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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 July 2008 (24.07.2008)

PCT

(10) International Publication Number
WO 2008/087466 A1

(51) International Patent Classification:
G06T 9/00 (2006.01) *H03M 7/46* (2006.01)
H04N 7/26 (2006.01)

(21) International Application Number: PCT/IB2007/000173

(22) International Filing Date: 17 January 2007 (17.01.2007)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant and
(72) Inventor: STEFANOV, Rosen [BG/BG]; Obelja 2, Bl. 238,, Entrance 3, floor 8, Ap 70, 1326 Sofia (BG).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(54) Title: RUN-LENGTH ENCODING OF BINARY SEQUENCES FOLLOWED BY TWO INDEPENDENT COMPRESSIONS

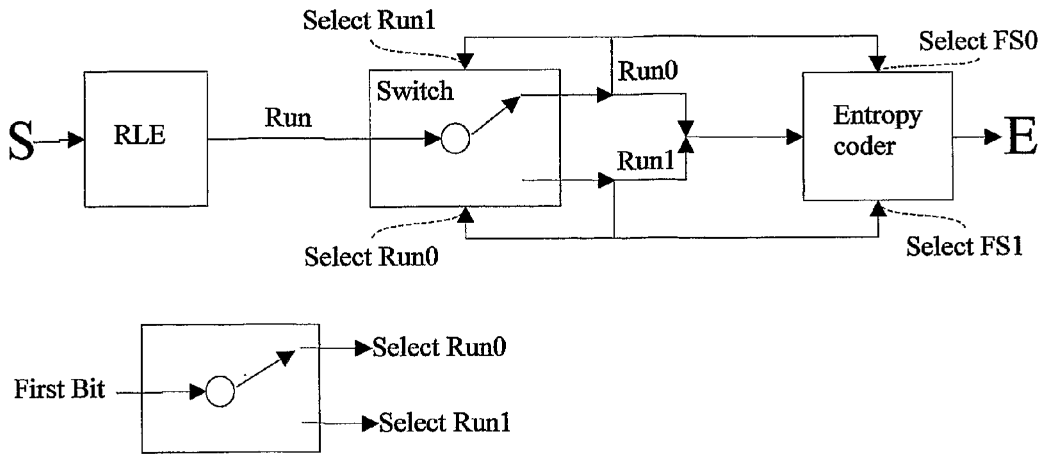

Fig.1

(57) **Abstract:** The enclosed method compresses a binary sequence $BS=\{b_1,b_2,b_3,...,b_N\}$ wherein a run-length encoder (*RLE*) and two independent compressions are used. Let the output sequence of *RLE* be denoted with *RLS*. If $b_1=0$ then *RLS* is 0, $x_0,y_0,x_1,$.... If $b_1=1$ then *RLS* is 1, $y_0,x_0,y_1,$ .... Where x_i is the length of *i*-th run with zeros and y_j is the length of j-th run with ones in *BS*. *RLS* without b_1 can be divided in two sequences RLS_0 and RLS_1 where: • $RLS_0=\{x_0,x_1,\dots\}$. • $RLS_1=\{y_0,y_1,\dots\}$. Let us denote the number of symbols in sequence X with $|X|$, and the entropy of X with $H(X)$. Then $|X|H(X)$ is the size of encoded X sequence. The inequalities: • $|RLS_0|H(RLS_0)+|RLS_1|H(RLS_1)\leq|BS|H(BS)$ • $|RLS_0|H(RLS_0)+|RLS_1|H(RLS_1)\leq|RLS|H(RLS)$ are proved.

WO 2008/087466 A1

Run-length encoding of binary sequences followed by two independent compressions

Technical Field

The present invention relates to compression of sequences of symbols - encoding /decoding method.

Background Art

The following terminology will be used:

