



[12] 发明专利申请公开说明书

[21] 申请号 03814902.8

[43] 公开日 2005年8月31日

[11] 公开号 CN 1663249A

[22] 申请日 2003.6.23 [21] 申请号 03814902.8

[30] 优先权

[32] 2002.6.24 [33] JP [31] 182506/2002

[32] 2002.11.1 [33] JP [31] 319756/2002

[32] 2002.11.1 [33] JP [31] 319757/2002

[32] 2002.11.19 [33] JP [31] 334831/2002

[86] 国际申请 PCT/JP2003/007908 2003.6.23

[87] 国际公布 WO2004/002144 日 2003.12.31

[85] 进入国家阶段日期 2004.12.24

[71] 申请人 松下电器产业株式会社

地址 日本大阪府

[72] 发明人 小林正明 酒井启行 松井谦二

桑野裕康 下田代雅文 安方满

远藤充

[74] 专利代理机构 永新专利商标代理有限公司

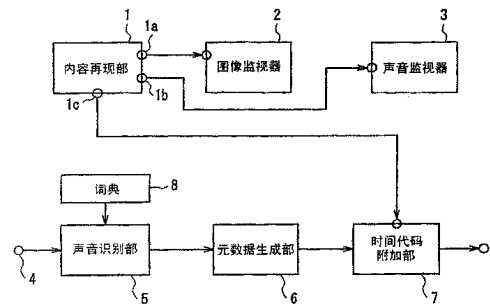
代理人 胡建新

权利要求书 3 页 说明书 14 页 附图 13 页
按照条约第 19 条的修改 3 页

[54] 发明名称 元数据生成装置、其生成方法以及检索装置

[57] 摘要

本发明涉及一种元数据生成装置，具有再现内容并输出的内容再现部(1)、用于监视内容再现部再现的内容的监视器(3)、声音输入部(4)、识别声音输入部输入的声音信号的声音识别部(5)、将声音识别部识别的信息转换成元数据的元数据生成部(6)，以及识别信息附加部(7)，该识别信息附加部(7)从内容再现部供给的再现后的内容获取用于识别所述内容中的各部分的识别信息，并赋予所述元数据；该元数据生成装置使生成的所述元数据与所述内容内的各部分建立关联。



1. 一种元数据生成装置，其特征在于，具有再现内容并输出的内容再现部、声音输入部、识别所述声音输入部输入的声音信号的声音识别部、将所述声音识别部识别的信息转换成元数据的元数据生成部，以及识别信息附加部，该识别信息附加部从所述内容再现部供给的再现后的内容，获取用于识别所述内容中的各部分的识别信息，并将其附加到所述元数据上；该元数据生成装置使生成的所述元数据与所述内容内的各部分相关联。

2. 如权利要求1所述的元数据生成装置，其特征在于，还具有与所述内容相关的词典，在通过所述声音识别部识别由所述声音输入部输入的声音信号时，与所述词典相关联地进行识别。

3. 如权利要求2所述的元数据生成装置，其特征在于，所述声音识别部以单词为单位，与所述词典相关联地识别所述声音信号。

4. 如权利要求1或3所述的元数据生成装置，其特征在于，还具有包含键盘的信息处理部，可以通过从所述键盘输入，由所述信息处理部修正元数据。

5. 如权利要求1~5中任一项所述的元数据生成装置，其特征在于，使用附在所述内容上的时间代码信息，作为所述识别信息。

6. 如权利要求1~6任一项所述的元数据生成装置，其特征在于，使用附在所述内容上的内容的地址、编号或帧号码，作为所述识别信息。

7. 如权利要求1所述的元数据生成装置，其特征在于，所述内容为静态图像内容，使用所述静态图像内容的各个地址，作为所述识别信息。

8. 如权利要求1所述的元数据生成装置，其特征在于，
所述内容再现部由内容数据库构成；
所述声音输入部按照与所述内容数据库供给的同步信号同步的时钟，将所输入的关键词的声音信号转化为数据并供给所述声音识别部；
所述声音识别部根据在所述声音输入部转化为数据的声音信号数据，识别所述关键词；

所述元数据生成部构成文件处理部，该文件处理部将表示所述内容

中包含的图像信号的时间位置的时间代码用作所述识别信息，并将所述声音识别部输出的关键词与所述时间代码相结合，来生成元数据文件。

9、如权利要求 8 所述的元数据生成装置，其特征在于，还具有记录部，将所述内容数据库供给的内容与所述元数据文件一道作为内容文件进行记录。

10、如权利要求 9 所述的元数据生成装置，其特征在于，还具有生成控制文件的内容信息文件处理部，所述控制文件用于管理应记录所述内容文件的记录位置和所述元数据文件之间的关系，

在所述记录部中，与所述内容文件和所述元数据文件一道记录所述控制文件。

11、如权利要求 8 所述的元数据生成装置，其特征在于，还具有词典数据库，所述声音识别部可以从多种类型词典中选择适合于所述内容的种类词典。

12、如权利要求 11 所述的元数据生成装置，其特征在于，可以将与内容相关的关键词供给所述声音识别部，所述声音识别部优先识别所述关键词。

13、一种元数据生成方法，其特征在于，对与内容相关的信息进行声音输入，由声音识别装置对输入的声音信号进行声音识别，将所述声音识别后的信息转换成元数据，将用于识别所述内容内的各部分的、附加在所述内容上的识别信息附加到所述元数据上，使生成的所述元数据与所述内容内的各部分相关联。

14、如权利要求 13 所述的元数据生成方法，其特征在于，再现所述内容并在监视器上显示，同时对与内容相关的信息进行声音输入。

15、如权利要求 13 所述的元数据生成方法，其特征在于，利用与所述内容相关的词典，并由所述声音识别装置与所述词典相关联地识别所述输入的声音信号。

16、如权利要求 13 所述的元数据生成方法，其特征在于，使用附在所述内容上的时间代码信息，作为所述识别信息。

17、如权利要求 13 所述的元数据生成方法，其特征在于，使用静态

图像内容作为所述内容，并使用所述静态图像内容的各个地址作为所述识别信息。

18、一种元数据检索装置，其特征在于，包括：

再现内容并输出的内容数据库；

声音输入部，按照与再现的所述内容的同步信号同步的时钟，将所输入的关键词的声音信号转化为数据；

声音识别部，根据由所述声音输入部转化为数据的声音信号数据识别关键词；

文件处理部，将所述声音识别部输出的关键词与表示所述内容中包含的图像信号的时间位置的时间代码相结合，来生成元数据文件；

内容信息文件处理部，生成控制文件，所述控制文件用于对内容文件的记录位置和所述元数据文件之间的关系进行管理；

记录部，记录所述内容文件、所述元数据文件和所述控制文件；

检索部，确定包含所输入的检索关键词的所述元数据文件，并参照所述控制文件提取与所述内容文件的所述关键词对应的记录位置；

所述内容文件的记录位置是所述记录部中的记录位置。

19、如权利要求 18 所述的元数据检索装置，其特征在于，所述内容信息文件处理部输出的控制文件是记载与所述内容的记录时间对应的、所述记录部中所述内容的记录位置的表，可以根据所述时间代码检索所述内容的记录位置。

20、如权利要求 18 所述的元数据检索装置，其特征在于，还具有词典数据库和向所述声音识别部供给与内容相关的关键词的关键词供给部，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类词典，并且优先识别所述关键词。

21、如权利要求 18 所述的元数据检索装置，其特征在于，还具有词典数据库，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类的词典，所述检索部利用从所述声音识别部使用的通用词典中选定的关键词进行检索。

元数据生成装置、其生成方法以及检索装置

技术领域

本发明涉及用于生成与制作好的图像、声音等内容相关联的元数据（metadata）的元数据生成装置及元数据生成方法。而且涉及利用生成的元数据检索内容的检索装置。

背景技术

近年来，人们向制作好的图像、声音内容附加与这些内容相关联的元数据。

但是，现有的元数据附加操作一般采用的方法是，以制作好的图像、声音内容的脚本或解说原稿为基础，使制作好的图像、声音内容再现并确认应作为元数据的信息，然后手工进行计算机输入来生成。因此，在元数据生成过程中需要付出相当多的劳动。

特开平 09—130736 号公报中记载了一种在照相机摄影时利用声音识别附加标记的系统。但该系统在摄影的同时使用，无法适用于向已经制作好的内容附加元数据。

发明内容

本发明解决上述问题，其目的在于，提供一种可以通过声音输入容易地针对制作好的内容生成元数据的元数据生成装置和元数据生成方法。

本发明的另一目的在于，提供一种利用上述生成的元数据可容易检索内容的检索装置。

本发明的元数据生成装置，具有再现内容并输出的内容再现部、声音输入部、识别所述声音输入部输入的声音信号的声音识别部、将所述声音识别部识别的信息转换成元数据的元数据生成部，以及识别信息附加部，该识别信息附加部从所述内容再现部供给的再现后的内容，获取用于识别所述内容中的各部分的识别信息，并将其附加到所述元数据上；

该元数据生成装置使生成的所述元数据与所述内容内的各部分相关联。

本发明的元数据生成方法是，对与内容相关的信息进行声音输入，由声音识别装置对输入的声音信号进行声音识别，将所述声音识别后的信息转换成元数据，将用于识别所述内容内的各部分的、附加在所述内容上的识别信息附加到所述元数据上，使生成的所述元数据与所述内容内的各部分相关联。

