

公告本

申請日期	87. 3. 16.
案 號	90>17797(由871038492(第44))
類 別	H03M7/30

(以上各欄由本局填註)

發 明 新 專利說明書 M252197		
一、發 明 新 名稱	中 文	實施適應性資料壓縮之裝置
	英 文	APPARATUS FOR PERFORMING ADAPTIVE DATA COMPRESSION
二、發 明 創 人	姓 名	大衛 約翰 卡瑞福
	國 籍	英國
	住、居所	美國德州奧斯丁市博拉夫園區114號
三、申請人	姓 名 (名 稱)	美商萬國商業機器公司
	國 籍	美國
	住、居所 (事務所)	美國紐約州阿蒙市
	代 表 人 名 姓	費羅普

煩請委員明示本局改請新案 倘是請變更原實質內容

裝 訂 線

(由本局填寫)

承辦人代碼：
大 類：
I P C 分類：

C6
D6

本案已向：

國（地區） 申請專利，申請日期： 案號： ，☐有 ☐無主張優先權
美 國 1997年9月19日 08/934,234 ☒有 ☐無 主張優先權

（請先閱讀背面之注意事項再填寫本頁各欄）

裝

訂

線

經濟部中央標準局員工消費合作社印製

五、創作說明(1)

創作背景

1. 技術領域

本創作係大致有關一種壓縮資料之方法及裝置，尤係有關一種執行適應性資料壓縮之方法及裝置。更具體而言，本創作係有關一種改良Lempel-Ziv 1變體的資料壓縮效率之方法及裝置。

2. 習用技術之說明

J. Ziv 及 A. Lempel 於 1977年出版的 IEEE Trans. Inform. Theory, Vol. IT-23, No.3, 第337-343頁所發表之論文 "A Universal Algorithm for Sequential Data Compression" 中首度提出了著名的Lempel-Ziv 編碼架構。J. Storer 及 T. Symanski 於1982年出版的J. ACM, Vol.29, No.4, 第928-951頁所發表之論文 "Data Compression via Textual Substitution" 中首度提出利用原始Lempel-Ziv編碼架構的一變體之古典實施例，後來 T. Bell 又以一種稱為LZSS編碼架構的實施例實施了利用Lempel-Ziv編碼架構的變體之實施例。

LZSS架構可讓一Lempel-Ziv 壓縮器的輸出為代碼符號 LITERAL_POINTER 及 COPY_POINTER 的一任意混合。LITERAL_POINTER 包含一個單一的嵌入資料符號，而 COPY_POINTER 則包含兩個元素：一 SYMBOL_COUNT 及一位移(DISPLACEMENT)(或偏移量(OFFSET))。

只利用一解壓縮器來提取嵌入資料符號，即可處理每一

五、創作說明(2)

LITERAL_POINTER。然後自該解壓縮器輸出所提取的資料符號，且亦將所提取的資料符號拷貝回一過去記錄緩衝區之現有位置，以供更新。提取並解碼 SYMBOL_COUNT及位移值，而處理各COPY_POINTER。然後自該解壓縮器的過去記錄緩衝區拷貝SYMBOL_COUNT所指定的某一數目之符號。以一次一個符號，自該位移值所指定的一過去記錄緩衝器位置開始之方式，循序拷貝這些符號。當將每一符號拷貝到該解壓縮器之輸出時，也以與LITERAL_POINTER相同之方式，將該符號拷貝回該過去記錄緩衝區，以供更新。

如同大部分的此類Lempel-Ziv 1(此種方式與也是Lempel 及Ziv提出的一種字典型架構Lempel-Ziv 2有所不同)變體之實施例，LZSS利用一種可變長度的SYMBOL_COUNT編碼架構，這是因為較短的計數值比較長的計數值更常出現。此外，LZSS利用一個單一旗標位元來區分經過編碼的各LITERAL_POINTER及COPY_POINTER。

在最近幾年中，有一種稱為適應性無耗損資料壓縮(Adaptive Lossless Data Compression；簡稱ALDC)的適應性壓縮演算法廣泛應用於電腦及相關聯周邊裝置內的一般用途之資料壓縮。在Quarter-Inch Cartridge Drive Standards股份有限公司於1994年3月10日出版的"QIC Development Standard QIC-154" Rev. A中，對也是一種Lempel-Ziv 1變體的ALDC有更詳細的說明。也可以上網際網路的網址"<http://www.qic.org>"取

五、創作說明(3)

得上述文件，本創作特此引用該文件以供參照。

如果ALDC碰到無法壓縮的資料(例如由隨機位元組所構成的輸入資料流)，則輸出中有相當大的部分包含若干LITERAL_POINTER，而且輸入隨機資料位元組將不太可能大量重複目前儲存在一過去記錄緩衝區的已出現資料序列。如同在LZSS編碼架構中，ALDC在一LITERAL_POINTER及一COPY_POINTER的開始處利用單一旗標位元來將這些指標相互區分。此種方式將使ALDC擴充了約12.5%，這是因為必須將每一個(8位元的)輸入隨機資料位元組值編碼成一個LITERAL_POINTER，ALDC中之該LITERAL_POINTER是一個"0"位元及接續的8個文字符號資料位元。因此，對於在過去記錄緩衝區中找不到相符資料位元組串的任何資料而言，ALDC將每8個輸入的資料位元擴充為9個資料位元。

本創作之揭露事項說明了一種無須利用任何旗標位元即可執行適應性資料壓縮之改良式方法及裝置，因而本創作之壓縮率將優於諸如前文所述的LZSS或ALDC等需要一旗標位元以供區分一LITERAL_POINTER及一COPY_POINTER的習用技術之壓縮率。