- **"codec/coder"**: method of encoding/decoding information aiming at shortening the result sequence's length.
- **"Finite scheme"**: the pair *FS=(A,P)* where $A=\{a_1, a_2, \ldots, a_N\}$ is an alphabet and $P=\{p(a_1), p(a_2), \ldots, p(a_N)\}$ consists of probability of occurrence to each letter [2, §1].
- If S is a sequence of symbols from an alphabet A, then **"finite scheme of sequence *S*"** is *FS(S)=(A,P(S))* where *P(S)* is the set of probabilites of occurrence of each symbol from *A* in *S*.
- **"Entropy of a finite scheme *FS=(A,P)*"**: the number $H(FS)=-\sum_{i=1}^{N} p(a_i)\log_2(p(a_i))$ [2, §1].
- If *S* is a sequence of symbols from an alphabet A, then **"Entropy of sequence *S*"**, or $H(S)$, is the entropy of *FS(S)=(A,P(S))*.[4, §23]
- If *S* is a sequence of symbols from an alphabet A, then **"entropy codec/entropy coder"** is a method and/or means to encode a symbol ***a*** using only *FS(S)* from the sequence *S* with an average size close to $-\log_2(p(a))$. As usual, $p(a)$ is probability of occurrence of ***a*** in S.(for example Huffman coding [4, §2.1], arithmetic coding [4, §4]- US patent 4,122,440)
- If *S* is a sequence of symbols from an alphabet A, then with $|S|$ the number of letters from *A* in *S* will be denoted.
- $|S|H(S)$ is the size of compressed *S* with an entropy coder.

The run length encoding is old and simple, but used only in very special cases. For example see [4, page 2,Pattern-Finding Approaches]: "For instance, fax machines send simple black and white images. These are easily compressed with a solution known as run length encoding, which counts the number of times a black or white pixel is repeated. ..." and "... Run length encoding is the most common example of solutions based on identifying patterns. In most of the cases, patterns are just too complicated for a computer to find regularly." O' Brien found and patented *RLE* followed by LZ77(US patent 4,988,998 "Data compression system for successively applying at least two data compression methods to an input data stream"). See [5,§3.3] for the usage of RLE in JPEG.

The enclosed method uses *RLE* to construct better entropy compression system, as compared to existing entropy coders. This means, that if *S* is a sequence of symbols from an alphabet *A* (*A* and *S* are fixing the statistical structure, i.e. finite scheme), then the length of encoded *S* with the enclosed method is less or equal to $|S|H(S)$ (the best possible achievement of a pure entropy coder).

Disclosure of Invention

Let two sets of numbers be given $\{c_k, k=1,2,...,\infty\}$ and $\{d_l, l=1,2,...,\infty\}$, where . $\sum_{k=1}^{\infty} c_k<\infty$ and $\sum_{l=1}^{\infty} d_l<\infty$.

The following denotations will be used:

- $C=\sum_{k=1}^{\infty} c_k$
- $D=\sum_{l=1}^{\infty} d_l$
- $I=\sum_{k=1}^{\infty} k\,c_k$
- $O=\sum_{l=1}^{\infty} l\,d_l$
- $B=I+O$
- $p_k=\frac{c_k}{C}, q_l=\frac{d_l}{D}$
- $|C-D|\leq 1$
- $p=\frac{I}{B}, q=\frac{O}{B}$

If a binary source *BS* is given, then $c_k, d_l, C, D, ...$ can be interpreted as:

- c_k is the number of 1-runs with size k
- d_l is the number of 0-runs with size l
- I is the number of ones in the *BS*
- O is the number of zeros in *BS*
- C is the number of runs with ones
- D is the number of runs with zeros
- c_k+d_k is the number of runs with size k

After every run with ones, there is a run with zeros (and vice versa), except the last run in the sequence. $|C-D|\leq 1$ is always true. If C and D are both even (or odd) then $C=D$. If C is even and D is odd (or vice versa), then $|C-D|=1$. But in the latter case the last run (a few bits) can be skipped (stored and loaded separately). So, from now on we will assume that ***C is equal to D***.

Let the output sequence of *RLE* applied on $BS=\{b_1, b_2, b_3, ..., b_{|BS|}\}$ be denoted with *RLS*. If $b_1=0$ then *RLS* is $0, x_0, y_0, x_1, y_1, \cdots$. If $b_1=1$ then *RLS* is $1, y_0, x_0, y_1, x_1, \cdots$. Where x_i is the length of i-th run with zeros in *BS* and y_j is the length of j-th run with ones in *BS*. *RLS* without the first symbol (which is b_1) can be devided in two sequences RLS_0 and RLS_1 where:

- $RLS_0=\{x_0, x_1, ...\}$, x_i - is the length of i-th run with zeros in *BS*.
- $RLS_1=\{y_0, y_1, ...\}$, y_j - is the length of j-th run with ones in *BS*.

