



(12) 发明专利

(10) 授权公告号 CN 101061489 B

(45) 授权公告日 2011.09.07

(21) 申请号 200580039609.9

HO4N 5/225(2006.01)

(22) 申请日 2005.09.20

(56) 对比文件

(30) 优先权数据

60/611,878 2004.09.21 US

US 6711278 B1, 2004.03.23, 说明书第1栏第7-10行, 第29-55行, 第5栏第18-21行, 第12栏第27-41行.

(85) PCT申请进入国家阶段日

2007.05.18

US 2003/0194134 A1, 2003.10.16, 全文.

(86) PCT申请的申请数据

PCT/US2005/033733 2005.09.20

US 2003/0235341 A1, 2003.12.25, 第1页第5段, 第5页第64段.

(87) PCT申请的公布数据

W02006/034308 EN 2006.03.30

US 2004/0135788 A1, 2004.07.15, 摘要, 正文512页第2.1节.

(73) 专利权人 欧几里得发现有限责任公司

地址 美国马萨诸塞州

Paul Viola, Michael Jones. Rapid Object Detection using a Boosted Cascade of SimpleFeatures. IEEE. 2001, I511-I518.

审查员 冯丹琼

(72) 发明人 查尔斯·保罗·佩斯

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华 陈姗姗

(51) Int. Cl.

G06K 9/00(2006.01)

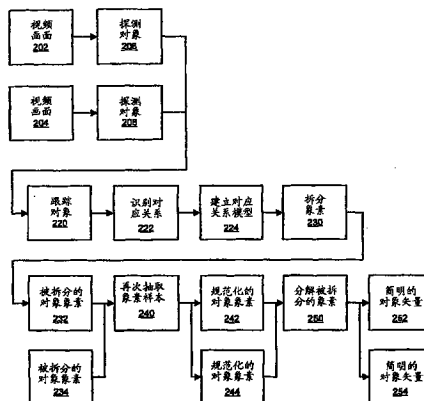
权利要求书 2 页 说明书 12 页 附图 6 页

(54) 发明名称

用来处理视频数据的装置和方法

(57) 摘要

这项发明描述用来处理视频数据的装置和方法。本发明提供一种视频数据的表达,该数据表达能用来评定所述数据和用于所述数据的特定参数表达的拟合模型之间的一致性。这允许比较不同的参数化技术和选择继续处理该特定视频数据的最佳技术。所述表达能以中间形式作为较大程序的一部分或作为反馈机制被用于处理视频数据。当以它的中间形式被利用的时候,本发明能被用于视频数据的储存、增强、提炼、特征提取、压缩、编码和传输的程序。本发明可用来以强健有效的方式提取显著的信息,同时确定通常与视频数据来源相关联的问题的地址。



1. 一种用以生成来自为数众多的视频画面的视频信号数据的编码形式的计算机装置，所述装置包括：

用于对视频信号数据的视频画面序列执行混合空间规范化压缩的装置，所述用于执行混合空间规范化压缩的装置包括：

探测装置，用于探测所述视频画面中的至少一个对象；

跟踪装置，用于通过所述视频画面跟踪所述对象；

识别装置，用于识别所述对象在与所述对象相关联的多幅视频画面之间的对应元素；

建模装置，用于对所述对应元素建模以产生对应关系模型；

再次抽样装置，用于在所述视频画面中再次抽取与所述对象相关联的像素数据的样本，所述再次抽样装置利用所述对应关系模型来形成再次抽样的像素数据；

解除规范化装置，用于对所述再次抽样的像素数据的空间位置解除规范化，所述解除规范化装置利用所述对应关系模型来形成恢复的像素数据；以及

所述再次抽样的像素数据是所述对象的原始像素数据的中间形式以用于视频处理。

2. 根据权利要求 1 的装置，其中所述对象的探测装置和跟踪装置包括 Viola/Jones 面部探测装置。

3. 根据权利要求 1 的装置，其中所述用于执行混合空间规范化压缩的装置包括用于将所述对象从视频画面中分割出来的分割装置，所述分割装置包括：

空间分割装置，用于在与所述对象相关联的所述视频画面中对与所述对象相关联的像素数据和其他像素数据进行空间分割；

用于将所述恢复的像素数据与空间分割的与所述对象相关联的像素数据组装在一起以形成原始视频画面的装置；以及

所述空间分割装置包括临时整合装置。

4. 根据权利要求 1 的装置，其中所述用于执行混合空间规范化压缩的装置还包括因式分解装置，用于将所述对应关系模型因式分解成若干整体运动模型，其中包括：

用于将对应关系整合成整体运动模型的装置；

所述对应关系的建模装置包括对二维仿射运动模型的解的强健的抽样共识装置，以及所述对应关系的建模装置使用由对两幅或多幅视频画面之间基于区段的运动进行评估的装置所产生的有限差分为基础的抽样总体。

5. 根据权利要求 1 的装置，其中所述用于执行混合空间规范化压缩的装置进一步包括用于对所述再次抽样的像素数据进行进一步编码的装置，其中包括：

分解装置，用于将所述再次抽样的像素数据分解成编码表达；

重组装置，用于重组来自编码表达的所述再次抽样的像素数据；

所述分解装置包括用于执行主要成份分析的装置，以及

所述重组装置包括用于执行主要成份分析的装置。

6. 根据权利要求 2 的装置，其中所述跟踪装置进一步包括用于跟踪多个对象以便识别对应元素的装置。

7. 根据权利要求 5 的装置，其中所述用于执行混合空间规范化压缩的装置进一步包括用于对所述视频画面的非对象像素的像素数据进行建模的装置，当所述对象被除去的时候，所述视频画面的非对象是残留物。