創作概述

有鑒於前文所述，因而本創作之一目的在於提供一種壓縮資料的改良式方法及裝置。

本創作之另一目的在於提供一種執行適應性資料壓縮之改良式方法及裝置。

五、創作說明(4)

本創作之又一目的在於提供一種可增強Lempel-Ziv 1變體的資料壓縮效率之改良式方法及裝置。

根據本創作的一方法及裝置，一壓縮器將一來源符號輸入資料流之起始部分編碼成一LITERAL_POINTER。一LITERAL_POINTER包含該資料流的至少一個資料位元組。該輸入資料流的一後續部分被編碼成一COPY_POINTER。該COPY_POINTER包含指向該壓縮器內的一過去記錄緩衝區之一計數及一位移。該輸入資料流的所有後續資料位元組都被編碼成以交替方式出現的若干LITERAL_POINTER及COPY_POINTER，因而該壓縮器輸出的一個經過編碼之輸出序列包含一個交替出現若干LITERAL_POINTER及若干COPY_POINTER之指標串。因此，無須用到一旗標位元即可如同習用技術般地區分一LITERAL_POINTER及一COPY_POINTER。

若參照下文中之詳細說明，將可易於了解本創作之所有目的、特徵、及優點。

附圖簡述

若參照下文中對一實施例之詳細說明，並配合各附圖，將可易於了解本創作、及其較佳使用模式、其他目的、及優點，這些附圖有：

圖1a是可採用本創作一較佳實施例的一壓縮器單元之方塊圖；

圖1b是可採用本創作一較佳實施例的一資料解壓縮器單元之方塊圖；

五、創作說明(6)

時，移出最舊的資料位元組。因此，在該過去記錄緩衝區中必然可以找到未經壓縮的資料流的最新部分之一循序拷貝。

該壓縮過程也包含對輸入資料流作一次檢查，以便識別已經存在於該過去記錄緩衝區內的任何資料位元組串或序列。如果在該過去記錄緩衝區內有一個相同的資料位元組串，則可將該相符的資料位元組串編碼成一個兩元素之COPY_POINTER，其中包含該過去記錄緩衝區內的一個位元組符號計數及一個位移(或偏移量)。因此，當一COPY_POINTER可將各資料位元組串以資訊位元數小於正常所需位元數之方式編碼時，即達到了壓縮之目的。然而，如果一輸入資料位元組序列無法與過去記錄緩衝區內的任何位置之內容相符，則應將該輸入資料位元組序列編碼成可直接代表該輸入資料位元組序列的一文字符號(LITERAL)。

現在請參閱圖1b，圖中示出可採用本創作一較佳實施例的一資料解壓縮器單元之方塊圖。如圖所示，解壓縮器單元(15)係耦合到一控制器(16)及一RAM/CAM(17)。與RAM/CAM(12)類似，係在RAM/CAM(17)內存放解壓縮器單元(15)的所有資料結構。在作業中，解壓縮器單元(15)首先自資料來源(19)接收一個經過壓縮的資料流。在經過資料解碼之後，然後將一個未經壓縮的資料流自解壓縮器單元(15)傳送到一資料接收器(18)。

另一方面，先將一COPY_POINTER的所指定位元組計

五、創作說明(7)

數及位移解碼，而處理該COPY_POINTER。然後以一次一個位元組的方式，自解壓縮器單元(15)內的過去記錄緩衝區讀取該資料位元組串。然後在存取次一資料位元組值之前，先將每一資料位元組值拷貝到該過去記錄緩衝區，並將該資料位元組值輸出作為一個經過解碼的資料位元組。一旦以此種方式處理了全部的資料位元組串之後，解壓縮器單元(15)內過去記錄緩衝區之內容將與壓縮器單元(10)內過去記錄緩衝區之內容相同。

II. 編碼架構

本創作與許多Lempel-Ziv 1變體間之者要不同之處在於用來表示各LITERAL與COPY_POINTER之編碼架構。下文中將說明根據本創作一較佳實施例的一種改良式編碼架構：

(1)一壓縮器單元的經過編碼之輸出必然以一 LITERAL_POINTER作為開始。該LITERAL_POINTER接續有交替出現的若干COPY_POINTER及LITERAL_POINTER序列。例如，< LITERAL_POINTER> < COPY_POINTER > < LITERAL_POINTER > < COPY_POINTER >... <LITERAL_POINTER > < COPY_POINTER > < LITERAL_POINTER > < COPY_POINTER >。

(2)一LITERAL_POINTER包含一LITERAL_COUNT欄位及接續的一對應數目的LITERAL_DATA符號(通常為位元組的形式)，其形式如下：

五、創作說明(8)

<LITERAL_POINTER>=[<LITERAL_COUNT>
<LITERAL_DATA>]

(3) — LITERAL_COUNT 欄位句有三個容許值 0、1、及 2。這三個值決定了接續的 LITERAL_DATA 符號的數目。LITERAL_COUNT 欄位的編碼格式示於表 I。

(4) — 一個 0 的 LITERAL_COUNT 值表示一 NULL_LITERAL，且並無接續的 LITERAL_DATA 符號(如表 I 的第 1 列所示)。此種編碼格式係用來隔離兩個連續的 COPY_POINTER，因而不會與上述的規則(1)衝突。

(5) — 一個 1 的 LITERAL_COUNT 值表示一個 LITERAL_DATA 符號及一個接續的 COPY_POINTER (如表 I 的第 2 列所示)。

