



(12) 发明专利

(10) 授权公告号 CN 101789082 B

(45) 授权公告日 2013. 11. 20

(21) 申请号 201010105363. 3

G06T 7/00 (2006. 01)

(22) 申请日 2010. 01. 26

(56) 对比文件

(30) 优先权数据

CN 1705376 A, 2005. 12. 07, 全文.

0901262. 6 2009. 01. 26 GB

审查员 唐楹琰

(73) 专利权人 三菱电机株式会社

地址 日本东京都

(72) 发明人 保罗·布拉斯尼特

斯塔夫罗斯·帕斯卡勒基斯

米罗斯瓦夫·博贝尔

(74) 专利代理机构 北京三友知识产权代理有限公司

11127

代理人 李辉

(51) Int. Cl.

G06K 9/64 (2006. 01)

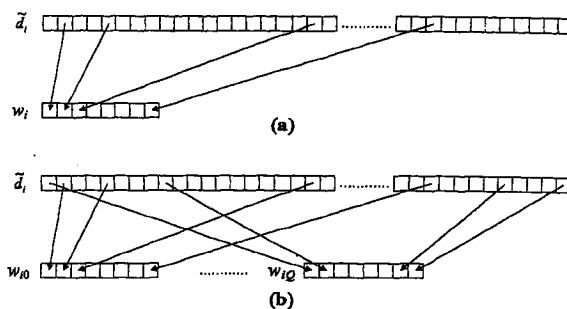
权利要求书4页 说明书10页 附图3页

(54) 发明名称

视频识别

(57) 摘要

本文公开了一种对第一图像序列和第二图像序列进行处理以比较第一序列和第二序列的方法和装置。通过以下操作来处理第一序列中的多个图像中的每一个以及第二序列中的多个图像中的每一个:(i) 针对该图像中的多个像素邻居中的每一个来处理图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及(ii) 根据该图像的所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合。第二序列的字是根据与第一序列的字相同的相应描述符元素比特组合而生成的。进行通过将针对第一序列中的所述多个图像生成的字与针对第二序列中的所述多个图像生成的字进行比较来比较第一序列和第二序列的处理。



1. 一种对图像的第一序列和图像的第二序列进行处理以比较第一序列和第二序列的方法,所述方法包括以下步骤:

(a) 通过以下操作对第一序列中的多个图像中的每一个以及第二序列中的多个图像中的每一个进行处理:

对该图像中的多个像素邻居中的每一个的图像数据进行处理,以生成每个所述像素邻居的至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及

根据该图像的所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合;

其中,第二序列的字是根据与第一序列的字相同的相应描述符元素比特组合来生成的;以及

(b) 进行通过将针对第一序列中的所述多个图像生成的字与针对第二序列中的所述多个图像生成的字进行比较来比较第一序列和第二序列的处理。

2. 根据权利要求1所述的方法,其中所述比较第一序列和第二序列的处理包括对第一序列和第二序列中不同字值的出现频率进行比较的处理。

3. 根据权利要求1所述的方法,其中所述比较第一序列和第二序列的处理包括:
确定第一序列中每个字的每个可能字值的出现次数,以生成第一序列的出现数据;
确定第二序列中每个字的每个可能字值的出现次数,以生成第二序列的出现数据;以及

将第一序列的所述出现数据与第二序列的所述出现数据进行比较。

4. 根据权利要求3所述的方法,其中对第一序列的所述出现数据与第二序列的所述出现数据进行比较,以生成针对每个字的相应比较结果,并且合并这些相应比较结果以生成整体比较结果。

5. 根据权利要求3所述的方法,其中所述比较第一序列和第二序列的处理包括使用Hough变换来估计时间偏移和帧速率的处理。

6. 根据前述任一项权利要求所述的方法,该方法还包括以下步骤:
识别匹配的图像对,每个图像对都包括来自第一序列的图像和来自第二序列的图像;
以及

执行进一步的对应性测试以确定至少一个所识别的图像对之间的相似度。

7. 一种对限定了图像的图像数据进行处理以生成至少一个表示该图像的值得方法,所述方法包括以下步骤:

针对所述图像中的多个像素邻居的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及

根据所述描述符元素来形成至少一个字,使得每个字都包括来自一个或更多个描述符元素的一个或更多个比特,并且使得所有字中的总比特数小于所有描述符元素中的总比特数。

8. 一种对限定了图像的图像数据进行处理以生成多个表示该图像的值得方法,所述方法包括以下步骤:

针对所述图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以

及

根据所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合。

9. 根据权利要求 7 或 8 所述的方法,其中:

所述方法还包括量化每个描述符元素的步骤;并且

根据量化后的描述符元素来形成每个字。

10. 根据权利要求 9 所述的方法,其中量化每个描述符元素的处理包括:量化每个描述符元素以形成相应的一比特二元结果。

11. 根据权利要求 7 或 8 所述的方法,其中:

以多种不同的像素分辨率针对所述图像来生成描述符元素;并且

通过将来自在不同像素分辨率下生成的描述符元素的一个或更多个比特进行组合而生成至少一个字。

12. 一种对限定了图像的图像数据进行处理以生成至少一个表示该图像的值得方法,所述方法包括以下步骤:

针对所述图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

量化每个描述符元素;以及

根据量化后的描述符元素来形成至少一个表示所述图像的值。

13. 根据权利要求 12 所述的方法,其中量化每个描述符元素的处理包括量化每个描述符元素以形成相应的一比特二元结果。

14. 一种对图像序列的表示进行存储的方法,所述方法包括以下步骤:

通过执行权利要求 7 或 8 中阐述的方法,针对所述序列中的每个图像来生成多个描述符元素以及至少一个表示所述序列中的每个图像的字,

生成在所述图像中能看到的字的直方图数据;以及

针对所述序列中的所述图像来存储所述直方图数据、字和描述符元素。

15. 根据权利要求 14 所述的方法,其中:

所述方法还包括量化每个描述符元素的步骤,并且每个字都是从量化后的描述符元素中得到的;

所述方法还包括量化所述直方图数据的步骤;并且

量化后的直方图数据、字和量化后的描述符元素被存储起来。

16. 一种对图像序列的表示进行存储的方法,所述方法包括以下步骤:

通过执行权利要求 12 中阐述的方法,针对所述序列中的每个图像来生成多个描述符元素以及至少一个表示所述序列中的每个图像的值,

生成直方图数据,所述直方图数据限定了哪些表示所述图像的值出现了以及出现频率;

量化所述直方图数据;以及

针对所述序列中的所述图像,对量化后的直方图数据、表示所述图像的值以及量化后的直方图数据、字和量化后的描述符元素进行存储。

17. 根据权利要求 14-16 中任一项所述的方法,其中,所述图像序列包括视频序列的片段。

18. 根据权利要求 14 所述的方法,其中,所述直方图数据、表示所述图像的值和描述符元素被存储在比特流中。

19. 一种对图像的第一序列和图像的第二序列进行处理以比较第一序列和第二序列的设备,所述设备包括:

描述符元素生成装置,其被设置为通过对第一序列中的多个图像中的每一个以及第二序列中的多个图像中的每一个中的多个像素邻居中的每一个的图像数据进行处理来对该图像进行处理,以生成每个所述像素邻居的至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

字生成装置,其被设置为根据每个图像的所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合,其中所述字生成装置被设置为根据与第一序列的字相同的相应描述符元素比特组合来生成第二序列的字;以及

序列比较装置,其被设置为进行通过将针对第一序列中的所述多个图像生成的字与针对第二序列中的所述多个图像生成的字进行比较来比较第一序列和第二序列的处理。

20. 一种对限定了图像的图像数据进行处理以生成至少一个表示该图像的值的设备,所述设备包括:

描述符元素生成装置,其被设置为针对所述图像中的多个像素邻居的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及

字生成装置,其被设置为根据所述描述符元素来形成至少一个字,使得每个字都包括来自一个或更多个描述符元素的一个或更多个比特,并且使得所有字中的总比特数小于所有描述符元素中的总比特数。

21. 一种对限定了图像的图像数据进行处理以生成多个表示该图像的值的设备,所述设备包括:

描述符元素生成装置,其被设置为针对所述图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及

字生成装置,其被设置为根据所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合。

22. 一种对限定了图像的图像数据进行处理以生成至少一个表示该图像的值的设备,所述设备包括:

描述符元素生成装置,其被设置为针对所述图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

量化装置,其被设置为量化每个描述符元素;以及

表示值形成装置,其被设置为根据量化后的描述符元素来形成至少一个表示所述图像的值。

23. 一种对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:

描述符元素生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多

个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

字生成装置,其被设置为,对于所述序列中的每一个图像,根据该图像的所述描述符元素来形成至少一个字,使得每个字都包括来自一个或更多个描述符元素的一个或更多个比特,并且使得该图像的所有字中的总比特数小于该图像的所有描述符元素中的总比特数;

直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些字出现在所述图像序列中以及出现次数;以及

数据写入装置,其被设置为针对所述序列中的图像将所述直方图数据、字和描述符元素写入存储器。

24. 一种对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:

描述符元素生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

字生成装置,其被设置为,对于所述序列中的每一个图像,根据该图像的所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合;

直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些字出现在所述图像序列中以及出现次数;以及

数据写入装置,其被设置为针对所述序列中的图像将所述直方图数据、字和描述符元素写入存储器。

25. 一种对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:

描述符元素生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;

描述符元素量化装置,其被设置为量化每个描述符元素;

表示值形成装置,其被设置为,对于每一个图像,根据该图像的量化后的描述符元素来形成至少一个表示该图像的值;

直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些表示所述图像的值出现在所述图像序列中以及出现次数;

直方图数据量化装置,其被设置为量化所述直方图数据;以及

数据写入装置,其被设置为将量化后的直方图数据、表示所述图像的值和量化后的描述符元素写入存储器。

视频识别

技术领域

[0001] 本发明涉及一种进行快速且可靠的视频识别的方法、装置和计算机程序产品，其对于诸如摄像捕捉、模/数转换、再压缩等常用编辑操作是鲁棒的。

背景技术

[0002] 专业人士和消费者正在存储数量激增的视频。对于专业人士来说，创建、存储和编辑视频内容的成本和技术难度已经在下降。对于消费者来说，机顶盒、个人录像机、视频摄像机和计算机的选择的增加以及成本的降低已经造成视频内容的增加。在过去数年间，因特网上可获得的合法和非法内容均呈爆炸性增长。索引、搜索和监控该内容的能力已经成为越发重要的问题。MPEG-7 标准是基于内容的搜索和检索领域中的早期工作。该标准原始版本中缺失的部分是近似副本视频检测 (near-duplicate video detection)。

[0003] 近似副本视频检测可以被如下定义：给定查询视频序列，在数据库中找到所有副本。（近似）副本的注解和解释是变化的。然而，对于本发明来说，副本被视为已经通过在原本 (original) 上应用普通视频编辑/处理操作而创建的序列。这样的操作的例子包括颜色改变、压缩、转码 (transcoding)、格式改变、帧速率改变、模拟 VCR 再捕获 (recapture) 和摄像再捕获。本发明还解决了副本部分仅可形成查询序列的一部分的问题。

[0004] 在本领域的早期工作 (T.Hoad and J.Zobel.Video similarity detection for digital rights management.In Proceedings of Australasian Computer Science Conference,pages 237-245,Adelaide,Australia,2003.) 中，镜头切换和边界被用来形成视频序列的标志 (signature)。这提供了非常紧凑的视频呈现，但是其对短序列却表现不佳，并且对于所用的镜头检测算法非常敏感 (T.Hoad and J.Zobel., Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, pages 262-269, Berkeley, US, 2003.)。

[0005] 在 [J.Sivic,A.Zisserman,Efficient Visual Search for Objects in Videos, Proceedings of the IEEE, April 2008,96(4), pages 548-566.] 中给出了现有技术中典型的对视频中（近似）副本检测的特征点途径，并且可以被概括为：i) 检测关键帧，ii) 检测该帧中的关键点，iii) 从该点周围的区域中提取特征，iv) 使用特征来匹配序列，v) 对序列中的对象的空间相关性进行测试。所概述的途径存在着数种弱点。首先，关键帧的使用意味着该方法可能在短片 (short clip) 上表现不那么好。特征的提取 (iii) 是一种在计算上昂贵的方法，并且导致很大的存储需求。对于步骤 (iv)，使用了从聚类数据 (clustering data) 习得的视觉辞典。这可能由于无法归纳 (generalise) 而导致对特定数据集的过度拟合 (over-fitting)。相关方法，例如 [Ondřej Chum, James Philbin, Michael Isard and Andrew Zisserman, Scalable near identical image and shot detection, Proceedings of the 6th ACM international conference on Image and video retrieval, pages 549-556, Amsterdam, The Netherlands, 2007] 提供了以针对所使用哈希表的高存储器需求为代价的快速搜索。尽管这对于一些场景可能是适用的，但是其在存储器资源通常非常有