Next statements are obvious:

- p_k is probability of occurrence of c_k (1-runs with size k) in RLS_1
- q_l is probability of occurrence of d_l (0-runs with size l) in RLS_0
- $\frac{p_k+q_k}{2}$ is probability of occurrence of c_k+d_k (runs with size k) in *RLS*

Tree finite schemes can be formed:

- $FS(RLS_1)=\left(A_1=\{c_1,c_2,...\},P_1=\{p_1,p_2,...\}\right)$
- $FS(RLS_0)=\left(A_0=\{d_1,d_2,...\},P_0=\{q_1,q_2,...\}\right)$
- $FS(RLS)=\left(\{c_1+d_1,c_2+d_2,...\},P=\left\{\frac{p_1+q_1}{2},\frac{p_2+q_2}{2},...\right\}\right)$.

Example 1: Let $BS=\{0110001001101110110010\}$. The runs from left to right are $(0)(11)(000)(1)(00)(11)(0)(111)(0)(11)(00)(1)(0)$ with coresponding lengths $RLS=\{1,2,3,1,2,2,1,3,1,2,2,1,1\}$.

The run-length sequences of *BS* are $RLS_1=\{2,1,2,3,2,1\}$ and $RLS_0=\{1,3,2,1,1,2,1\}$. Then $c_1=2,c_2=3,c_3=1$ are found after counting the equal elements in RLS_1 (if *n* is met *N* times then $c_n=N$). By analogy $d_1=4,d_2=2,d_3=1$. Now $C=2+3+1=6$ and $D=4+2+1=7$. Because $C\neq D$ *BS* is reduced to $\{011000100110111011001\}$ (last 0 is skipped). Then $RLS_0=\{1,3,2,1,1,2\}$, $d_1=3,d_2=2,d_3=1$ and $C=D=6$. The rest will follow $I=1c_k+2c_2+3c_3=2+6+3=11$, $O=1d_1+2d_2+3d_3=3+4+3=10$, $p=11/21$, $q=10/21$, $p_1=2/6$, $p_2=3/6$, $p_3=1/6$, $q_1=3/6$, $q_2=2/6=1/3$, $q_3=1/6$.

$$FS(RLS_0)=FS(RLS_1)=\left(\{1,2,3\},\left\{\frac{1}{6},\frac{1}{3},\frac{1}{2}\right\}\right)$$

$RLS=\{1,2,3,1,2,2,1,3,1,2,2,1\}$ without the last 1. The $FS(RLS)=\left(\{1,2,3\},\left\{\frac{5}{12},\frac{5}{12},\frac{1}{6}\right\}\right)$ and $|RLS|H(RLS)=17.800269059780$.

The size in bits of *BS* is 21(without the skipped last 0). The entropy is $H(BS)=0.99836$. The overall size of encoded *BS* with an entropy coder is $|BS|H(BS)=B\,H(BS)=21*0.99836=20.9655637$. The sum of sizes of encoded RLS_0 and RLS_1 is $-\sum_{k=1}^{3}c_k\log_2(p_k)-\sum_{l=1}^{3}d_l\log_2(q_l)=17.5097$ ($17.50<17.80<20.96$).

Example 1 is a regular case and is the object of the invention, as indicated in claim 1 to compress binary sequences better than the other entropy coders do. **BS22RLS** and **RLS22RLS Inequalities** explain the reasons, which will be proven below.

Lemma: Let $x_1,x_2,...;\ y_1,y_2,...$ be arbitrary positive numbers with $\sum_{i=1}^{\infty}x_i=1$, $\sum_{i=1}^{\infty}y_i=1$ then $-\sum_{i=1}^{\infty}x_i\log_2(x_i)\leq-\sum_{i=1}^{\infty}x_i\log_2(y_i)$.

The lemma was proven in [3 Lemma 1.4.1 page 16].

BS22RLS Inequality $-\sum_{k=1}^{\infty}c_k\log_2(p_k)-\sum_{l=1}^{\infty}d_l\log_2(q_l)\leq-I\log_2(p)-O\log_2(q)$

Proof: Let us use the *lemma* two times, having in mind that $\sum_{k=1}^{\infty} p_k = \sum_{l=1}^{\infty} q_l = 1$.