(6) — 一個 2 的 LITERAL_COUNT 值表示兩個 LITERAL_DATA 符號及另一個接續的 LITERAL_POINTER(如表 I 的第 3 列所示)，且係按照上述規則(2)到(5)來編碼。

(7) — COPY_POINTER 包含一 SYMBOL_COUNT 及一位移欄位，其編碼格式完全與 ALDC 相同，係利用相同的保留控制字組及 End_Marker 值，但是不用虛耗一個前導的旗標位元。

編碼	值	接續符號或指標
0	0	<COPY_PTR>
10	1	<LITERAL_DATA> <COPY_PTR>
11	2	<LITERAL_DATA><LITERAL_DATA>

五、創作說明(11)

出資料位元組，將 LITERAL_DATA 儲存在 L_DATA[0]，並設定 L_CTR 等於 1。然而，如果 L_CTR 值不等於 2，則將該資料位元組儲存在以 L_CTR 為索引的 L_DATA，並將 L_CTR 值遞增 1。

對於將一 COPY_POINTER 編碼而言，必須首先編碼並輸出先前儲存的 LITERAL_DATA 位元組。如果 L_CTR 值等於 2，則輸出 "11"，並進入步驟 (3)。如果 L_CTR 值等於 1，則輸出 "10"，並進入步驟 (3)。如果 L_CTR 值等於 0，則輸出 "0"，並進入步驟 (4)。在步驟 (3) 中，自 L_DATA[0] 輸出資料位元組；如果 L_CTR > 1，則自 L_DATA[1] 輸出資料位元組；輸出 "0"，並進入步驟 (4)。在步驟 (4) 中，設定 L_CTR 等於 0；然後編碼及輸出 COPY_POINTER。

如前文所述，本創作提供了一種在一壓縮器單元內執行適應性資料壓縮之改良式方法及裝置。在本創作下，將一輸入資料流個各資料位元組編碼成交替出現的若干 LITERAL_POINTER 及 COPY_POINTER。因此，該壓縮器單元的輸出序列包含一個交替出現若干 LITERAL_POINTER 及 COPY_POINTER 的指標串，且係以一 LITERAL_POINTER 作為開頭。一 LITERAL_POINTER 現在可以不包含或者包含某一數目的文字符號資料位元(或符號)、及一個前導的 LITERAL_COUNT。可利用所容許的最大 LITERAL_POINTER 值，而重複地連結所需要數目的 LITERAL_POINTER，以便將一個任

五、創作說明(12)

意長度的 LITERAL_DATA 符號序列編碼。此外，可利用一個 2 位元固定長度或 1 或 2 位元可變長度之編碼架構，而將一 LITERAL_POINTER 內之一 LITERAL_COUNT 編碼。

請務必注意，雖然已在一硬體壓縮器的環境中說明了本創作，但是熟悉本門技術者當可了解，亦可以利用軟體來實施本創作之各項機制，且可以各種形式的程式產品之方式配售本創作之該等機制，而且不論用來實際進行配售的信號承載媒體為何種特定類型，本創作都可同樣適用。信號承載媒體的例子包括(但不限於)：諸如軟碟、光碟等可記錄型媒體、以及諸如類比或數位通訊鏈路等傳輸型媒體。

雖然已參照一較佳實施例而詳細示出並說明了本創作，但是熟悉本門技術者當可了解，在不脫離本創作的精神及範圍下，尚可對本創作之形式及細節作出各種改變。

四、中文創作摘要（創作之名稱：實施適應性資料壓縮之裝置）

本創作揭露了一種一來源符號輸入資料流編碼而產生一輸出指標序列之方法。一壓縮器將該輸入資料流之起始部分編碼成一 LITERAL_POINTER。一 LITERAL_POINTER 包含該資料流的至少一個資料位元組。該輸入資料流的一後續部分被編碼成一 COPY_POINTER。該 COPY_POINTER 包含指向該壓縮器內的一過去記錄緩衝區之一計數及一位移。該輸入資料流的所有後續資料位元組都被編碼成以交替方式出現的若干 LITERAL_POINTER 及 COPY_POINTER，因而該壓縮器輸出的一個經過編碼之輸出序列包含一個交替出現若干 LITERAL_POINTER 及若干 COPY_POINTER 之指標串。

英文創作摘要（創作之名稱：APPARATUS FOR PERFORMING ADAPTIVE DATA COMPRESSION）

A method for encoding an input data stream of source symbols to produce an output sequence of pointers is disclosed. An initial part of the input data stream is encoded as a LITERAL_POINTER by a compressor. A LITERAL_POINTER includes at least one data byte from the data stream. A subsequent part of the input data stream is encoded as a COPY_POINTER. The COPY_POINTER includes a count and a displacement pointing to a history-buffer within the compressor. All succeeding data bytes from the input data stream are encoded as LITERAL_POINTERS and COPY_POINTERS in an alternating fashion, such that an encoded output sequence output by the compressor includes a string of pointers alternating between LITERAL_POINTERS and COPY_POINTERS.

