

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-506355

(P2004-506355A)

(43) 公表日 平成16年2月26日(2004.2.26)

(51) Int. Cl.⁷

H04N 7/30

H03M 7/30

F I

H04N 7/133

H03M 7/30

テーマコード (参考)

5C059

5J064

審査請求 未請求 予備審査請求 未請求 (全 22 頁)

(21) 出願番号 特願2002-518087 (P2002-518087)
 (86) (22) 出願日 平成13年7月23日 (2001.7.23)
 (85) 翻訳文提出日 平成14年4月5日 (2002.4.5)
 (86) 国際出願番号 PCT/EP2001/008539
 (87) 国際公開番号 W02002/013536
 (87) 国際公開日 平成14年2月14日 (2002.2.14)
 (31) 優先権主張番号 00402246.3
 (32) 優先日 平成12年8月8日 (2000.8.8)
 (33) 優先権主張国 欧州特許庁 (EP)
 (81) 指定国 EP (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), CN, JP, KR

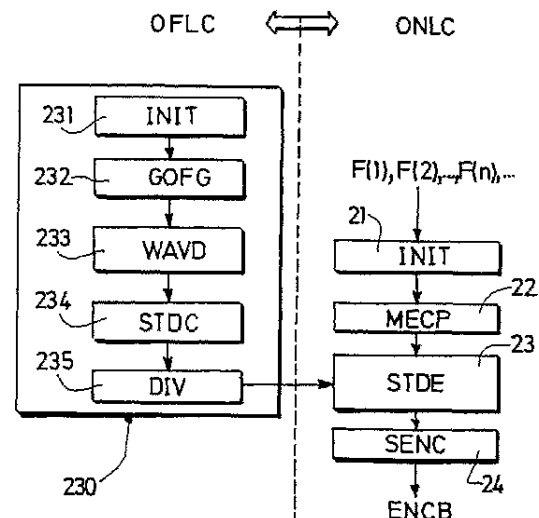
(71) 出願人 590000248
 コーニンクレッカ フィリップス エレクトロニクス エヌ ヴィ
 Koninklijke Philips Electronics N. V.
 オランダ国 5621 ペーアー アインドーフエン フルーネヴァウツウェッハ 1
 Groenewoudseweg 1, 5621 BA Eindhoven, The Netherlands
 (74) 代理人 100087789
 弁理士 津軽 進
 (74) 代理人 100114753
 弁理士 宮崎 昭彦

最終頁に続く

(54) 【発明の名称】 ウェーブレット分解に基づくビデオ符号化方法

(57) 【要約】

本発明は、フレーム群に再分割されるビデオシーケンスを符号化する方法に関する。この方法は、三次元ウェーブレット分解が最適の重み付け定数を使用してリフティング方式で双直交フィルタバンクに関連して実施されるオンラインプロシーダを含む。これらの定数は、いかなる重み付け定数も用いずに同様の分解が実施される付加的なオフラインプロシーダにより決定される。



【特許請求の範囲】

【請求項 1】

フレーム群に再分割されるビデオシーケンスに適用されるとともに、
符号化プロセスに関して入力パラメータを規定するための初期化ステップと、
連続するフレームの対の間で動き推定及び補償を実施するための動き推定／補償ステップと、
前記初期化ステップで規定される重み付け定数を使用していわゆるリフティング方式で双直交フィルタバンクを使用して三次元ウェーブレット分解を実行するための空間的・時間的分解ステップと、
このように得られた変換係数を符号化するための符号化ステップと、
を含むオンラインプロシージャに基づく符号化方法であって、前記符号化方法が更に、
前記重み付け定数を除き、オフラインプロシージャに関して同じタイプの入力パラメータを規定するサブステップと、
生成されるフレーム群が、元のシーケンスの表現に適応化された平均値及び標準偏差の白色ガウスノイズを含む、ランダム的なフレーム群生成サブステップを実行するサブステップと、
前記リフティング方式に基づき、いかなる重み付け定数ももたない同じフィルタバンクを使用して、三次元ウェーブレット分解を実現するサブステップと、
前記三次元分解に起因する空間的・時間的サブバンドの標準偏差を計算するサブステップと、
前記標準偏差をノイズの標準偏差で除算するサブステップと、
を含むオフライン計算からなるオフラインプロシージャを含み、最適な重み付け定数は、前記除算サブステップの出力で利用可能であり、前記空間的・時間的サブバンドの出力を重み付けすることを可能にするために前記フィルタバンクに向けて送られることを特徴とする、符号化方法。

10

20

【請求項 2】

前記符号化方法がいわゆる S P I H T アルゴリズムである、請求項 1 に記載の符号化方法。

【請求項 3】

0 から 255 までの値をもつ 8 ビットで表現される元のビデオシーケンスに関して、輝度成分について 128 の平均値及び 32 の標準偏差が使用され、クロミナンス成分について 128 の平均値及び 20 の標準偏差が使用される、請求項 1 又は 2 に記載の符号化方法。