8. 根据权利要求 1 的装置,其中所述用于执行混合空间规范化压缩的装置进一步包括分割装置,用于将与所述对象相关联的像素数据从与所述对象相关联的视频画面中的其他像素数据分割出来;以及组合装置,用于利用传统的视频压缩/解压缩装置来组合分割的与所述对象相关联的像素数据和所述再次抽样的像素数据,包括:

用于将所述再次抽样的像素数据作为标准的视频数据供应给所述传统的视频压缩/解压缩装置的装置;

用于连同对应的编码视频数据一起储存和传输对应关系模型的装置;以及借此所述组合装置能使传统的视频压缩方法提高压缩效率。

9. 根据权利要求 4 的装置,其中所述用于执行混合空间规范化压缩的装置进一步包括整合装置,用于对运动参数进行整合以使运动参数的紧凑参数表达变得容易。

10. 根据权利要求 3 的装置,其中所述空间分割装置进一步包括对所述像素数据进行分割以便对所述像素数据进行再次抽样的装置。

## 用来处理视频数据的装置和方法

[0001] 这份申请要求 2004 年 9 月 21 日以“System and Method for Video Compression Employing Principal Component Analysis”为题申请的美国专利临时申请第 60/611, 878 号的优先权。这份申请是 2005 年 7 月 28 日申请的美国专利申请第 11/191, 562 号的部分继续申请。上述申请都在此通过引证被全部并入。

### 技术领域

[0002] 本发明一般地涉及数字信号处理领域,更具体地说涉及用来有效地表达和处理信号或图像数据(最具体地说,视频数据)的计算机装置和计算机实现的方法。

### 背景技术

[0003] 本发明存在于其中的现有技术的一般系统描述能用图 1 表示。在这里方框图显示典型的现有技术视频处理系统。这样的系统通常包括下列几级:输入级 102、处理级 104、输出级 106 和一个或多个数据储存机制 108。

[0004] 输入级 102 可能包括若干元素,例如,照相机敏感元件,照相机敏感元件阵列、测距敏感元件、或从储存机制取回数据的方法。输入级提供表达人造的和/或自然发生的现象的时间相关序列的视频数据。该数据的显著成份可能被噪音或其它不想要的信号掩盖或污染。

[0005] 视频数据可以依照预先定义的转移协议以数据流、阵列或数据包的形式直接地或通过中间的储存元素 108 提交给处理级 104。处理级 104 可以采用专用模拟或数字器件或可编程器件(例如,中央处理器(CPU)、数字信号处理器(DSP)、或现场可编程门阵列(FPGA))的形式来执行所需要的一组视频数据处理操作。处理级 104 通常包括一个或多个 CODEC(编码/解码器)。

[0006] 输出级 106 产生能够影响使用者或外部装置的信号、显示或其它响应。通常,输出器件被用来产生指示信号、显示、硬拷贝、处理过的数据在存储器中的表达,或开始向远程站点传输数据。它也可以用来提供在后面的处理操作中使用的中间信号或控制参数。

[0007] 存储器在这个系统中是作为非必选的元素出现的。在使用时,储存元素 108 可以是非易失的,例如,只读储存媒体,或易失的,例如,动态随机存取存储器(RAM)。单一的视频处理系统包括若干种储存元素并非是罕见的,这些元素对输入级、处理级和输出级有各种不同的关系。这样储存元素的例子包括输入缓冲器,输出缓冲器和处理高速缓冲存储器。

[0008] 图 1 所示视频处理系统的主要目的是处理输入数据,产生对特定应用意义深长的输出。为了实现这个目标,可以利用多种处理操作,包括减少或消除噪音、特征提取、对象拆分和/或规范化、数据分类、事件探测、编辑、数据选择、数据重新编码和代码变换。

[0009] 产生受不良约束的数据的许多数据来源(尤其是声音和可视图像)对人是重要的。在大多数情况下,这些来源信号的基本特征对有效数据处理的目标有不利的影响。来源数据固有的易变性是在不引进起因于在推导工程假定中使用的未试过的经验的和探索性的方法的误差的情况下以可靠且有效的方式处理数据的障碍。这种易变性对于某些应用

当输入数据被自然地或故意地限制在定义狭窄的特征组（例如，一组有限的符号值或狭窄的带宽）之中的时候减轻。这些限制时常导致商业价值低的处理技术。

[0010] 信号处理系统的设计受该系统的预期用途和作为输入使用的来源信号的预期特征的影响。在大多数情况下，所需的完成效率也将是重要的设计因素。完成效率反过来受与可得的储存数据相比较将被处理的数据的数量以及与可得的计算能力相比较该应用程序的计算复杂性的影响。

[0011] 传统的视频处理方法因具有许多低效率性质而蒙受损害，这些低效率性质是以数据通信速度慢、存储需求大和干扰感性假象的形式出现的。这些可能是严重的问题，因为人们希望使用和操纵视频数据的方法是多样的和人们对某些形式的可视信息有先天的敏感性。

[0012] “最佳的”视频处理系统在完成一组预期的处理操作方面是有效的、可靠的和强健的。这样的操作可以包括数据的储存、传输、显示、压缩、编辑、加密、增强、分类、特征探测和识别。次要的操作可以包括这种经过处理的数据与其它信息来源的整合。在视频处理系统的情况下，同样重要的是，输出应该通过避免引进感性假象与人类视觉相容。

[0013] 视频处理系统如果它的速度、效率和质量不强烈地取决于输入数据的任何特定特征的特性可以被描述为“强健的”。强健也与一些输入出错时完成操作的能力有关。许多视频处理系统无法强健到足以考虑到一般的应用类别，仅仅适用于与在研发那个系统时所使用的数据一样受狭窄约束的数据。

