolutions; and d) means for entropy coding the resulting information in a progressive way, beginning from a first reference layer, called base-layer. 32. Encoder for coding a video signal, comprising: means for acquiring a first digital signal (SI) representative of a video signal, said digital signal comprising a sequence of numerical images; means for operating a transformation (TS) of that first digital signal (SI) in order to generate at least a second digital video signal (S2) having a different spatial resolution with respect to that of said first digital signal (SI); means for carrying out a transformation (TT) of the at least second digital video signal (s2) generating a third digital signal (S3), said transformation (TT) leaving substantially unchanged said spatial resolution of the second digital signal video (S2); means for carrying out an operation of compression (CE) of said third digital signal (S3) in order to generate a first digital compressed signal (Sc); means for operating a transformation (CE^{-1}) of said first digital compressed signal (Sc) in order to generate a reconstructed signal (Sr) equivalent to that obtainable by means of a decompression of the said first digital compressed signal (Sc); means for carrying out a processing of the first digital video signal (SI) comprising a transformation

(TT) of the kind of the transformation generating said third digital signal (S3) in order to generate a digital signal (St); means for carrying out a transformation (TS) of at least a portion of said digital signal (St) to obtain a digital signal (S4) comprising a first component (10) and a second component (11), where said first and second components represent in combination the same informative content of the digital signal (St), said transformation (TS) being adapted to generate said first component (10) so that it has substantially the same spatial resolution of the digital signal (S3); means for generating a prediction error signal (Sp) representative ' of the difference between said first component (10) and at least one portion of the reconstructed signal (Sr) corresponding to the signal portion (St); and means for carrying out an operation of compression

(CE) of said prediction error signal (Sp) and of said second component (11) in order to generate a second digital compressed signal (Sc2).

33. Decoder for decoding a video signal coded according to a coding method of any one of the previous claims, comprising: - means for acquiring a digital signal comprising at least the first and the second digital compressed signals (Sc, SC2; SC, Sc2'); means for carrying out a decompression operation (CE^{-1}) of the first said compressed digital signal (Sc; Sc') in order to obtain a reconstructed signal (Sr; Sr'); means for carrying out a transformation (TT^{-1}) of said reconstructed signal (Sr, Sr') in order to generate a decoded signal (Sd) representative of the second digital video signal (S2); - means for carrying out an operation of decompression (CE^{-1}) of said second digital compressed signal (Sc2, Sc2') in order to reconstruct a signal (Sp') representative of the prediction error signal (Sp) and a signal (H') representative of the second component (11) of the digital signal (S4); means for carrying out a sum between said signal (Sp') and at least one portion of said reconstructed signal (Sr; Sr') in order to reconstruct a signal (10') representative of the first component (10) of the digital signal (S4); means for carrying out a transformation (TS^{-1}) of said signals (10', 11') in order to reconstruct at least one portion of a signal (St') representative of the digital signal (St); and - means for carrying out a transformation (TT^{-1}) of said digital signal (St') in order to generate a decoded signal (SdI) representative of the first digital video signal (SI). 34. Extractor device comprising: - means for carrying out the extraction of streams according to the partitioning representation system of claims 28-30;

phases: [0055] a) acquiring a first digital signal (SI) representative of a video signal, said digital signal comprising a sequence of numerical images; [0056] b) operating a transformation (TS) of said first digital signal (SI) in order to generate at least a second digital video signal (S2) having a different spatial resolution with respect to that of said first digital signal (SI) ,- [0057] c) carrying out a transformation (TT) of the at least second digital video signal (s2) generating a third digital signal (S3) , said transformation (TT) leaving substantially unchanged said spatial resolution of the second digital signal video (S2) . As substantially unchanged spatial resolution is intended that she signal endured a variation of spatial information which remains inferior, as an example, to 20%;

- means for updating and storing the information needed for further extractions starting from an already extracted stream.