30

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、フレーム群（G O F）に再分割されるビデオシーケンスに適用される符号化方法であって、
一符号化プロセスに関して入力パラメータを規定するための初期化ステップと、
一連続したフレームの対の間で動き推定及び補償を実施するための動き推定／補償ステップと、
一初期化ステップで規定された重み付け定数を使用していわゆるリフティング方式で双直交フィルタバンクを使用して 3 次元（3 D）ウェーブレット分解を実行するための空間的・時間的分解ステップと、
一こうして得られた変換係数を、3 D 分解に適応化された符号化方法によって符号化するための符号化ステップと、
を含むオンラインプロシージャに基づく符号化方法に関する。

40

【0002】

【従来の技術】

ビデオ圧縮方式で、時間冗長性の低減は、主に 2 つのタイプの方法によって達成される。
第 1 の方法であるいわゆる「ハイブリッド」又は予測的方法では、現在のフレームの予測

50

は、以前に伝送されたフレームに基づいて計算され、予測誤差のみがイントラ符号化され、伝送される。第2の方法において、時間冗長性は、時間的変換である3D（又は2D+t）方法によって利用される。この方法は、冗長性を取り除くための空間的技法と同様である。この後者の方法によれば、フレームのシーケンスは、3Dボリュームとして処理される、こうして、画像符号化において多く使用される古典的なサブバンド分解は、分離可能な変換（例えばフィルタバンクにより実現されるウェーブレット又はウェーブレットパケット変換）を使用することにより3次元の空間的・時間的データに拡張されることができ、3D構造には異方性があることは明らかであるが、これは、時間的及び空間的な方向において異なるフィルタバンクを使用することにより考慮することができる（通常、より長いフィルタを使用することの付加的な遅延は望ましくないため、ハール（Haar）フィルタが時間フィルタリングのために使用される。更にこれらは2タップフィルタであり、境界効果を示さない唯一の完全な再現直交フィルタである）。

10

【0003】

この3D符号化方式の符号化効率は、時間分解のそれぞれのレベルで、低い時間的サブバンドで動き推定／補償を実施することにより改善されることができ、3Dサブバンド分解は、補償されたフレーム群に適用される（このフレーム群は、2の累乗のフレーム数を含まなければならない、通常は16である）。最後の時間分解レベルでは、最も低い時間的サブバンド内に2つのフレームがある。時間的サブバンドのそれぞれのフレームでは、空間分解が実施される。

【0004】

20

データの3次元構造をサブバンド符号化することは、空間的なサブバンド符号化技法の拡張として実現されることができ、画像圧縮のための最も効果的なウェーブレットに基づく方式の1つは、A. Said及びW. A. Pearlmanによる「A new, fast, and efficient image codec based on set partitioning in hierarchical trees (= SPIHT)」(IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, June 1996, pp. 243-250)に詳細に記載された2D SPIHTアルゴリズムに基づくものであり、最近では3D構造にまで拡張されている。3D符号化技法において使用される基礎的な概念は、以下の通りである。同じ位置に対応する空間的・時間的ツリーが、ウェーブレットドメインにおいて形成され、これらのツリーにおけるウェーブレット変換係数は、それらの大きさのビットプレーン表現における最高位ビットのレベルにより規定される組に区分される。最終的に、最上位の残りのビットプレーンが符号化され、結果として得られたビットが伝送される。

30

【0005】

元のSPIHTアルゴリズムは、直交分解の仮説に基づくものであり、この仮説によれば、再現誤差は、サブバンド歪みの和として測定される量子化誤差に等しい（こういうわけで、各サブバンドのエネルギーに基づきビットバジェットを分散することが可能である）。更に、画像及びビデオ符号化における最善の結果は、直交フィルタを使用して達成されるのではなく、双直交フィルタを使用して達成されることが確認された（これは、双直交フィルタの対称性がペアレントと同じ空間位置にウェーブレット係数を生じさせるという事実による）。しかしながら、双直交フィルタは量子化誤差の L^2 ノルムを保たないので、直交変換について導き出されるビット配分（repartition）は最小の再現誤差をもたらさない。

40

【0006】

更に、双直交フィルタは乗算定数まで決定される。従って、完全な再現条件だけが課されるときには無限数のフィルタバンクが設計されうる。フィルタバンクのリフティングインプリメンテーションが使用される場合、ポリフェーズマトリクス $P(z)$ は常にファクタ（因数）に分解することが可能であることが知られている。

【数1】

50

$$P(z) = \prod_{i=1}^m \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix}$$

ここで、 t_i 及び s_i は、L a u r e n t 多項式であり、 K は実定数である。この場合、合成部分に対応するデュアルポリフェーズマトリクスは次式により与えられる。

【数 2】

$$\tilde{P}(z) = \prod_{i=1}^m \begin{bmatrix} 1 & 0 \\ -s_i(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -t_i(z^{-1}) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/K & 0 \\ 0 & K \end{bmatrix}$$