本发明的元数据检索装置，包括：再现内容并输出的内容数据库；声音输入部，按照与再现的所述内容的同步信号同步的时钟，将所输入的关键词的声音信号转化为数据；声音识别部，根据由所述声音输入部转化为数据的声音信号数据识别关键词；文件处理部，将所述声音识别部输出的关键词与表示所述内容中包含的图像信号的时间位置的时间代码相结合，来生成元数据文件；内容信息文件处理部，生成控制文件，所述控制文件用于对内容文件的记录位置和所述元数据文件之间的关系进行管理；记录部，记录所述内容文件、所述元数据文件和所述控制文件；检索部，确定包含所输入的检索关键词的所述元数据文件，并参照所述控制文件提取与所述内容文件的所述关键词对应的记录位置；所述内容文件的记录位置是所述记录部中的记录位置。

附图说明

图 1 所示为本发明实施例 1 的元数据生成装置的结构框图。

图 2 所示为本发明实施例 1 的附有时间代码的元数据的一个例子。

图 3 所示为本发明实施例 2 的元数据生成装置的结构框图。

图 4 所示为相同装置中静态图像内容·元数据显示部的一个例子。

图 5 所示为本发明实施例 2 的元数据生成装置的其它结构框图。

图 6 所示为本发明实施例 3 的元数据生成装置的结构框图。

图 7 所示为同一实施例的装置中词典 DB 的一个例子的结构图。

图 8 所示为作为同一实施例的装置中使用的内容脚本的一个例子的食谱。

图 9 所示为表示同一实施例的装置生成的元数据文件的一个例子的 TEXT 形式的数据图。

图 10 所示为本发明实施例 4 的元数据生成装置的结构框图。

图 11 所示为同一实施例的装置生成的信息文件的一个例子的结构图。

图 12 所示为本发明实施例 5 的元数据检索装置的结构框图。

图 13 所示为本发明实施例 6 的元数据生成装置的结构框图。

具体实施方式

根据本发明的元数据生成装置，在进行与内容相关的元数据的生成或添加标记时，利用声音识别并通过声音输入，生成元数据或标记，同时使元数据或标记与内容的时间或场景等建立关联。从而，可以通过声音输入，自动生成以往通过键盘输入来生成的元数据。所谓元数据是指标记的集合体，本发明中称为元数据的场合还包括指标记自身的场合。另外，所谓内容，是指制作好的图像、声音内容，静态图像内容，数据库化的图像、声音内容等一般被称为内容的所有对象。

本发明的元数据生成装置最好还具有与内容相关的词典，在通过所述声音识别部识别由所述声音输入部输入的声音信号时，与所述词典关联起来进行识别。根据该结构，将事先从制作好的内容的脚本等当中提取的关键词作为声音信号输入，并根据脚本设定词典领域和赋予关键词优先次序，从而可以高效、准确地利用声音识别单元生成元数据。

所述声音识别部也可以以单词为单位与所述词典相关联地识别所述声音信号。而且，该元数据生成装置最好还具有包含键盘的信息处理部，可以利用从所述键盘输入，通过所述信息处理部对所述元数据进行修正。可以使用附在所述内容上的时间代码信息，作为所述识别信息。或者也可以使用附在所述内容上的内容的地址、编号或帧号码，作为所述识别信息。另外，所述内容可以为静态图像内容，使用所述静态图像内容的各个地址作为识别信息。

作为本发明的一个应用例，可以构成如下的元数据生成装置。即，所述内容再现部由内容数据库构成。所述声音输入部按照与所述内容数据库供给的同步信号同步的时钟，将所输入的关键词的声音信号转化为数据并供给所述声音识别部。所述声音识别部根据在所述声音输入部转

化为数据的语音信号数据，识别所述关键词。所述元数据生成部构成文件处理部，该文件处理部将表示所述内容中包含的图像信号的时间位置的时间代码用作所述识别信息，并将所述声音识别部输出的关键词与所述时间代码相结合，来生成元数据文件。

根据该结构，在数秒单位的间隔内也可以高效地附加元数据。从而可以在依靠以往的键输入很困难的短时间间隔内生成元数据。

该结构最好还具有记录部，将所述内容数据库供给的内容与所述元数据文件一道作为内容文件进行记录。而且，最好还具有生成控制文件的内容信息文件处理部，所述控制文件用于管理应记录所述内容文件的记录位置和所述元数据文件之间的关系，在所述记录部中，与所述内容文件和所述元数据文件一道记录所述控制文件。另外，最好还具有词典数据库，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类的词典。最好还可以将与内容相关的关键词供给所述声音识别部，所述声音识别部优先识别所述关键词。

本发明的元数据生成方法最好是，再现所述内容并在监视器上显示，同时对与内容相关的信息进行声音输入。而且，最好利用与所述内容相关的词典，并由所述声音识别装置与所述词典相关联地识别所述输入的声音信号。另外，最好使用附在所述内容上的时间代码信息作为所述识别信息。另外，可以使用静态图像内容作为所述内容，并使用所述静态图像内容的各个地址作为所述识别信息。

根据本发明的元数据检索装置，通过使用表示内容记录位置的控制文件和表示元数据以及时间代码等的元数据文件，可以根据元数据快速检索内容的期望部分。

本发明的元数据检索装置中，最好所述内容信息文件处理部输出的控制文件是记载与所述内容的记录时间对应的、所述记录部中的所述内容的记录位置的表，可以根据所述时间代码检索所述内容的记录位置。

而且，最好还具有词典数据库和向所述声音识别部供给与内容相关的关键词的关键词供给部，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类词典，并且优先识别所述关键词。

另外，最好还具有词典数据库，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类的词典，所述检索部利用从所述声音识别部使用的通用词典中选定的关键词进行检索。

以下参照附图具体说明本发明的实施例。

(实施例 1)

图 1 所示为本发明实施例 1 的元数据生成装置的结构框图。内容再现部 1 是用于在元数据生成时确认所制作内容的部件。内容再现部 1 的输出供给图像监视器 2、声音监视器 3 和时间代码附加部 7。设置麦克风 4，作为用于生成元数据的声音输入部。从麦克风 4 输入的声音被输入到声音识别部 5。声音识别部 5 上连接有声音识别用词典 8，并且可以参照其中的数据。声音识别部 5 的识别输出供给元数据生成部 6，生成的元数据供给时间代码附加部 7，然后可以从时间代码附加部 7 输出到外部。

可以使用例如 VTR、硬盘装置、光盘装置等图像·声音信号再现装置，以半导体存储器等存储单元为记录介质的图像·声音信号再现装置，或者对通过传输、广播供给的图像·声音信号进行再现的图像·声音再现装置等，作为内容再现部 1。

再现的图像信号从内容再现部 1 的图像信号输出端子 1a 供给图像监视器 2。再现的声音信号从声音信号输出端子 1b 供给声音监视器 3。再现的时间代码从时间代码输出端子 1c 供给时间代码附加部 7。再有，图像监视器 2 和声音监视器 3 作为元数据生成装置的部件不是必须的，可以根据需要连接并使用。

生成元数据时，操作人员确认图像监视器 2 和声音监视器 3 中的一个或两者，同时根据情况参照脚本或解说原稿等，读出应输入的元数据，并通过麦克风 4 输入。从麦克风 4 输出的声音信号供给声音识别部 5。另外，声音识别部 5 根据需要参照声音识别用词典 8 的数据。由声音识别部 5 识别出的声音数据被供给元数据生成部 6，并转换成元数据。

如上生成的元数据上，为了附加使内容各部分的时间或场景的关系相对应的信息，由时间代码附加部 7 附加从再现的内容中取得的、由内容再现部 1 供给的时间代码信息。

为了更加具体地说明以上动作，假设内容为料理说明的情况下，一个场面为例。这种情况下，操作者确认图像监视器 2 的显示画面，同时通过麦克风 4 发出“盐 1 勺”的声音，声音识别部 5 参照词典 8 识别为“盐”、“1 勺”，然后元数据生成部 6 分别将其转换为“盐”、“1 勺”的标记。再有，声音识别部 5 的结构没有特别限定，可以利用平常使用的各种声音识别单元来进行声音识别，只要识别为“盐”、“1 勺”等数据即可。一般情况下，元数据是指这些标记的集合体。声音识别的结果如图 2 所示，从元数据生成部 6 输出元数据 9a，然后供给时间代码附加部 7。

时间代码附加部 7 根据内容再现部 1 供给的时间代码信号 9b，生成由附加了时间代码的、附有时间代码的元数据 10 构成的信息包数据。生成的元数据可以直接输出，也可以存储在硬盘等记录介质中。

本例中表示的是生成信息包形式的元数据的情况，但不限于此。

（实施例 2）

图 3 所示为本发明实施例 2 的元数据生成装置的结构框图。本实施例以静态图像内容为元数据生成对象。为了识别静态图像内容，利用相当于动态图像场合下的时间代码的内容地址，使生成的元数据与静态图像内容相关联。

在图 3 中，照相机 11 是用于制作静态图像内容的元件。由静态图像内容记录部 12 对照相机 11 的输出附加地址信息并进行记录。为了生成元数据，此处记录的静态图像内容和地址信息被供给静态图像内容·元数据记录部 13。地址信息还被供给元数据地址附加部 19。