[0058] d) operating a compression (CE) of said third digital signal (S3) in order to generate a first digital compressed signal (Sc) ; [0059] e) operating a transformation $(CE^{"1})$ of said first digital compressed signal (Sc) in order to generate a reconstructed signal (Sr) equivalent to that obtainable by means of a decompression of the said first digitally compressed signal (Sc) ; [0060] f) carrying out a processing of the first digital video signal (SI) , which comprises a transformation (TT) of the kind of the transformation of phase c) , in order to generate a digital signal (St) ; [0061] g) carrying out a transformation (TS) of said signal (St) of the kind of the transformation of phase b) in order to obtain a digital signal (S4) which comprises a first component (10) and a second component (11) , wherein said first and second components represent in their combination the same informative content of the digital signal (St) , said transformation (TS) being adapted to generate said first component (10) so that it has substantially the same spatial resolution of the digital signal (S3) ; [0062] h) generating a prediction error signal (Sp) representative of the difference between said first component (10) and the reconstructed signal (Sr) ; and [0063] i) carrying out an operation of compression (CE) of said prediction error signal (Sp) and of said second component (11) in order to generate a second digital compressed signal (Sc2) . [0064] In the case of figure 5a, wherein at step b) , two or more video signals are generated (s2, s2(l), ..., s2(i) , ..., s2 (L-I),) having a spatial resolution differing from that of said first digital signal (si) , the reconstructed signal (Sr) used in step h) is obtained from the sequence of the following steps, repeated for each signal (s2 (i)) at an intermediate spatial resolution (i) : a(i)) carrying out a transformation (TT) of the video digital signal at intermediate spatial resolution (S2(i)) generating a third digital signal (S3(i)) , said transformation leaving said spatial resolution of the video digital signal at intermediate spatial resolution (s2(i)) substantially unchanged; b(i)) carrying out a transformation (TS) of at least one portion of said signal at intermediate spatial layer (S3(i)) of the kind of the transformation of the phase b) in order to obtain a digital signal (S4(i)) comprising a first component (10 (i)) and a second component (ll(i)), where said first and second components represent combined the same informative content of the at least one portion of said digital signal (S3 (i)) , said transformation (TS) being adapted to generate said first component (10 (i)) so that it has substantially the same spatial resolution as the digital signal (s3(i-l), where S3(O)=S3) ; c (i)) generating a prediction error signal (Sp(i)) representative of the difference between said first component (10 (i)) and at least one portion of the reconstructed signal (Sr(i-l)), where Sr(O)=Sr) corresponding to the portion of signal (S3 (i)) ; d(i)) carrying out an operation of compression (CE) of said prediction error signal (Sp(i)) and of said second component (11 (i)) in order to generate a digital compressed signal at intermediate layer (Sc (i)) ; e(i)) operating a transformation $(CE^{"1})$ of said digital compressed signal at intermediate layer (Sc (i)) in order to generate a residual reconstructed signal (Srr(i)) equivalent to that obtainable by means of a decompression of said digital compressed signal at intermediate layer (Sc (i)) ; f(i)) operating a combination (C) of said reconstructed residual signal Srr(i) and of the reconstructed signal at a previous spatial layer Sr(i-l) which preserves the layers of spatial resolution of the above signals for generating a combined signal Sb(i) and g(i)) operating a transformation $(TS^{'1})$ of said combined signal Sb(i) , said transformation $(TS^{"1})$ being able to generate a reconstructed signal at intermediate layer Sr(i) so that it substantially has the same spatial resolution as the digital signal (s3 (i)) . [0065] The decoding method shown in figure 5 and associated to the above described coding method comprises the following phases:

[0066] a) acquiring a digital signal comprising at least the first and the second digital compressed signals (Sc,

Sc2) ;

[0067] b) carrying out an operation of decompression (CE'

^{1}J of said first digital compressed signal (Sc) in order to obtain the reconstructed signal (Sr) ; [0068] c) carrying out a transformation $(TT^{'1})$ of said reconstructed signal (Sr) in order to generate a decoded signal (Sd) representative of the second digital video signal (S2) ;

[0069] d) carrying out an operation of decompression (CE" ▪"▪) of said second digital compressed signal (Sc2) in order to reconstruct a signal (Sp') representative of the prediction error signal (Sp) and a signal 11' representative of the second component 11 of the digital signal (S4) ; [0070] e) carrying out a sum between said signal (Sp') and said signal (Sr) in order to reconstruct a signal 10' representative of the first component 10 of the digital signal (S4) ;

[0071] f) carrying out a transformation $(Ts^{'1})$ of said signals 10', 11' in order to reconstruct a signal (St') representative of the digital signal (St) ; and [0072] g) carrying out a transformation $(TT^{"1})$ of said signal (St') in order to generate a decoded signal

(SdI) representative of the first digital video signal (SI) . [0073] In the case illustrated in figure 5b, wherein at step a) of the decoding method described above, besides the first signal (Sc; Sc'), two or more compressed signals are acquired (Sc2, Sc2(I) , ..„, Sc2(i) ; Sc2', Sc2(I)', ..., Sc2(i) ') relating to different spatial resolutions and according to a predetermined order of such spatial resolutions, steps from d) to g) are repeated for each signal (Sc2 (i) ; Sc2(i)') at intermediate spatial resolution (i) . The reconstructed signal used in step e) to be added to the signal representative of the prediction error Sp' (i) is obtained through a composition of the decompressed signals (Sp' , 11') belonging to all lower spatial resolutions and of the reconstructed signal Sr at the base layer. [0074] In other words, starting from the original resolution, also said "final" because corresponding to the last refinement stage, different resolution video streams are extracted through a spatial transformation (TS) . The lower resolution constitutes the "base" layer to which the coded stream, said "base-layer", is associated. The intermediate layers can be 0, 1 or more than one .

[0075] A temporal transform (TT) is applied for every layer of spatial resolution to the aim of uncorrelating data along the temporal direction. [0076] Starting from the base layer an inter-layer relationships architectures is instantiated that foresee a prediction mechanism (P) of some data at the current layer of spatial resolution starting from an inferior layer of spatial resolution. [0077] The block (CE) in Figures 5, 5a carries out the signal compression according to scalability features in terms of quality. In block (CE) there may be comprised spatiotemporal transformation operations, quantization and entropic coding of the coefficients produced at the different layers of spatial resolution.