1) Substitute: $x_i = p_i$ and $y_i = qp^{i-1}$

$$-\sum_{k=1}^{\infty} p_k \log_2(p_k) \le -\log_2(p)\left(\sum_{k=1}^{\infty}(k-1)p_k\right) - \log_2(q)\left(\sum_{k=1}^{\infty} p_k\right)$$

$$-\sum_{k=1}^{\infty} p_k \log_2(p_k) \le -\log_2(p)\left(\sum_{k=1}^{\infty}(k-1)p_k\right) - \log_2(q)\left(\sum_{l=1}^{\infty} q_l\right)$$

2) Substitute: $x_i = q_i$ and $y_i = pq^{i-1}$

$$-\sum_{l=1}^{\infty} q_l \log_2(q_l) \le -\log_2(q)\left(\sum_{l=1}^{\infty}(l-1)q_l\right) - \log_2(p)\left(\sum_{l=1}^{\infty} q_l\right)$$

$$-\sum_{l=1}^{\infty} q_l \log_2(q_l) \le -\log_2(q)\left(\sum_{l=1}^{\infty}(l-1)q_l\right) - \log_2(p)\left(\sum_{k=1}^{\infty} p_k\right)$$

Summing 1) and 2):

$-\sum_{k=1}^{\infty} \frac{c_k}{C} \log_2(p_k) - \sum_{l=1}^{\infty} \frac{d_l}{D} \log_2(q_l) \le -\frac{I}{C}\log_2(p) - \frac{O}{D}\log_2(q)$ and because $C = D$

$$-\sum_{k=1}^{\infty} c_k \log_2(p_k) - \sum_{l=1}^{\infty} d_l \log_2(q_l) \le -I \log_2(p) - O \log_2(q) \cdot$$

BS22RLS Inequality is proved.

Equality in *BS22RLS* is reached, when $p = q = 0.5$. It is colorary from [3, Lemma 1.4.1].

If a binary source *BS* is given, then symbolize $-\sum_{k=1}^{\infty} c_k \log_2(p_k) - \sum_{l=1}^{\infty} d_l \log_2(q_l)$ by $BS22RLS(BS)$.

Using *BS22RLS*, it is possible to design multiple hybrid methods.

The following inequality explains why it is better to use two independent compressions(RLS_0, RLS_1) than just one (*RLS*).

RLS22RLS Inequality

$$|RLS_0| H(RLS_0) + |RLS_1| H(RLS_1) \le |RLS| H(RLS)$$

Proof:

Because function $x\log_2(x)$ is continuous and convex then [2, page 4 or page 6]

$\frac{p_k \log_2(p_k)}{2} + \frac{q_k \log_2(q_k)}{2} \ge \left(\frac{p_k + q_k}{2}\right)\log_2\left(\frac{p_k + q_k}{2}\right)$ is true for all k.

$$-p_k \log_2(p_k) - q_k \log_2(q_k) \le -(p_k + q_k)\log_2\left(\frac{p_k + q_k}{2}\right)$$

$$-\frac{c_k}{C}\log_2(p_k) - \frac{d_k}{C}\log_2(q_k) \le -\left(\frac{c_k + d_k}{C}\right)\log_2\left(\frac{p_k + q_k}{2}\right)$$

$$-c_k\log_2(p_k)-d_k\log_2(q_k)\leq-(c_k+d_k)\log_2\left(\frac{p_k+q_k}{2}\right)$$

Now after summing for all k

$$-\sum_{k=1}^{\infty}c_k\log_2(p_k)-\sum_{l=1}^{\infty}d_l\log_2(q_l)\leq-\sum_{k=1}^{\infty}(c_k+d_k)\log_2\left(\frac{p_k+q_k}{2}\right)$$

or

$$|RLS_0|H(RLS_0)+|RLS_1|H(RLS_1)\leqslant|RLS|H(RLS)$$

RLS22RLS Inequality is proved.

Brief Description of Drawinfgs

An example for carrying out the invention is shown in the attached drawings and is described in detail as follows:

Fig.1 shows a simplified hardware implementation of an encoder according to claim 1. It consist of:

- "RLE" - run length encoder, its input receives the bits of a binary sequence and its outputs are 0 or 1 runs.
- "Switch" - It has one input and two outputs ("Run0" and "Run1"). The input is going directly to the active output. The output "Run0" can be activated by activating "Select run 0". The output "Run1" can be activated by activating "Select run 1". It is necessary to activate an output before starting the encoding process. The first bit of encoded binary sequence can be used to activate an output. The first bit is needed for initialization of the decoder as well.
- "Entropy coder" (e.g. Huffman[4, §2.1] or Arithmetic [4, §4]) consists of two finite schemes ("FS0" and "FS1") and only one of them is active at a given moment. The "Entropy coder" encodes the input symbol depending on the active finite scheme. "FS0" can be activated by activating "Select FS0". "FS1" can be activated by activating "Select FS1". Activation of an finite scheme must be done before receiving a symbol. "FS0" and/or "FS1" are found earlier or are updated after every symbol (adaptive compression[4, §5]).
- Line "Run" is used to move current run from "RLE" to the "Switch"
- Line "Run0" is used to move runs with zeros from "Switch" to the "Entropy coder". The line is responsible to activate the "Select FS0" and "Select run 1" before sending the run to the "Entropy coder".
- Line "Run1" is used to move runs with ones from "Switch" to the "Entropy coder". Also the line is responsible to activate the "Select FS1" and "Select run 0" before sending the run to the "Entropy coder".
- "First bit": The first bit of encoded binary sequence and it is used to initialize the device.
- "S": the input of the device.
- "E": the output of the device.