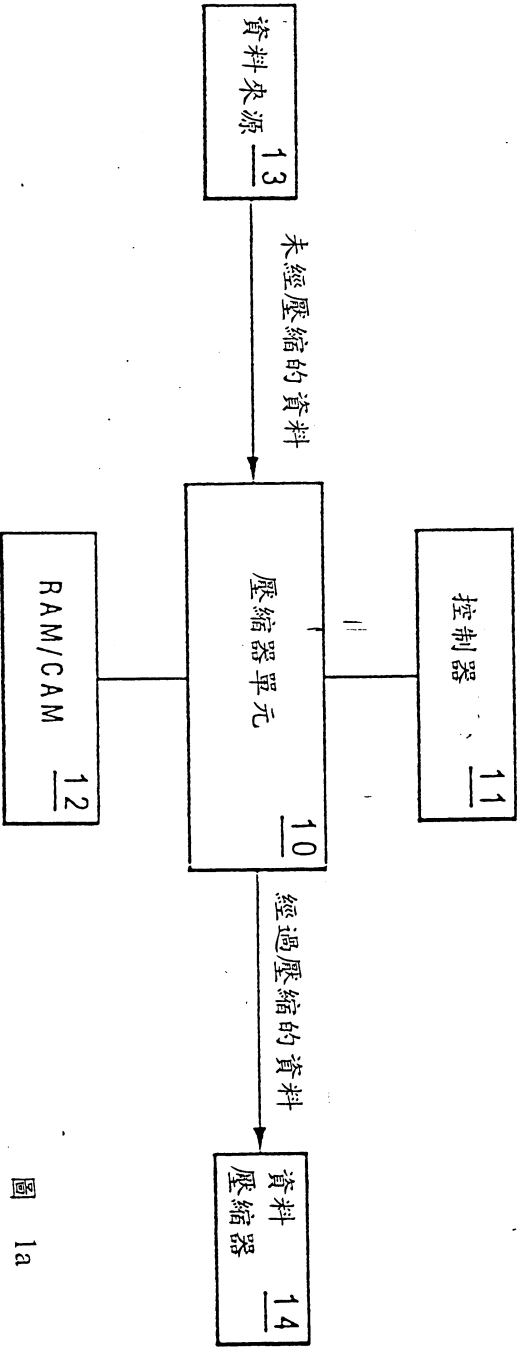


圖 1a

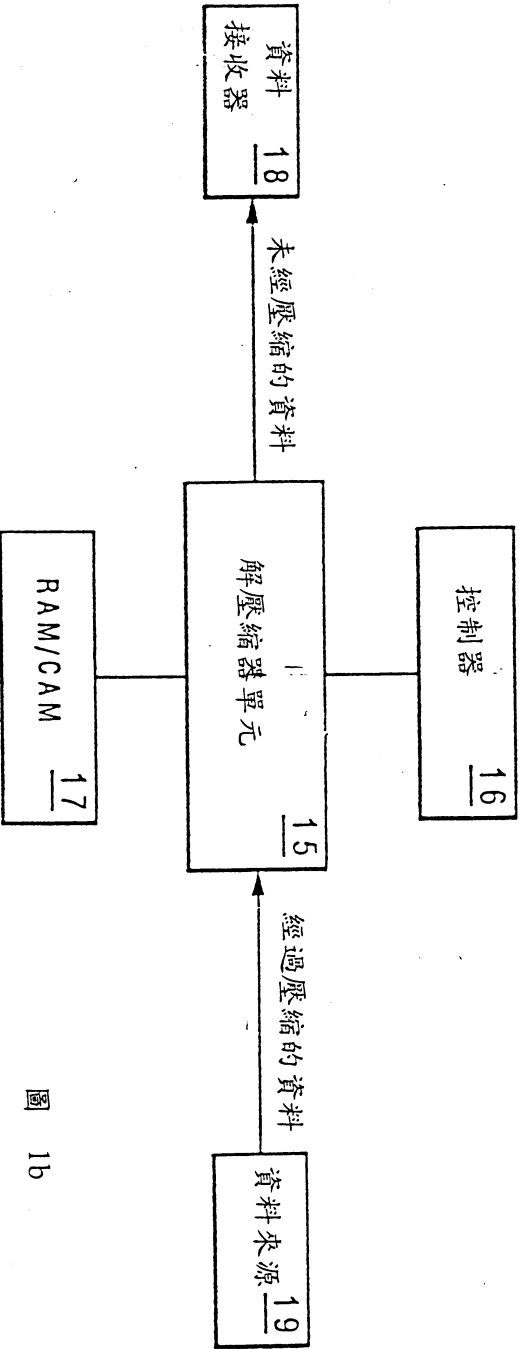


圖 1b

五、創作說明(5)

圖 2(a) 及 (b) 是根據本創作一較佳實施例的一固定長度 LITERAL_ COUNT 編碼架構之高階虛擬碼；以及

圖 3(a) 及 (b) 是根據本創作一較佳實施例的一可變長度 LITERAL_ COUNT 編碼架構之高階虛擬碼。

較佳實施例之詳細說明

可在一資料壓縮器及一資料解壓縮器內實施本創作。熟悉本門技術者當可了解，亦可在硬體或軟體中實施本創作。

I. 編碼器及解碼器

現在請參閱各圖示，尤其請參閱圖 1a，圖中示出可採用本創作一較佳實施例的一資料壓縮器單元之方塊圖。如圖所示，壓縮器單元(10)係耦合到一控制器(11)、及一隨機存取記憶體(Random-Access Memory；簡稱RAM)或內容可定址的記憶體(Content-Addressable Memory；簡稱CAM)(12)。係在RAM/CAM(12)內存放與諸如過去記錄緩衝區等壓縮演算法相關聯的所有資料結構。因此，RAM/CAM(12)的最佳容量有相當的程度係取決於過去記錄緩衝區的容量。在作業中，壓縮器單元(10)首先自一資料來源(13)接收一個未經壓縮的資料流。在經過資料編碼之後，然後將經過壓縮的資料流傳送到一資料接收器(14)。

壓縮器單元(10)以一次一個資料位元組之方式處理未經壓縮的資料流。也循序將經過處理的每一資料位元組拷貝到一過去記錄緩衝區，並且在該過去記錄緩衝區已填滿

五、創作說明(9)