[0078] To the aim to guarantee a correct updating (A) of the prediction during the decoding phase, it must be noted that the data coming from the lower spatial resolution layer and used for the prediction (P) are equivalent to those that will be available during the decoding (that is those decoded through CE^{-1}) . [0079] The coded bit-streams produced by the coding method and containing the compressed signals (Sc, Sc2) can be sent to a block called "Extractor" that extracts and recomposes a new coded stream in which, according to the technologies which are used, the various spatial layers can go through temporal resolution (frame rate) and/or quality scaling.

[0080] The decoder inverts the whole coding processing chain by giving back the decoded video streams at the various spatial resolutions according to the streams provided by the extractor.

[0081] The spatial scalability is just obtained by discarding, starting from a certain layer of interest (that becomes the final layer) , all the coded streams of the layer at higher spatial resolution.

[0082] It must be noted that, to the aim to execute a prediction operation, it is necessary to confront data at the same layer of spatial resolution. This can happen, as an example, in two ways:

[0083] 1) through an interpolation of the data to an higher spatial layer,

[0084] 2) through a decimation of data at an higher resolution while activating a mechanism of conservation of the information discarded in phase of decimation (signals 10 and 11) .

[0085] With particular reference to figure 5a, it should be noted that predictors P at higher layers of resolution

(not strictly close to the base layer) will have to use all the decoded information from the lower layers and manage the appropriate inverse spatial transformations to obtain a signal Sr usefully comparable with signal 10. [0086] A further aspect of the invention, schematically illustrated in figure 6, relates to a mechanism for the non-synchronised management of the prediction in coding and updating in decoding operations to improve the performance of an SVC system according to the basic idea of the present invention. As mentioned above, the reconstructed signal Sr, that is, the signal S3 coded (block CE) and decoded (block CE^{-1}) in the coding step and used as base prediction layers of a higher spatial resolution layer, could not be exactly that (Sr') used in decoding to update the prediction error. The described degree of freedom can be useful for example to carry out a distribution of the lengths of the coded sub-streams at the various layers of resolution according to overall length constraints and with the aim of maximising, according to heuristic or quantitative methods, the quality of the reconstructed data relative to a set of one or more working points.

[0087] In accordance with an advantageous embodiment, the described mechanism is exemplified in Figure 6 through a combined action of an SVC system like the one of Figure 5 and an extractor block (E) . The figure shows a two layer architecture (one prediction layer) similar to that shown in Figure 5. For the stream coded at the base layer there are indicated three decoding points or length: Dmax greater than or equal to DP in turn greater than or equal to DA. Dmax is the length of the coded bit stream for the base layer corresponding to the maximum reconstruction quality. DP is the length corresponding to the stream portion of the base layer that is decoded (block CE^{-1}) to provide the signal Sr used by the encoder for the crossed prediction. It should be noted that DP does not necessarily coincide with Dmax. DA is the length corresponding to the stream portion of the base layer that is decoded to provide the signal Sr' used by the decoder for the crossed update. It should be noted that DA does not necessarily coincide with DP. Moreover, in the example, the selection of DA is taken from the extractor block according to an overall length constraint DA+DS or a set of constraints relating to various working points and comprising the constraint D=DAH-DS for the specific working points considered herein. DS is the length corresponding to the stream portion of the higher layer that is decoded. In general, the decisional processes of distribution of the DA and DS

values according to constraint D can be implemented also in a distributed way in the encoder and/or in the extractor and/or in the decoder and they can be realised through heuristics and/or in tabular form and/or through computational methods, without interfering with the nature of the further aspect of the invention just described. [0088] The example can be easily generalised in the case of N layers of prediction, where each layer at a spatial resolution n is considered as a base layer for the prediction of the layer with spatial resolution n-1, where n coincides with the number of spatial transformations carried out in the initial process step. [0089] According to a preferred embodiment, illustrated in Figures 8-10, of an SVC architecture according to the basic idea of the present invention, the block of temporal transform (TT) is realized with a motion- compensated temporal filtering (MCTF) technique. The operation of motion estimation produces for every pair of images, defined as "reference frame A - current frame B", a motion field, that can be expressed by motion vectors (in Figure 9 the motion vectors that describe the movement of an object in motion are represented) . The estimated motion field is coded to the aim of its compressed representation. The motion field is used from the wavelet filtering MCTF that with the aid of such information carries out a temporal decorrelation much more effective with respect to the single application of the normal wavelet filtering along the temporal direction. The data (frames) created from a MCTF wavelet transform, called temporal subbands, can be organized in layers of temporal resolution and it is such organization that enable the coding with temporal scalability. The principle of an MCTF filtering is represented in Figure 10, where the curve (c) represents the trajectory followed by the filtering. Mathematically, two pixels taken at successive time instants (frame A and B) will come visited by the filter along the sequence of motion compensated coordinates (x, y, tA)->(x+k, y+1, tB) , rather than (x, y, tA)->(x, y, tB) . Details on occlusion issues and uncovered areas generated by objects in motion can be found in literature (e.g. [J. R. Ohm, "Three- Dimensional Subband Coding with Motion Compensation" , IEEE Trans. On Image Processing, vol. 3, no. 5, pp. 572- 588, Sept. 1994; S. Choi and J. Woods, "Motion- compensated 3-D subband coding of video," IEEE Trans Image Processing, vol. 8, no. 2, pp. 155-167, Feb. 1999] . In the proposed scheme MCTF filtering is carried out on every layer of resolution so that motion fields at different spatial resolutions are generated and coded. Scalable coding mechanisms of the motion vectors can be implemented so as to increase the performance of the SVC system. [0090] In accordance with an advantageous embodiment of the coding and decoding method, illustrated in Figure 7, the information at a predetermined layer of spatial resolution can be predicted starting from the lower spatial layer according to an crossed prediction mechanism that operates according to two degrees of freedom:

[0091] 1) crossed prediction on a subset of temporal frames [0092] 2) crossed prediction on a subset of the temporal decomposition layers.