[0014] 显著的信息可能由于输入元素的抽样速率与感知现象的信号特征不匹配在连续取值的数据来源的离散化中丢失。另外，当信号强度超过敏感元件的极限，从而导致饱和的时候也有丢失。同样，当用一组不连续的数值表示完整的输入数据数值范围，借此降低数据表达的准确性之时在任何量化程序中发生输入数据的准确性降低的时候，信息被遗失。

[0015] 总体易变性指的是数据或信息来源的类别的任何无法预测性。因为可视信息通常是不受约束的，所以代表可视信息的数据有程度非常大的整体易变性。视觉数据可以表达任何能借助在敏感元件阵列上入射的光线形成的空间阵列序列或时空序列。

[0016] 在建立视觉现象模型时，视频处理器通常把一些限制和 / 或结构设定强加在数据的表达或解释方式上。结果，这样的方法可能引进系统误差，这些系统误差将影响输出的质量、可能与输出有关的置信度和能在该数据上可靠地完成的后续处理任务的类型。

[0017] 量化方法在试图保持那组数据的统计变化的同时降低视频画面数据的准确性。通常，视频数据是这样分析的，以致数据值的分布被收集到概率分布之中。另外，有一些方法，这些方法为了把数据的特色表现为空间频率的混合把数据映射到相空间中，借此允许准确性下降以较少引起反对的方式散布。这些量化方法在被大量利用的时候往往导致知觉上难以置信的颜色而且能在视频画面的原本平滑的区域中引起突然的怪僻状态。

[0018] 差分编码通常也用来利用数据的局部空间相似性。在画面的一个部分中的数据倾向于聚集在那个画面中的相似数据周围和后续画面中的相似位置。然后，根据它的空间毗连数据表达该数据能与量化组合起来，而最终结果是对于给定的准确性表达差分比使用数据的绝对值更精确。这个假定在原始视频数据的光谱分辨率有限的时候（例如，在黑白图像或颜色少的图像中）很好地工作。随着图像的光谱分辨率逐渐增加，相似性假定被严重破坏。这种破坏是由于没有能力有选择地保护视频数据准确性造成的。

[0019] 残差编码与差分编码类似,因为这种表达的误差被进一步差分编码,以便把原始数据的准确性恢复到预期的准确性水平。

[0020] 这些方法的变化尝试把视频数据变换成把数据相关关系暴露在空间相位和刻度之中的替代表达。一旦视频数据已经以这些方式变换,量化和差分编码的方法就能适用于被变换的数据,从而导致增加显著图像特征的保存。这些变换视频压缩技术中最普遍两种是离散余弦变换(DCT)和离散子波变换(DWT)。DCT变换的误差表明在视频数据数值方面有广泛的变化,因此,DCT通常被用在视频数据的区段上,为的是使这些错误的相关关系定位。来自这种定位的假象往往沿着这些区段的边界出现。就DWT而言,更复杂的假象在基础函数和某些纹理之间有误配的时候发生,而且这引起模糊效应。为了抵消DCT和DWT的负面效应,提高表达的准确性以便以宝贵的带宽为代价减少失真。

### 发明内容

[0021] 本发明是一种在计算和分析方面均优于现有顶级技术的方法的计算机实现的视频处理方法。原则上本发明的方法是线性分解法,空间拆分法和空间规范化法的整合。从空间上限制视频数据大大提高线性分解法的强健性和适用性。此外,数据的空间拆分在其它的高变异性数据在空间上与正在分析的数据毗连的时候能减轻所引起的非线性。

[0022] 具体地说,本发明提供一种方法,采用该方法能把信号数据有效地处理成一个或多个有益的表达。本发明在处理许多普遍发生的数据组时是有效的而且在处理视频和图像数据时是特别有效的。本发明的方法分析该数据并且提供那个数据的一种或多种简洁表达以使它的处理和编码变得容易。对于许多应用(包括但不限于:视频数据的编码、压缩、传输、分析、储存和显示),每种新的比较简洁的数据表达都允许减少计算处理、传输带宽和储存需求。本发明包括用来识别和提取视频数据的显著成份的方法,从而允许区分数据的处理和表达的优先次序。信号中的噪音和其它多余部分被看作是优先权比较低的,以致进一步的处理能集中在分析和表达视频信号中优先权比较高的部分上。结果,视频信号的表达比先前可能的表达更简洁。而且把准确性的损失集中在视频信号中知觉上不重要的部分。

### 附图说明

[0023] 图1是举例说明现有技术视频处理系统的方框图。

[0024] 图2是提供本发明的概观的方框图,它展示用来处理影像的主要组件。

[0025] 图3是举例说明本发明的运动评估方法的方框图。

[0026] 图4是举例说明本发明的整体配准方法的方框图。

[0027] 图5是举例说明本发明的规范化方法的方框图。

[0028] 图6是举例说明混合式空间规范化压缩方法的方框图。

### 具体实施方式

[0029] 在视频信号数据中,视频画面被组装成通常描绘投影到二维成像表面上的三维现场的图像序列。每个画面(或图像)都由代表响应抽样信号的成像敏感元件的象素组成。时常,抽样信号对应于用二维敏感元件阵列抽样的一些反射的、折射的或发射的电磁能。连续的顺序抽样导致时空数据流,每个画面的两个空间维度和一个时间维度对应于该画面在