限的消费类电子环境中却是不适用的。

[0006] 本发明旨在解决这些现有技术的方法的至少一项或多项限制。

发明内容

[0007] 在所附权利要求书中阐述了本发明的某些方面。其他方面在下面的实施方式中描述,并且本领域技术人员将通过对本说明书的阅读而理解到这些方面。

[0008] 通过总结的方式,本发明提供了对第一图像序列和第二图像序列进行处理以比较所述第一和第二序列的设备,所述设备包括:描述符元素生成装置,其被设置为通过对第一序列中的多个图像中的每一个以及第二序列中的多个图像中的每一个中的多个像素邻居(neighbourhood)中的每一个的图像数据进行处理来对该图像进行处理,以生成每个所述像素邻居的至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;字生成装置,其被设置为根据每个图像的所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合,其中所述字生成装置被设置为根据与所述第一序列的字相同的相应描述符元素比特组合来生成所述第二序列的字;以及序列比较装置,其被设置为进行通过将针对所述第一序列中的所述多个图像生成的字与针对所述第二序列中的所述多个图像生成的字进行比较来比较所述第一序列和第二序列的处理。

[0009] 所述序列比较装置可以被设置为对所述第一序列和第二序列进行比较的处理,所述处理包括对所述第一序列和第二序列中的不同字值的出现频率进行比较的处理。

[0010] 所述序列比较装置可以被设置为进行对所述第一序列和第二序列进行比较的处理,所述处理包括:确定所述第一序列中每个字的每个可能字值的出现次数,以生成所述第一序列的出现数据;确定所述第二序列中每个字的每个可能字值的出现次数,以生成所述第二序列的出现数据;以及将所述第一序列的出现数据与所述第二序列的出现数据进行比较。

[0011] 所述序列比较装置可以被设置为对所述第一序列的出现数据与所述第二序列的出现数据进行比较,以生成针对每个字的相应比较结果,并且合并这些相应比较结果以生成整体比较结果。

[0012] 所述设备还可以进包括:用于针对每个字来确定在两个序列中均出现的字值的装置;用于针对每个字来确定字值在两个序列中出现的时间顺序的装置;以及用于对出现的时间顺序进行比较的装置。

[0013] 所述设备还可以进包括:用于识别匹配图像对的装置,每个图像对都包括来自所述第一序列的图像和来自所述第二序列的图像;以及用于执行进一步的对应性(correspondence)测试以确定至少一个所识别的图像对之间的相似度的装置。

[0014] 本发明还提供了用于对限定了图像的图像数据进行处理以生成至少一个表示该图像的值的设备,所述设备包括:描述符元素生成装置,其被设置为针对所述图像中的多个像素邻居的每一个来处理所述图像数据,以针对每一个所述像素邻居来生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及字生成装置,其被设置为根据所述描述符元素来形成至少一个字,使得每个字都包括来自一个或更多个描述符元素的一个或更多个比特,并且使得所有字中的总比特数小于所有描述符元素中的总比特数。

[0015] 本发明还提供了用于对限定了图像的图像数据进行处理以生成多个表示该图像

的值的设备,所述设备包括:描述符元素生成装置,其被设置为针对所述图像中多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;以及字生成装置,其被设置为根据所述描述符元素来形成多个字,使得每个字都包括唯一的描述符元素比特组合。

[0016] 所述设备还可以包括用于量化每个描述符元素的量化装置;并且所述字生成装置可以被设置为根据量化后的描述符元素来形成每个字。

[0017] 所述量化器可以被设置为量化每个描述符元素以形成相应的一比特二元结果。

[0018] 所述描述符元素生成装置可以被设置为以多种不同的像素分辨率针对所述图像来生成所述描述符元素;并且所述字生成装置可以被设置为通过将来自在不同的像素分辨率下生成的描述符元素的一个或更多个比特进行组合而生成至少一个字。

[0019] 本发明还提供了用于对限定了图像的图像数据进行处理以生成至少一个表示该图像的值的设备,所述设备包括:描述符元素生成装置,其被设置为针对所述图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;量化装置,其被设置为量化每个描述符元素;以及表示值形成装置,其被设置为根据量化后的描述符元素来设计(format)至少一个表示所述图像的值的形成。

[0020] 所述量化器可以被设置为量化每个描述符元素以形成相应的一比特二元结果。

[0021] 本发明还提供了用于对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:描述符元素生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;字生成装置,其被设置为,对于所述序列中的每一个图像,根据该图像的所述描述符元素来形成至少一个字,使得每个字都包括来自一个或更多个描述符元素的一个或更多个比特,并且使得该图像的所有字中的总比特数小于该图像的所有描述符元素中的总比特数;直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些字出现在所述图像序列中以及出现次数;以及数据写入装置,其被设置为针对所述序列中的图像将所述直方图数据、字和描述符元素写入存储器。

[0022] 所述设备还可以包括被设置为量化每个描述符元素的描述符元素量化装置;并且所述字生成装置可以被设置为根据量化后的描述符元素来形成每个字。另外,所述设备还可以包括被设置为对所述直方图数据进行量化的直方图数据量化装置,并且所述数据写入装置可以被设置为将量化后的直方图数据、字和量化后的描述符元素写入存储器。

[0023] 所述设备可以运转从而对包括视频序列片段的图像序列进行处理。

[0024] 所述数据写入装置可以被设置为将所述直方图数据、字和描述符元素存储在比特流中。

[0025] 本发明还提供了用于对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:描述符元素生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;字生成装置,其被设置为,对于所述序列中的每一个图像,根据该图像的所述描述符元

素来形成多个字,使得每个字都包括唯一的描述符元素比特组合;直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些字出现在所述图像序列中以及出现次数;以及数据写入装置,其被设置为针对所述序列中的图像将所述直方图数据、字和描述符元素写入存储器。

[0026] 所述设备还可以包括被设置为量化每个描述符元素的描述符元素量化装置;并且所述字生成装置可以被设置为根据量化后的描述符元素来形成每个字。另外,所述设备还可以包括被设置为量化所述直方图数据的直方图数据量化装置,并且所述数据写入装置可以被设置为将量化后的直方图数据、字和量化后的描述符元素写入存储器。

[0027] 所述设备可以运转从而对包括视频序列片段的图像序列进行处理。

[0028] 所述数据写入装置可以被设置为将所述直方图数据、字和描述符元素存储在比特流中。

[0029] 本发明还提供了用于对限定了图像序列的图像数据进行处理以生成并存储所述序列的表示的设备,所述设备包括:描述符生成装置,其被设置为,对于所述序列中的每一个图像,针对该图像中的多个像素邻居中的每一个来处理所述图像数据,以针对每一个所述像素邻居生成至少一个相应描述符元素,每个描述符元素都包括一个或更多个比特;描述符元素量化装置,其被设置为量化每个描述符元素;表示值形成装置,其被设置为,对于所述序列中的每一个图像,根据量化后的描述符元素来形成至少一个表示该图像的值;直方图数据生成装置,其被设置为生成直方图数据,所述直方图数据限定了哪些表示图像的值出现在所述图像序列中以及出现次数;直方图数据量化装置,其被设置为量化所述直方图数据;以及数据写入装置,其被设置为将量化后的直方图数据、表示所述图像的值和量化后的描述符元素写入存储器。