[0093] Figure 7 shows, for two different layers of spatial resolution (by way of an example, CIF and QCIF) , the result of a temporal decomposition (by way of an example, an MCTF) on 3 layers. There are shown 8 video frames (V) at an initial temporal resolution of M fps and thereafter, three temporal decomposition layers (FTi) . After the first decomposition, 8 temporal frames FT3 are obtained, of which 4 frames (or temporal subbands) at a temporal resolution M/2 and 4 detail frames (or temporal subbands) at layer of detail 3. The temporal decomposition is iterated only on the 4 frames at a resolution M/2 to obtain 4 temporal frames FT2 of which 2 frames at a temporal resolution M/4 and 2 detail frames at layer 2; similarly, an iteration is carried out on the 2 frames at a resolution M/4 to obtain one frame at a resolution M/8, called layer 0, and one detail frame at layer 1. The frame at layer 0 and the detail frames at layer 1, 2 and 3 together allow reconstructing signal V. To better understand the temporal references of the single transformed frames, on each frame there is shown a black rectangle representing a moving object. Such object is found on all frames at a low resolution and the corresponding detail information on all detail frames. The representation being discussed allows exemplifying the two degrees of freedom mentioned above and referred to the crossed prediction mechanism at the basis of the present invention:

[0094] 1) the crossed prediction mechanism may relate to all the frames needed to reconstruct the video signal V or only a subset thereof; in the example of Figure 7, the rectangle dashed as line-dots comprises, for the resolutions QCIF and CIF, the subsets of frames that could be involved in the crossed prediction. [0095] 2) The crossed prediction mechanism can be implemented also in the case the number of temporal decomposition layers for a lower spatial resolution layer is higher (in the example in Figure 7, QCIF with 3 temporal layers, that is, all those shown) than that of a higher spatial resolution layer (in the example in Figure 7, CIF with 1 temporal layer, that is, excluding those included in the rectangle dashed as line-line) . Two solutions are possible in this case:

[0096] a. The temporal decomposition of the higher spatial resolution layer is anyway continued to implement an crossed prediction adapted to the temporal frames of the temporal decomposition layer of the lower spatial resolution. In the example in Figure 7, the rectangle dashed as line-line shows the additional decomposition layers to reach the three layers required for the resolution QCIF.

[0097] b. The temporal decomposition of the lower spatial resolution layer is partly inverted to obtain an crossed prediction adapted to the temporal frames of the temporal decomposition layer of the higher spatial resolution. In the example in Figure 7, the rectangle dashed as line- line shows the temporal decomposition layers to be inverted to reach the single layer required for the resolution CIF. [0098] In accordance with a favoured realization form of the method of coding and decoding, illustrated in Figures 11 and 12, the information to a determined layer of spatial resolution in the transformed domain is obtained through a spatial processing of the transformed information of a different layer of spatial resolution, homologous to the processing used for the generation of the video to the same different layer of spatial resolution.

[0099] More in detail, the reference video streams at the various layers of spatial resolution are obtained by the use of a spatial discrete wavelet transform (DWT, Figure 11) . In the three-layers scheme of figure 11, the predictive structure is still evidenced (on two prediction layers) where the inter-layer relations are intercepted by a predictive video encoder (CVP) , detailed (along with the base-layer video encoder CV) in Figure 12. Notice that the same TWD used for the spatial resolution reduction is used within the CVP (figure 12) in order to create low resolution spatiotemporal subbands (signal 10) directly comparable with the decoded temporal subbands at the lower resolution layer (signal Sr) . In this way one can realize, in a DPCM way (that is coding the only difference) , a prediction between homologous subbands. [00100] It should be noted that the signal Sr is obtained by immediate decoding if it results from the base layer, whereas it is obtained through multiple decoding with inverse spatial transformations in the presence of intermediate layers. [00101] As already mentioned above, the prediction mechanism can be applied on all or part of the temporally transformed signal (signal St) . In the latter case, the St portion not involved in the prediction mechanism is directly addressed to the entropic quantization and coding block. Subbands between which the difference is executed are thought to be homologous in that:

[00102] 1) they substantially have the same layer of spatial resolution,

[00103] 2) they have undergone the same kind of space-temporal transformation, even if in the two cases the spatial and temporal transforms have been executed in a reversed order.