10

下式により、すべての K について完全な再現が保証される。

【数 3】

$$P(z)\tilde{P}(z^{-1})' = I$$

【0 0 0 7】

【発明が解決しようとする課題】

この分解は、図 1 及び図 2 に示されるデュアル方式を通して実現される。従って、異なる因数分解及び定数 K の異なる値に対応して無限数のインプリメンテーションが存在する。符号化の観点から、これらのフィルタバンクがすべて同等というわけではない。この乗算定数を選ぶための通常の基準は、1 に等しいポリフェーズマトリクスの行列式を課すことであるが、この選択は、符号化性能にとって最善ではないことがある（こういうわけで、大部分の通常のアルゴリズムは概ね直交する双直交変換に基づく）。

20

【0 0 0 8】

従って本発明の目的は、3 D ウェーブレット分解及び自動ビット割当て機構に基づいて、ビデオ符号化方式に関して高度な非直交ウェーブレット変換の適用可能性を改善するための技法を提案することである。

【0 0 0 9】

【課題を解決するための手段】

30

このために、本発明は、この説明の冒頭部分に述べたような符号化方法であって、この符号化方法が更に、

一重み付け定数を除いて、オフライン計算に関して同じタイプの入力パラメータを規定し、

一生成される G O F が元のシーケンスの表現に適応化される平均値及び標準偏差の白色ガウスノイズを含む、ランダム的な G O F 生成サブステップを実行し、

一前記リフティング方式に基づき且ついかなる重み付け定数ももたない同じフィルタバンクを使用して、3 D ウェーブレット分解を実現し、

一前記 3 D 分解に起因する空間的時間のサブバンドの標準偏差を計算し、

一前記標準偏差をノイズの標準偏差で除算する、サブステップを含むオフライン計算からなるオフラインプロシージャを含み、最適の重み付け定数が、前記除算サブステップの出力において利用可能であり、空間的時間のサブバンドの出力を重み付けすることを可能にするために前記フィルタバンクに向けて送られることを特徴とする、符号化方法に関する。

40

【0 0 1 0】

こうして、サブバンドにおけるエネルギー配分に基づいて、3 D 分解における双直交フィルタに関する最適の重みが決定される。画像符号化のフレームワークにおいてこれらの定数を選ぶために、緩和アルゴリズムに基づくいくつかの複雑な方法が文献において既に使用されており、例として P. M o u l i n による「A m u l t i s c a l e r e l a x a t i o n a l g o r i t h m f o r S N R m a x i m i z a t i o n i n n o

50

n-orthogonal subband coding」(IEEE Transactions on Image Processing, vol. 4, n° 9, September 1995, pp. 1269-1281)に記載されている方法がある。再現誤差を主観的により許容可能なものにするために人間の視覚システムを利用する技法が存在し、例えばP. Desarte他による「Signal-adapted multi resolution transform for image coding」(IEEE Transactions on Information Theory, vol. 38, n° 2, March 1992, pp. 897-904)に記載されているような技法がある。しかし、すべてのこれらの既存のアルゴリズムは、画像符号化に関連している。ここで提案される方法、つまり以前の研究の3Dウェーブレット分解のケースへの拡張は、ビデオ符号化のためのものである。この方法は、いかなるタイプの線形フィルタに適用されてもよく、具体的には動き補償された時間フィルタリングのフレームワークにおいて適用されてもよい。

【0011】

【発明の実施の形態】

以下、本発明について添付の図面を参照して例示により説明する。

【0012】

図1には、SPIHTのような符号化アルゴリズムがあとに続く3Dウェーブレット分解に基づく従来のビデオ符号化方式が示されている。連続したフレームF(1)、F(2)、…、F(n)、…を含む元のビデオシーケンスは、ガウスの白色ノイズ(通常の法則に従う独立同分布の確率変数)を含む、圧縮する際に使用されるものと同じサイズのフレーム群(GOF、group of frames)に再分割される。入力パラメータが、初期化ステップ11(INIT)で規定される(GOFフレームの数、ウェーブレットタイプ、サブ最適の重み)。符号化方式は更に、動き推定/補償ステップ12(MECP, motion estimation/compensation)と、サブ最適定数をもつフィルタバンクを使用する3D空間的時分解ステップ13(STDE, spatio-temporal decomposition)と、SPIHTのような符号化ステップ14(ENCL)と、を連続して含む。符号化ステップ14の出力において、符号化されたビットストリーム(ENCB)が利用可能となる。

【0013】

本発明による包括的なビデオ符号化方式が図2に示されている。このビデオ符号化方式は、上述したステップ11乃至14と同様のステップ21、22、23、24を含み、以下のやり方で変更されている。図1の連続したステップに従って実行されるオンライン計算ONLCに関して、3D空間的時分解ステップで使用される重み付け係数の計算処理を最適化するために、オフライン計算が導入される。この分解ステップ23は、最適の定数をもつ重み付けされたフィルタバンクを使用する。図2の左側部分のモジュール230に表されているこれらのオフライン計算OFFLCについてここで説明する。