		<LITERAL PTR>
--	--	---------------

表 I

請注意，在本創作一較佳實施例下的該編碼架構並不需要一旗標位元，就可如同習用技術的Lempel-Ziv 1變體將一 LITERAL_POINTER 及一 COPY_POINTER 相互區分。此外，此種特定的編碼架構在壓縮器單元內也只需要最小的緩衝空間，例如兩個符號(形式為位元組)的空間。

上述編碼架構採用一種可變長度的編碼架構，LITERAL_COUNT可以是一個或兩個位元，提供了0到2的數值範圍。此外，COPY_POINTER內的SYMBOL_COUNT的長度為2到12個位元，因而總共可表示286個值的範圍。該範圍中的最後16個連續的代碼值都是以12位元數來表示，可保留這16個代碼值作為控制之用。因此，共有270個可使用的長度代碼，可對自2到271個位元組的相符位元組串進行編碼。不得將任何相符的單一位元組串編碼為COPY_POINTERS。

此外，可在一作業開始時清除過去記錄緩衝區。必然以升序將資料儲存在該過去記錄緩衝區，亦即自零位址的位置開始儲存。每當填滿整個過去記錄緩衝區時，必須將此位置(或位移值)重定為零，且並不將更新位置用於啟動或繼續位元組串的比對作業。

III. 實施例

現在請參閱圖2(a)及(b)，圖中示出根據本創作一較佳實施例的一固定長度LITERAL_COUNT編碼架構之高階虛擬碼。例

五、創作說明(10)

如，在此實施例中，使用了一個2位元的計數器L_CTR及三個8位元的L_DATA[]緩衝區。對於處理一LITERAL_DATA符號而言，如果L_CTR值等於3，則輸出"11"，自L_DATA[0]輸出資料位元組，自L_DATA[1]輸出資料位元組，自L_DATA[2]輸出資料位元組，將LITERAL_DATA位元組儲存在L_DATA[0]，並設定L_CTR等於1。然而，如果L_CTR值不等於3，則將該位元組儲存在以L_CTR為索引的L_DATA，並將L_CTR值遞增1。

對於將一COPY_POINTER編碼而言，必須首先編碼並輸出先前儲存的LITERAL_DATA位元組。如果L_CTR值等於3，則輸出"11"，並進入步驟(3)。如果L_CTR值等於2，則輸出"10"，並進入步驟(3)。如果L_CTR值等於1，則輸出"01"，並進入步驟(3)。如果L_CTR值等於0，則輸出"00"，並進入步驟(4)。在步驟(3)中，自L_DATA[0]輸出資料位元組；如果L_CTR>1，則自L_DATA[1]輸出資料位元組；如果L_CTR>2，則自L_DATA[2]輸出資料位元組；輸出"00"，並進入步驟(4)。在步驟(4)中，設定L_CTR等於0；然後編碼及輸出COPY_POINTER。

現在請參閱圖3(a)及(b)，圖中示出根據本創作一較佳實施例的一可變長度LITERAL_COUNT編碼架構之高階虛擬碼。例如，在此實施例中，使用了一個2位元的計數器L_CTR及兩個8位元的L_DATA[]緩衝區。對於將一LITERAL_DATA符號編碼而言，如果L_CTR值等於2，則輸出"11"，自L_DATA[0]輸出資料位元組，自L_DATA[1]輸

五、創作說明(13)

元件符號說明

- 10 壓縮器單元
- 11 控制器
- 12 內容可定址的記憶體
- 13 資料來源
- 14 資料接收器
- 15 解壓縮器單元
- 16 控制器
- 17 RAM/CAM
- 18 資料接收器
- 19 資料來源

91年8月22日

六、申請專利範圍

1. 一種利用一過去記錄緩衝區將一來源符號輸入資料流編碼而產生一輸出指標序列之資料壓縮器，該資料壓縮器包含：

編碼成 LITERAL_POINTER 之裝置，用以將該輸入資料流之一起始部分編碼成一 LITERAL_POINTER，其中該 LITERAL_POINTER 包含該資料流的至少一個資料位元組且該 LITERAL_POINTER 包含一 LITERAL_COUNT 欄位、及接續的數個 LITERAL_DATA 符號及該 LITERAL_COUNT 為一個位元或兩個位元；

編碼成 COPY_POINTER 之裝置，用以將該輸入資料流的一後續部分編碼成一 COPY_POINTER，其中該 COPY_POINTER 包含指向該過去記錄緩衝區之一計數欄位及一位移；以及

編碼成交替出現的 LITERAL_POINTER 及 COPY_POINTER 之裝置，用以將該輸入資料流的各後續資料位元組編碼成以交替方式出現的若干 LITERAL_POINTER 及 COPY_POINTER，因而該資料壓縮器的該輸出序列包含一個交替出現若干 LITERAL_POINTER 及若干 COPY_POINTER 之指標串。

2. 根據申請專利範圍第 1 項之資料壓縮器，其中該 LITERAL_COUNT 為固定長度。
3. 根據申請專利範圍第 1 項之資料壓縮器，其中該 LITERAL_COUNT 為兩個位元。
4. 根據申請專利範圍第 1 項之資料壓縮器，其中該 LITERAL_COUNT 為可變長度。

六、申請專利範圍

5. 根據申請專利範圍第1項之資料壓縮器，其中該計數欄位為可變長度。
6. 根據申請專利範圍第5項之資料壓縮器，其中該計數欄位為2到12個位元。
7. 一種利用一過去記錄緩衝區將一來源符號輸入資料流編碼而產生一輸出指標序列之程式軟體產品，該程式軟體產品包含：