麦克风 16 用于对静态图像的相关信息的声音输入，其输出被输入到声音识别部 17。声音识别部 17 与声音识别用词典 20 相连，并且可以参照其数据。声音识别部 17 的识别输出供给元数据生成部 18，生成的元数据供给元数据地址附加部 19。

静态图像内容·元数据记录部 13 中记录的静态图像内容和元数据，由静态图像内容·元数据再现部 14 进行再现，然后在静态图像内容·元数据显示部 15 上显示。

以下详细说明上述结构的元数据生成装置的动作。

由照相机 11 拍摄的静态图像内容，由静态图像内容记录部 12 被记录到记录介质（未图示）上，并被附加地址信息而地址信息也记录到记录介质上。记录介质一般由半导体存储器构成，但不限于半导体存储器，也可以使用例如磁存储器、光记录介质、光磁记录介质等各种记录介质。所记录的静态图像内容通过输出端子 12a 和输入端子 13a，地址信息同样通过地址输出端子 12b 和输入端子 13b 被供给静态图像内容·元数据记录部 13。地址信息还通过输出端子 12b 和输入端子 19b，被供给元数据地址附加部 19。

另一方面，由照相机 11 拍摄的静态图像的相关信息，通过麦克风 16 输入到声音识别部 17。与静态图像相关的信息包括标题、拍摄时间、拍摄者、拍摄场所（哪里）、被拍摄者（谁）以及被拍摄物（什么）等。另外，还可以根据需要对声音识别部 17 供给声音识别用词典 20 的数据。

由声音识别部 17 识别的声音数据被供给元数据生成部 18，并被转换成元数据或标记。一般情况下，元数据是指作为与内容相关的信息的标题、拍摄时间、拍摄者、拍摄场所（哪里）、被拍摄者（谁）、被拍摄物（什么）等标记的集合体。为了附加与静态图像内容的内容或场景的关系相对应的信息，上述生成的元数据或标记被供给元数据地址附加部 19。元数据地址附加部 19 将静态图像内容记录部 12 供给的地址信息附加到元数据上。如上附加了地址信息的附地址元数据，通过输出端子 19c 和输入端子 13c 被供给静态图像内容·元数据记录部 13。在静态图像内容·元数据记录部 13，将同一地址的静态图像内容和同一地址的元数据关联起来进行记录。

为了更具体地说明附有地址的元数据，图 4 中示出了由静态图像内容·元数据再现部 14 再现静态图像内容·元数据记录部 13 记录的静态图像内容和元数据、并由静态图像内容·元数据显示部 15 显示结果的一例。

图 4 中的静态图像内容·元数据显示部 15 的画面是一个例子，由静态图像内容显示部 21、地址显示部 22 和元数据显示区域 23 构成。元数

据显示区域 23 由例如 1) 标题记述部 23a、2) 日期时间记叙部 23b、3) 拍摄者记叙部 23c、4) 拍摄场所记叙部 23d 等构成。这些元数据由上述声音识别部 17 识别的声音数据生成。

在上述动作中，在静态图像内容拍摄前、几乎与拍摄同时或拍摄之后等不必确认所拍摄的静态图像内容的情况下，生成元数据。

以下参照图 5 说明在静态图像内容制作好后进行元数据生成的情况下，再现静态图像内容并针对所监视的静态图像内容生成元数据的情况。与图 3 相同的部件使用相同的标记，并省略对其功能等的说明。这种情况下，静态图像内容记录部 12 与静态图像内容·元数据记录部 13 之间设有静态图像内容·地址再现部 24。并且，设有供给静态图像内容·地址再现部 24 的输出的监视器 25。

由照相机 11 拍摄并供给静态图像内容记录部 12 的静态图像内容，被记录在记录介质（未图示）上并被附加了地址，地址也记录在记录介质中。这样的记录介质被供给静态图像内容·地址再现部 24。这样，在用于再现制作完毕的静态图像内容并针对所监视的静态图像内容生成元数据的元数据生成装置中，照相机 11 和静态图像内容记录部 12 不是必须部件。

由静态图像内容·地址再现部 24 再现的静态图像内容供给监视器 25。同样，再现的地址信息通过输出端子 24b 和输入端子 19b 供给元数据地址附加部 19。元数据生成的负责人员在确认监视器 25 上显示的静态图像内容的基础上，通过麦克风 16 声音输入对生成元数据所必须的词汇。这样，与照相机 11 拍摄的静态图像相关的信息通过麦克风 16 被输入到声音识别部 17。静态图像的相关信息包括标题、拍摄时间、拍摄者、拍摄场所（哪里）、被拍摄者（谁）以及被拍摄物（什么）等。此后的动作与对图 3 所示结构的说明相同。

（实施例 3）

图 6 所示为本发明实施例 3 的元数据生成装置的结构框图。本实施例是以一般的数字数据内容为元数据生成对象的例子。为了识别数字数据内容，利用内容的地址或编号，使数字数据内容与所生成的元数据建

立关联。

在图 6 中, 31 为内容数据库(以下称为内容 DB), 由内容 DB31 再现的输出被供给声音输入部 32、文件处理部 35 和记录部 37。声音输入部 32 的输出被供给声音识别部 33。另外还可以将词典数据库(以下称为词典 DB) 34 的数据供给声音识别部 33。元数据由声音识别部 33 输出, 并输入到文件处理部 35。文件处理部 35 利用内容 DB31 供给的时间代码值, 在声音识别部 33 输出的元数据上附注规定数据来对其进行文件化处理。文件处理部 35 输出的元数据文件供给记录部 37, 与内容 DB31 输出的内容一道进行记录。声音输入部 32 和词典 DB34 上分别设有声音输入端子 39 和词典领域选择输入端子 40。内容 DB31 的再现输出和记录部 37 的再现输出可以通过图像监视器 41 进行显示。

内容 DB31 具有这样的功能: 针对例如 VTR、硬盘装置、光盘装置等图像·声音信号再现装置, 以半导体存储器等存储单元为记录介质的图像·声音信号再现装置, 或者对通过传输、广播供给的图像·声音信号进行一次性记录并再现的图像·声音再现装置等的制作好的内容, 产生对应于内容的时间代码, 并进行再现。

以下说明上述元数据生成装置的动作。内容 DB31 再现的附有时间代码的图像信号被供给图像监视器 41 并放映。操作者利用麦克风根据图像监视器 41 上放映的内容输入解说的声音信号后, 声音信号通过声音输入端子 39 输入到声音输入部 32。

此时, 操作者最好确认图像监视器 41 上放映出的内容或时间代码, 读出在脚本、解说原稿或内容等的基础上提取的内容管理用关键词。通过使用预先由脚本等限定的关键词作为如上输入的声音信号, 可以改善其后的声音识别部 33 的识别率。

声音输入部 32 利用与内容 DB1 输出的垂直同步信号同步的时钟, 使声音输入端子 39 输入的声音信号转化为数据。由声音输入部 32 转化为数据后的声音信号数据输入到声音识别部 33, 同时, 由词典 DB34 供给声音识别所必要的词典。词典 DB34 中使用的声音识别用词典可以由词典领域选择输入端子 40 设定。

例如图 7 所示，假定词典 DB34 按各领域构成，则所使用的领域由词典领域选择输入端子 40（例如可以进行键输入的键盘端子）设定。例如在料理节目的场合下，可以由端子 40 将词典 DB34 的领域设定成料理—日本料理—烹饪法—蔬菜炒法。通过如此设定词典 DB34，可以限制所使用的单词和可以声音识别的单词，从而改善声音识别部 33 的识别率。

另外，可以通过图 6 中的词典领域选择端子 40，输入从脚本、脚本原稿或内容的内容当中提取的关键词。例如，在内容为料理节目的场合下，通过端子 40 输入图 8 所示的食谱。考虑到内容，食谱中记载的单词作为声音信号输入的可能性高，因而词典 DB34 中明示出从端子 40 输入的食谱单词的识别优先度，并优先进行声音识别。例如，“西红柿”和“貝之牡蛎”在词典中，而从端子 40 输入的食谱单词仅有“貝之牡蛎”时，赋予“貝の牡蛎”优先次序 1。声音识别部 33 在识别到“牡蛎”这样的声音时，识别为词典 DB34 中设定的单词的优先次序记载为 1 的“貝之牡蛎”。

这样，词典 DB34 中按从端子 40 输入的领域限定单词，并且在从端子 40 输入脚本后明示单词的优先度，从而可以改善声音识别部 33 的识别率。

图 6 中的声音识别部 33 根据词典 DB34 供给的词典，识别从声音输入部 32 输入的声音信号数据并生成元数据。声音识别部 33 输出的元数据输入到文件处理部 35。如前所述，声音输入部 32 与内容 DB31 再现的垂直同步信号同步地将声音信号转化为数据。这样，文件处理部 35 利用来自声音输入部 32 的同步信息和内容 DB31 供给的时间代码值，在例如前述的料理节目的场合下，生成图 9 所示的 TEXT 形式的元数据文件。即，文件处理部 35 在声音识别部 33 输出的元数据上，附加作为开始归档（file）之后的每 1 秒的基准时间的 TM_ENT（秒）、表示从基准时间开始的偏移帧数的 TM_OFFSET 和时间代码，以这样的形式进行归档（file）处理。

记录部 37 记录文件处理部 35 输出的元数据文件和内容 DB31 输出

的内容。记录部 37 由 HDD、存储器、光盘等构成，内容 DB31 输出的内容也以文件形式记录。

(实施例 4)