视频序列中的次序。

[0030] 本发明如同图 2 举例说明的那样分析信号数据和识别显著成份。当信号由视频数据组成的时候,时空流分析揭示时常作为特定对象(例如,面部)的显著成份。识别程序限定显著成份的存在和重要性并且选择那些显著成份之中最重要的一个或多个显著成份。这不限在现在描述的处理之后或同时识别和处理其它较为不显著的成份。然后,上述的显著成份被进一步分析,以便识别易变的和不变的子成份。不变的子成份的识别是建立该成份的某个方面的模型的程序,借此揭示该模型的参数表达法,以允许将该成份合成到预期的准确性水平。

[0031] 在本发明的一个实施方案中,探测和跟踪前景对象。该对象的像素被识别并且从每个视频画面中拆分出来。基于区段的运动评估被应用于从多个画面中拆分出来的对象。然后,这些运动评估被整合成一个高级的运动模型。该运动模型用来把该对象的例证隐藏到公用的空间配置中。对于特定的数据,在这个配置中,该对象更多的特征被对准。这种规范化允许将该对象的像素在多个画面上的数值的线性分解被紧凑地表达。属于该对象的外观的显著信息包含在这个紧凑的表达之中。

[0032] 本发明的优选实施方案详细描述前景视频对象的线性分解。该对象在空间上被规范化,借此得出紧凑的线性外观模型。此外,进一步的优选实施方案在空间规范化之前把前景对象从视频画面的背景中拆分出来。

[0033] 本发明的优选实施方案将本发明应用于一个人对着摄像机边说话边进行少量运动的影像。

[0034] 本发明的优选实施方案将本发明应用于影像中能通过空间转换被很好地表达的任何对象。

[0035] 本发明的优选实施方案明确地使用基于区段的运动评估来确定两个或多个视频画面之间的有限差分。为了要提供更有效的线性分解,高级运动模型是依据那些有限差分因式分解的。

[0036] 探测 & 跟踪

[0037] 一旦已经确定信号的显著构成成份,这些成份就可以被保留,而所有其它的信号成份可以被减少或去除。探测显著成份的程序展示在图 2 中,在那里视频画面(202)是用一个或多个探测对象(206)程序处理的,从而导致一个或多个对象被识别并且随后被跟踪。保留的成份代表视频数据的中间形式。然后,可以使用对于现有的视频处理方法通常不可得的技术给这个中间数据编码。因为该中间数据以几种形式存在,所以标准的视频编码技术也能用来给这些中间形式中的一些编码。对于每个例证,本发明都先确定然后使用最有效的编码技术。

[0038] 在一个优选实施方案中,特征分析程序完成显著信号模式的探测和分类。这个程序的一个实施方案使用专门为产生强度与在视频画面中探测到的对象特征有关的响应信号而设计的空间过滤器的组合。该分类程序是以不同的空间刻度应用于视频画面的不同位置的。来自分类程序的响应的强度指出显著信号模式出现的可能性。在把中心置于十分显著的对象上的时候,该程序用对应的强烈响应给它分类。显著信号模式的探测通过激活对视频序列中的显著信息的后续处理和分析来辨别本发明。

[0039] 给出显著信号模式在一个或多个视频画面中的探测位置,本发明分析显著信号模

式的无变化特征。此外,对于无变化的特征,本发明分析该信号的残值,“较少显著的”信号模式。无变化特征的识别提供用来减少多余信息和拆分(即,分隔)信号模式的基础。

#### [0040] 特征点跟踪

[0041] 在本发明一个实施方案中,在一个或多个画面中的空间位置是通过空间强度场梯度分析确定的。这些特征对应于“线”的一些交点,这些交点能被宽松地描述为“拐角”。这样的实施方案进一步选择一组这样的拐角,这些拐角是强壮的而且在空间上是彼此完全不同的,在此称之为特征点。此外,使用光流的分层次的多分辨率评估允许确定随着时间流逝特征点的平移位移。

[0042] 在图 2 中,跟踪对象(220)程序是为了把来自探测对象程序(208)的探测例证拉到一起和进一步识别一个或多个被探测对象的特征在许多视频画面(202 和 204)上的对应关系。

[0043] 特征跟踪的非限制性实施方案能被这样使用,以致这些特征被用来限定更规则的梯度分析法(例如,基于区段的运动评估)。

[0044] 另一个实施方案期待以特征跟踪为基础的运动评估的预测。

#### [0045] 基于对象的探测和跟踪

[0046] 在本发明的一个非限制性实施方案中,强健的对象分类程序被用来跟踪视频画面中的面部。这样的分类程序以已经在那些面部上训练过的定向边缘的级联响应为基础。在这个分类程序中,边缘被定义为一组基本的 Haar 特征和那些特征的 45 度旋转。级联分类程序是 AdaBoost 算法的变体。此外,响应计算能通过使用总面积表优化。

#### [0047] 局部配准

[0048] 配准包括在两个或多个视频画面中被识别对象的元素之间的对应关系的分配。这些对应关系变成建立视频数据中时间点截然不同的视频数据之间的空间关系模型的基础。

[0049] 为了根据广为人知的算法和那些算法的富有创造性的派生算法举例说明特定的实施方案和与它们相关联的实践缩减量,现在描述用于本发明的各种不同的非限制性的配准方法。

[0050] 在时空序列中建立明显的光流模型的一种方法可以通过从视频数据的两个或多个画面产生有限差分域实现的。如果对应关系在空间和强度双重意义上符合特定的恒定不变的限制,光流场能被稀疏地评估。