Fig.2 shows a simplified hardware implementation of a decoder according to claim 1. It consist of:

- "RLE" - run length decoder, its input receives 0 or 1 runs. Its outputs are bits of a binary sequence.
- "Switch" - It has one input and two outputs ("Run0" and "Run1"). The input is going

directly to the active output. The output "Run0" can be activated by activating "Select run 0". The output "Run1" can be activated by activating "Select run 1". It is necessary to activate an output before starting the decoding process. The first bit of decoded sequence can be used to activate an output.

- "Entropy coder" (for example Huffman[4, §2.1] or Arithmetic [4, §4]) - consists of two finite schemes ("FS0" and "FS1") and only one of them is active at a given moment. The "Entropy coder" decodes the input symbol depending on the active finite scheme. "FS0" can be activated by activating "Select FS0". "FS1" can be activated by activating "Select FS1". Activation of an finite scheme must be done before receiving a symbol. "FS0" and/or "FS1" are found earlier or are updated after every symbol (adaptive compression[4, §5]).
- Line "Run" is used to move current run from active output of "Switch" to "RLE".
- Line "Run0" is used to move runs with zeros from "Switch" to the "Run". The line is responsible to activate the "Select FS0" and "Select run 1" before sending the run to the "Run".
- Line "Run1" is used to move runs with ones from "Switch" to the "Run". The line is responsible to activate the "Select FS1" and "Select run 0" before sending the run to the "Run".
- "First bit": The first bit of encoded binary sequence and it is used to initialize the device.
- "S": the input of the device.
- "E": the output of the device.

The required explanation of the invention by means of two drawings is attached.

Modes for Carrying Out the Invention (Advanced Entropy Coders)

An advantageous embodiment of the invention is indicated in claim 2 . The further development according to claim 2: it is possible to compress any sequence better than other entropy coders by compressing three sequences, one of which is binary.

Let a sequence S of source symbols be given. The number of its symbols is $|S|$ and the symbols are from an alphabet A . Let $S=\{s_1, s_2, \dots, s_{|S|}\}$ then, the set $S_B=\{b_1, b_2, \dots, b_{|S|}\}$ will denote the sequence of first bits[1] or b_i is the first bit of s_i . S_B can be seen as a binary source and as a random variable associated with S . Substitute (X,Y) with S and X with S_B in main equation for conditional uncertainty $H(X,Y)=H(X)+H(Y/X)$ [3, theorem 1.4.4] .Then $H(S)=H(S_B)+H(Y/S_B)$ where:

- $H(Y/S_B)=-p(\sum_{i=1}^{|I|} p(\acute{a}_i/1)\log_2(p(\acute{a}_i/1)))-q(\sum_{o=1}^{|O|} p(\mathring{a}_o/0)\log_2(p(\mathring{a}_o/0)))$.
- p is the probability of 1 in S_B
- $q=1-p$ is the probability of 0 in S_B
- $p(\acute{a}_i/1)=\frac{p(a_i)}{p}$ where $a_i \in A$. First bit of a_i is 1 and $\acute{a}_i$ is a_i without the first bit: $a_i=1\,\acute{a}_i$
- $p(\mathring{a}_o/0)=\frac{p(a_o)}{q}$ where $a_o \in A$. First bit of a_i is 0 and $\mathring{a}_i$ is a_i without the first bit: $a_i=0\,\mathring{a}_i$

1Can be last bit or some other bit.

- If S_I is the sequence of all elements from S starting with 1, and S_O is the sequence of all elements from S starting with 0, then next equality is true.

Base Equality: $|S|H(S)=|S_B|H(S_B)+|S_I|H(S_I)+|S_O|H(S_O)$

Proof: Because $H(S)=H(S_B)+H(Y/S_B) \Rightarrow H(S)=H(S_B)+pH(S_I)+qH(S_O)$. $|S|=|S_B|$, $|S_I|=p|S|$, $|S_O|=q|S|$.