图 10 所示为本发明实施例 4 的元数据生成装置的结构框图。本实施例的装置与实施例 3 的结构相比，增加了内容信息文件处理部 36。内容信息文件处理部 36 生成表示记录部 37 中记录的内容的记录位置关系的控制文件，并记录在记录部 37 中。

即，内容信息文件处理部 36 以内容 DB31 输出的内容和记录部 37 输出的内容的记录位置信息为基础，生成该内容所具有的时间轴信息和表示记录部 37 中记录的内容的地址关系的信息，并转化为数据后作为控制文件输出。

例如，如图 11 所示，针对表示内容记录位置的记录介质地址，使表示内容的时间轴基准的 $TM_ENT\#j$ 指向等时间轴间隔。例如，将 $TM_ENT\#j$ 每隔 1 秒（NTSC 信号的情况下为 30 帧）指向记录介质地址。通过这样的映射，即使内容以 1 秒为单位进行分散记录，也可以根据 $TM_ENT\#j$ 唯一地求出记录部 37 的记录地址。

而且，如图 9 所示，元数据文件中以 TEXT 的形式记录有作为文件开始之后的每 1 秒的基准时间的 TM_ENT （秒）、表示从基准时间开始的偏移帧数的 TM_OFFSET 、时间代码和元数据。因此，只要在元数据文件中指定元数据 1，就可以知道时间代码、基准时间和帧偏移值，从而可以立刻从图 11 所示的控制文件得知记录部 37 中的记录位置。

$TM_ENT\#j$ 的等时间轴间隔不限于上述每隔 1 秒进行指向，也可以依照 MPEG2 压缩等中使用的 GOP 单位等进行描述。

电视图像信号的 NTSC 中垂直同步信号为 60/1.001Hz，因此，可以使用 2 种类型，即，使用依照丢帧（drop-frame）模式的时间代码，以便与绝对时间一致；以及，使用依照垂直同步信号（60/1.001Hz）的非遗漏时间代码（nondrop-timecode）。这种情况下，例如以 $TM_ENT\#j$ 表示非遗漏时间代码，以 $TC_ENT\#j$ 表示丢帧对应时间代码。

而且，控制文件的数据化也可以利用 SMIL2 等现有语言进行数据

化，如果利用 SMIL2 功能，则可以与相关内容和元数据文件的文件名对应地进行数据化后，存储在控制文件中。

图 11 所示为直接表示记录部的记录地址的结构，但是也可以表示从内容文件的开头到时间代码的数据容量，来代替记录地址，根据数据容量和文件系统的记录地址计算并检测出记录部中时间代码的记录地址。

另外，如上所述，也可以不采取将 TM_ENT#j 和时间代码的对应表存储在元数据文件中的形式，而是将 TM_ENT#j 和时间代码的对应表存储在控制文件中，也可以取得同样的效果。

（实施例 5）

图 12 所示为本发明实施例 5 的元数据检索装置的结构框图。本实施例的装置与实施例 4 的结构相比，附加了检索部 38。检索部 38 从进行声音识别而检测出元数据时使用的同一词典 DB34 中，选择想要检索的场景的关键词并进行设定。

然后，检索部 38 检索元数据文件的元数据项目，并显示与关键词一致的标题名称和内容场景的位置（时间代码）一览。一览显示中设定有一个特定场景时，根据元数据文件的基准时间 TM_ENT（秒）和偏移帧数 TM_OFFSET，自动地检测出控制文件中的记录介质地址并设定在记录部 37 中，由记录部 37 再现该记录介质地址中记录的内容场景并显示在监视器 41 上。通过上述结构可以在检测出元数据后立即检测出想要看到的场景。

若备有与内容链接的缩略图（thumbnail）文件，则可以在显示与前述关键词一致的内容名一览时，再现并显示内容的代表性缩略图。

（实施例 6）

前述实施例 3~5 描述的是向预先记录的内容附加元数据的装置，本实施例涉及将本发明扩展到照相机等在拍摄时附加元数据的系统，特别涉及将本发明扩展到在内容被预先限定的风景拍摄时附加拍摄位置作为元数据的装置的例子。图 13 所示为本发明实施例 6 的元数据生成装置的结构框图。

照相机 51 的摄像输出作为图像内容记录在内容 DB54 中。同时，

GPS52 检测出照相机所拍摄的场所，其位置信息（经纬度数值）被声音合成部 53 转化为声音信号，然后作为位置信息记录在内容 DB54 的声音频道中。照相机 51、GPS52、声音合成部 53 和内容 DB54 可以一体构成为带有记录部的照相机 50。内容 DB54 将声音频道中记录的声音信号的位置信息，输入到声音识别部 56。词典 DB55 将词典数据供给声音识别部 56。词典 DB55 可以通过从端子 59 的键盘输入等，选择并限制地域名和地标等，然后输出到声音识别部 56。

声音识别部 56 利用所识别的经纬数据和词典 DB55 的数据，检测出地域名和地标并输出到文件处理部 57。文件处理部 57 将从内容 DB54 输出的时间代码和声音识别部 56 输出的地域名、地标文本（TEXT）化为元数据，然后生成元数据文件。元数据文件被供给记录部 58，记录部 58 记录该元数据文件和从内容 DB54 输出的内容数据。

通过上述结构可以自动地向所拍摄的每个场景附加地域名、地标等元数据。

上述实施例描述的是将声音识别部识别的关键词与时间代码一起文件化为元数据文件的结构，也可以在声音识别部所识别的关键词的基础上追加相关关键词后进行文件化。例如，通过声音识别为“淀川”（注：日本的河流名称）的情况下，另外附加地形、河流等一般属性的关键词后进行文件化。这样在检索时，也可以使用附加的地形、河流等关键词，从而可以提高检索性能。

