



## [12] 发明专利申请公开说明书

[21] 申请号 03803243.0

[43] 公开日 2005 年 6 月 15 日

[11] 公开号 CN 1628302A

[22] 申请日 2003.1.21 [21] 申请号 03803243.0

[30] 优先权

[32] 2002. 2. 5 [33] EP [31] 02075498.2

[86] 国际申请 PCT/IB2003/000217 2003.1.21

[87] 国际公布 WO2003/067466 英 2003.8.14

[85] 进入国家阶段日期 2004.8.4

[71] 申请人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 J·A·海特斯马

A·A·C·M·卡克

S·M·希梅尔

[74] 专利代理机构 中国专利代理(香港)有限公司

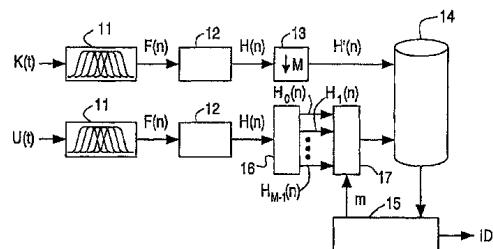
代理人 吴立明 王忠忠

权利要求书 2 页 说明书 5 页 附图 2 页

[54] 发明名称 指纹的有效存储器

[57] 摘要

本发明公开一种稳固的取指纹系统。该系统识别未知多媒体内容( $U(t)$ )，这是通过从所述内容提取指纹(一系列散列字)并且在其中存储了多个已知内容( $K(t)$ )的指纹的数据库中检索近似的指纹来完成的。为了更有效地在数据库中存储指纹和加快检索，已知信号( $K(t)$ )的散列字( $H(n)$ )在存储到数据库(14)之前通过因子  $M$  子采样。该已知信号( $K(t)$ )的散列字( $H(n)$ )被分成  $M$  个交叉的子系列( $H_0(n) \dots H_{M-1}(n)$ )。该交叉的子系列(17)在计算机(15)的控制下有选择地被应用于数据库(14)。只要子系列中存在一个充分地匹配一个存储的指纹，就识别了所述信号。



1. 用于为每个音频—图象信号在数据库中存储识别音频—图象媒体信号的指纹的方法，包括步骤：

- 5      —将所述音频—图象媒体信号分成一序列的帧；  
      —用因子  $M$  子采样所述帧序列以获得子采样的帧序列；  
      —为所述子采样的帧序列的每一个帧提取表示所述帧内基于感知的基本特性的散列字，以获得散列字的各个子采样的序列；  
      —将散列字的所述子采样的序列作为指纹存储在所述数据库中。

10     2. 如权利要求 1 所述方法，其中所述连续的帧是重叠的。

3. 用于在数据库中存储识别音频—图象媒体信号 ( $K(t)$ ) 的指纹的装置，该装置包括：

—组帧装置 (11)，用于将所述音频—图象媒体信号分成一序列的帧；

15     —子采样装置 (13)，通过用因子  $M$  子采样所述帧序列来获得子采样的帧序列；

      —装置 (12)，用于为所述子采样的帧序列的每一个帧提取表示所述帧内基于感知的基本特性的散列字，以获得散列字的各个子采样的序列；

20     —数据库 (14)，用于将散列字的所述子采样的序列作为指纹存储在所述数据库中。

4. 识别未知音频—图象媒体信号的方法，该方法包括步骤：

—将至少一部分未知音频—图象媒体信号分成一系列帧；  
      —为每个帧提取表示所述帧内基于感知的基本特性的散列字，以  
25     获得各个散列字系列；

      —将所述散列字系列分成  $M$  个交错的散列字子系列；  
      —将所述  $M$  个子系列连续地应用于其中已经为多个多媒体信号存  
      储了散列字的子采样的序列的数据库；

30     —将未知信号识别为多媒体信号，该多媒体信号的存储的散列字  
      的子采样的序列的至少一部分基本上与  $M$  个所应用的散列字的子系列  
      的至少一个匹配。