[00104] In accordance with a preferred embodiment, the motion information used to determine the transformed domain of the video representation is obtained in a synchronous way and by refinement between at least two layers of the considered spatial resolutions. In this way, a temporal scalability of the motion vectors is obtained which exhibits advantages from the coding performance point of view. In order to obtain a good decorrelation of the temporal coefficients it is essential that the used wavelet functions are applied along sequences of homologous pixels, that is pertaining to a same object. The operation that allows to identify such sequences of pixel is the motion field estimation (block 20 in Figure 13) , that is the estimation of the parameters that describe the spatiotemporal evolution of a block of homologous pixel in the sequence S of images that must be coded.

[00105] The motion estimator to use in the base-layer coding phase can be chosen between the most performing available technical solutions, such as those used in the H.264 standard. In the higher spatial resolution layer, with respect to the base-layer, coding phase, an opportunely modified estimator that could support multi- layer motion coding, base layer plus enhancement layer, should be used. In particular, the feedback block which modifies the estimator parameters must not only take into account the distortion introduced on the original image but also the distortion introduced on the image difference in the MCTF domain, as evidenced by the arrow

(f) in Figure 13. To the aim to fulfil the above mentioned constraints, the motion field (generally consisting of a partitioning of the frame into macro blocks and of the relevant motion vector values) of the predicted spatial layer will have therefore to be coherent with the corresponding motion field estimated on the base layer.

[00106] According to another aspect of the invention, the progressive coding is obtained through a mechanism of entropic coding of the symbols obtained from the approximation of transformed frames which are related to a same layer of temporal resolution, so as to guarantee an adequate stream coding temporal scalability and in terms of quality. [00107] After the temporal transform, the coefficients are temporally decorrelated. It can be observed that, on the whole, that is after both the temporal and spatial transformation (with or without the prediction mechanism) , sets of decorrelated coefficients are generated which can be classified according their space- temporal resolution location. For the realization of a scalable video encoder either in a spatial and in a temporal and quality resolution sense it is necessary that the coded description of the coefficients is organized according to an a-priori defined prioritization. Typically, as far as the temporal scalability, the encoder will have to guarantee an exclusive access to the data at a certain temporal resolution, while as far as the quality scalability the encoder will have to guarantee a re-definable and refinable coefficient quantization layer according to the desired reproduction quality. Moreover it is desirable that this quality refinement could be carried out finely and progressively. [00108] In this perspective, the following proposed preferential realization of the entropic coding comprises a coding system that acts jointly on same temporal resolution data and it is capable to realize a progressive refinement of the quality of the aforesaid data.

[00109] To such aim it is possible to use any still image encoder and preferably, a encoder adapted to manage groups of frames. This adaptation can regard the simple management of more than one frame or modifications made on the coefficients scanning and coding mechanisms that would benefit of the structural and statistical coherence of the frame data.

[00110] According to a preferred implementation, the above entropy coding is obtained through a symbols coding mechanism that works on a bit-plane basis on the same temporal resolution layer transformed frames. [00111] In a more detail, the transformed coefficients (spatially, temporally or that undergo both the transformations) typically present portions of greater activity (in an energetic sense) where the aforesaid operations of linear decorrelation do not succeed to act effectively. On video data this is typically verified in correspondence of spatial or temporal discontinuities which typically characterize the objects and the relative motion properties. The use of motion-compensated temporal transforms allows increasing the temporal decorrelation degree, limiting the residual temporal discontinuities and therefore decreasing the energy of the

transformed coefficients. This occurs provided that the motion model used is sufficiently accurate. In the practice, for coding efficiency and simplicity reasons, it is not possible to adopt a too complex and accurate motion model. Motion estimate systems based on the transfer of pixel blocks are typically used. [00112] The morphologic coherence of the objects represented in the video and the relative local motion regularity determine the fact that the areas of greater activity of the transformed coefficients are typically concentrated to form spatiotemporal agglomerates (clusters) (clusterization hypothesis) at the spatial discontinuities, at the non motion-compensable residual temporal discontinuities and at the temporal discontinuities located in the areas where an accurate motion estimate cannot be obtained. Therefore, the description of the position and the value of the coefficients that is necessary to make (implicitly or explicitly) through an entropy coding algorithm, will be able to usefully take advantage of this a-priori entropic information (clusterization hypothesis) in order to increase the coding performances. It is moreover decided to take advantage of the clusterization hypothesis on the

"active" coefficients according to various thresholds of activity, typically determined through the bit-planes of the binary representation of the data itself. The proposed encoder takes the name of Progressive Encoder based on Morphological Multilayer Dilation on Groups of

Frames at Homogenous Temporal Resolution: CPDMM-GFRTO and is an extension to the case of video-streams of the EMDC

(Embedded Morphological Dilation Coding) still image coding technique.

[00113] The input data are typically coefficients pertaining to groups of frames of same temporal layer

(block "GdF") . The reason of this resides in the possibility to simply enable the temporal scalability in an effective way by simply discarding the bit-streams associated to undesired temporal layers. It is also important to group frames together so as to be able to take advantage of the statistical similarities of same temporal layer frames in order to being able to construct more reliable probability tables to the end of an eventual adaptive entropic coding of the information streams.

