

**Internal distribution code:**

- (A) [ - ] Publication in OJ  
(B) [ - ] To Chairmen and Members  
(C) [ - ] To Chairmen  
(D) [ X ] No distribution

**Datasheet for the decision  
of 27 September 2017**

**Case Number:** T 0602/12 - 3.5.07

**Application Number:** 02253733.6

**Publication Number:** 1263145

**IPC:** H03M7/30

**Language of the proceedings:** EN

**Title of invention:**

Method, apparatus, computer program and storage medium for data compression

**Applicant:**

Canon Kabushiki Kaisha

**Headword:**

Data compression/CANON

**Relevant legal provisions:**

EPC Art. 123(2)

**Keyword:**

Amendments - added subject-matter (yes)



**Beschwerdekammern**  
**Boards of Appeal**  
**Chambres de recours**

European Patent Office  
D-80298 MUNICH  
GERMANY  
Tel. +49 (0) 89 2399-0  
Fax +49 (0) 89 2399-4465

Case Number: T 0602/12 - 3.5.07

**D E C I S I O N**  
**of Technical Board of Appeal 3.5.07**  
**of 27 September 2017**

**Appellant:** Canon Kabushiki Kaisha  
(Applicant) 30-2, 3-chome, Shimomaruko,  
Ohta-ku  
Tokyo (JP)

**Representative:** Hitching, Peter Matthew  
Canon Europe Ltd  
European Patent Department  
3 The Square  
Stockley Park  
Uxbridge Middlesex UB11 1ET (GB)

**Decision under appeal:** Decision of the Examining Division of the  
European Patent Office posted on 27 October 2011  
refusing European patent application No.  
02253733.6 pursuant to Article 97(2) EPC.

**Composition of the Board:**

**Chairman** R. Moufang  
**Members:** R. de Man  
M. Jaedicke

## Summary of Facts and Submissions

I. The applicant (appellant) appealed against the decision of the Examining Division refusing European patent application No. 02253733.6.

II. The decision refers to the following documents:

D1: EP 0 688 104 A2; and

D2: Ogihara T. et al.: "Compression Method Using LZW Coding and a Sliding Window", Electronics and Communications in Japan, Part III: Fundamental Electronic Science, Vol. 77, No. 8, August 1994.

The Examining Division decided that claims 1 and 5 of the then main request violated Articles 84 and 123(2) EPC and that their subject-matter lacked inventive step over document D2 combined with document D1. The then auxiliary request was not admitted into the proceedings under Rule 137(3) EPC.

III. With the statement of grounds of appeal, the appellant submitted a new set of claims 1 to 10 and requested that the decision under appeal be set aside and that a patent be granted on the basis of "the claims as now presented".

IV. In a communication accompanying a summons to oral proceedings, the Board expressed *inter alia* the preliminary view that the sole substantive request did not comply with Article 123(2) EPC.

V. Oral proceedings were held on 27 September 2017. At the end of the oral proceedings, the chairman pronounced the Board's decision.

VI. The appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of the claims of the sole request filed with the statement of grounds of appeal.

VII. Claim 1 of the sole request reads as follows:

"A data compression method of reading current input data to be compressed from a window buffer, searching for previous input data that matches current input data, each of the previous input data and the current input data consisting of leading data and auxiliary data, the leading data being a predetermined number of bits of the input data and the auxiliary data constituting the next data after the leading data, and generating coded data based on position information (412, 422, 432, 1012, 12012) representing a position in said buffer of the previous input data, comprising the steps of:

reserving a dictionary constructed with a first storage area (40, 1101) and a plurality of second storage areas (41, 42, 43, 44, 101, 1102-1107, 1201),

wherein the first storage area (40, 1101) is addressed by leading data of the current input data, and the first storage area stores a pointer to a current one of the second storage areas (41),

wherein each second storage area (41, 42, 43, 44, 101, 1201) stores the auxiliary data (411, 421, 431, 1011, 12011) of a corresponding one of the previous input data, and also stores an offset (412, 422, 432, 1012, 12012) representing a position of the corresponding previous input data, and also stores link information (413, 423, 1014, 12014) for linking to another second storage area (42, 43);