【0014】

まず、図1の方式に関して上述したように、入力パラメータが、規定サブステップ231(INIT)の間に規定され(GOFフレーム数、ウェーブレットタイプ、…)、そののちランダムなGOF生成サブステップ232(GOFG)が続く。ガウスのホワイトノイズに関して、元の画像は、8ビットで(すなわち0から255までの範囲の値により)表現されるので、128に等しい平均値及び32に等しい標準偏差をもつガウス分布が選択された。従って、それぞれのフレーム内の値は、以下の方程式を満たす。

$$F(m, n) = (32 \times N(m, n)) + 128$$

ここで、N(m, n)は、ゼロ平均値及び一標準偏差をもつガウスのホワイトノイズである。この条件は、ほぼすべての生成された値が係数の入力値が位置する間隔(0, 255)の中にあることを確実にする(それにもかかわらず、係数のダイナミックレンジのほぼ全体を便利な信頼区間を含める条件が満たされるとすぐに、分布の平均値及び標準偏差の正確な値は重要でなくなることに注意されたい)。

【0015】

そののうちウェーブレット分解サブステップ233 (WAVD) が与えられる。このステップは、例えばI. Daubechies及びW. Sweldensによる「Factoring wavelet transforms into lifting steps」(Bell Laboratories technical report, Luce nt Technologies, 1996) に記載されているリフティング方式を使用する。ランダム的に生成されたGOFは、いかなる重み付け定数ももたないフィルタ ($K=1$) を使用して分解される。そののうちサブステップ234 (STDC) は、結果として得られる空間的時間的サブバンドの標準偏差を計算することを可能にする。これらの標準偏差は、除算サブステップ235 (DIV) の出力において分解ステップ23に送信されるべき最適の重み付けを得るために、32 (ノイズの標準偏差) で除算される (実際に、各サブバンドにおける係数を当該サブバンドに対応する標準偏差で除算することにより、それぞれのサブバンドの一標準偏差を与える。これは、正規直交分解の場合のように、各サブバンドにおけるエネルギーが1に等しいことを意味する)。

【0016】

この最適のエネルギー配分は、最終的に、正規直交分解によって暗示されるものと同じ仮説の下で、符号化ステップ24においてSPIHTアルゴリズムを使用することを可能にする。係数は、ビットプレーン毎に送信され、再現誤差は、量子化誤差に等しい。これは、所与のフィルタバンクについて最大の符号化利得をもたらす。

【0017】

上述したように、本技法は、重み付け係数を計算するために、動き推定及び補償をもたない元のランダムGOFに適用された。これらの定数が一旦得られると、空間的時間的分解が動き補償されたGOFに適用される。この技法は、時間分解が動きの方向において直交する必要があるという事実により正当化される。実際に、動き補償の後に時間フィルタリングを実施することにより、同じ軌跡に属するポイントのみが共にフィルタリングされる。

【図面の簡単な説明】

【図1】 SPIHTのような符号化アルゴリズムがあとに続く3Dウェーブレット分解に基づく包括的なビデオ符号化方式を示す図。

【図2】 本発明により、3D空間的時間的分解に関して最適の重み付け定数を計算するために実行される連続する演算に再分割されるフィルタバンク最適化ステップを含む包括的なビデオ符号化方式を示す図。

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/13536 A2

- (51) International Patent Classification: H04N 7/26 (74) Agent: LANDOUSY, Christian; Internationaal Octroibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (21) International Application Number: PCT/EP01/08539
- (22) International Filing Date: 23 July 2001 (23.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 00402246.3 8 August 2000 (08.08.2000) EP
- (71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. (NL/NL); Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).
- (72) Inventor: PESQUET-POPESCU, Beatrice; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (81) Designated States (national): CN, JP, KR.
- (84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/13536 A2

(54) Title: VIDEO ENCODING METHOD BASED ON A WAVELET DECOMPOSITION

(57) Abstract: The invention relates to a method for encoding a video sequence subdivided into groups of frames. This method comprises an on-line procedure in which a three-dimensional wavelet decomposition is performed involving a biorthogonal filter bank in a lifting scheme using optimal weighting constants. These constants are determined thanks to an additional off-line procedure in which a similar decomposition is performed but without any weighting constants.

WO 02/13536

PCT/EP01/08539

Video encoding method based on a wavelet decomposition

FIELD OF THE INVENTION

The present invention relates to an encoding method applied to a video sequence subdivided into groups of frames (GOFs) and based on an on-line procedure that comprises the steps of :

- 5 - an initializing step, provided for defining for the encoding process its input parameters ;
- a motion estimation/compensation step, provided for performing a motion estimation and compensation between pairs of successive frames ;
- a spatio-temporal decomposition step, provided for carrying out a three-
10 dimensional (3D) wavelet decomposition using a biorthogonal filter bank in a so-called lifting scheme itself using weighting constants defined in the initialization step ;
- an encoding step, provided for encoding the transform coefficients thus obtained, by means of an encoding method adapted to 3D decompositions.