[00114] The following (rectangular) functional blocks are visible in the reference diagram of Figure 14: [00115] - Dm: morphological dilation of the clusters ,

[00116] - Pm: multiscale prediction,

[00117] - Dme: morphological dilation of the cluster boundaries, [00118] - CpCs: coding of the significant coefficient position,

[00119] - R: refinement of the coefficients relative to the current bit plane,

[00120] - CEsp: lossless entropic coding (e.g. adaptive and context-based arithmetic coding) ,

[00121] - mCST: spatiotemporal clusterization engine,

[00122] and the following data (parallelograms) :

[00123] - GdF: group of frames with homogeneous temporal resolution having undergone the same spatiotemporal transformation,

[00124] - LCS: set of significant coefficient lists divided by each subband (or data portion related to a same spatiotemporal resolution) , [00125] - LCC: set of boundary coefficient lists divided by each subband (or data portion related to a same spatiotemporal resolution) ,

[00126] - Bit stream: is the product of CPDMM-GFRTO as regards the GdF considered. [00127] The coding method CPDMM-GFRTO and the action of the various blocks shown in Figure 14 shall now be described.

[00128] - Initialisation and main loop: The CPDMM-GFRTO coding has a structure that iterates itself on the bit- planes so as to realize, at the same time, the progressive scanning of the map of significant coefficients and their quantization and entropic coding. To such aim two lists are managed. The main one is the list of the coefficients found significant "LCS" with respect to the threshold determined by the current bit- plane. A second list keep memory of the non-significant coefficients in the neighbourhood of significant coefficients clusters "LCC". Such lists are divided by data blocks at a homogeneous spatiotemporal decomposition layer and the algorithm CPDMM-GFRTO scans, at each bit plane, all such blocks of data according to an order deemed appropriate and determined in advance or dynamically. The lists are initialized to be empty. For the purposes of the prediction relations and of the coding uniqueness, and if not already determined, an order regarding the scan of the data blocks (typically subbands) with the same spatiotemporal resolution is decided. [00129] - Morphological dilation block ("Dm") : the coefficients of the LCS list are analyzed and a morphological dilation operation is realized around them with the objective to find clustered coefficients. Every new found significant coefficient comes added to the LCS and the analysis continues until reaching the end of the LCS. The non-significant found coefficients are necessarily on the boundaries of the analyzed clusters and are placed in the LCC list. The significance state of the coefficients and their sign information are coded. [00130] - Multiresolution prediction block ("Pm") : in the multiresolution representations they are usually present inter-band (inter-scale) dependencies from which to take advantage for a more effective compression. To this end filiation

relationships can be constructed with respect to "father" coefficients at an inferior layer of resolution. In this case, once the task of the previous block is finished, new significant coefficients are searched with a priority criterion, among the possible sons of already found significant fathers. This makes the search of new (clusters of) significant coefficients more effective as compared to a simple sequential or random search. As shown in the scheme, once a new significant coefficient is found in the "Pm" step, the procedure of morphologic dilation Dm (only relating to the cluster being examined rather than the entire list LCS) is immediately recalled, in that it is that one that turns out to be more effective in terms of coding performance. [00131] - Block of morphological dilation of the borders ("Dmc") : always with reference to multiresolution representations and in particular to wavelet representations, it has been noticed that the cluster of significant coefficients often have an "archipelago" structure, that is they are agglomerate of connected coefficients (cluster) with some peripheral parts that, even not necessarily "strictly connected", are however located in the immediate vicinities of the main or considered cluster. This evidence suggests a second strategy for the search of new significant coefficients. Therefore, when the actions of the previous blocks are exhausted, new significant coefficients are not searched arbitrarily or sequentially but the boundaries of the already found clusters are first considered (the positions of such boundary coefficients are listed in the LCC) and a morphologic dilation is executed on them. Again, when a significant coefficient is found, the procedure of morphological dilation Dm (only relating to the cluster being examined rather than to the entire list LCS) is immediately launched.

[00132] - Coding block of the significant coefficient position ("CpCs") : when the capability of prediction of the algorithm is exhausted, or at the beginning of the coding, it is necessary to explicitly send the coordinates of the significant coefficients. Again, when the position of a significant coefficient is coded, the procedure of morphological dilation Dm (only relating to the cluster being examined rather than to the entire list LCS) is immediately launched.

[00133] - Bit-plane refinement block ("R") : once the whole significant coefficients have been determined for the current bit plane, all the coefficients found to be significant at the previous bit plane to return them to the accuracy of the current bit plane.

[00134] - Lossless entropic coding block ("CEsp") : the CSsp entropic coding associated to each processing block has the role of eliminating the statistical redundancies within the chains of symbols generated by every block. Generally, given the spatiotemporal data structure, contexts based adaptive arithmetic coding techniques (selection of the probability tables according to coding state and to the state of a causal neighbourhood of coefficients of the coefficient under examination) are adopted. Each symbol issued by the described algorithm is coded through the CEsp. Even though such block can be theoretically omitted, it serves to increase the coding performance. If the block CEsp is an entropic encoder (adaptive) based on contexts, besides generating the bits relating to significance map and to the refinement values, the functional blocks must generate the information relating to the context. [00135] - STOP condition: The process finishes once caught up a STOP condition. This can be determined by the completion of a given bit-plane (ending oriented to the precision) or the achievement of the amount of assigned bit (ending oriented to the amount of available bits) . Even if not explicitly indicated in the scheme, it is also possible to constantly supervise the quality of the decoded data (or an estimation of it) and to arrest the coding process on the base of a decoded data quality criterion. [00136] Once assessed that the operation Dm is the most advantageous one, the order of the operations Pm, Dmc, CpCs and R can be modified globally or within each bit plane based on predetermined empirical observations, as described in [F. Lazzaroni, R. Leonardi and A. Signoroni - High-performance embedded morphological wavelet coding - IWDC 2002, pp.319-326, Capri, Italy, Sep.2002] , or based on an analysis on the specific data to be coded. [00137] - Spatiotemporal clusterization engine block ("mCSP") : EMDC coding technique has been up to now only conceived for still (or fixed) data. On the contrary, through an analysis made by the block "mCSP" , the CPDMM- GFRTO technique:

[00138] 1) manages problems of joint coding of groups of any number of temporal frames, preferably of a same temporal layer, [00139] 2) can adapt its algorithmic structure to the characteristics of the data to be coded, as regards the type and number of spatiotemporal transformations undergone, [00140] 3) manages spatiotemporal prediction problems of significance and can adapt its algorithmic structure to the complexity requirements and to spatial scalability requirements, in particular:

[00141] 3a) it can manage simple mechanisms of prediction of significant coefficient clusters among different layers of spatial resolution starting from lower spatial resolution layers,

[00142] 3b) it can manage simple mechanisms of prediction of significant coefficient clusters among temporal frames starting from frames at a lower or equal spatial resolution.

[00143] The algorithm cycles on each bit-plane and considers for each functional step all the subbands according to a scanning order that can be decided freely on the group of frames, respecting the prediction principles to be used. [00144] In the video coding, temporal subbands that might be subject to spatial transforms are coded. [00145] The two dashed functional blocks (rectangular) (Pm and Dmc) in Figure 14 typically intervene in case of incoming coefficients organised in a multi-resolution structure (such as if a wavelet transform is used for the spatial decorrelation, as indicated in the preferred embodiment) . For representations that have not a spatial multi resolution structure, for which the clusterization hypothesis of the transformed coefficients can be applied

(e.g. coefficients generated by a temporal transformation) it is possible to simply eliminate the dashed blocks Pm and Dmc.

[00146] In general, the block Pm will be eliminated when there are no inter-resolution relations to exploit. [00147] In general, the block Dmc will be eliminated when the hypothesis of archipelago clusterization (around a main cluster there are peripheral clusters not strictly connected) is not useful for an efficient prediction and thus for an improvement of the coding performance.

[00148] An ulterior aspect of the invention regards a coding stream partitioning representation system that could be adequate to being submitted to the decoding method of the present invention. [00149] The above system comprises indexing mechanisms that permit a coding stream partitioning with the aim of actually allowing to determine the stream that must be associated to a certain layer of quality, for a given working point in term of spatial and temporal resolution, and that possibly allow to satisfy some given maximum coding-decoding delay requirements.

[00150] In a scalable coding system, a certain portion of the higher working point related coded stream, must be "extracted" and transmitted to the decoder. The extraction operation must be simple (does not have to require decoding or transcoding operations) in that the scalability property of the coded stream must guarantee an immediate extraction with the single aid of simple heading information for the navigation and the structural scanning of the coded video stream.

[00151] Said that, at least two typologies of extraction should be considered: 1) single extraction; 2) multiple extraction. [00152] 1) single extraction: the above described extraction operation is intended as single extraction, where the extracted stream is decodable and corresponds the desired working point. From the extracted stream it is not normally possible to execute others optimal extractions to inferior working points. [00153] 2) multiple extraction: this case is referred to an extraction which preserves the necessary- information in order to allow a later extraction of the widest possible subset, or alternatively, of a predetermined set, of inferior working points starting from the extracted bit-stream.

[00154] It will be. now described an example of a construction and extraction mechanism of a scalable coded video bit-stream. In particular, in the scheme of Figure 15 an example of organization of a coded stream is proposed (block "GOP") according to a GOP (Group of Pictures) based subdivision of a whole video sequence. The first characterizing subdivision (at the GOP or whole video sequence level) is that one corresponding to the spatial resolution (block "Rs" in the example in figure, in which three spatial resolution layers are evidenced: QCIF, CIF and 4CIF) . Every spatial layer sub-stream is in turn subdivided in temporal layers (block "Lt") with the coded information relative to the motion fields associated to each of them (the example illustrates a subdivision of the CIF layer into four temporal layers) . Every temporal sub-stream can be cut with a criterion directed to obtain a certain global quality or a certain global bit-rate (more common situation) . While the headings for the access to the GOP and to the spatial layers only contain the positions of the beginning of stream fragments, in the case of the temporal layers, beyond to the access information, an additional information can turn out necessary to transmit in order to inform the extractor on how and where cutting the several temporal layer bit-streams to the aim to optimize the reconstruction quality for an amount of assigned bit

(bit budget) .