5. 如权利要求 4 所述方法，其中所述连续的帧是重叠的。

6. 用于识别未知音频—图象媒体信号的装置，该装置包括：

—组帧装置（11），用于将至少一部分未知音频—图象媒体信号  
(U(t)) 分成一系列帧；

5 特性的散列字，以获得各个散列字系列；

—交错装置（12），用于将所述散列字系列分成 M 个交错的散列  
字子系列；

—选择装置（17），用于将所述 M 个子系列连续地应用于其中已经  
为多个多媒体信号存储了散列字的子采样的序列的数据库；

10 —计算机装置（15），用于将未知信号识别为多媒体信号，该多  
媒体信号存储的散列字的子采样的序列的至少一部分基本上与 M 个所  
应用的散列字的子系列的至少一个匹配。

7. 识别未知音频—图象媒体信号的方法，该方法包括步骤：

15 —从远程工作站接收一系列散列字，这些散列字是通过将至少一  
部分未知音频—图象媒体信号分成一系列帧来生成的，以及为每个帧  
提取表示所述帧内基于感知的基本特性的散列字。

—将所述散列字系列分成 M 个交错的散列字子系列；

—将所述 M 个子系列连续地应用于其中已经为多个多媒体信号存  
储了散列字的子采样的序列的数据库；

20 —将未知信号识别为多媒体信号，该多媒体信号存储的散列字的  
子采样的序列的至少一部分基本上与 M 个所应用的散列字的子系列的  
至少一个匹配。

8. 如权利要求 5 所述方法，其中所述连续的帧是重叠的。

## 指纹的有效存储器

### 发明领域

5 本发明涉及用于将识别音频—图象媒体信号的指纹存储到数据库的方法和装置。本发明还涉及用于识别未知音频—图象媒体信号的方法和装置。

### 发明背景

10 指纹（在文献中也被称之为签名或散列）是信息信号的数字提要。在密码学中，散列在很长时间内被用作验证大型文件的正确接收。最近，为识别多媒体内容而引入散列的概念。识别诸如音频或视频剪辑的未知内容，这是通过将从所述剪辑提取的指纹和存储在数据库中的指纹集合比较来完成的。与密码学的极其脆弱（在一个大型文件中翻  
15 转一个比特将导致完全不同的散列）的散列相比，从音频—图象内容中提取的指纹是稳固的。对于大的范围，诸如压缩或解压缩、A/D 或 D/A 转换的处理是恒定的。

一个现有的取指纹系统在 Content-Based Multimedia Indexing(CBMI) conference in Brescia(Italy), 2001 由 Haitsma et  
20 al. 发表的：Robust Hashing for Content Identification 中公开。如这篇论文描述的，指纹从内容的基于感知的基本特性，即是从音频频谱频带中的能量分布中提取。对于视频信号，已经提出将视频图象中的亮度等级分布作为构建稳固指纹的基础。

通过将信号分成一系列（可能重叠的）帧，并提取每个帧中表示  
25 信号的基于感知的基本特性的散列字以获得散列字的各个系列来创建指纹。为了识别未知剪辑，数据库接收有关的散列字系列，并检索最近似的所存储的散列字系列。通过确定系列的多少个比特与数据库中的散列字系列匹配来测量相似性。如果 BER（比特错误率，不匹配比特的百分比）在特定的阈值之下，该剪辑被识别为源于数据库中最类  
30 似的散列字系列的歌曲或电影。

现有的取指纹方法的一个问题是数据库的大小。在 Haitsma et al.  
的论文中，音频信号被分成带有 31/32 重叠的 0.4 秒的帧。这样每

11.6ms (=0.4/32) 产生一个新帧。对于每个帧，提取 32 比特的散列字。由此，5 分钟的歌曲需要大概 100k 字节，即 5(分钟) × 60(秒) × 4(字节每散列字) /0.0116(秒每散列字)。更不必说数据库需要巨大的容量以允许识别歌曲的大量清单。类似的考虑适用于视频取指纹系统。