本发明的声音识别部采用以单词为单位进行声音识别的单词识别方式，并通过限制声音输入的单词数和所使用的识别词典的单词数，可以改善声音识别率。

一般情况下，声音识别中可能产生任何误识别。上述各实施例中具有包含键盘的计算机等信息处理部，在产生误识别的情况下，可以通过键盘操作来修正所生成的元数据或标记。

产业上的可利用性

根据本发明的元数据生成装置，为了进行与内容相关的元数据的生成或标记的附加，利用声音识别，通过声音输入来生成元数据，并建立

元数据和内容的规定部分的关联，因而与现有的键盘输入相比，可以有效生成元数据或实施附加标记。

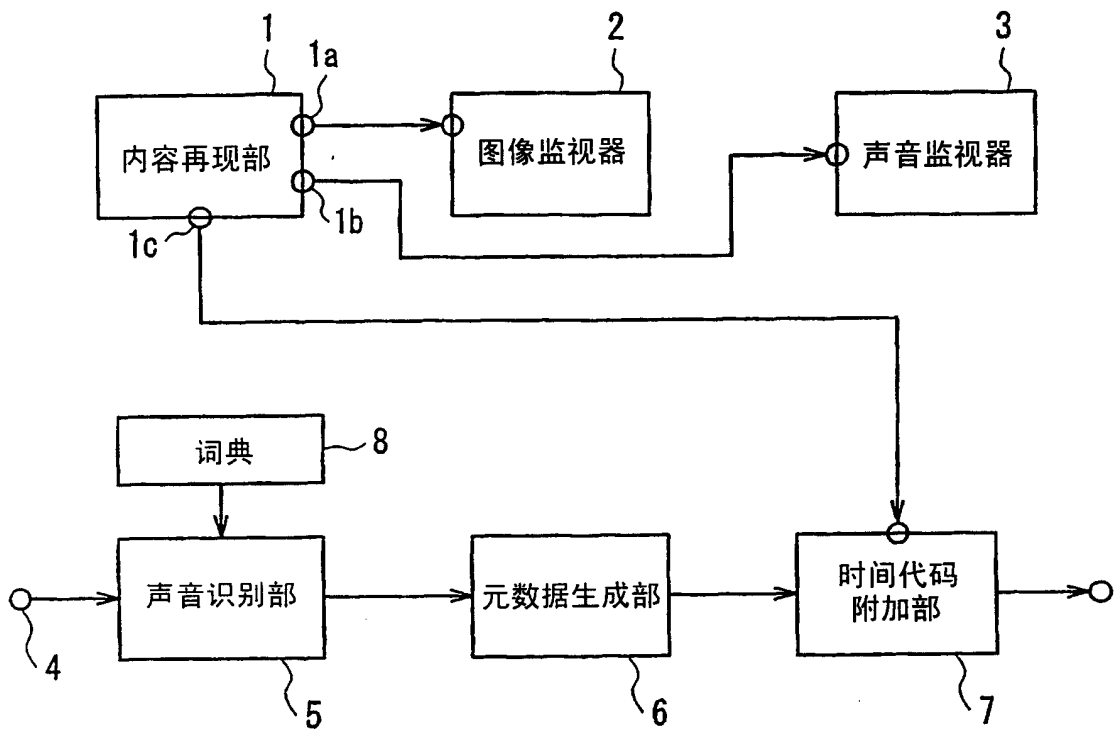


图1

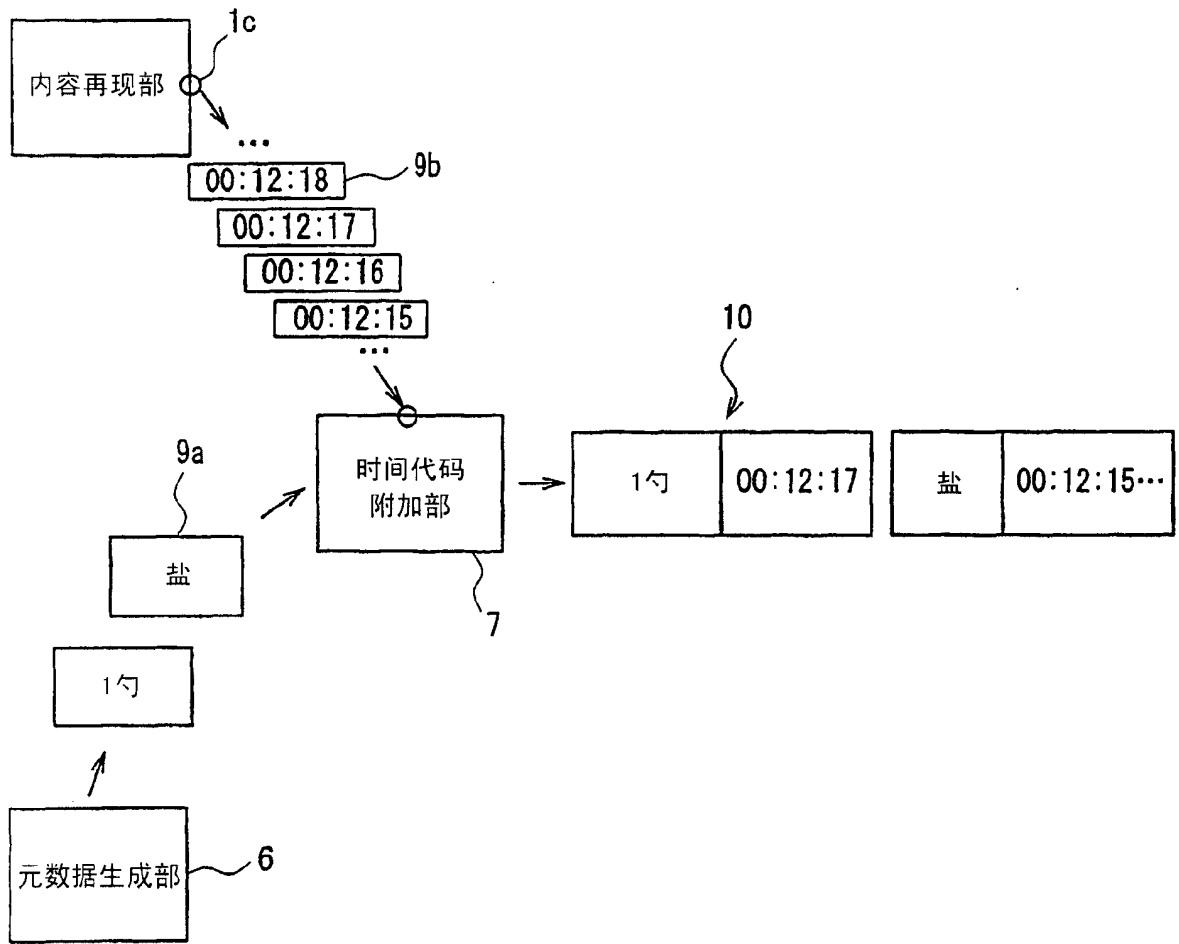