[0030] 所述设备可以运行从而对包括视频序列片段的图像序列进行处理。

[0031] 所述数据写入装置可以被设置为将所述直方图数据、表示所述图像的值和描述符元素存储在比特流中。

[0032] 本发明的实施方式提供了一种用于视频识别的新方法和设备,其

[0033] ●提供了非常紧凑的视频序列表示,

[0034] ●不需要依赖于训练数据,

[0035] ●提供了非常快速的搜索和匹配,

[0036] ●对于普通的编辑/处理操作来说是鲁棒的

[0037] ●并且提供了对匹配帧位置的精确定位。

附图说明

[0038] 现在将参照附图,仅通过示例性的方式来描述本发明的实施方式,在附图中:

[0039] 图1示出了一实施方式中用于生成描述符元素的处理;

[0040] 图2a和2b示出了一实施方式中用于根据描述符元素来生成字的处理;

[0041] 图3示出了一实施方式中的二分化;

[0042] 图4示出了一实施方式中的时间偏移和帧速率变化的确定;而

[0043] 图5示出了用于执行一实施方式的处理操作的处理设备的例子。

具体实施方式

[0044] 现在将描述本发明实施方式中的处理设备所执行的方法。该方法包括多个处理操作。如本说明书最后所说明的,这些处理操作可以由使用硬件、固件、根据计算机程序指令而运行的处理单元或其组合的处理设备来执行。

[0045] 考虑视频帧序列 $f_i(c, x, y)$, 其中, i 是帧索引, 例如 $i \in [0, T-1]$, 其中, T 是该序列中的总帧数; c 是 1 个颜色通道的索引, 例如 $c \in \{Y, U, V\}$, 其中, $1 = 3$; 而 x 和 y 是空间坐标, 例如 $x \in [0, M-1]$, $y \in [0, N-1]$, 其中, M 和 N 分别是水平帧和垂直帧的分辨率。

[0046] 在本发明的优选实施方式中,我们就每个帧的像素强度和差异来对其进行描述。这些描述符在我们的共同未决专利申请 EP 1640913 和 EP1640914 中有描述,本文通过引用并入这些专利申请的全部内容。在本发明的优选实施方式中,我们考虑 $m \times m$ 像素分辨率的空间重采样帧,其中, m 是 2 的幂。优选地, m 是较小的值,例如 $m = 16$ 或 $m = 32$,但是并不限于此。然后,将每个帧划分成多个不重叠的 2×2 像素邻居,并且在每个邻居中按照下式来计算描述符元素:

[0047]

$$\bar{d}_i^m(c, p, q) = (f_i(c, p, q) + f_i(c, p+1, q) + f_i(c, p+1, q+1) + f_i(c, p, q+1))/4 \quad (1)$$

[0048]

$$\bar{d}_i^m(c, p+1, q) = (f_i(c, p, q) - f_i(c, p+1, q))/2 \quad (2)$$

[0049]

$$\bar{d}_i^m(c, p+1, q+1) = (f_i(c, p+1, q) - f_i(c, p+1, q+1))/2 \quad (3)$$

[0050]

$$\bar{d}_i^m(c, p, q+1) = (f_i(c, p+1, q+1) - f_i(c, p, q+1))/2 \quad (4)$$

[0051] 针对 $m = 8$ 和 $c = Y$ 的情况且不失一般性地在图 1 中对此进行了例示。符合等式 (1) 的描述符元素是平均强度,并且对于整个帧,以 $(m/2) \times (m/2)$ 像素分辨率给出了该帧的重采样版本,根据等式 (1)-(4) 对所述重采样版本进行重新处理。该处理持续进行,直到计算出整个帧的平均强度为止。因此,帧 f_i 的完整描述符包含了该帧在不同像素分辨率 (即, $m \times m, m/2 \times m/2, \dots, 4 \times 4, 2 \times 2$) 下的描述符元素,于是由 $\mathbf{d}_i = \{\bar{d}_i^2, \bar{d}_i^4, \dots, \bar{d}_i^{m/2}, \bar{d}_i^m\}$ 给出。

在本发明的优选实施方式中,帧是用它们的亮度通道 Y 来表示的,并且描述符 $\mathbf{d}_i$ 是从该亮度信息中提取的。

[0052] 可选地,将描述符元素量化成期望的比特数,该比特数对于不同的元素、颜色通道等可以不同。在本发明的优选实施方式中,将描述符元素二元化如下:

[0053]

$$\tilde{d}_{i, \text{AVG}} = \begin{cases} 1 & \text{如果 } \bar{d}_{i, \text{AVG}} \geq r/2 \\ 0 & \text{其他情况} \end{cases} \quad (5)$$

[0054] 其中, r 是动态范围,例如,对于 8 比特像素值, $r = 256$

[0055]

$$\tilde{d}_{i, \text{DIFF}} = \begin{cases} 1 & \text{如果 } d_{i, \text{DIFF}} \geq 0 \\ 0 & \text{其他情况} \end{cases} \quad (6)$$

[0056] 其中, $d_{i, \text{AVG}}$ 和 $d_{i, \text{DIFF}}$ 分别表示符合等式 (1) 的平均值和符合等式 (2)-(4) 的差。(5) 和 (6) 的优点在于, 它们可以被实现, 从而它们变得等同于保留各个元素的整数部分的二元表示的 MSB (最高有效位)。

[0057] 在另选实施方式中, 还可以采用其他量化 / 二元化技术, 例如, 位置敏感哈希法。尽管本文未考查该方法, 但是在 Samet H., “Foundations of Multidimensional and Metric Data Structures”, Morgan Kaufmann, 2006 中有描述。

[0058] 帧的完整描述符随后被用来基于其描述符元素中的一个或多个描述符元素的一个或多个比特, 来形成至少一个紧凑字 (compact word)。图 2a 中对此进行了例示。应该注意的是, 本文所使用的术语“字”是指可以具有任意比特数而并不具体限于某些计算机文献中的特定比特数的字。更具体来讲, 在本发明的优选实施方式中, 对二元化的 $\tilde{d}_i$ 中的小的有序元素集进行串接从而产生字。对于要选择哪些元素是没有限制的, 并且一个字可以混合有来自不同颜色通道、来自不同粗度级别 (例如以 16×16 或 4×4 的分辨率所提取) 的元素, 它们是符合等式 (1) 的平均值或者是符合等式 (2)-(4) 的差。例如, 在本发明的一个实施方式中, 可以在所有可用元素中随机选择元素。在本发明的另一实施方式中, 可以根据一些几何图案 (例如, 仅覆盖视频帧的特定空间区域) 来选择元素。在本发明的又一实施方式中, 像素可能是按照它们的熵来排序的, 并且选择熵最高的像素来形成字。因此, 字形成的处理是从 Φ 维空间到 Ψ 维空间的投影, 其中优选的是 $\Psi \ll \Phi$ 。对于两个视频帧, 两个对应字之间的距离 (即, 顺序选择的对应元素的比特图案) 是全部帧描述符的距离的近似。