#### [0051] 菱形搜寻

[0052] 假定把一个视频画面分割成若干不重叠的区段,搜寻先前的与每个区段匹配的视频画面。以全面搜寻区段为基础(FSBB)的运动评估找出与当前画面中的区段相比较时在早先的视频画面中误差最小的位置。完成 FSBB 可能是计算费用十分浩大的,而且往往不产生比以局域化运动假设为基础的其它评估方案更好的匹配。以菱形搜寻区段为基础(DSBB)的梯度下降运动评估是 FSBB 的常见的替代品,它使用各种不同尺寸的菱形搜寻图案朝着对于某个区段最好的匹配的方向反复地横越误差梯度。

[0053] 在本发明的一个实施方案中,为了产生数值稍后被因式分解成高阶运动模型的有限差分,DSBB 被用于一个或多个视频画面之间的图像梯度域分析。

[0054] 熟悉这项技术的人知道基于区段的运动评估能被视为规则网孔顶点分析的同价物。



[0055] 基于相位的运动评估

[0056] 在现有技术中,基于区段的运动评估通常是作为导致一个或多个空间匹配的空间搜寻实现的。基于相位的规范化的互相关 (PNCC) 如同图 3 举例说明的那样把来自当前画面和先前画面的区段变换到“相空间”中,并且寻找那两个区段的互相关。这种互相关被表达为位置与两个区段之间的边缘的“相移”相对应的数值域。这些位置通过定阈值被隔离,然后被逆变换成空间坐标。这些空间坐标是截然不同的边缘位移,而且对应于运动矢量。

[0057] PNCC 的优势包括反差掩蔽,该反差掩蔽在视频流中预留增益 / 曝光调节的容许偏差。另外,PNCC 允许来自单一步骤的结果,该单一步骤或许处理来自基于空间的运动评估程序的许多迭代。此外,该运动评估是子像素精确的。

[0058] 本发明的一个实施方案在一个或多个视频画面之间的图像梯度域的分析中利用 PNCC,为的是产生其数值稍后被因式分解成高阶运动模型的有限差分。

[0059] 整体配准

[0060] 在一个实施方案中,本发明将来自有限差分评估的域的一个或多个线性模型因式分解。发生这样的抽样的域在此被称为有限差分的一般总体。所描述的方法使用与 RANSAC 算法类似的强健的评估。

[0061] 如图 4 所示,在建立整体运动模型的情况下,有限差分是集中在通过那些运动评估的随机抽样 (410) 迭代处理的一般总体库 (404) 之中的平移运动评估 (402),而且线形模型被因式分解,提取那些样本的公因子 (420)。然后,那些结果被用来调节总体 (404) 以便通过排除该模型的异己样本更好地阐明该线性模型,如同通过随机处理发现的那样。

[0062] 在线性模型评估算法的一个实施方案中,运动模型评估程序以线性最小二乘解为基础。这种相关性使该评估程序摆脱异己样本数据。基于 RANSAC,所揭示的方法是一种通过反复评估数据子集抵消异己样本的效应探查将描述重要的数据子集的运动模型的强健方法。每个探头产生的模型都对它所代表的的数据百分比进行测试。如果有足够的迭代次数,则将发现与最大的数据子集拟合的模型。

[0063] 如同图 4 设想和举例说明的那样,本发明揭示一些在算法变更形式上超过 RANSAC 算法的改革,包括有限差分的初始抽样 (样本) 和线性模型的最小二乘评估。综合误差是使用已解的线性模型对一般总体中的所有样本评估的。根据残差符合预先设定的阈值的样本的数目把一个等级分配给该线性模型。这个等级被看作是“候选的共识”。

[0064] 初始抽样、求解和归类是通过迭代完成的,直到终止判据得到满足为止。一旦该判据得到满足,等级最高的线性模型被看作是该总体的最后共识。

[0065] 非必选的改进步骤包括按照与候选模型拟合最好的次序反复地分析样本的子集和逐渐增加该子集的大小,直到再加一个样本将会超过整个子集的残留误差阈值。

[0066] 为了在将与某特定的线性模型相对应的另一个参数矢量空间中确定子空间的拓扑空间,所描述的本发明的非限制性实施方案可以作为对矢量空间 (前面被描述为有限差分矢量的域) 抽样的一般方法进一步推广。

[0067] 整体配准程序的进一步的结果是这个配准程序和局部配准程序之间的差异产生局部配准残差。这个残差是整体模型在近似局部模型时的误差。

[0068] 规范化

[0069] 规范化指的是朝着标准的或通常的空间配置方向再次抽取空间强度场样本。当这

些相关的空间配置是这样的配置之间可逆的空间变换的时候,象素的再次抽样和附带插值也是直到拓扑极限可逆的。本发明的规范化方法是用图 5 举例说明的。

[0070] 当两个以上空间强度场被规范化的时候,提高的计算效率可以通过保存中间的规范化计算结果来实现。

[0071] 为了配准的目的,或等效地为了规范化,用来再次抽取图像样本的空间变换模型包括总体模型和局部模型。总体模型有从平移变换到影射变换逐渐增加的阶次。局部模型是有限差分,该有限差分暗示在基本上用区段或更复杂地用分段线性网孔确定的关于邻近象素的内插式。

[0072] 原始强度场向规范化强度场的插值增加基于强度场子集的 PCA 外观模型的直线性。

[0073] 如图 2 所示,对象象素 (232 和 234) 能被再次抽样 (240) 以便得到所述对象象素的规范化版本 (242 和 244)。

[0074] 三维规范化

[0075] 本发明进一步的实施方案把特征点镶嵌到基于三角形的网孔中,跟踪该网孔的顶点,并且使用每个三角形的顶点的相对位置来估计与那三个顶点一致的平面的三维表面法线。当该表面法线与摄影机的投影轴相符的时候,成像象素能提供与该三角形相对应的对象的扭曲最小的透视图。创造倾向于支持正交表面法线的规范化图像能产生保存中间数据类型的象素,这将提高后来以外观为基础的 PCA 模型的直线性。