图2

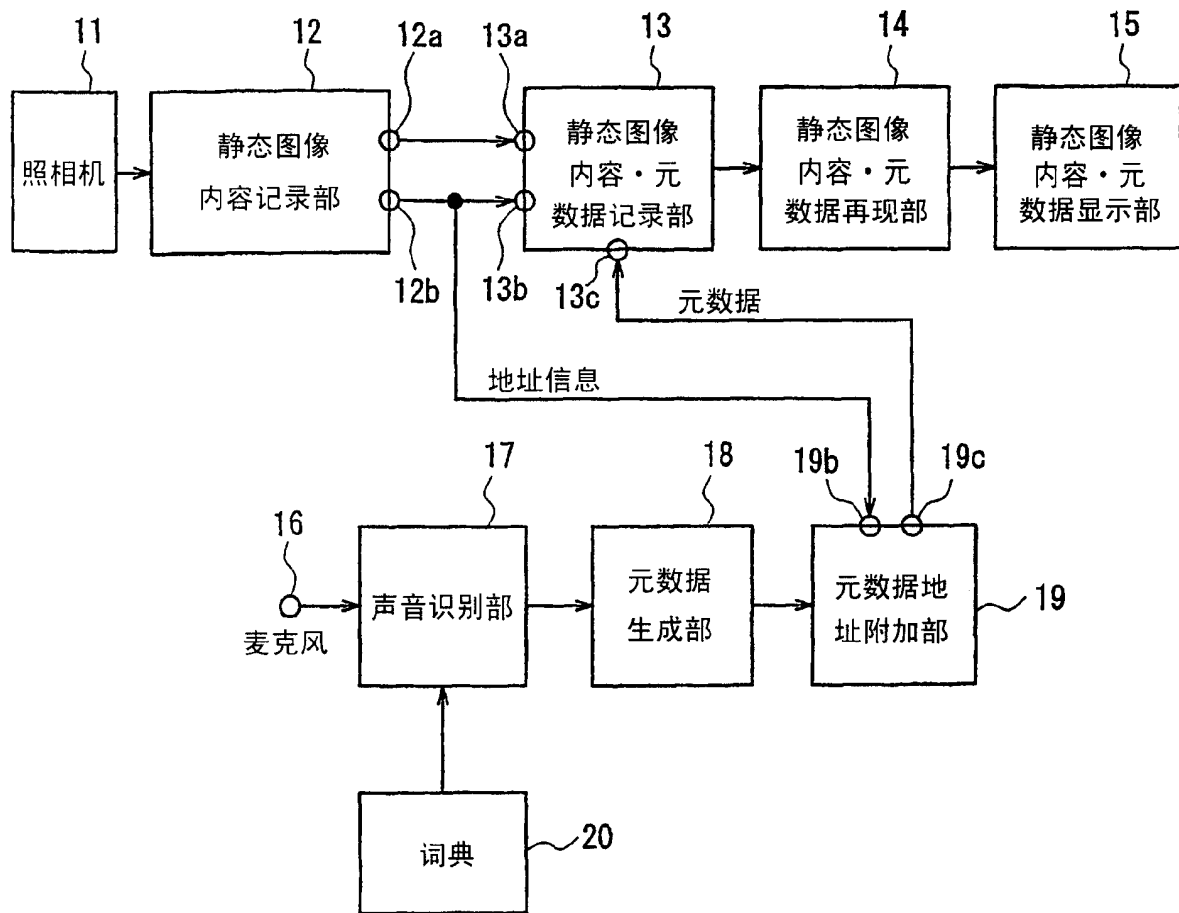


图3

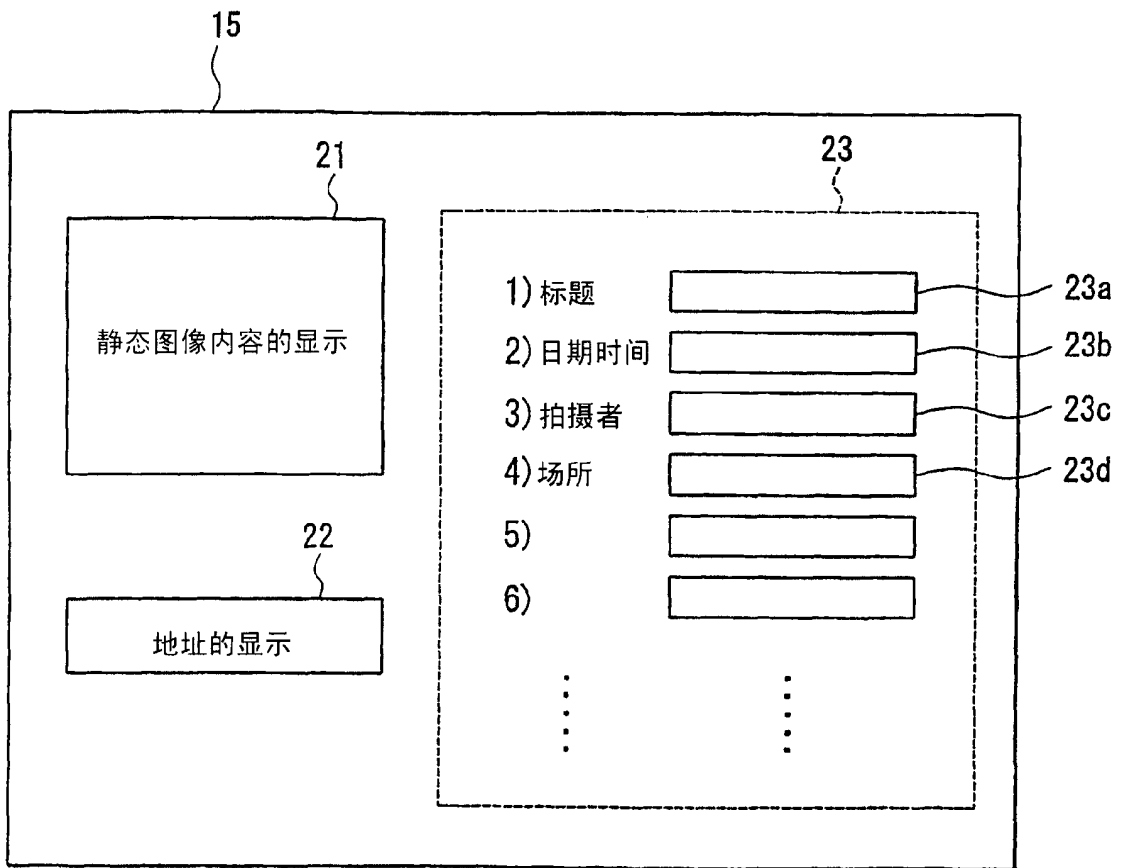


图4

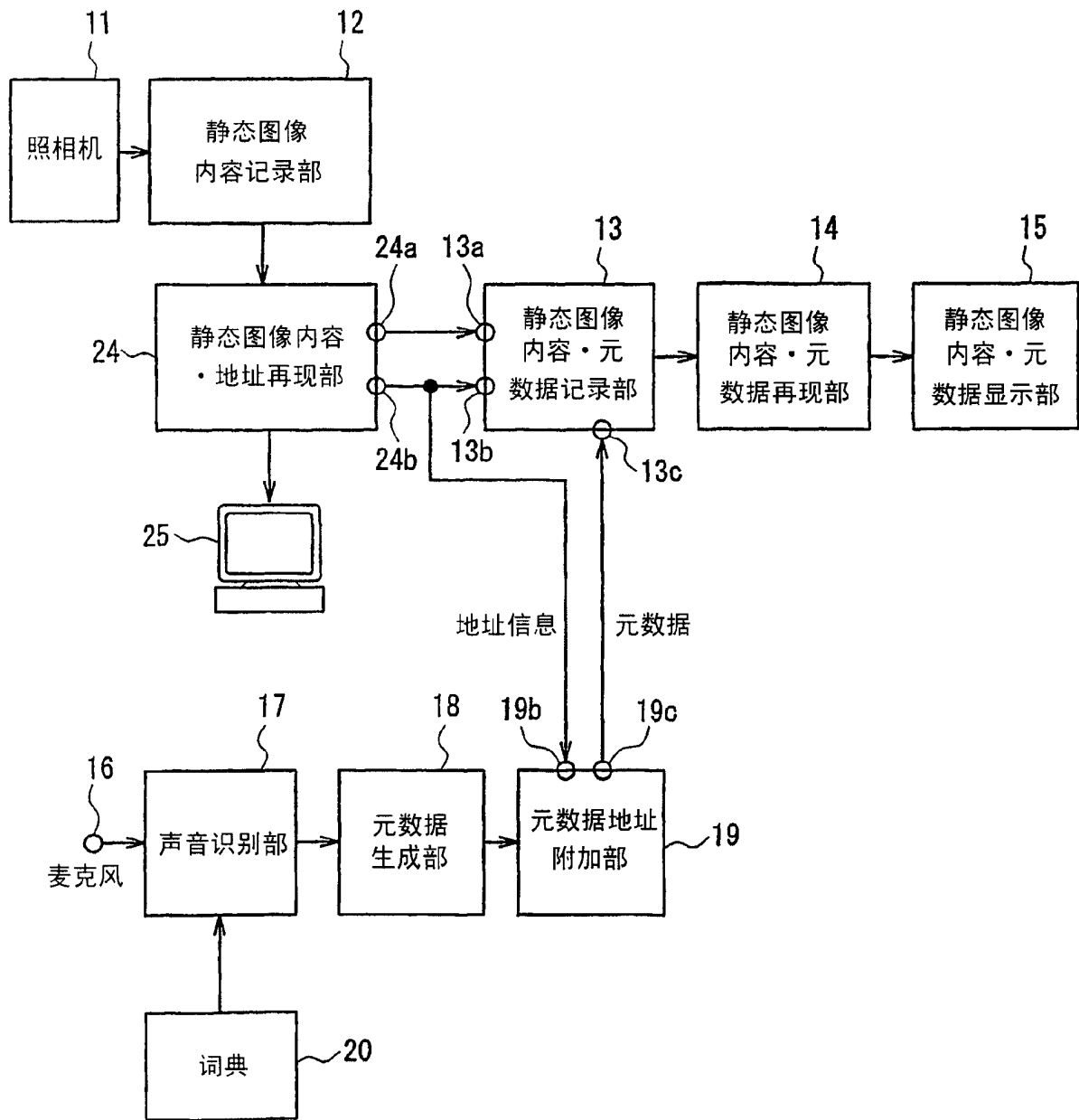


图5

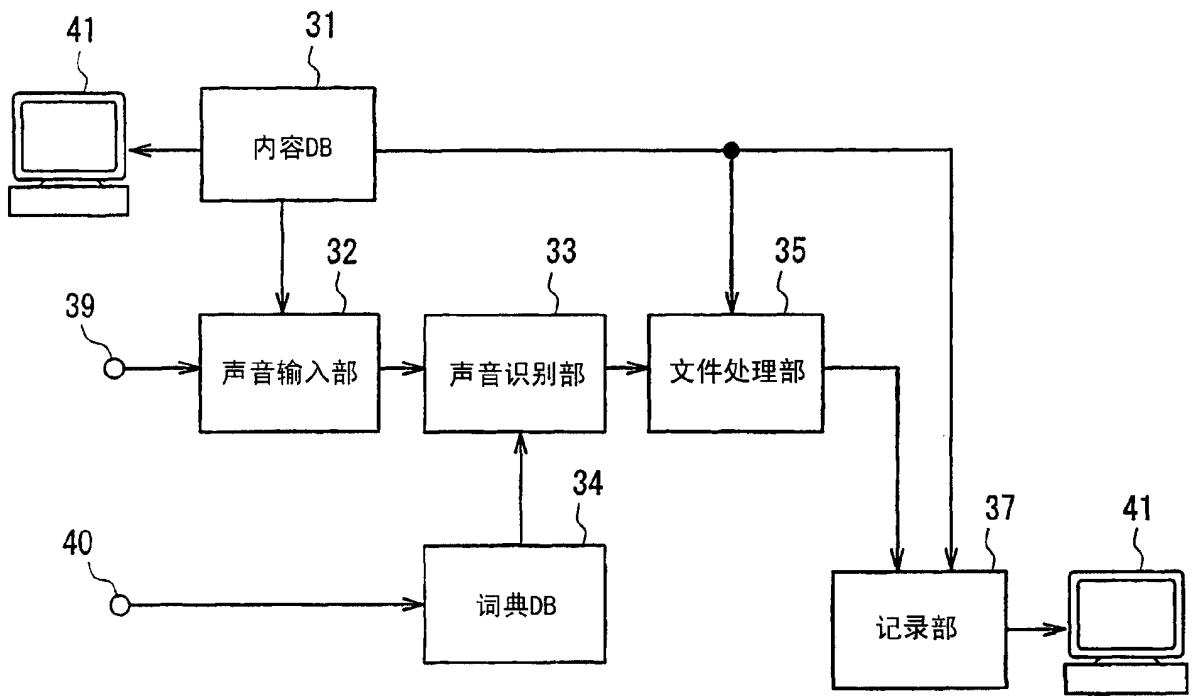


图6

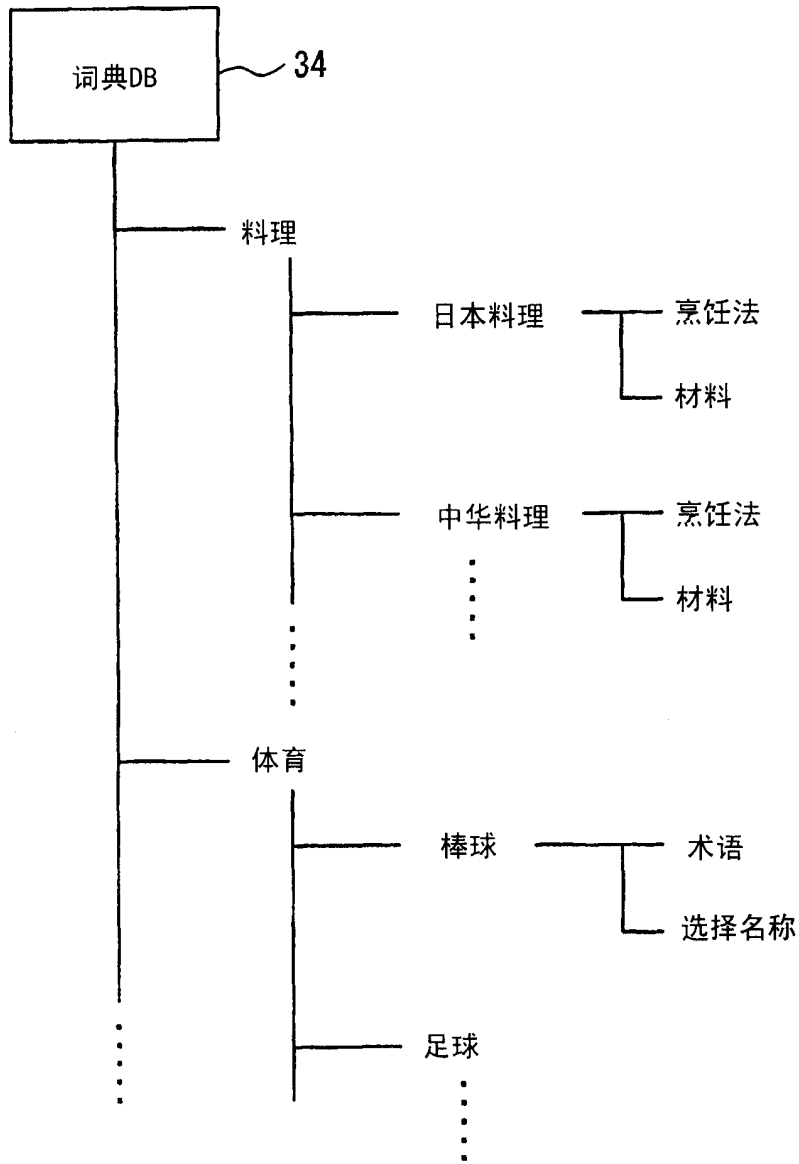


图7

食谱	
<菜单> 西红柿煮蛤子 白葡萄酒口味	
<材料> 4人份	
材料	分量
蛤子	50个
牡蛎	10个
洋葱	1个
西红柿（热水剥皮并切成小块）	1个
大蒜（切碎）	2瓣
白葡萄酒	1杯
时间	少量
香菜（切碎）	1/2杯
橄榄油	1/4杯