Example 2: If $A=\{0,1,2,3\}$, $S=\{1,1,3,2,1,0,2,3,1,2\}$ then

$S_B=\{0,0,1,1,0,0,1,1,0,1\}$ -first bits from the binary representation of elements of *S*.

$S_I=\{1,0,0,1,0\}$ - second bits from the binary representation of elements of *S*, but only if the first bit is 1.

$S_O=\{1,1,1,0,1\}$ - second bits from the binary representation of elements of *S*, but only if the first bit is 0.

Now $p=0.5$, $p_0=0.1, p_1=0.4, p_2=0.3, p_3=0.2$.

Some calculations follow

$H(S)=-p_0\log_2(p_0)-p_1\log_2(p_1)-p_2\log_2(p_2)-p_3\log_2(p_3)=1.8464393$

$|S|H(S)=1.8464393446710*10=18.464393$

$H(S_B)=-0.5\log_2(0.5)-0.5\log_2(0.5)=1$ then $|S_B|H(S_B)=10$

$$H(Y/S_B)=q(-\frac{p_0}{q}\log_2\frac{p_0}{q}-\frac{p_1}{q}\log_2\frac{p_1}{q})+p(-\frac{p_2}{p}\log_2\frac{p_2}{p}-\frac{p_3}{p}\log_2\frac{p_3}{p})$$

$$H(Y/S_B)=-0.1\log_2\frac{0.1}{0.5}-0.4\log_2\frac{0.4}{0.5}-0.3\log_2\frac{0.3}{0.5}-0.2\log_2\frac{0.2}{0.5}$$

$H(Y/S_B)=0.84643934467102$, $|S|H(Y/S_B)=8.4643934467102$

$$|S_I|H(S_I)=p|S|(-\frac{p_2}{p}\log_2\frac{p_2}{p}-\frac{p_3}{p}\log_2\frac{p_3}{p})=4.8547529722734$$

$$\frac{p_2}{p}=\frac{0.3}{0.5}=\frac{3}{5}=\frac{\textit{number of zeros}\in S_I}{\textit{All bits}\in S_I}$$

$$\frac{p_3}{p}=\frac{0.2}{0.5}=\frac{2}{5}=\frac{\textit{number of ones}\in S_I}{\textit{All bits}\in S_I}$$

$$|S_O|H(S_O)=q|S|(-\frac{p_0}{q}\log_2\frac{p_0}{q}-\frac{p_1}{q}\log_2\frac{p_1}{q})=3.6096404744368$$

$$\frac{p_0}{p}=\frac{0.1}{0.5}=\frac{1}{5}=\frac{\textit{number of zeros}\in S_O}{\textit{All bits}\in S_O}$$

$$\frac{p_1}{p}=\frac{0.4}{0.5}=\frac{4}{5}=\frac{\textit{number of ones}\in S_O}{\textit{All bits}\in S_O}$$

$8.4643934467102=4.8547529722734+3.6096404744368$

$H(S_B)+pH(S_I)+qH(S_O)=10+4.85475297+3.60964=18.464393=|S|H(S)$

An advantageous embodiment of the invention is indicated in claim 3. The further development according to claim 3: it is possible to compress any sequence better than other entropy coders by compressing several binary sequences.

AEC Equality: $|S|H(S)=\sum_{u=1}^{U}|(S_B(u))|H(S_B(u))$

Proof: Base equality 1 can be applied for S_I and S_O also. And so on.

AEC Inequality: $\sum_{u=1}^{U} BS22RLS(S_B(u)) \le |S| H(S)$

Proof: Because $BS22RLS(S_B(u)) \le |S_B(u)| H(S_B(u))$ for every u.

In *example 2* S_I and S_O can be compressed further with *BS22RLS* but there is no further compression for S_B because $p=q=0.5$.

Industrial Applicability

The invention can be used in: digital communication, digital television, digital photography, computers. Especially in JPEG, MPEG: "The JPEG algorithm, for instance, can use either Huffman coding or arithmetic coding to compress the coefficients"[4,§4.4], "Lossy JPEG compression can be described in six main steps: .... 5.Run length coding- in order to make the best possible use of the long series of zeros ... 6. Variable length coding(Huffman coding)..."[5,§3.3].