[0076] 另一个实施方案利用传统的以区段为基础的运动评估来含蓄地建立整体运动模型。在一个非限制性实施方案中,该方法将来自传统的以区段为基础的运动评估 / 预测所描述的运动矢量的整体仿射运动模型因式分解。

[0077] 渐进的几何规范化

[0078] 空间间断点的分类被用来对准镶嵌的网孔,以便在它们与网孔边缘一致的时候含蓄地建立间断点模型。

[0079] 同种区域的边界是用多角形轮廓近似的。为了确定每个初始顶点的优先权,该轮廓被逐次接近。为了保护共享顶点的顶点优先权,顶点优先权在各个区域上传播。

[0080] 图像配准用强烈变化的图像梯度向优先权高的顶点倾斜。由此产生的变形模型倾向于保护与成像对象的几何形状有关的空间间断点。

[0081] 拆分

[0082] 通过进一步描述的拆分程序识别的空间间断点是通过它们各自边界的几何参数表达法(被称为空间间断点模型)有效地编码的。这些空间间断点模型可以不断地考虑到与编码子集相对应的更简洁的边界描述的渐进方式编码。渐进式编码提供一种在保留空间间断点的许多显著方面的同时区分空间几何学优先次序的强健方法。

[0083] 本发明的优选实施方案把多分辨率拆分分析和空间强度梯度域分析结合起来并且进一步使用时间稳定性限制,为的是实现强健的拆分。

[0084] 如图 2 所示,一旦已经随着时间的流逝跟踪对象的特征的对对应关系 (220) 并且建立了模型 (224),遵守这个运动 / 变形模型能用来拆分与那个对象相对应的象素 (230)。可以对画面 (202 和 204) 中已探测到的许多对象 (206 和 208) 重复这个程序。

[0085] 本发明使用的无变化特征分析的一种形式被集中在空间间断点的识别上。这些间

断点是作为边缘、阴影、遮蔽、线、拐角或任何其它的在一个或多个视频成像画面中引起像素之间突然的可辨别的分离的可见特征出现的。此外，在颜色和 / 或纹理类似的对象之间的细微的空间间断点可能仅仅出现在视频画面中各个对象的像素相对于那些对象本身正在经历粘附运动而相对于其它对象正在经历不同的运动之时。本发明利用频谱拆分、纹理拆分和运动拆分的组合强健地识别与显著信号模式有关的空间间断点。

#### [0086] 时间拆分

[0087] 把平移运动矢量或在空间强度场中等价的有限差分测量结果按时间整合成高阶运动模型是现有技术描述的一种运动拆分形式。

[0088] 在本发明的一个实施方案中，产生运动矢量的稠密域，表现视频画面中对象运动的有限差分。这些导数是通过规则地分割瓦片或借助某种初始化程序（例如，空间拆分）按空间集合的。每个集合的“导数”使用线性最小二乘评估程序整合成一个高阶运动模型。然后，由此产生的运动模型作为矢量在运动模型空间中使用 k 一方法群集技术群集。这些导数是基于与它们拟合最好的群分类的。然后，群标是作为空间分割的演化按空间群集的。该程序一直继续到空间分割稳定为止。

[0089] 在本发明的进一步的实施方案中，适合给定的孔径的运动矢量被内插到一组与该孔径相对应的像素位置。当用这种内插定义的区域横越与对象边界相对应的像素时候，由此产生的分类是该区段的某种不规则的对角线分割。

[0090] 在现有技术中，用来整合导数的最小二乘评估程序对异常值是非常敏感的。这种敏感性能产生使运动模型的群集方法严重地向迭代结果大大发散的点倾斜的运动模型。

[0091] 在本发明中，运动拆分方法通过分析两个以上视频画面上明显的像素运动识别空间间断点。明显的运动是针对这些视频画面上的一致性分析的并且被整合成参数运动模型。与这种一致的运动相关联的空间间断点被识别出来。运动拆分也可以被称为时间拆分，因为时间变化可能是由运动引起的。然而，时间变化也可能是由一些其它的现象（例如，局部变形、照明变化，等等）引起的。

[0092] 通过所描述的方法，与规范化方法相对应的显著信号模式能被识别而且能通过几种背景减法之一与环境信号模式（背景或非对象）分开。时常，这些方法从统计上建立背景模型，因为像素在每个时间样本都呈现最小的变化量。变化能被视为像素数值差异。作为替代，运动拆分能在给出显著图像模式的探测位置和数值范围的情况下实现。距离变换能用来确定每个像素距探测位置的距离。如果与最大距离相关联的像素数值被保留，合理的背景模型能被求解。换句话说，环境信号能使用信号差异度量标准按时间再次抽样。

[0093] 给出环境信号的模型，就能按每个时间样本使完全的显著信号模式有差别。这些差别每个都能通过再次抽样变成空间规范化的信号差异（绝对差异）。然后，这些差异相互对准和累积。由于这些差异相对于显著信号模式已按空间规范化，所以，差异的峰通常将对应于与显著信号模式相关联的像素位置。