图8

标题：西红柿煮蛤子

TM_ENT#j	TM_OFFSET*1	TC	元数据
.	.	.	.
930	+8	00 15 30 08	洋葱 1个
990	+5	00 16 30 05	大蒜 2瓣
.	.	.	.
.	.	.	.
1230	+10	00 20 30 10	炒蛤子、牡蛎
.	.	.	.

* 1: 以偏移帧数表示

图9

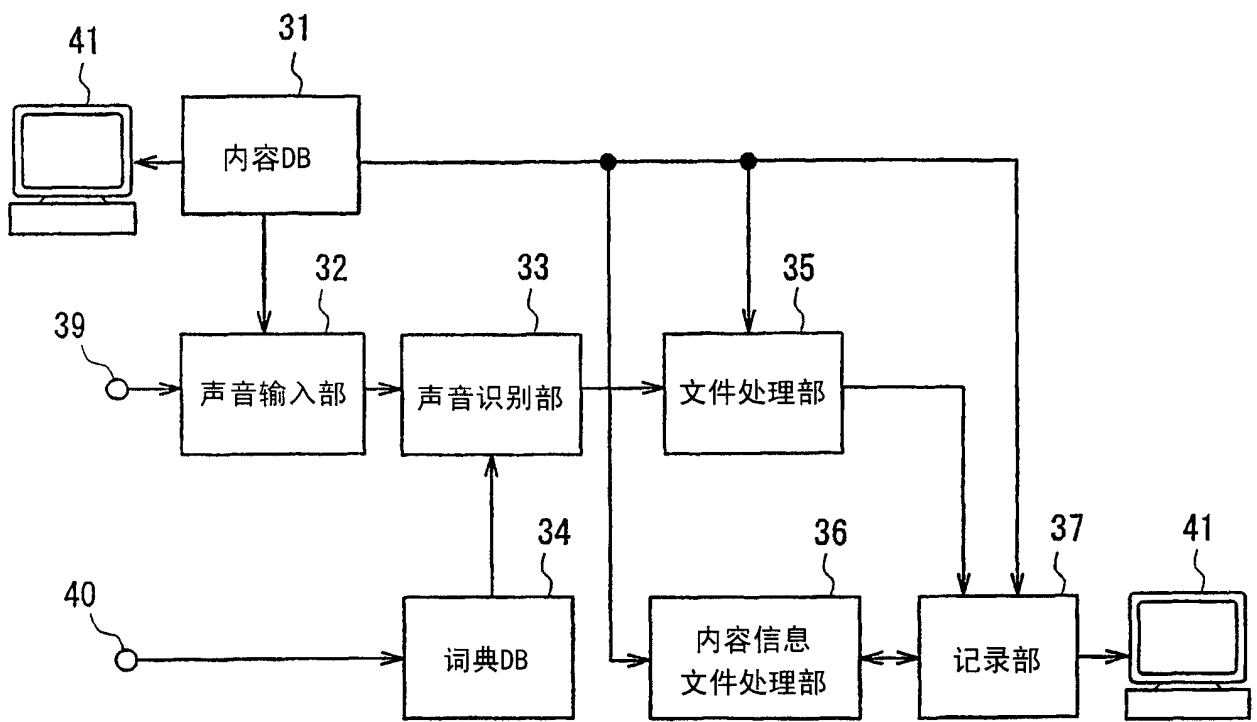


图10

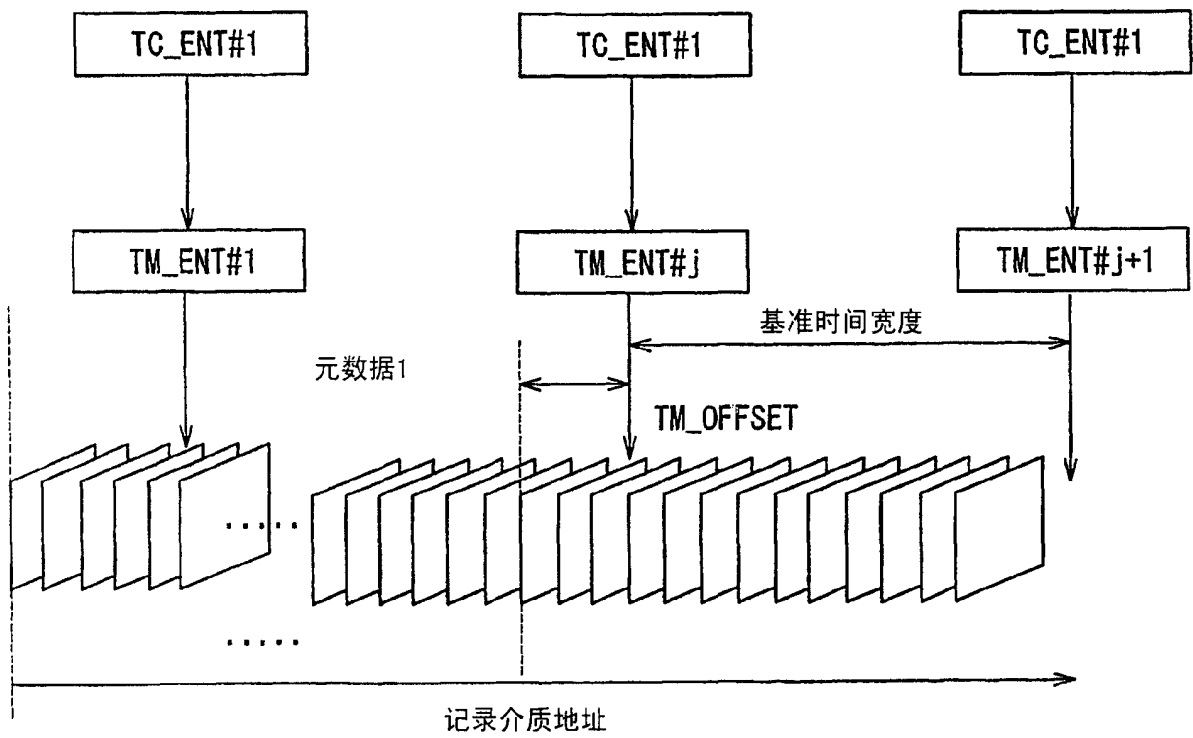


图 11

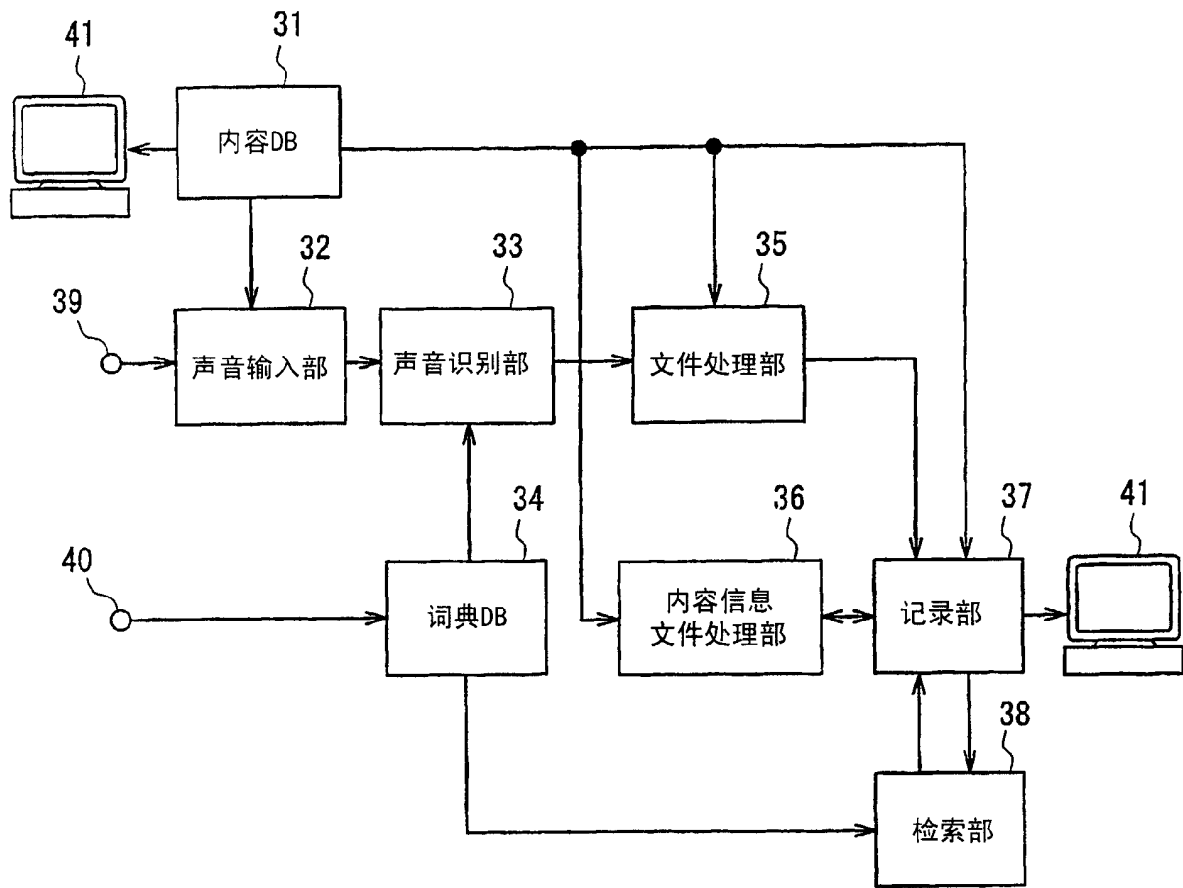


图12

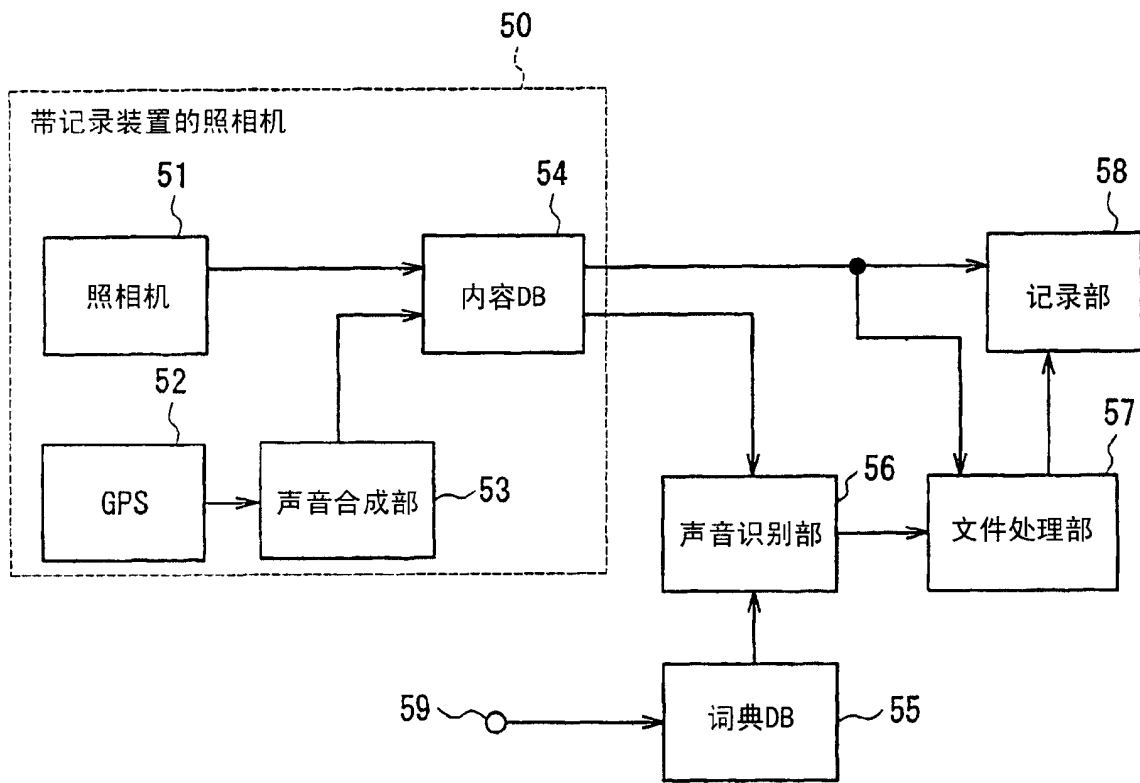


图13

1、(修改后)一种元数据生成装置,其特征在于,具有:再现内容并输出的内容再现部;声音输入部;识别从所述声音输入部输入的声音信号的声音识别部;将所述声音识别部识别的信息转换成元数据的元数据生成部;识别信息附加部,从所述内容获取用于识别所述内容中的各部分的识别信息,并附加到所述元数据上;以及根据所述内容限定的词典,

该元数据生成装置使生成的所述元数据与所述内容内的各部分相关联;所述声音识别部在识别由所述声音输入部输入的声音信号时,与所述词典相关联地进行识别。

2、(删除)

3、(修改后)如权利要求1所述的元数据生成装置,其特征在于,所述声音识别部以单词为单位,与所述词典相关联地识别所述声音信号。

4、如权利要求1或3所述的元数据生成装置,其特征在于,还具有包含键盘的信息处理部,可以通过从所述键盘输入,由所述信息处理部修正元数据。

5、(修改后)如权利要求1以及2~5中任一项所述的元数据生成装置,其特征在于,使用附在所述内容上的时间代码信息作为所述识别信息。

6、(修改后)如权利要求1以及2~6中任一项所述的元数据生成装置,其特征在于,使用附在所述内容上的内容的地址、编号或帧号码,作为所述识别信息。

7、如权利要求1所述的元数据生成装置,其特征在于,所述内容为静态图像内容,使用所述静态图像内容的各个地址,作为所述识别信息。

8、如权利要求1所述的元数据生成装置,其特征在于,

所述内容再现部由内容数据库构成;

所述声音输入部按照与所述内容数据库供给的同步信号同步的时钟,将所输入的关键词的声音信号转化为数据并供给所述声音识别部;

所述声音识别部根据在所述声音输入部转化为数据的声音信号数

据, 识别所述关键词;

所述元数据生成部构成文件处理部, 该文件处理部将表示所述内容中包含的图像信号的时间位置的时间代码用作所述识别信息, 并将所述声音识别部输出的关键词与所述时间代码相结合, 来生成元数据文件。

9、如权利要求 8 所述的元数据生成装置, 其特征在于, 还具有记录部, 将所述内容数据库供给的内容与所述元数据文件一道作为内容文件进行记录。