5

### 发明目的和概要

本发明的一个目的是提供用于在数据库中存储指纹的方法和系统，以缓解上述的问题。本发明的另一个目的是提供用于在数据库中识别未知音频—图象信号的方法和系统。

10 为此，本发明提供在独立权利要求 1 中定义的用于在数据库中存储指纹的方法。该方法与现有技术的不同之处在于，只有散列字的子采样序列被存储在数据库中。用在该权利要求中的单词“序列”被称之为完整长度的信号（歌曲或电影）。通过因子 M 达到存储量的缩小。

15 在该数据库中识别未知音频—图象的方法在独立权利要求 4 中定义。因为存在 M 个可能的子采样散列字序列存储在数据库中的不确定性，散列字的完整（即非子采样）系列根据该方法从未知剪辑中提取。在这里使用单词“系列”以引用可能的未知信号的短的剪辑或片段。散列字重叠的子系列被连续地应用于数据库，以便与保存在数据库中的子采样序列相匹配。如果所应用的子系列的其中至少一个具有在特定阈值以下的 BER，信号就被识别了。

20 利用本发明达到了在维持已知技术识别方法的稳固性和可靠性的同时，减少（通过 M 因子）了对存储容量的要求。

在从属权利要求中定义方法的更多优点的实施例。

25

### 附图简述

图 1 根据本发明显示用于在数据库中存储和识别音频—图象信号的指纹的装置的示例图。

图 2 是说明如图 1 中所示的装置的第一操作模式。

图 3 是说明如图 1 中所示的装置的第二操作模式。

30

图 4 是由图 1 中所示的计算机执行的操作步骤的流程图。

### 实施例的描述

将为音频信号描述本发明。图 1 根据本发明显示装置的示意图。该装置用作在数据库中存储未知音频信号的指纹（第一操作模式），以及用于识别未知音频信号（第二操作模式）。

将首先描述装置的第一操作模式（存储）。在该模式中，装置接收完整长度的音乐歌曲  $K(t)$ 。该信号在组帧电路 11 中被分成时间间隔或大概具有 0.4 秒长度的帧  $F(n)$ ，并由 Hanning 窗口用 31/32 的重叠加权。该重叠用作引入连续帧之间大的相关性。对于音频信号，这成为必要条件，因为应用于待识别的未知信号的组帧可能是不同的。

组帧电路 11 每  $11.6\text{ms} (=0.4/32)$  生成一个新的帧。散列提取电路 112 为每个帧生成 32 比特的散列字  $H(n)$ 。该散列提取电路的一个实际实施例在发明背景一章中的 Haitsma et al. 的论文中描述。简单概括，电路将每个音频信号帧的频谱分成若干频带并为每个频带产生一个指明在所述频带中的能量是高于还是低于给定阈值的散列位。图 2 显示了如此获得的散列字 21 的序列。

根据本发明，散列字的序列由子采样器 13 用因子  $M$  子采样，以生成一个子序列  $H'(n)$ 。散列字的子序列和诸如歌名、艺人名字等的识别数据一起组成已知音乐歌曲的指纹。在图 2 中显示了这样的指纹，其中数字 22 表示散列字的子序列，数字 23 表示识别歌曲的名称、艺人等。在计算机 15 的控制下该指纹存储在数据库 14 中。在这个例子中，子采样因子  $M=4$  被当作例子，5 分钟的歌曲需要大概  $6,000 \times 32$  比特的存储容量。与没有采用子采样的现有技术相比这可节约 75%。在实际上，可为巨大数量的已知音乐歌曲执行上述的存储操作。可以理解，散列字提取（12）和子采样（13）的操作顺序可以颠倒。