[0059] 构成字的有序元素的每个可能值的所有可能的组合给出了该字的词汇表 (vocabulary)。因此, 一个字是描述符元素的唯一组合, 而字的词汇表是该字可以采取的不同值的集合。例如, 如果一个字包含两个描述符元素, 并且每个描述符元素都为一比特长, 且每个比特都自由地取值 0 或 1, 则该字的词汇表为 00、01、10 和 11。

[0060] 在本发明的优选实施方式中, 从每个帧中提取多个字, 如图 2b 中所示, 每个字都是从二元化的 $\tilde{d}_i$ 中的小的有序元素集给出的。优选的是, 每个字的各个元素都是根据任何适当方法 (即, 随机投影、几何约束、熵排序等) 来选择的。每个字均有词汇表, 并且得到每个字的比特组合是唯一的, 但并不排除部分重叠的可能性, 例如, 一些元素在不止一个字中出现。此外, 每个字都可以与任何其他字有相同或不同的长度, 即, 可使维度从 Φ 减小到 Ψ_1 、 Ψ_2 等。

[0061] 在本发明的另选实施方式中, 可以根据任何版本的描述符 d_i 来形成各个字, 所述描述符 d_i 处于其原始分辨率和动态范围, 并被量化或二元化。在这些实施方式中, 一个元素可以为一个字贡献不止一比特, 向不同的字贡献不同数量的比特, 并且一个字可以包含来自不同元素的不同数量的比特。

[0062] 按照前述方式, 通过视频序列的多个帧中的一个或多个帧描述了该视频序列。然后, 我们计算不同字在一个或多个词汇表中的每一个中的出现频率。更具体来讲, 在本发明的优选实施方式中, 对序列中的多个帧进行处理, 并且从每个帧 f_i 中提取 Φ 维完整描述

符和多个字 W_{ik} , $k \in [0, Q-1]$, 每个字都对应于 Q 个词汇表中的一个。然后, 对于每个词汇表 k , 绘制在帧序列中找到的字的直方图 h_k 。简言之, 该直方图示出了哪些字出现在帧序列中以及出现的频率。更具体来讲, 对于每个字, 生成这样的直方图, 其在一个轴上具有该字可以取的每个值 (即, 该字的词汇表中的每个值) 的相应区间 (bin), 并在另一轴上记录了每个值的出现次数。

[0063] 对于两个帧序列 f_i^1 和 f_j^2 , 可以通过针对每个词汇表 k 比较直方图 h_k^1 和 h_k^2 并合并结果以达成判定来估计它们的相似性。该比较可以使用诸如直方图交集 (histogram intersection) 这样的适当距离测量来实现。在本发明的优选实施方式中, 直方图被二元化。然后, 这种二元化的直方图示出了大量存在 (significant presence) 的字。二元化可以通过多种适当方式中的任意一种来实现, 例如, 简单的门限化、自适应门限化等。对于二元化的直方图 $\tilde{h}_k^1$ 和 $\tilde{h}_k^2$, 优选地通过下式给出的 Jaccard 距离测量来测量它们的距离:

$$[0064] \quad D_j(\tilde{h}_k^1, \tilde{h}_k^2) = 1 - \frac{\tilde{h}_k^1 \cap \tilde{h}_k^2}{\tilde{h}_k^1 \cup \tilde{h}_k^2} \quad (7)$$

[0065] 简言之, 等式 (7) 根据给定词汇表中的帧序列 f_i^1 和 f_j^2 所共有的不同字以及它们联合包含的所有不同字来测量它们的距离。

[0066] 对于 Q 个词汇表, 我们得到了 Q 个 Jaccard 距离: $D_0, D_1, \dots, D_{Q-1}$ 。在本发明的优选实施方式中, 这些距离被融合以给出复合距离 D_j 为:

$$[0067] \quad D_j = \sum_{k=0}^{Q-1} D_k \quad (8)$$

[0068] 然后, 可以通过对 D_j 进行处理 (优选地, 进行门限化) 来达成关于帧序列是否相似的判定。在这种情况下, 如果 D_j 小于特定阈值, 则断定这些序列相匹配, 否则断定它们不匹配。

[0069] 显然, 等式 (8) 存在许多另选方案, 例如, 截尾均值 (trimmed mean)、中值、最大值等。此外, 另选实施方式可以不通过距离融合而通过判定融合 (即, 通过用合适的阈值单独地门限化每个 D_k , 然后例如通过多数表决来融合这些单独的决定) 来达成关于帧序列的相似度的判定。在本发明的另选实施方式中, 还可以将距离融合与判定融合进行组合, 即, 不仅要通过计算复合 D_j 随后进行门限化来达成判定, 而且还要求通过至少一组多个单独的 D_k 距离来支持相匹配的判定。

[0070] 可选地, 通过考虑视频序列中的匹配字对于每个词汇表出现的时间顺序来验证所检测到的匹配。在优选实施方式中, 对于每个词汇表 k , 找到了出现在视频序列 f_i^1 和 f_j^2 二者中的一组字。优选地, 通过识别在直方图 h_k^1 和 h_k^2 二者中非零的元素, 即二元直方图 $\tilde{h}_k^1$ 和 $\tilde{h}_k^2$ 的交集, 来找到这些字。上面提到, 每个直方图的每个元素都对应于词汇表中的一个字, 该交集生成了一组具有 z 个元素的字 $W_k^{12} = \{w_{\lambda k}, \lambda \in [0, z-1]\} = \{w_a, \dots, w_z\}$, 该 z 个元素对于两个帧是公共的, 并且优选地根据词汇表来进行排序, 例如, 数字或字母顺序。然后, 对于每个视频序列, 生成一组帧号 $\Omega_k = \{\omega_{\lambda k}, \lambda \in [0, z-1]\} = \{f_a, \dots, f_z\}$, 来记录有序集合 W_k^{12} 中每个字的首次出现。

[0071] 分别针对 f_i^1 和 f_j^2 比较这两组帧号 Ω_k^1 和 Ω_k^2 , 以确定它们之间的距离。在优选实施方式中, Ω_k^1 和 Ω_k^2 被二元化, 并且通过归一化的 Hamming 距离来给出它们之间的距离。在优选实施方式中, 如图 3 中所示, 通过评估该集合的相邻元素之间的差来执行二元化:

[0072]

$$\tilde{\omega}_\lambda = \begin{cases} 0 & \omega_\lambda < \omega_{\lambda+1} \\ 1 & \text{其他情况} \end{cases} \quad (9)$$