15

BACKGROUND OF THE INVENTION

- In video compression schemes, the reduction of temporal redundancy is mainly achieved by two types of approaches. In the first one, the so-called "hybrid" or predictive approach, a prediction of the current frame is computed based on the previously
20 transmitted frames, and only the prediction error is intra-coded and transmitted. In the second one, the temporal redundancy is exploited by means of a temporal transform, the 3D (or 2D+t) approach, which is similar to spatial techniques for removing redundancies. According to this last approach, the sequence of frames is processed as a 3D volume, and the classical subband decomposition often used in image coding can thus be extended to 3D spatio-
25 temporal data by using separable transforms (for example, wavelet or wavelet packets transforms implemented by means of filter banks). Obviously, there is, in the 3D structure, an anisotropy, which can however be taken into account by using different filter banks in the temporal and spatial directions (usually, Haar filters are used for temporal filtering since the added delay of using longer filters is undesirable ; furthermore, they are two-tap filters and

WO 02/13536

PCT/EP01/08539

2

are the only perfect reconstruction orthogonal filters which do not present the boundaries effect).

The coding efficiency of this 3D coding scheme can be improved by performing motion estimation/compensation in the low temporal subbands, at each level of the temporal decomposition. The 3D subband decomposition is applied on the compensated group of frames (this group of frames must contain a power of two number of frames, usually 16), and, at the last temporal decomposition level, there are two frames in the lowest temporal subband. In each frame of the temporal subbands, a spatial decomposition is performed.

Subband coding the three-dimensional structure of data can be realized as an extension of the spatial subband coding techniques. One of the most effective wavelet-based scheme for image compression is based on the 2D SPIHT algorithm, described in a detailed manner in "A new, fast, and efficient image codec based on set partitioning in hierarchical trees (= SPIHT)", by A. Said and W.A. Pearlman, IEEE Transactions on Circuits and Systems for Video Technology, vol.6, June 1996, pp.243-250, and recently extended to the 3D structures. The basic concepts used in the 3D coding technique are the following : spatio-temporal trees corresponding to the same location are formed in the wavelet domain, then the wavelet transform coefficients in these trees are partitioned into sets defined by the level of the highest significant bit in a bit-plane representation of their magnitudes, and, finally, the highest remaining bit planes are coded and the resulting bits are transmitted.

The original SPIHT algorithm is based on the hypothesis of an orthogonal decomposition, according to which the reconstruction error is equal to the quantization error measured as the sum of subband distortions (that is why it is possible to distribute the bit budget based on the energy of each subband). It has also been shown that the best results in image and video coding are not achieved using orthogonal filters, but biorthogonal ones (this is due to the fact that the symmetry of biorthogonal filters yields wavelet coefficients at the same spatial location as their parents). However, as biorthogonal filters do not preserve the L^2 norm of the quantization error, the bit repartition deduced for an orthogonal transform would not lead to the minimum reconstruction error.

Moreover, biorthogonal filters are determined up to a multiplicative constant and therefore an infinity of filter banks may be designed when only the perfect reconstruction condition is imposed. If a lifting implementation of the filter bank is used, it is known that the polyphase matrix $P(z)$ can always be decomposed into factors

WO 02/13536

PCT/EP01/08539

3

$$P(z) = \prod_{i=1}^m \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix},$$

where t_i and s_i are Laurent polynomials and K is a real constant. The dual polyphase matrix, corresponding to the synthesis part, is given in this case by :

5

$$\tilde{P}(z) = \prod_{i=1}^m \begin{bmatrix} 1 & 0 \\ -s_i(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -t_i(z^{-1}) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/K & 0 \\ 0 & K \end{bmatrix}.$$

The perfect reconstruction is insured for all K , since :

10

$$P(z)\tilde{P}(z^{-1})^t = I$$

This decomposition is implemented through the dual schemes illustrated in Figs 1 and 2.

There exists therefore an infinity of implementations, corresponding to different

factorizations and different values of the constant K . From the coding point of view, these

15 filter banks are not all equivalent. A usual criterion for choosing this multiplicative constant

is to impose the determinant of the polyphase matrix to be equal to 1, but this choice may not

be the best for coding performances (that is why most of the usual algorithms are based on

biorthogonal transforms that are nearly orthogonal).

20

SUMMARY OF THE INVENTION

It is therefore an object of the invention to propose a technique for improving the applicability of highly non-orthogonal wavelet transforms for video coding schemes, based on a 3D wavelet decomposition and an automatic bit allocation mechanism.