determining an address of the current second storage area (41) by referring to the pointer, in the

first storage area, represented by the leading data of the current input data;

comparing the auxiliary data (411) of the previous input data, stored in the current second storage area (41), with data after the leading data of the current input data;

if the auxiliary data (411) of the previous input data does not match with the data after the leading data of the current input data, referring to another second storage area addressed by the link information of the current second storage area, and making that other second storage area the current second storage area, and repeating said comparing step; and

if the auxiliary data (411) of the previous input data matches with the data after the leading data of the current input data, encoding the offset stored in the current second storage area, and a longest-match length between the current input data and the previous input data stored, in the buffer, at the position represented by the offset stored in the current second storage area so as to generate the coded data;

characterised, in a case where the auxiliary data matching the data after the leading data of the current input data is not found in said comparing steps, by further comprising:

generating a new second storage area for the current input data; and

adding the new second storage area into the dictionary so that each auxiliary data, of the previous input data and the current input data, of the second storage areas, classified by the same leading data, is arranged in a predetermined order."

VIII. The appellant's arguments as relevant to this decision are discussed in detail below.

## **Reasons for the Decision**

1. The appeal complies with the provisions referred to in Rule 101 EPC and is therefore admissible.

*Added subject-matter - Article 123(2) EPC*

2. Claim 1 is directed to a method of lossless data compression. The method reads input data and generates compressed "coded data". It employs a "window buffer" which slides over the input data that is being compressed, and it constructs a "dictionary" data structure that keeps track of phrases present in the sliding window. If the compression method detects that the next input data to be compressed starts with initial data corresponding to a phrase in the dictionary, it generates as "coded data" position information representing the position in the window buffer of this initial data together with the length of the match (which may be longer than the length of the initial phrase).
3. The claim specifies in some detail how the next input data to be compressed is matched with phrases in the dictionary. The initial data to be compared consists of "leading data being a predetermined number of bits of the input data" and "auxiliary data constituting the next data after the leading data". The leading data is used to address a "first storage area", which stores a pointer to a "second storage area". The second storage area is essentially a linked list, each element of the list containing auxiliary data, an offset into the sliding window, and a pointer to the next element of the list. If the auxiliary-data part of the input data matches the auxiliary data stored in one of the

elements of the second storage area, the offset stored in that element corresponds to a position in the sliding window buffer where earlier input data with the same initial part (leading and auxiliary data) is present.

4. In its decision, the Examining Division argued that the feature of claim 1 specifying that the leading data is "a predetermined number of bits" offended against Article 123(2) EPC, because the application as filed only disclosed, in paragraph [0035] of the A2 publication, leading data consisting of two characters.

In the statement of grounds of appeal and at the oral proceedings, the appellant argued that the application as filed disclosed both leading data consisting of two characters (in paragraph [0035]) and leading data consisting of one character (in paragraph [0053]). These two disclosures formed a sufficient basis for the term "a predetermined number of bits". At the oral proceedings, the appellant confirmed that it relied on no other passages from the application as filed, nor on the originally filed claims.

5. The Board agrees with the appellant that the application as filed discloses both the use of a single character and the use of two characters as "leading data" for addressing a first storage area. But the presence of these two possibilities alone is not yet sufficient to conclude that the skilled person would directly and unambiguously infer that the leading data for addressing the first storage area may consist of any predetermined number of bits.
6. The detailed description of the present application describes a number of quite specific embodiments.

- 6.1 In the "First Embodiment" disclosed in paragraphs [0024] to [0050] of the A2 publication of the present application, the first storage area is an array "X[i]" having  $2^{16} = 65,536$  elements (paragraph [0027]). In this embodiment, phrases of three characters/bytes of input data are registered in the dictionary, whereby the first two characters (corresponding to a 16-bit value from 0 to 65,535) are used to address the array forming the first storage area, and the third character is stored in the second storage area pointed to by the addressed array element (paragraphs [0029] and [0036]). Thus, in this embodiment the "leading data" (referred to as "representative data" in the description) consists of two characters and the "auxiliary data" of one character.
- 6.2 In the "Second Embodiment" disclosed in paragraphs [0051] to [0055] of the A2 publication, the first storage area is an array of 256 elements (paragraph [0052]). The phrases registered in the dictionary are still three characters long, but now the first character of input data is used as "leading data" to address the array forming the first storage area, and the second and third characters serve as "auxiliary data" (paragraph [0053]).
- 6.3 In a variation of the second embodiment described in paragraphs [0056] to [0059], the array forming the first storage area is addressed by a two-character value, which is calculated by applying a hash function to the three-character string. Paragraph [0059] explains that in this variation, "an element of a header array ... expresses the first byte, and two element areas linked from the element of the header array express the second and third bytes respectively".