[0073] 二元化集合被表示为 $\tilde{\Omega}_k = \{\tilde{\omega}_{\lambda k}, \lambda \in [0, z-1]\}$, 而归一化的 Hamming 距离被表示为 $\bar{D}_H^\Omega$ 。简言之, 这种二元化的集合记录了有序集合 W_k^{12} 中的每个字在视频序列中的首次出现是在有序字集 W_k^{12} 中的下一字之前还是之后。

[0074] 在本发明的另选实施方式中, 可以在未二元化的情况下使用适当的距离测量来对 Ω_k^1 和 Ω_k^2 进行比较, 所述适当距离测量例如为 L1, 其将给出各个字在一个序列中的出现不同于其他序列的总帧数。

[0075] 对于 Q 个词汇表, 我们得到 Q 个归一化的 Hamming 距离 $\bar{D}_H^{\Omega_0}, \bar{D}_H^{\Omega_1}, \dots, \bar{D}_H^{\Omega_{Q-1}}$ 。在本发明的优选实施方式中, 这些距离被融合来给出复合距离 $\bar{D}_H$:

$$\bar{D}_H^\Omega = \sum_{k=0}^{Q-1} \bar{D}_H^{\Omega_k} \quad (10)$$

[0077] 然后, 可以通过对 $\bar{D}_H^\Omega$ 进行处理 (优选地进行门限化) 来达成关于帧序列是否相似的判定。在这种情况下, 如果 $\bar{D}_H^\Omega$ 小于特定阈值, 则断定这些序列相匹配, 否则就断定它们不匹配。

[0078] 显然, 等式 (10) 存在许多另选方案, 例如, 截尾均值、中值、最大值等。此外, 另选实施方式可以不通过距离融合而通过判定融合 (即, 通过用合适的阈值单独地门限化每个 $\bar{D}_H^{\Omega_k}$, 然后例如通过多数表决来融合这些单独的判定) 来达成关于帧序列的相似度的判定。在本发明的另选实施方式中, 还可以将距离融合与判定融合进行组合, 即, 不仅要通过计算复合 $\bar{D}_H^\Omega$ 随后进行门限化来达成判定, 而且还要求通过至少一组多个单独的 $\bar{D}_H^{\Omega_k}$ 距离来支持相匹配的判定。

[0079] 可选地, 应用进一步的改进和定位阶段以非常高的准确度来判定视频序列是否相匹配, 并确定两个序列中的对应帧位置。来自两个序列的匹配字的集合被用来确定可能的帧对应。然后可以估计序列之间的时间变化。在优选实施方式中, 序列之间的时间偏移和帧速率差被确定为如图 4 中所示。然后可以使用来自多个可能的对应帧的描述符来确定对应的有效性, 和 / 或改进时间变化参数。

[0080] 更具体来讲, 对于每个词汇表 k, 交集 $W_k^{12} = \{w_{\lambda k}, \lambda \in [0, z-1]\} = \{w_\alpha, \dots, w_\zeta\}$ 中多个字中的每一个都对应于每个视频帧中的一个或更多个帧, 其中“一个或更多个”对于每个序列可以是不同的数量, 例如, 一个字可以通过一个序列中的仅一个帧以及第二序列中的许多帧来产生。对于一对视频序列、对于每个词汇表 k 以及对于 W_k^{12} 中 z 个字中的每一个, 我们构建一集合, 其包含序列 f_i^1 和 f_j^2 中对应于该字的帧号对。因此, 如果 W_k^{12} 中 z 个字中的

一个由 f_i^1 和 f_j^1 中的单对帧生成,则该字的集合将仅包含一对帧号。如果另一个字由 f_i^1 和 f_j^2 中的多对帧生成,则该字的集合将包含数量相同的多对帧号。

[0081] 然后,对于 w_k^{12} 中的每个字以及对于每个词汇表 k ,将与该字相对应的帧号对绘制在单个 2 维直方图 G_{ij}^{12} 中,, $i \in [0, T_1-1]$, $j \in [0, T_2-1]$, 其中, T_1 和 T_2 分别是序列 f_i^1 和 f_j^2 中的帧数。简言之,在 G_{ij}^{12} 中,区间值 0 表示讨论中的帧对 (即 (i, j)) 在 Q 个词汇表的任何一个中都不产生匹配字;区间值 $k < Q$ 表示讨论中的帧对仅在 Q 个词汇表的 k 个中产生了匹配字;而区间值 Q 表示讨论中的帧对在所有 Q 个词汇表中均产生了匹配字。

[0082] 如果 f_i^1 和 f_j^2 中的帧在一个或多个词汇表中具有相同的字,则认为它们是相对应的,并且这可以视为是弱的对应性测试,即它们是弱对应的。对于弱对应的多个帧对中的每一对,通过根据更详细的描述符对它们进行比较来进行更强的对应性测试。优选地,使用之前描述的从 f_i^1 和 f_j^2 中提取的二元描述符 $\tilde{d}_i^1$ 和 $\tilde{d}_j^2$ 。这些二元描述符之间的距离优选地被计算为 Hamming 距离 $D_H^{\tilde{d}}$ 。

[0083] 在本发明的另选实施方式中,可以使用等式 (1)-(4) 中所描述的描述符 d 的任何版本,或者使用该描述符的适当子集,该描述符处于其原始分辨率和动态范围,并被量化或二元化,并且 Hamming 距离 $D_H^{\tilde{d}}$ 被诸如 L1 这样的合适的距离测量所取代。

[0084] 为了在存在大量弱对应的帧对时有利地缩短处理时间,可以仅将它们中有限数量的帧对确认为强对应的帧对。优选地,以具有最多公共字 (most common words) 的顺序来检查这些帧对。即,具有 Q 个公共字的帧对被首先检查,然后检查具有 $Q-1$ 个公共字的帧对,以此类推。这种类型的根据公共字的数量进行的排序容易通过 2D 直方图 G_{ij}^{12} 来提供。如果帧描述符之间的距离满足特定判据 (即,优选地,低于预定阈值),则弱对应的帧对被接受为强对应的帧对。重复该过程,直到没有剩余的弱对应的帧对,或者已经找到了特定数量的强对应的帧对。根据上面的讨论得出:具有少于 Q 个或者甚至仅有 1 个匹配字的帧对可能被确认为强对应的帧对,而在 Q 个词汇表中具有 Q 个匹配字的帧对也可能未被确认为强对应的帧对,这是因为根据它们的详细描述符,它们具有较低相似度。因此,一般来说,将大量弱对应的帧对减少为少量强对应的帧对。