編碼成LITERAL_POINTER之程式碼裝置，用以將該輸入資料流之一起始部分編碼成一LITERAL_POINTER，其中該LITERAL_POINTER包含該資料流的至少一個資料位元組且該LITERAL_POINTER包含一LITERAL_COUNT欄位、及接續的數個LITERAL_DATA符號及該LITERAL_COUNT為一個位元或兩個位元；

編碼成COPY_POINTER之程式碼裝置，用以將該輸入資料流的一後續部分編碼成一COPY_POINTER，其中該COPY_POINTER包含指向該過去記錄緩衝區之一計數欄位及一位移；以及

編碼成交替出現的LITERAL_POINTER及COPY_POINTER之程式碼裝置，用以將該輸入資料流的各後續資料位元組編碼成以交替方式出現的若干LITERAL_POINTER及COPY_POINTER，因而該資料壓縮器的該輸出序列包含一個交替出現若干LITERAL_POINTER及若干COPY_POINTER之指標串。

8. 根據申請專利範圍第7項之程式軟體產品，其中該

六、申請專利範圍

LITERAL_COUNT為固定長度。

9. 根據申請專利範圍第7項之程式軟體產品，其中該
LITERAL_COUNT為可變長度。

93年6月9日 修正
補充

LITERAL-DATA

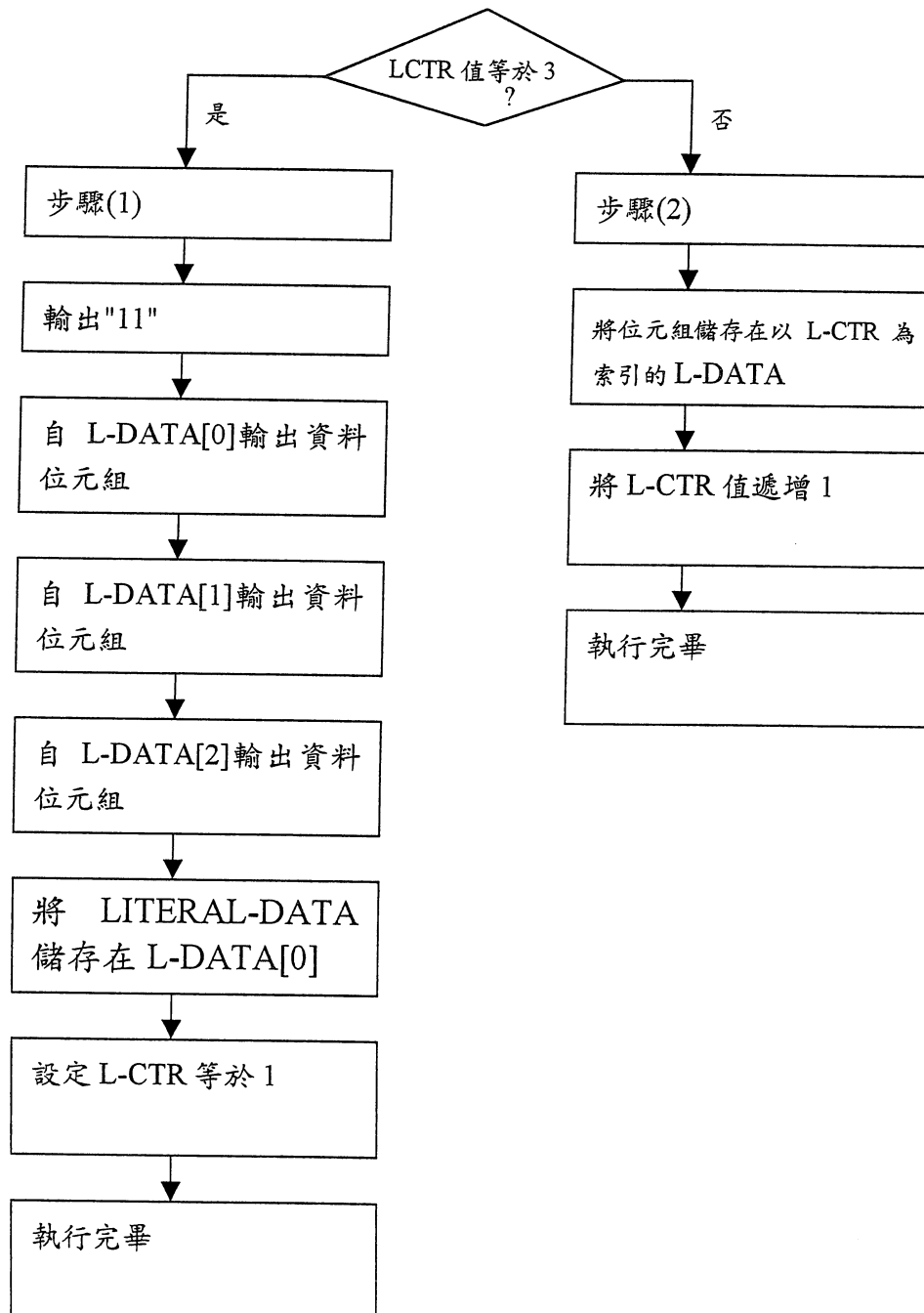


圖 2(a)