现在将描述装置的第二操作模式（识别）。在这种模式中，装置接收一部分（比如，3 秒）的未知歌曲，即音频剪辑  $U(t)$ 。该剪辑通过类似（或相同）的上述的组帧电路 11 和散列提取电路 12 处理。散列提取抽取电路 13 抽取剪辑的完整的散列块（未子采样）。对于 3 秒的剪辑，该操作生成一系列大概 256 个散列字  $H(n)$ 。这个表示未知音频剪辑的散列字系列也被称为散列块。在另一个实施例中，散列块从远程工作站抽取并只被所述装置接收。

散列块被应用于交错电路 16，将其分成  $M$  个交错的子系列或子块  $H_0(n), H_1(n), \dots, H_{M-1}(n)$ ，其中  $M$  是与上述子采样器 13 中使用的相同的

整数。图 3 说明对于  $M=4$  的交错过程。在该图中，数字 31 表示散列块连续的散列字，数字 32 表示字块  $H_0(n)$ ，数字 33 表示  $H_1(n)$ ，数字 34 表示  $H_{M-1}(n)$ 。

子块被应用于选择电路 17 的各个输入。在计算机 15 的控制下，  
5 子块  $H_0(n), H_1(n), \dots, H_{M-1}(n)$  连续地应用于数据库 14 以用作识别。如果在数据库中找到散列字系列，对于该散列字系列，比特错误率（即，在所述系列和所使用的子块之间不匹配的百分比）在特定的阈值之下，那么包括所述散列字系列的指纹识别出未知音频剪辑。

图 4 是由计算机执行的该识别操作的流程图。在步骤 41 中，索引  
10  $m$  获得初始值 0。该索引  $m$  被应用于选择电路 17，使得选择散列字的第一交错子块  $H_0(n)$  用于识别。在步骤 42 中，所选择的子块  $H_m(n)$  被应用于数据库。在步骤 43 中，检测是否在数据库中找到近似的散列字系列。单词“近似”被理解为引用具有最低 BER 的散列字系列，条件是，所述 BER 小于给定的阈值  $T$ 。在数据库中检索最近似散列字系列的策略的一个实际例子在前面提到的 Haitsma et al. 的论文中公开。检索策略的优选实施例还在申请人所附的未公开的欧洲专利申请  
15 01200505.4 (PHNL010110) 和 01202720.7 (PHNL010510) 中提出。

如果 BER 在阈值之下，识别了音频剪辑。在步骤 44 中，存储在数据库（图 2 中的 32）中歌曲的名称和制作者被传递给用户。如果不是这样，将索引  $m$  加一（步骤 45），使得另一个交错子块中的一个被应用于数据库。如果已经检索了所有  $M$  个交错子块而没有成功（步骤 46），则不能识别该音频剪辑。在步骤 47 中，这个结果被传递给用户。

使用本法达到利用因子  $M$  减小了数据库容量。需要注意，相同的容量需求的减少能通过有效地选择不同的帧重叠达到，在本例中即为  
25 7/8。只要涉及第一操作模式这就是正确的。但是，如果在识别过程中选择 7/8 相同的重叠而不用交错，那么将严重地影响识别的稳固性和可靠性。由此达到从一系列帧得到交错子块的至少一个，该一系列帧基本上（按时）与从其中得到所存储的散列字的一系列帧匹配。根据本发明的识别过程使用 31/32 的重叠得到基本上和现有技术方法相同的稳固性和可靠性。现在给出其数学背景。  
30

当应用使用因子  $M$  的子采样并且如果散列块中的比特是随机 i. i. d (独立而同一分布的)，BER 的标准偏差按因子  $M$  的开方增加。这表

明影响了稳固性和/或可靠性。如果 BER 上的阈值保持不变，没有影响稳固性但可靠性下降。另一方面，如果将阈值减小适当的量，可靠性保持不变但稳固性下降。

但是，音频一图象媒体信号的散列块中的比特沿时间轴有大的相关性，这是通过组帧的大的重叠和音乐中固有的相关性引入的。因此，当应用使用因子 M 的子采样时，标准偏差 s 不随因子 M 的开方而增加。实验显示，对于小的值 M，标准偏差的增加根本不显著。在没有子采样的实际系统中，BER 上的阈值被设置成 0.35。如果应用使用因子 M=4 的子采样，那么阈值只需要低于 0.342。因此，稳固性的减少不显著，同时数据库中所需的存储量按 4 的因子降低。而且，检索散列数据库所需的时间将单调的降低，因为在数据库中有少 4 倍的散列值。