[0085] 然后,使用强对应的帧对来估计帧序列之间的时间参数的变化。在优选实施方式中,时间参数是时间偏移和帧速率变化。这些是线性属性,因此可以使用两个强对应的帧对来进行估计。优选地,使用 Hough 变换来估计这些时间参数。Hough 空间中的峰对应于最可能的时间参数。尽管本文未考查 Hough 变换和 Radon 变换,但是在文献 van Ginkel, M., Hendriks, C.L., van Vliet, L.J., "A short introduction to the Radon and Hough transforms and how they relate to each other", Number QI-2004-01 in the Quantitative Imaging Group Technical Report Series, Delft University of Technology 中有专门描述。

[0086] Hough 变换通常用来给出参数的粗略估计。使用这些参数,在两个视频序列之间建立多个帧对应。在优选实施方式中,这些帧对应可以被更准确地确定。单个描述符提供了一定的准确度,而通过组合多个描述符就达到了更高的准确度。在优选实施方式中,使用描述符之间的 Hamming 距离,然后针对这些描述符将 Hamming 距离加在一起。优选地,选择来自两个序列的多个对应帧,并通过检查来自这些帧的描述符的组合之间的距离来确定序列

之间的距离。在优选实施方式中,从第一序列的开始位置选择一组(例如,7个)连续帧,并将这些帧与来自第二序列的对应帧进行比较。另选地,这些帧可以来自序列的不同部分(中间、结尾等),和/或这些帧可以是非连续的。在优选实施方式中,通过保持对应帧对中的一个帧固定,并在其对应帧的位置搜索更优的对应帧,来改进帧对应。

[0087] 在一实施方式中,有效的描述符存储方案被用来形成视频序列的紧凑表示。在优选实施方式中,将序列划分成如1秒的片段。对于每个片段,将Q个二元直方图 $\tilde{h}_k$ 存储在比特流中。依照(following)该二元直方图来存储片段中的每个帧的描述符 $\tilde{d}_i$ 。这样,就将序列中的所有片段连续存储在比特流中。可选地,可以将以下信息存储在比特流的头部,例如,序列的帧速率、帧数、片段数等等。

[0088] 在另选实施方式中,可以在比特流中用多个帧表示来表示序列。其中,每个帧用Q个字和二元描述符 $\tilde{d}_i$ 来表示。这样,可以存储序列中的所有帧。在另一另选实施方式中,可能希望首先存储所有帧中的所有字,之后存储所有描述符。显然,可以存储直方图、字和描述符的量化版本或者原始版本。

[0089] 在本发明的另选实施方式中,可以将从视频帧中提取的描述符变为合适的帧描述符,例如,颜色直方图或边缘强度直方图,它们可以以原始、量化或二元化形式被存储,并且可以根据它们来形成字。在本发明的不同实施方式中,可以省略不同的判定阶段。例如,基于视频序列中对于每个词汇表的匹配字出现的时间顺序的处理和判定阶段可以被整体省略,处理直接前进到对应帧的改进、定位和确定阶段。在本发明的另选实施方式中,根据每个视频序列中找到的匹配字的判定阶段可以被省略,处理可以直接前进到基于视频序列中匹配字出现的时间顺序的处理与判定阶段。

[0090] 图5中示出了用于执行上述处理操作的数据处理设备1。例如,设备1可以是个人台式计算机或便携式计算机。

[0091] 设备1包括技术人员熟知的数据处理设备的常规部件,从而不必进行详细描述。简言之,图5的设备1包括输入数据接口3,用于接收来自诸如存储介质5这样的计算机程序产品的计算机程序指令或信号7以及要处理的视频数据。例如,通过用总线15连接起来的CPU 9、随机存取存储器11和只读存储器13来提供处理系统。CPU 9控制整体操作。RAM 11是CPU 9用来执行程序和控制ROM 4的工作存储器,ROM 4存储有程序和其他数据。设备1的处理设备被配置为执行如本文上面所描述的那样对限定了图像的图像数据进行处理的方法。该处理的结果由输出接口17输出。

[0092] 尽管上面描述的处理设备1根据计算机程序指令来执行处理,但是另选处理设备可以以任何合适或期望的方式被实现为硬件、软件或硬件和软件的任意合适的组合。还应该注意的是,本发明还可以被实施为计算机程序,当所述计算机程序被加载到可编程处理设备中并在其上运行时,其执行上述多种处理图像数据的方法中的一种,并且本发明还可以被实施为计算机程序产品,例如,存储这样的计算机程序的数据载体。

[0093] 已经出于说明和描述的目的给出了以上对于本发明实施方式的描述。其并不旨在穷举或者将本发明限制在所公开的精确形式。可以进行替换、修改和变化,而不偏离本发明的范围。