25

To this end, the invention relates to an encoding method such as defined in the introductory part of the description and which is moreover characterized in that it also comprises an off-line procedure that consists of off-line computations including the sub-steps of :

- defining for these off-line computations the same type of input parameters,

30

with the exception of the weighting constants ;

WO 02/13536

PCT/EP01/08539

4

- carrying out a random GOF generation sub-step, the generated GOF containing a white, Gaussian noise of mean and standard deviation adapted to the representation of the original sequence ;
- implementing a 3D wavelet decomposition based on said lifting scheme and using the same filter bank without any weighting constants ;
- computing the standard deviations of the spatio-temporal sub-bands resulting from said 3D decomposition ;
- dividing said standard deviations by the standard deviation of the noise, optimal weighting constants being available at the output of said division sub-step and sent towards said filter bank in order to allow to weight the output of the spatio-temporal sub-bands.

Based on the energy repartition in subbands, the optimal weights for the biorthogonal filters in a 3D decomposition are thus determined. Some complicated methods based on relaxation algorithms are already used in the literature for choosing these constants in the framework of image coding, for instance the method described in "A multiscale relaxation algorithm for SNR maximization in non-orthogonal subband coding", by P. Moulin, IEEE Transactions on Image Processing, vol.4, n°9, September 1995, pp.1269-1281. Techniques exploiting the human visual system in order to make the reconstruction error subjectively more acceptable also exist, such as the one described in "Signal-adapted multiresolution transform for image coding", by P. Desarte and al, IEEE Transactions on Information Theory, vol.38, n°2, March 1992, pp.897-904. However, all these existing algorithms refer to image coding, while the method here proposed - an extension of previous works to the case of a 3D wavelet decomposition - is dedicated to video coding. This method may be applied to any type of linear filter, in particular in the framework of a motion compensated temporal filtering.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawings in which :

Fig. 1 is a global video coding scheme based on a 3D wavelet decomposition followed by a SPIHT-like coding algorithm ;

Fig. 2 gives the global video coding scheme, which, according to the

WO 02/13536

PCT/EP01/08539

5

invention, includes a step of optimization of the filter bank, itself subdivided into successive operations carried out for computing optimal weighting constants for the 3D spatio-temporal decomposition.

5

DETAILED DESCRIPTION OF THE INVENTION

A conventional video coding scheme based on a 3D wavelet decomposition followed by a SPIHT-like coding algorithm is shown in Fig.1. The original video sequence, comprising successive frames $F(1), F(2), \dots, F(n), \dots$, is subdivided into groups of frames (GOFs) of the same size as those used in compressing, containing a Gaussian white noise (random variables independent, identically distributed following a normal law). The input parameters being defined (number of GOF frames, wavelet type, sub-optimal weights) in an initialization step 11 (INIT), the coding scheme also comprises successively a motion estimation/compensation step 12 (MECP), a 3D spatio-temporal decomposition step 13 (STDE), using filter banks with sub-optimal constants, and a SPIHT-like encoding step 14 (SENC), at the output of which the encoded bitstream (ENCB) is available.

The global video coding scheme according to the invention, shown in Fig.2 and which comprises steps 21, 22, 23, 24 similar to the steps 11 to 14 is modified in the following manner : with respect to the on-line computations ONLC carried out according to the successive steps of Fig.1, off-line computations are introduced in order to optimize the computation operation of the weighting factors used in the 3D spatio-temporal decomposition step. This decomposition step, now referenced 23, uses weighted filter banks with optimal constants. These off-line computations OFLC, appearing in Fig.2 in the module 230, in the left part of the figure, are now described.

First, as previously for the scheme of Fig.1, the input parameters are defined (number of GOF frames, wavelet type, ...) during a definition sub-step 231 (INIT), followed by a random GOF generation sub-step 232 (GOFG). Concerning the Gaussian white noise, as the original images are represented with 8 bits (i.e. by values ranging from 0 to 255), a Gaussian distribution with a mean equal to 128 and a standard deviation equal to 32 has been chosen. The values inside each frame therefore satisfy the following equation :

$$F(m, n) = (32 \times N(m, n)) + 128$$

where $N(m, n)$ is a Gaussian white noise with zero mean and unit standard deviation. This condition ensures that almost all the generated values are inside the interval (0,255) in which the input values of the coefficients are located (it may nevertheless be noted that the exact

WO 02/13536

PCT/EP01/08539

6

values of the mean and standard deviation of the distribution are not essential, as soon as the condition to include almost the whole dynamic range of the coefficients into a convenient confidence interval is satisfied).

5 A wavelet decomposition sub-step 233 (WAVD) is then provided, that uses
for instance the lifting scheme described in "Factoring wavelet transforms into lifting steps",
I. Daubechies and W. Sweldens, Bell Laboratories technical report, Lucent Technologies,
1996. The random generated GOF is decomposed using the filters without any weighting
constants ($K = 1$). A sub-step 234 (STDC) then allows to compute the standard deviations of
the resulting spatio-temporal sub-bands, and these standard deviations are divided by 32 (the
10 standard deviation of the noise) in order to obtain at the output of a division sub-step 235
(DIV) the optimal weights to be sent to the decomposition step 23 (indeed, dividing the
coefficients in each sub-band by the standard deviation corresponding to the sub-band leads
to a unit standard deviation of each sub-band, which means that the energy in each sub-band
is equal to the unity, as it is the case with an orthonormal decomposition).