[00155] Moreover, it is possible to add an ulterior level of subdivision in groups of frames (block "GFn") of same temporal layer. Such subdivision, slightly disadvantageous in terms of coding efficiency, could reveals more useful in applications that demand a low delay between the absolute coding and decoding timings (i.e. the temporal delay, not considering transmission and computational contributions, between the frame coding time and the time when the same frame is available after its decoding) . The number of subdivisions for every temporal layer depends on the desired characteristics in terms of data access and the adopted type of MCTF. [00156] In Figure 16 an exemplification of the extractor achievement is shown. The example regards an extraction of the spatial resolution CIF at temporal frame-rate of 15 frame per second (fps) and assigned bit- rate, starting from a global bit-stream related to a 4CIF and 60 fps coding. The portions of bit-stream that are completely eliminated and those, associated to the layers to maintain, which are cut can be noticed as those crossed out. The dashed cross on the headers that contain the truncating instructions can be noticed. In fact, such elimination only happens in the case of single extraction modality.

[00157] According to a preferred embodiment, in the case of the previously described entropic coding the information that can be useful to insert in the temporal layer headings, as regards the truncation instructions, can consist in the partial lengths of the coded stream for every CPDMM-GFRTO coding step and every bit-plane. In fact, with these information, the extractor is capable to cut the sub-streams in an homogeneous and optimized way according to some distortion minimization criterion.

[00158] Figure 17 shows a possible structure (as a three-dimensional array, that is, a set of bi-dimensional tables) of the truncation instructions referred to the coding method CPDMM-GFRTO: the coding algorithm can create the generic element l $\pm_{jk}$ of the table, where

[00159] • "i" indexes the set (or a subset A) of the coding steps of the algorithm CPDMM-GFRTO (or of another algorithm compatible with the present preferred embodiment) [00160] ˙ "j" indexes the bit-planes (or sets of bit-plane or similar division based on a coefficient amplitude criterion) starting from the most significant ones to the least significant ones (in the example j is from M to 0) , [00161] • "k" determines the temporal layer of the GoF (if for each temporal layer there are multiple GoF, the index k will take a suitable vectorial form) on a range from 0 (lower temporal layer) to T (maximum temporal layer relating to the current extraction) . [00162] The generic element li_{jk} contains, in a convenient format for a compact description, the

information of the length of the coded bit-stream relating to the GoF k and to the coding step determined by the two values (i,j) . In particular, we shall consider the value of li_{ij} equal to the difference $L(i,j),k- L(i,j)-l,k$ between the length of the bit-stream of the k- th GoF at the coding step (i,j) and the same length at the immediately preceding coding step "(i,j)-l" (note that the tables of each GoF in Figure 17 are filled by lines starting from top) . When the extractor has to cut on more GoF' s according to an overall length constraint LC for a certain layer of spatial resolution, an advantageous method for respecting such constraint while obtaining information on the cutting lengths LT,k of the bit-streams is as follows: [00163] L=O

[00164] Flag=0

[00165] for j from M to 0

[00166] for i on the elements of A contained in {Dm, Pm, Dmc , CpCs , R}

[00167] for k from 0 to T

[00168] $L=L+l_{ijk}$

[00169] If Flag = 1 then l_{ijk}=0

[00170] If L greater than or equal to LC then [00171] $li_{jk} = l_{ijk} - (L-LC)$

[00172] Flag = 1

[00173] end If

[00174] end for k

[00175] end for i [00176] end for j

[00177] for k from 0 to T

[00178] LT, k= summation for each (i , j) of the li_{jk}

[00179] end for k

[00180] Following a cutting operation, the portion of tables corresponding to the cut zones can be deleted so as to minimise the occupation of the information itself .

[00181] Finally, it should be noted that with multiple extractions , or if it is necessary to make further extractions on an already extracted stream, the headers relating to the partial lengths of the coded stream are updated and stored.

NON-PATENT CITATIONS

Reference

1	*	See references of WO2006040793A2

* Cited by examiner

CLASSIFICATIONS

International Classification	H04N7/26
Cooperative Classification	H04N19/36, H04N19/63, H04N19/132, H04N19/615, H04N19/31, H04N19/13, H04N19/146, H04N19/46, H04N19/61, H04N19/647, H04N19/187, H04N19/33
European Classification	H04N19/00C4, H04N19/00C1, H04N7/26A8Y, H04N7/26E6, H04N7/26A6E, H04N7/26H50A, H04N7/26A4Z, H04N7/26H30H6, H04N7/26A10S

LEGAL EVENTS

Date	Code	Event	Description
25 Jul 2007	AK	Designated contracting states:	**Kind code of ref document:** A2 **Designated state(s):** AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR
25 Jul 2007	17P	Request for examination filed	**Effective date:** 20070514
23 Jan 2008	DAX	Request for extension of the european patent (to any country) deleted	
16 Jul 2014	17Q	First examination report	**Effective date:** 20140612
23 Jul 2014	RAP1	Transfer of rights of an ep application	**Owner name:** ADAMI, NICOLA **Owner name:** LEONARDI, RICCARDO

Owner name: SIGNORONI, ALBERTO

3 Jun 2015	18D	Deemed to be withdrawn	**Effective date:** 20141222
14 Oct 2015	R18D	Ep-application deemed to be withdrawn: (correction)	**Effective date:** 20141223

Google Home - Sitemap - USPTO Bulk Downloads - Privacy Policy - Terms of Service - About Google Patents - Send Feedback

Data provided by IFI CLAIMS Patent Services