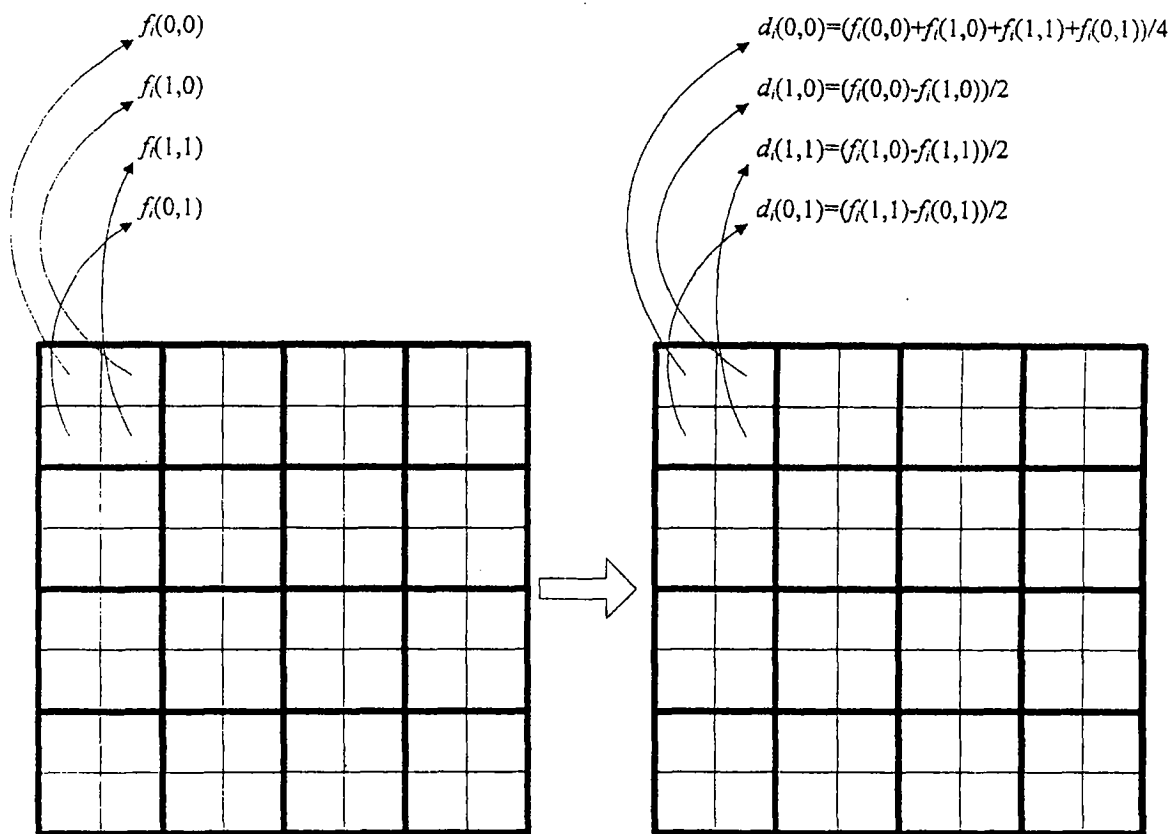


图 1

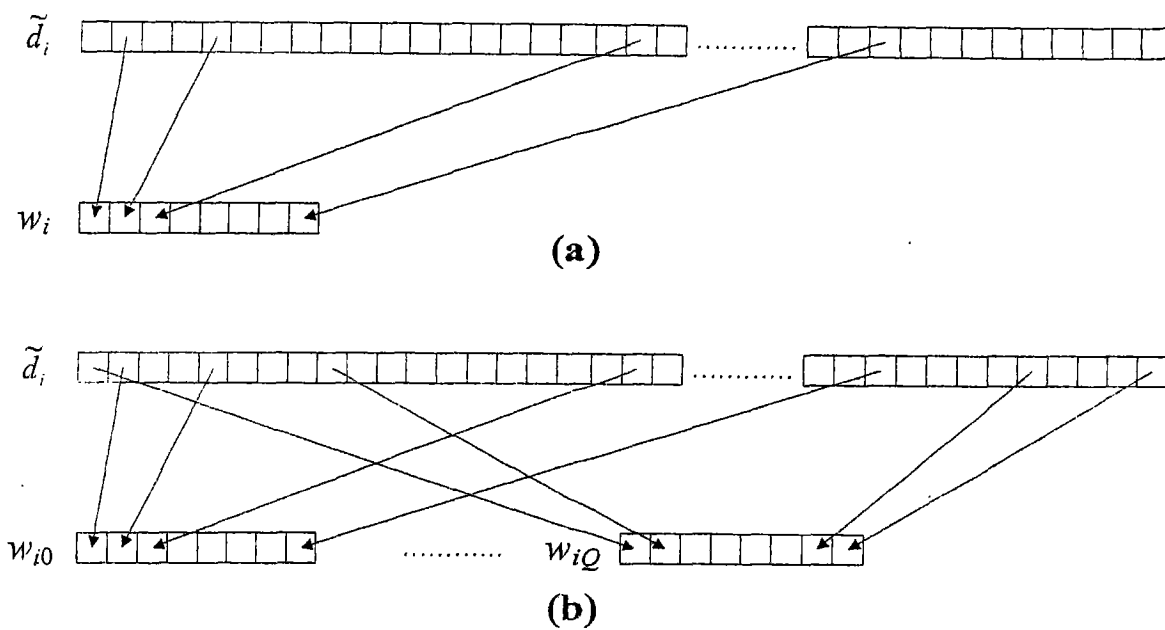


图 2

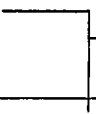
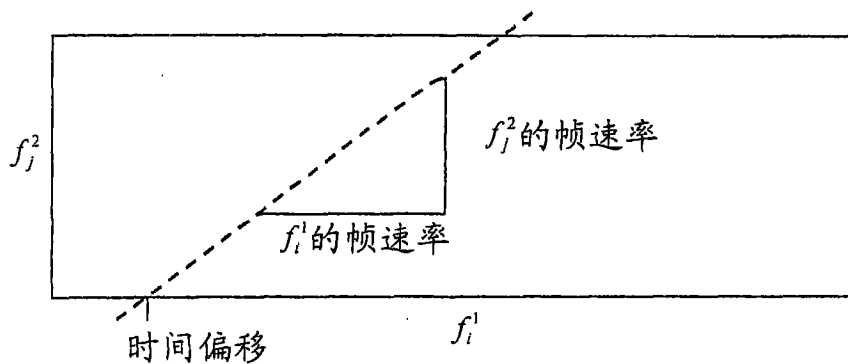
字	第一次出现的帧号	二元表示
w_α	ω_0 	$\tilde{\omega}_0 = \begin{cases} 0 & \omega_0 < \omega_1 \\ 1 & \text{其他情况} \end{cases}$
$w_{\alpha+1}$	ω_1 	$\tilde{\omega}_1 = \begin{cases} 0 & \omega_1 < \omega_2 \\ 1 & \text{其他情况} \end{cases}$
$\vdots$	$\vdots$	$\vdots$
w_ζ	ω_{z-1}	$\tilde{\omega}_{z-1} = \begin{cases} 0 & \omega_{z-1} < \omega_0 \\ 1 & \text{其他情况} \end{cases}$

图 3



$$\text{帧速率改变} = \frac{f_j^2 \text{ 的帧速率}}{f_i^1 \text{ 的帧速率}}$$

图 4

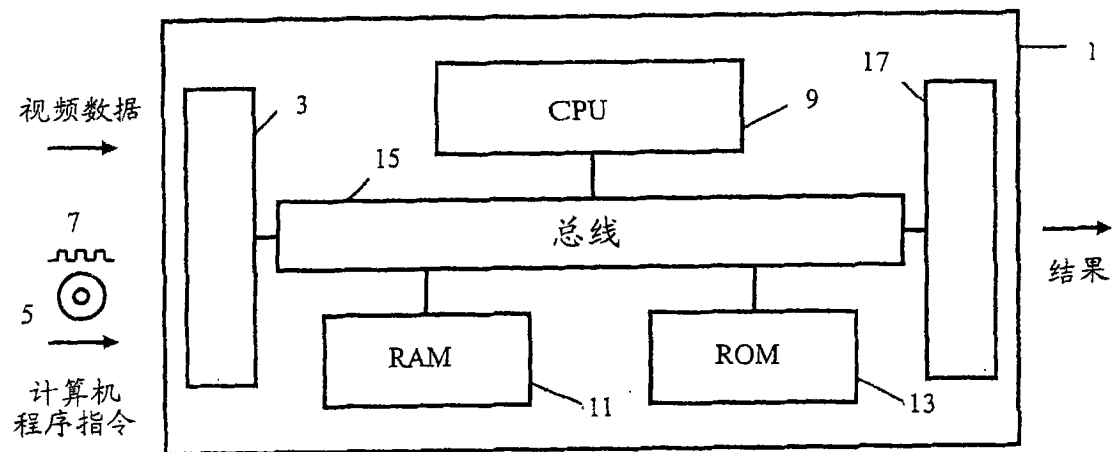


图 5