15 This optimal energy repartition finally allows to use - in the encoding step 24 -
the SPIHT algorithm under the same hypotheses as those implied by an orthonormal
decomposition : the coefficients are sent bitplane by bitplane, and the reconstruction error is
equal to the quantization error. This leads to maximum coding gains for a given filter bank.

As it has been seen, the present technique has been applied to the original
20 random GOF, without motion estimation and compensation, in order to compute the
weighting factors. Once these constants have been got, the spatio-temporal decomposition is
applied to the motion compensated GOF. This technique is justified by the fact that the
temporal decomposition has to be orthogonal in the direction of the movement. Indeed, by
performing the temporal filtering after the motion compensation, only points belonging to the
25 same trajectory are filtered together.

WO 02/13536

PCT/EP01/08539

7

CLAIMS:

1. An encoding method applied to a video sequence subdivided into groups of frames (GOFs) and based on an on-line procedure that comprises the steps of :
 - an initialization step, provided for defining for the encoding process its input parameters ;
 - 5 - a motion estimation/compensation step, provided for performing a motion estimation and compensation between pairs of successive frames ;
 - a spatio-temporal decomposition step, provided for carrying out a three-dimensional (3D) wavelet decomposition using a biorthogonal filter bank in a so-called lifting scheme itself using weighting constants defined in the initialization step ;
 - 10 - an encoding step, provided for encoding the transform coefficients thus obtained,
said method being characterized in that it also comprises an off-line procedure that consists of off-line computations including the sub-steps of :
 - defining for this off-line procedure the same type of input parameters, with
 - 15 the exception of the weighting constants ;
 - carrying out a random GOF generation sub-step, the generated GOF containing a white, Gaussian noise of mean and standard deviation adapted to the representation of the original sequence;
 - implementing a 3D wavelet decomposition based on said lifting scheme and
 - 20 using the same filter bank without any weighting constants ;
 - computing the standard deviations of the spatio-temporal sub-bands resulting from said 3D decomposition ;
 - dividing said standard deviations by the standard deviation of the noise, optimal weighting constants being available at the output of said division sub-step and sent
 - 25 towards said filter bank in order to allow to weight the output of the spatio-temporal sub-bands.
2. An encoding method according to claim 1, in which said encoding method is the so-called SPIHT algorithm.

WO 02/13536

PCT/EP01/08539

8

3. An encoding method according to anyone of claims 1 and 2, in which, for an original video sequence represented on 8 bits with values from 0 to 255, a mean of 128 and a standard deviation of 32 are used for the luminance component and a mean of 128 and a
5 standard deviation of 20 are used for the chrominance components.

WO 02/13536

PCT/EP01/08539

1/2

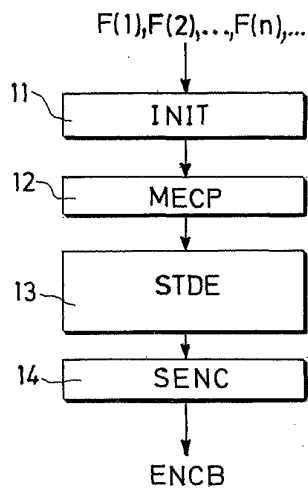


FIG.1

2/2

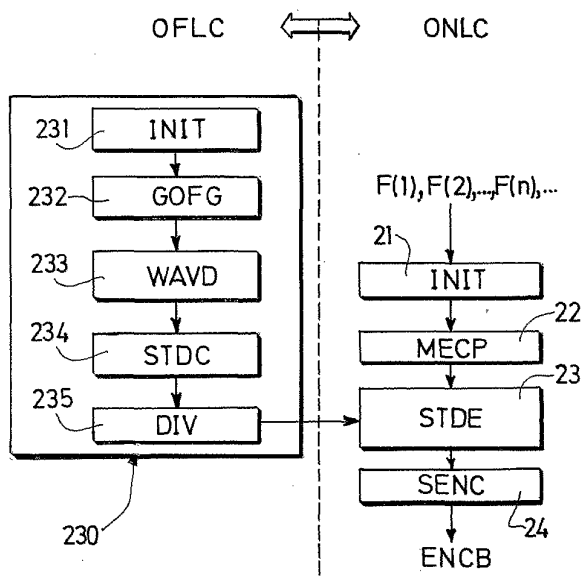


FIG.2

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 February 2002 (14.02.2002)