References:

[1] Claude E. Shannon, Warren Weaver, *The mathematical theory of communication*,University of Illinois Press, (1998)
[2] A.I. Khinchin, *Mathematical foundations of information theory*, Dover Publications, Inc., New York (1957)
[3] Robert B. Ash, *Information Theory*, Dover Publications, Inc., New York (1990)
[4] Peter Wayner, *Compression Algorithms for Real Programmers*, Morgan Kaufmann-Academic Press, a Harcourt Science and Technology Company (2000)
[5] H. Benoit, *Digital television:MPEG-1, MPEG-2 and principles of DVB system*,Focal Press (Second edition 2002)

Claims

What is claimed is:

1. Software and/or hardware machine-implemented compression method, in which a binary sequence *BS* is being compressed by compressing two sequences RLS_0, RLS_1 derived from *BS*. Where

- $RLS_0=\{x_0, x_1, ...\}$, x_i - is the length of *i*-th run with zeros in *BS*.
- $RLS_1=\{y_0, y_1, ...\}$, y_j - is the length of *j*-th run with ones in *BS*.

2. Software and/or hardware machine-implemented compression method, in which a sequence *S* is being compressed by compressing three sequences S_B, S_I and S_O derived from *S*. S_B is a binary sequence and it is compressed as in claim 1. The conection among the sequences as random variables is given by the equation:

$H(S)=H(S_B)+pH(S_I)+qH(S_O)$ where p is the probability of 1 in S_B and $q=1-p$ is the probability of 0 in S_B.

3. Software and/or hardware machine-implemented compression method in which a sequence *S* is being compressed by compressing *U* binary sequences $S_B(u), u \in 1,2,...,U$. The latter are derived from *S*. One, several or all of the binary sequences are compressed as in claim 1.

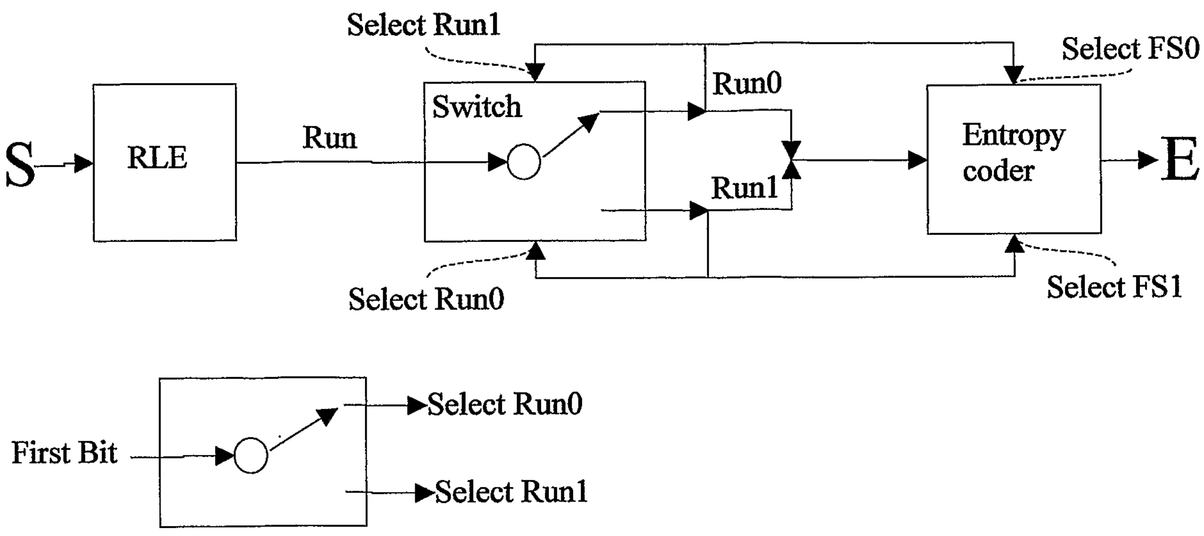
Select Run1
Select FS0
Switch
Run0
S
RLE
Run
Entropy coder
E
Run1
Select Run0
Select FS1
First Bit
Select Run0
Select Run1