93年6月29日 修正
補充

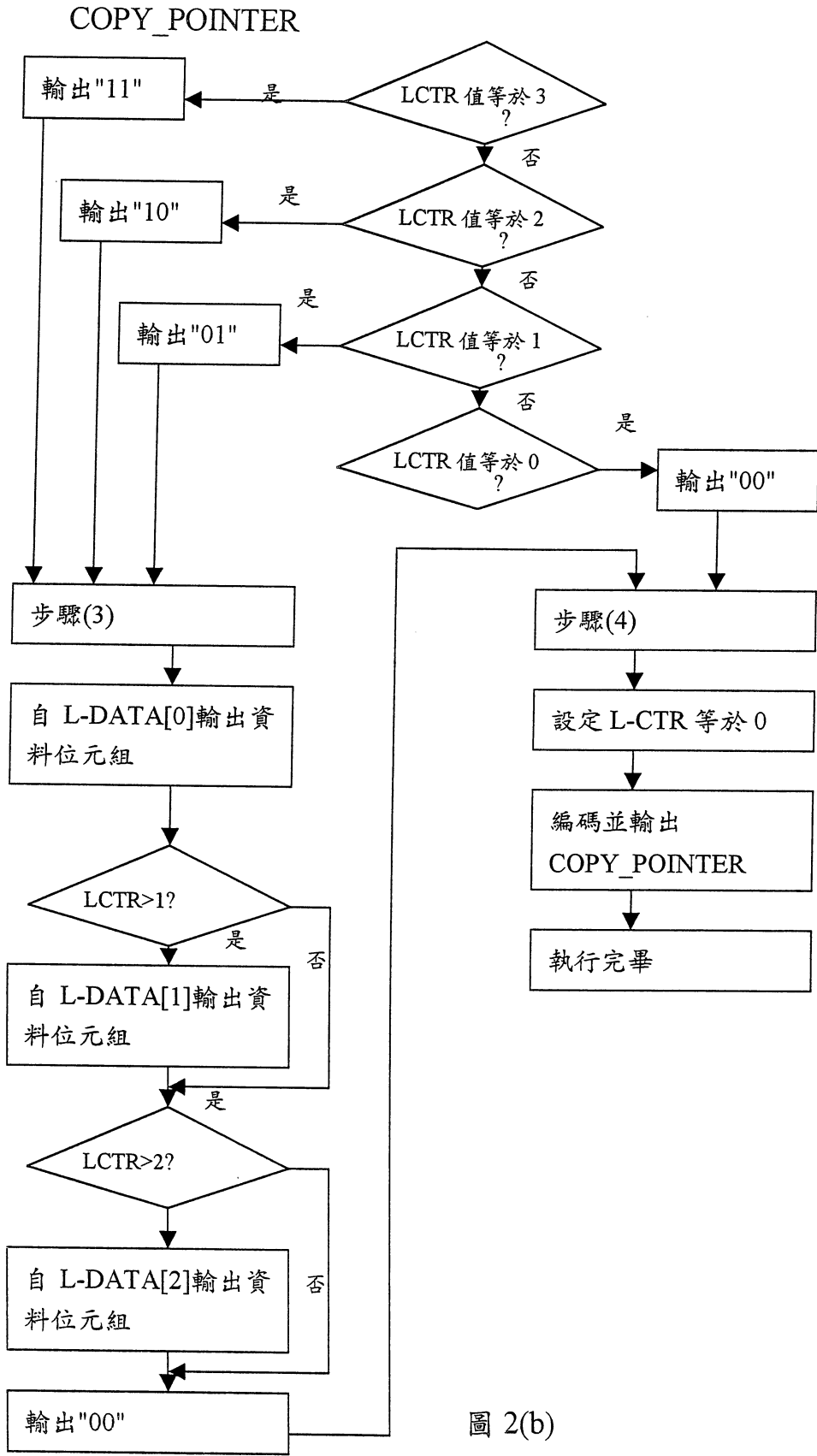


圖 2(b)