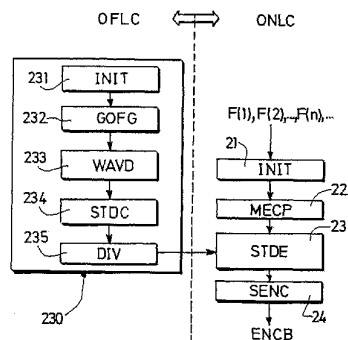
PCT

(10) International Publication Number
WO 02/13536 A3

- (51) International Patent Classification: H04N 7/26 (74) Agent: LANDOUSY, Christian; Internationaal Oetrouw Bureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (21) International Application Number: PCT/EP01/08539
- (22) International Filing Date: 23 July 2001 (23.07.2001) (81) Designated States (national): CN, JP, KR.
- (25) Filing Language: English (84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- (26) Publication Language: English Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (30) Priority Data: 00402246.3 8 August 2000 (08.08.2000) EP
- (71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). (88) Date of publication of the international search report: 11 April 2002
- (72) Inventor: PESQUET-POPESCU, Beatrice; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VIDEO ENCODING METHOD BASED ON A WAVELET DECOMPOSITION



(57) Abstract: The invention relates to a method for encoding a video sequence subdivided into groups of frames. This method comprises an on-line procedure in which a three-dimensional wavelet decomposition is performed involving a biorthogonal filter bank in a lifting scheme using optimal weighting constants. These constants are determined thanks to an additional off-line procedure in which a similar decomposition is performed but without any weighting constants.

WO 02/13536 A3

【国際調査報告】

INT ^N NATIONAL SEARCH REPORT		Internat. Application No PCT/EP 01/08539
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 H04N7/26		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, INSPEC, COMPENDEX		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MINAMI ET AL: "3-D wavelet coding of video with arbitrary regions of support" SIGNALS, SYSTEMS, AND COMPUTERS, 1999. CONFERENCE RECORD OF THE THIRTY-THIRD ASILOMAR CONFERENCE ON OCT. 24-27, 1999, PISCATAWAY, NJ, USA, IEEE, US, 24 October 1999 (1999-10-24), pages 1422-1425, XP010373867 ISBN: 0-7803-5700-0 the whole document	1-3
A	I DAUBECHIES AND W SWELDENS: "Factoring Wavelet Transforms into Lifting Steps" J. FOURIER ANAL. APPL., vol. 4, no. 3, 1998, pages 247-269, XP001051011 cited in the application the whole document	1-3
-/-		
☒ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claims or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search 4 February 2002		Date of mailing of the international search report 11/02/2002
Name and mailing address of the ISA European Patent Office, P.B. 5618 Patentaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2340, Tx: 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Georgiou, G

Form PCT/ISA(210) (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/EP 01/08539

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 946 417 A (MEADOWS HENRY E ET AL) 31 August 1999 (1999-08-31) column 3, line 60 -column 4, line 39 column 15, line 58 -column 16, line 4 -----	1-3
A	BEONG-JO KIM ET AL: "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)" DATA COMPRESSION CONFERENCE, IEEE COMPUTER SOCIETY PRESS, LOS ALAMITOS, CA, US, 25 March 1997 (1997-03-25), pages 251-260, XP002143648 the whole document -----	1-3
A	SAID A ET AL: "A NEW, FAST AND EFFICIENT IMAGE CODEC BASED ON SET PARTITIONING IN HIERARCHICAL TREES" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 6, no. 3, 1 June 1996 (1996-06-01), pages 243-250, XP000592420 ISSN: 1051-8215 cited in the application the whole document -----	1-3
A	MOULIN P: "A MULTISCALE RELAXATION ALGORITHM FOR SNR MAXIMIZATION IN NONORTHOGONAL SUBBAND CODING" IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE INC. NEW YORK, US, vol. 4, no. 9, 1 September 1995 (1995-09-01), pages 1269-1281, XP000533958 ISSN: 1057-7149 cited in the application the whole document -----	1-3
A	DESARTE P ET AL: "SIGNAL-ADAPTED MULTIRESOLUTION TRANSFORM FOR IMAGE CODING" IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE INC. NEW YORK, US, vol. 38, no. 2 PT 2, 1 March 1992 (1992-03-01), pages 897-904, XP000257730 ISSN: 0018-9448 cited in the application the whole document -----	1-3

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat. Application No.

PCT/EP 01/08539

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5946417	A	31-08-1999	US 5870502 A	09-02-1999
			EP 0893029 A1	27-01-1999
			JP 2000508488 T	04-07-2000
			WO 9738533 A1	16-10-1997
			US 6002794 A	14-12-1999

フロントページの続き

(74)代理人 100121083

弁理士 青木 宏義

(72)発明者 ベスクエトーポベスク ビュートリス

オランダ国 5 6 5 6 アーアー アインドーフエン プロフ ホルストラーン 6

Fターム(参考) 5C059 MA05 MA24 MC11 NN01 PP04 UA02

5J064 AA00 BA16 BB12 BC09 BC11 BC21 BC28 BD02