Hence, to the extent that this variation falls within the scope of the claim, the "leading data" corresponds to the first character of input data, and the "auxiliary data" corresponds to the second and third characters of input data.

- 6.4 The "Third Embodiment" described in paragraphs [0060] to [0065] is a variation of the second embodiment. The leading data is still the first character of input data, but a modification of the structure of the elements in the second storage area allows the auxiliary data to be of indefinite length (paragraph [0061]).
- 6.5 The "Fourth Embodiment" is based on the first embodiment; its description in paragraphs [0066] and [0067] does not specifically relate to the use of leading data to address the first storage-area array.
- 6.6 The final "Fifth Embodiment" disclosed in paragraphs [0068] to [0080] deviates from the other embodiments in that the array data structure is missing. Instead, the dictionary data structure (shown in Figure 13) consists of a hierarchy of linked lists, the elements of the top-level linked list storing the first character of dictionary phrases (paragraph [0069]; elements 1302, 1305, 1309) and lower-level linked lists storing further characters. Hence, in this embodiment the first character or characters of input data is/are not used as an index to address a first storage area.
7. The appellant relies on the first and second embodiments as providing a basis for leading data having "a predetermined number of bits". However, in the first embodiment the leading data consists of two characters (i.e. 16 bits) and the auxiliary data of one

character (i.e. 8 bits). In the second embodiment, the leading data consists of one character and the auxiliary data of two characters. In addition, the dictionary data structure of each embodiment is, at least to some extent, specifically designed for these lengths. In other words, these embodiments are not presented to the skilled reader as being readily generalisable to other lengths. In particular, they do not hint at using leading data consisting of a fractional number of characters/bytes (i.e. having a "predetermined number of bits" that is not a multiple of eight), which would require additional implementation considerations not touched upon in the description. And in view of the exponential growth of the number of elements of the first storage-area array, the Board considers that the skilled person would not even read into these embodiments an implicitly disclosed generalisation to leading data having three characters (corresponding to an array with over 16 million elements), four characters (corresponding to an array with over 4 billion elements) or more; such dictionary sizes are impractical for storing the phrases occurring in a sliding window having a typical size of 2-32 KB (see paragraph [0005]). Hence, the general concept of using leading data having any "predetermined number of bits" to address a first storage area is not directly and unambiguously derivable from the first two embodiments. Nor do the further embodiments described in the application disclose the concept.

8. The originally filed claims do not disclose this general concept, either.
- 8.1 Independent method claim 1 as filed states only that the dictionary is searched for previous input data that

matches the (current) input data. The use of leading or representative data to address a first storage area is mentioned neither in this claim nor in any of its dependent claims.

- 8.2 According to independent method claim 10 as filed, the first storage area stores representative data and is searched using representative data obtained from the input data as a key. Independent claim 10 as filed and its dependent claims hence do not correspond to embodiments as claimed now, in which the leading or representative data is used to address the first storage area (e.g. used as an index into a first storage-area array as in the first to third embodiments).
- 8.3 The features of the remaining computer program, storage medium and apparatus claims as filed correspond to features of the method claims.
9. In summary, the subject-matter of claim 1, in particular the feature specifying the use of leading data consisting of a predetermined number of bits to address the first storage area, is not directly and unambiguously derivable from the application as filed, contrary to Article 123(2) EPC.
10. Since the sole request is not allowable, the appeal is to be dismissed.

**Order**

**For these reasons it is decided that:**

The appeal is dismissed.

The Registrar:

The Chairman:



I. Aperribay

R. Moufang

Decision electronically authenticated