LITERAL-DATA

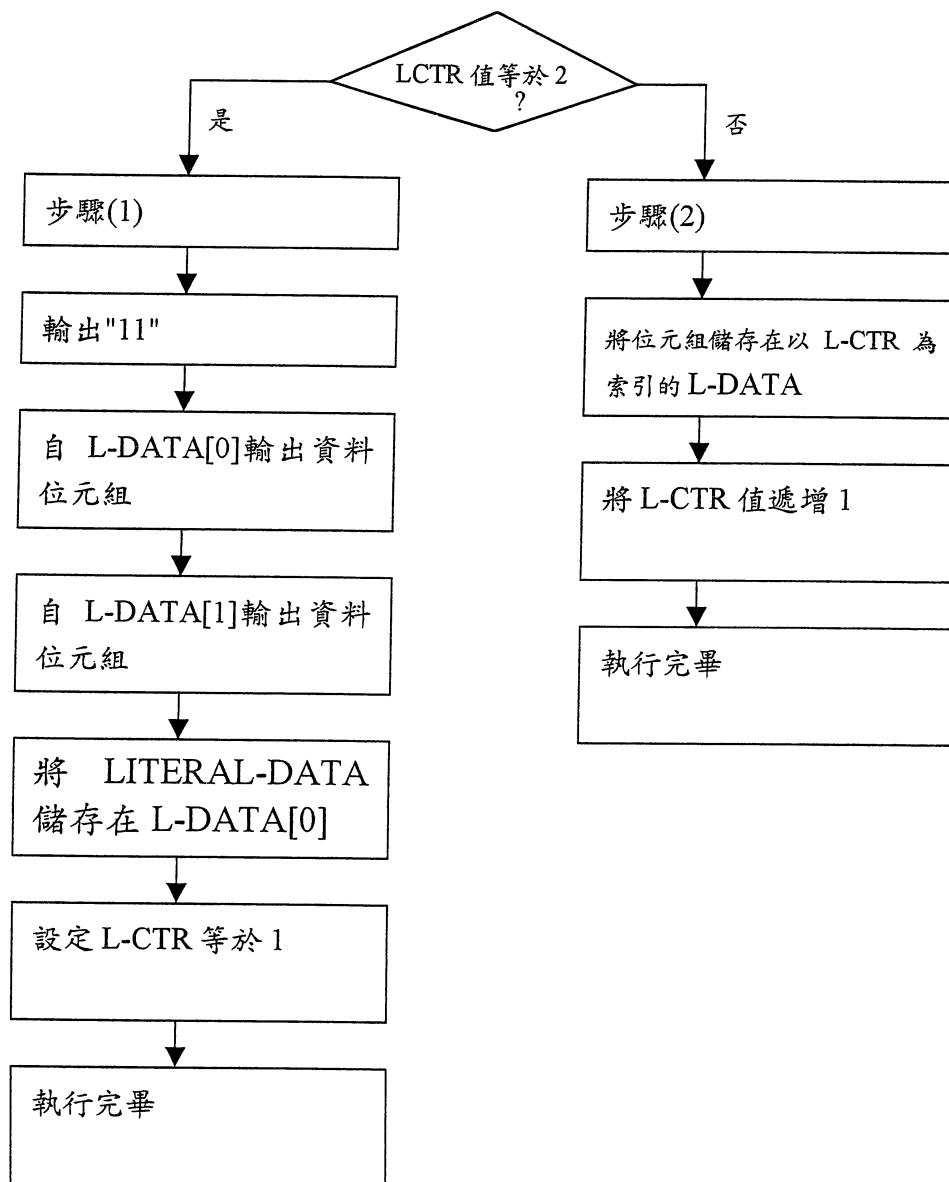


圖 3(a)

93年6月29日 修正
補充

COPY_POINTER

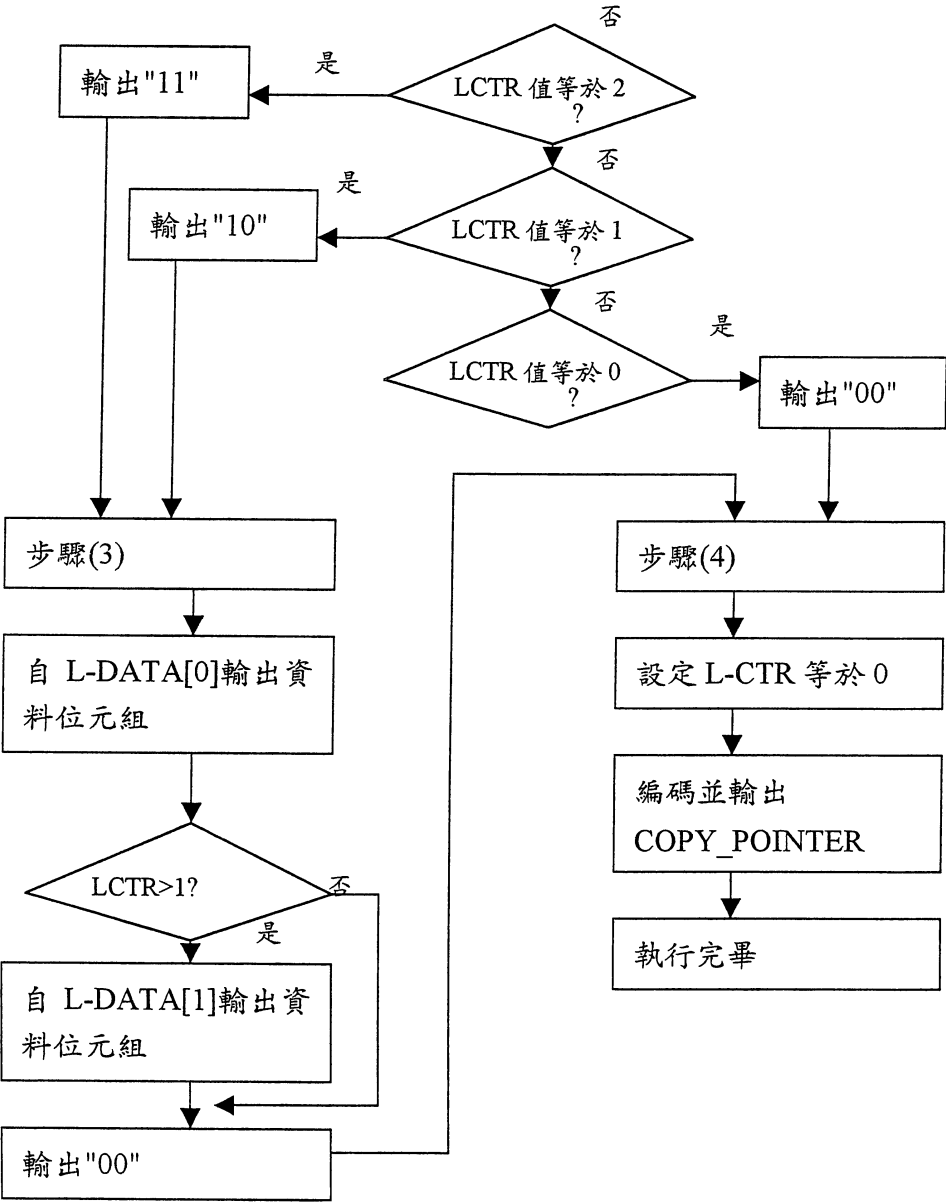


圖 3(b)