#### [0094] 非对象的分辨率

[0095] 给出清晰的背景图像，这个图像和当前画面之间的误差可以按空间规范化和按时间累积。这样的清晰背景图像是在“分辨率”部分中描述的。

[0096] 然后，由此产生的累积误差通过阈值检验提供初始轮廓。然后，该轮廓在空间上被扩展以使残留误差与轮廓变形平衡。

**[0097] 梯度拆分**

[0098] 纹理拆分方法或同义的强度梯度拆分分析像素在一个或多个视频画面中的局部梯度。梯度响应是一种表征空间间断点的统计尺度,其中所述空间间断点对于该视频画面中的像素位置是局部的。然后,使用几种空间群集技术之一把这些梯度响应组合成一些空间区域。这些区域的边界在识别一个或多个视频画面中的空间间断点方面是有用的。

[0099] 在本发明的一个实施方案中,来自计算机图形纹理生成的总面积表概念被用于加快强度场梯度计算的目的。累加值域的产生使通过与四次加法运算结合的四次查询计算任何长方形原始域的总和变得容易。

[0100] 进一步的实施方案使用对一幅图像产生的 Harris 响应,而每个像素的邻近区域被归类为同种的、边缘或拐角。响应数值是依据这个信息产生的并且指出画面中每种元素的边缘化或拐角化的程度。

**[0101] 多刻度梯度分析**

[0102] 本发明的实施方案通过以几种空间刻度产生图像梯度值进一步约束图像梯度支持。这个方法能帮助限定图像梯度的资格,以致在不同刻度下的空间间断点能用来彼此相互支持,只要“边缘”在几种不同的空间刻度下能被分辨,该边缘应该是“显著的”。更有资格的图像梯度将倾向于与更显著的特征相对应。

[0103] 在优选实施方案中,纹理响应区域是首先产生的,然后,这个区域的数值以 k-方法分区 / 分割为基础被量化成若干区间。然后,使用每个区间作为单一迭代能把分水岭拆分能应用于它的数值间隔渐进地处理最初的图像梯度数值。这种方法的好处是同种是在相对意义上用强烈的空间偏置定义的。

**[0104] 光谱拆分**

[0105] 光谱拆分方法分析视频信号中黑白像素、灰度像素或彩色像素的统计概率分布。频谱分类程序是通过完成关于那些像素的概率分布的群集操作构成的。然后,使用该分类程序把一个或多个像素分类,使之属于某个概率类别。然后,由此产生的概率类别和它的像素被赋予类别标签。然后,使这些类别标签在空间上合并成有截然不同的边界的像素区域。这些边界识别在一个或多个视频画面中的空间间断点。

[0106] 本发明可以利用基于光谱分类的空间拆分来拆分视频画面中的像素。此外,各个区域之间的对应关系可以是基于各个光谱区域与先前拆分的区域的重叠确定的。

[0107] 业已观察到当视频画面大体上由空间上被连接成与视频画面中的对象相对应的较大区域的连续彩色区域组成的时候,彩色(或光谱)区域的识别和跟踪能促进图像序列中对象的后续拆分。

**[0108] 背景拆分**

[0109] 本发明包括以每幅视频画面中的探测对象和每个个别像素之间的空间距离测量结果的瞬时最大值为基础建立视频画面背景模型的方法。给定探测到的对象位置,应用距离变换,产生适合画面中每个像素的标量距离数值。在所有的视频画面上每个像素的最大距离的映射图被保留。当最初分配最大数值的时候,或后来用不同的新数值更新该最大数值的时候,适合于那幅视频画面的对应的像素被保留在“清晰的背景”画面中。

**[0110] 建立外观模型**

[0111] 视频处理的共同目标往往是建立模型和保存视频画面序列的外观。本发明以允许

通过预处理的运用以强健的和广泛适用的方式应用强制性外观建模技术为目标。先前描述的配准、拆分和规范化明显地适合这个目的。

[0112] 本发明揭示建立外观变化模型的方法。建立外观变化模型的主要基础在线性模型的情况下是分析特征矢量,以揭示开发利用线性相关关系的坚实基础。表达空间强度场象素的特征矢量能被组装成外观变化模型。

[0113] 在替代实施方案中,外观变化模型是依据被拆分的象素子集计算的。此外,该特征矢量能被分成若干空间上不重叠的特征矢量。这样的空间分解可以用空间铺瓦来实现。计算效率可以通过处理这些临时总体来实现,而不牺牲更普遍的 PCA 方法的维数减少。

[0114] 在产生外观变化模型时,空间强度场规范化能用来减少空间变换的 PCA 建模。

[0115] PCA

[0116] 产生外观变化模型的优选方法是通过把视频画面作为图案矢量组装成一个训练矩阵或总体然后把主要成份分析 (PCA) 应用在该训练矩阵上。当这样的展开式被截取的时候,由此产生的 PCA 变换矩阵被用来分析和合成后面的视频画面。基于截取水平,改变象素的初始外观质量水平能实现。

[0117] 图案矢量的特定的构成和分解方法对于熟悉这项技术的人是广为人知的。

[0118] 给出来自环境信号的显著信号模式的空间拆分和这个模式的空间规范化,象素本身或同义的由此产生的规范化信号的外观能被因式分解成线性相关的成份,其中低级参数表达考虑到适合表达象素外观的近似值误差和比特率之间的直接交换。

[0119] 如图 2 所示,为了得到量纲上简明的数据版本 (252 和 254),规范化的对象象素 (242 和 244) 能投射到矢量空间中而且线性对应关系能使用分解程序 (250) 建立模型。

[0120] 连续的 PCA

[0121] PCA 使用 PCA 变换把图案编码成 PCA 系数。用 PCA 变换表达的图案越好,给该图案编码所需要的系数就越少。承认图案矢量可能随着时间在获得训练图案和待编码图案之间流逝降级,更新变换能帮助抵消这种降级。作为产生新变换的替代品,现有图案的连续更新在特定的情况下是计算上更有效的。