如果子块的其中一个（一般为第一个）呈现具有比另一个阈值（基本上大于阈值 T）大的 BER，检索速度甚至能通过抑制应用另一个子块到数据库来更进一步增加。因为子块之间大的相关性（由于帧重叠和音乐中固有的相关性），另一个子块不太可能具有显著的较低的 BER。

公开了稳固的取指纹系统。该系统能识别未知多媒体内容 ( $U(t)$ )，这是通过从所述内容提取指纹（一系列散列字）并且在其中存储了多个已知内容 ( $K(t)$ ) 的指纹的数据库中检索近似的指纹来完成的。为了更有效地在数据库中存储指纹和加快检索，已知信号 ( $K(t)$ ) 的散列字 ( $H(n)$ ) 在存储到数据库 (14) 之前通过因子 M 子采样。该已知信号 ( $K(t)$ ) 的散列字 ( $H(n)$ ) 被分成 M 个交叉的子系列 ( $H_0(n) \dots H_{M-1}(n)$ )。该交叉的子系列 (17) 在计算机 (15) 的控制下可选择地被应用于数据库 (14)。只要子系列中的一个充分地匹配一个存储的指纹，就识别了所述信号。

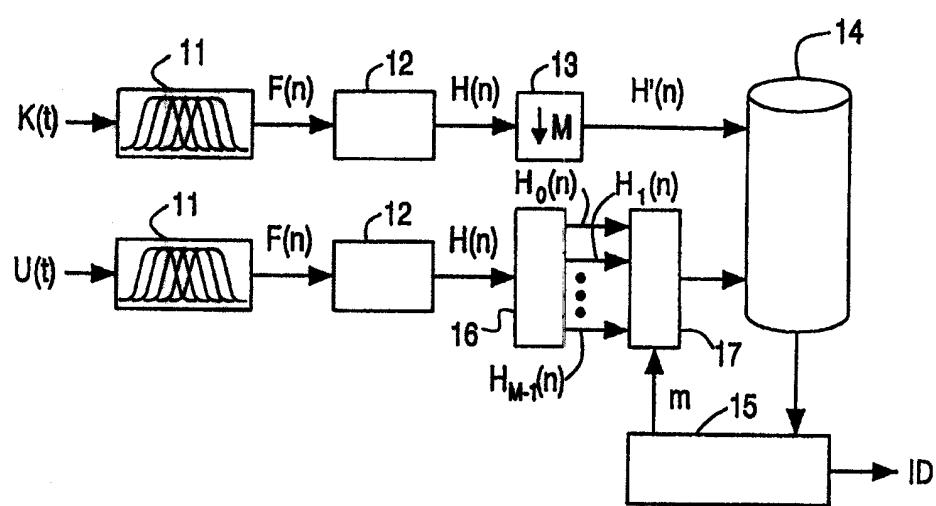


图 1

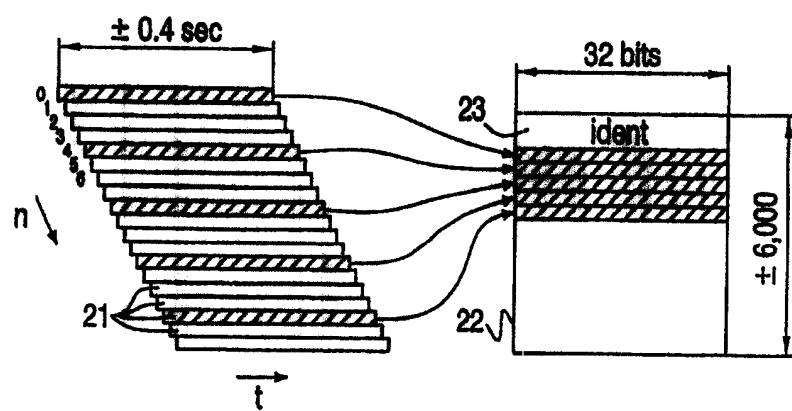


图 2

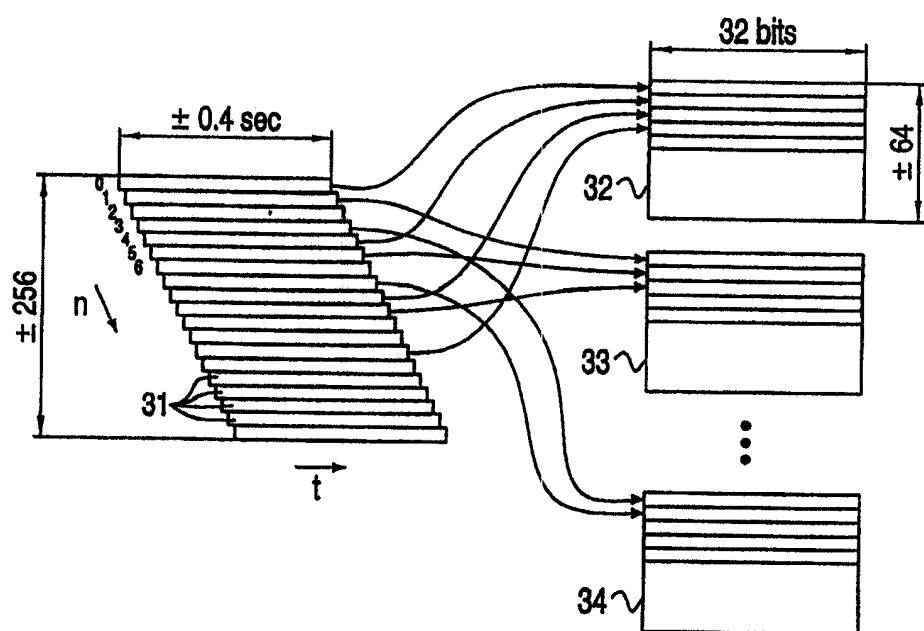


图 3

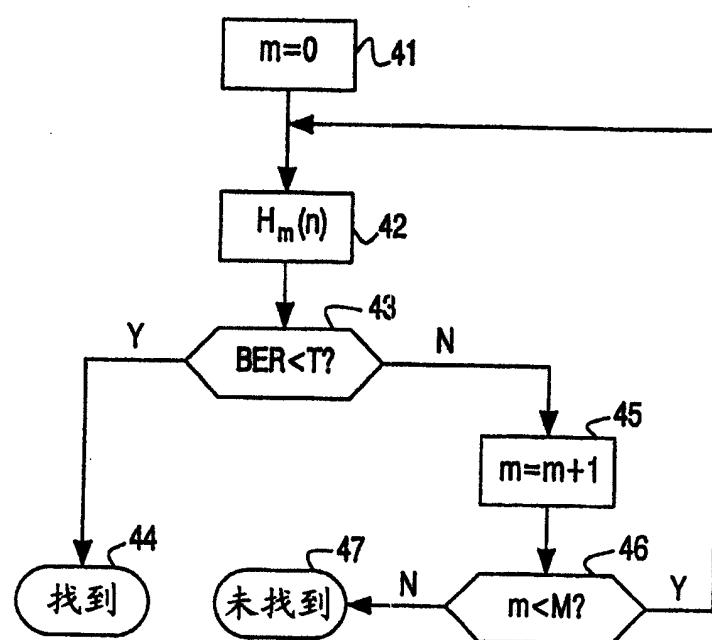


图 4