Fig.1

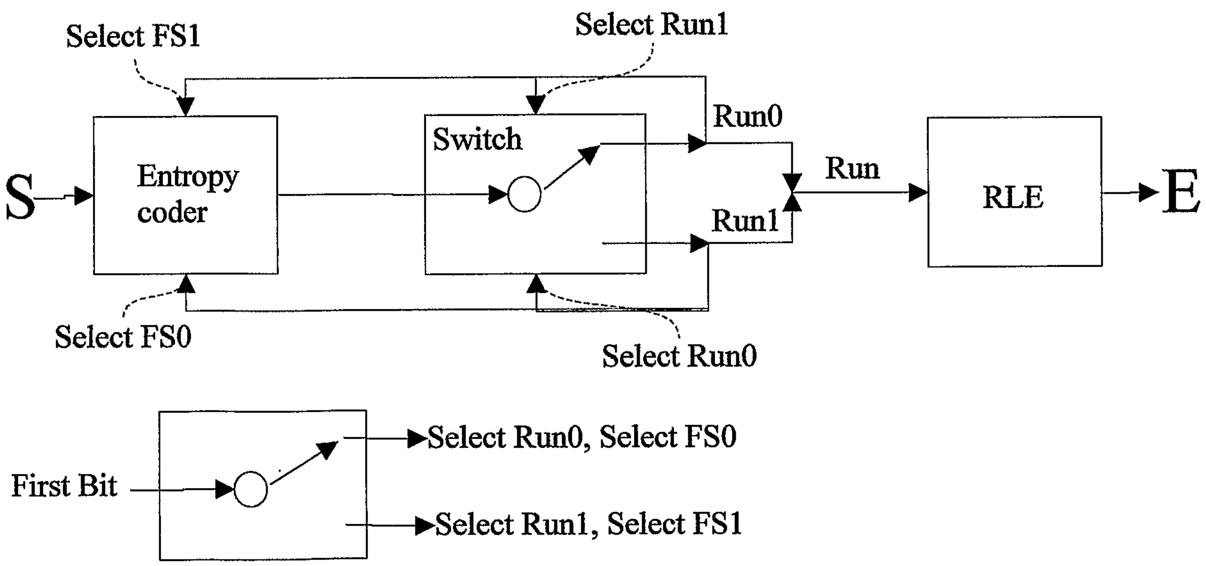
Select FS1
Select Run1
Switch
Run0
S
Entropy coder
Run
RLE
E
Run1
Select FS0
Select Run0
First Bit
Select Run0, Select FS0
Select Run1, Select FS1

Fig.2

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2007/000173

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06T9/00 H04N7/26 H03M7/46

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)
G06T H04N H03M

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, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BRUNO AIAZZI ET AL: "Lossless Image Compression by Quantization Feedback in a Content-Driven Enhanced Laplacian Pyramid" IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 6, no. 6, June 1997 (1997-06), XP011026160 ISSN: 1057-7149 abstract; figure 7 page 838, right-hand column, paragraph 2 - page 839, left-hand column, paragraph 1 -----	1-3
X	US 2006/039615 A1 (CHEN WEN-HSIUNG [US] ET AL) 23 February 2006 (2006-02-23)	1
A	abstract paragraphs [0018] - [0020], [0046] - [0054]; figure 1 -----	2,3
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family 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 claim(s) 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.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 January 2008	08/02/2008

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	CAKIROGLU GARTON, S

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2007/000173

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VRANKEN H ET AL: "Atpg padding and ate vector repeat per port for reducing test data volume" PROCEEDINGS INTERNATIONAL TEST CONFERENCE 2003. (ITC). CHARLOTTE, NC, SEPT. 30 - OCT. 2, 2003, INTERNATIONAL TEST CONFERENCE, NEW YORK, NY : IEEE, US, vol. 1, 30 September 2003 (2003-09-30), pages 1069-1078, XP010685309 ISBN: 0-7803-8106-8 the whole document	1-3
A	WURTENBERGER A ET AL: "A hybrid coding strategy for optimized test data compression" PROCEEDINGS INTERNATIONAL TEST CONFERENCE 2003. (ITC). CHARLOTTE, NC, SEPT. 30 - OCT. 2, 2003, INTERNATIONAL TEST CONFERENCE, NEW YORK, NY : IEEE, US, vol. 1, 30 September 2003 (2003-09-30), pages 451-459, XP010685236 ISBN: 0-7803-8106-8 the whole document	1-3
A	YE Y ET AL: "Improvements to FGS layer Variable Length Coder" INTERNET CITATION, [Online] 31 March 2006 (2006-03-31), XP002458086 Retrieved from the Internet: URL:http://ftp3.itu.int/av-arch/jvt-site/> [retrieved on 2007-11-06] the whole document	1-3

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2007/000173

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006039615 A1	23-02-2006	NONE	