10、如权利要求 9 所述的元数据生成装置, 其特征在于, 还具有生成控制文件的内容信息文件处理部, 所述控制文件用于管理应记录所述内容文件的记录位置和所述元数据文件之间的关系,

在所述记录部中, 与所述内容文件和所述元数据文件一道记录所述控制文件。

11、如权利要求 8 所述的元数据生成装置, 其特征在于, 还具有词典数据库, 所述声音识别部可以从多种类型词典中选择适合于所述内容的种类词典。

12、如权利要求 11 所述的元数据生成装置, 其特征在于, 可以将与内容相关的关键词供给所述声音识别部, 所述声音识别部优先识别所述关键词。

13、(修改后) 一种元数据生成方法, 其特征在于, 在监视器上显示内容的同时对与内容相关的信息进行声音输入, 由声音识别装置利用根据所述内容限定的词典, 对所输入的声音信号进行声音识别, 并将所述声音识别后的信息转换成元数据, 将用于识别所述内容内各部分的、附在所述内容上的识别信息附加到所述元数据上, 使生成的所述元数据与所述内容内的各部分相关联。

14、(删除)

15、(删除)

16、如权利要求 13 所述的元数据生成方法, 其特征在于, 使用附在所述内容上的时间代码信息, 作为所述识别信息。

17、如权利要求 13 所述的元数据生成方法, 其特征在于, 使用静态

图像内容作为所述内容，并使用所述静态图像内容的各个地址作为所述识别信息。

18、一种元数据检索装置，其特征在于，包括：

再现内容并输出的内容数据库；

声音输入部，按照与再现的所述内容的同步信号同步的时钟，将所输入的关键词的声音信号转化为数据；

声音识别部，根据由所述声音输入部转化为数据的声音信号数据识别关键词；

文件处理部，将所述声音识别部输出的关键词与表示所述内容中包含的图像信号的时间位置的时间代码相结合，来生成元数据文件；

内容信息文件处理部，生成控制文件，所述控制文件用于对内容文件的记录位置和所述元数据文件之间的关系进行管理；

记录部，记录所述内容文件、所述元数据文件和所述控制文件；

检索部，确定包含所输入的检索关键词的所述元数据文件，并参照所述控制文件提取与所述内容文件的所述关键词对应的记录位置；

所述内容文件的记录位置是所述记录部中的记录位置。

19、如权利要求 18 所述的元数据检索装置，其特征在于，所述内容信息文件处理部输出的控制文件是记载与所述内容的记录时间对应的、所述记录部中所述内容的记录位置的表，可以根据所述时间代码检索所述内容的记录位置。

20、如权利要求 18 所述的元数据检索装置，其特征在于，还具有词典数据库和向所述声音识别部供给与内容相关的关键词的关键词供给部，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类的词典，并且优先识别所述关键词。

21、如权利要求 18 所述的元数据检索装置，其特征在于，还具有词典数据库，所述声音识别部可以从多种类型的词典中选择适合于所述内容的种类的词典，所述检索部利用从所述声音识别部使用的通用词典中选定的关键词进行检索。