[0122] 许多最新技术的视频压缩算法依据一个或多个其它画面预测某视频画面。预测模型通常基于把每个预测画面分割成与在另一画面中对应的补丁相匹配的不重叠的瓦片和相关联的用偏移运动矢量参数化的平移位移。这个非必选地与画面索引耦合的空间位移提供瓦片的“运动预测”版本。如果预测的误差在特定的阈值以下,则瓦片的象素适合残差编码;而且在压缩效率方面有对应的增益。否则,瓦片的象素被直接编码。这种基于瓦片的换句话说基于区段的运动预测方法通过平移包含象素的瓦片建立影像模型。当影像中的成像现象坚持这种建模的时候,对应的编码效率增加。为了与在基于区段的预测中固有的平移假定一致,这个建模限制为了与在基于区段的预测中固有的平移假设一致假定特定的时间分辨率水平(或帧频)对于正在运动的成像对象是存在的。这种平移模型的另一个必要条件是对于特定的时间分辨率空间位移必须受到限制;换言之,用来推导预测结果的画面和被预测的画面之间的时间差必须是比较短的绝对时间。这些时间分辨率和运动限制使存在于视频流中的某些多余的视频信号成份的识别和建模变得容易。

[0123] 基于残差的分解

[0124] 在 MPEG 视频压缩中,当前的画面是通过先使用运动矢量对先前的画面进行运动

补偿,然后把残差更新应用于那些补偿区段,最后将任何没有充份匹配的区段作为新区段完成编码构成的。

[0125] 对应于残留区段的象素通过运动矢量映射到先前画面的象素上。结果是象素能通过连续应用残值合成的影像的瞬时路径。这些象素被确认为能使用 PCA 最明确地表达的象素。

[0126] 基于遮挡的分解

[0127] 本发明的进一步提高确定适用于多个区段的运动矢量是否将导致来自先前画面的任何象素被移动象素遮挡(覆盖)。对于每个遮挡事件,都把遮挡象素劈成新层。没有历史的象素也将暴露出来。暴露出来的象素被放到任何将在当前画面中与它们拟合而且历史拟合也能在那层上完成的层上。

[0128] 象素的时间连续性是通过象素对不同层的接合和移植得到支持的。一旦获得稳定的层模型,每层中的象素就能基于对条理分明的运动模型的从属关系编组。

[0129] 分波段时间量化

[0130] 本发明的替代实施方案使用离散余弦变换(DCT)或离散子波变换(DWT)把每个画面分解成分波段图像。然后,将主要成份分析(PCA)应用于这些“分波段”影像之中的每幅影像。概念是视频画面的分波段分解与原始视频画面相比较减少任何一个分波段中的空间变化。

[0131] 就移动对象(人)的影像而言,空间变化倾向于支配用 PCA 建模的变化。分波段分解减少任何一个分解影像中的空间变化。

[0132] 就 DCT 而言,任何一个分波段的分解系数都按空间安排在分波段影像之中。举例来说,DC 系数是从每个区段获取的并且被安排在看起来像原始影像的邮票版本一样的分波段影像之中。这将对所有其它的分波段重复,而且使用 PCA 处理每个由此产生的分波段影像。

[0133] 就 DWT 而言,分波段已经按针对 DCT 描述的方式排列好。

[0134] 在非限制性实施方案中,PCA 系数的截取是变化的。

[0135] 子波

[0136] 当使用离散子波变换(DWT)分解数据的时候,多个带通数据组以较低的空间分辨率为结果。变换程序能被递归地应用于导出数据直到仅仅产生单一的标量数值为止。在已分解的结构中标量元素通常以分等级的父母/孩子方式相关。由此产生的数据包含多分辨率的分等级结构以及有限差分。

[0137] 当 DWT 被应用于空间强度场的时候,许多自然发生的图像现象由于空间频率低是用第一或第二低带通导出数据结构以微不足道的知觉损失表达的。截短该分等级结构在高频空间数据不是不存在就是被视为噪音的时候提供简明的表达。

[0138] 尽管 PCA 可以用来以为数不多的系数实现精确的重建,但是这种变换本身可能是相当大的。为了减少这个“初始”变换的规模,可以使用子波分解的嵌零树(EZT)结构来建立变换矩阵的越来越精确的版本。

[0139] 子空间分类

[0140] 本发明把几种离散参数表达法表达成矢量的元素。这些参数以非限制性的方式包括在被拆分对象的规范化外表中的象素、运动参数和在二维空间或三维空间中特征或顶点

的任何结构位置。这些矢量每个都存在于矢量空间之中，而空间几何学分析被用来产生参数矢量的简洁表达。有益的几何学包括集中在矢量群上的子空间。当若干子空间混合成更复杂的子空间的时候，个别较简单的子空间可能难以辨别。有一些拆分方法考虑到通过在通过原始矢量的一些交互作用（例如，内积）产生的有较高维数的矢量空间中考查数据分离这样的子空间。

[0141] 一种拆分矢量空间的方法包括把矢量投射到表达多项式的 Veronese 矢量空间中。这个方法在现有技术中作为广义的 PCA 或 GPCA 技术是众所周知的。通过这样的投射，找到多项式的法线，而且能把与那些法线有关的矢量聚集在一起。这种技术的实用性的例子是把随着时间推移跟踪的二维空间点对应关系因式分解成三维结构模型和那个三维模型的运动。

[0142] 混合空间规范化压缩

[0143] 本发明通过把拆分视频流添加到“规范化”的视频流之中充分发挥以区段为基础的运动预测编码方案的效率。然后，这些视频流分开编码以允许传统的编码解码器的平移运动假设是有效的。在完成规范化视频流的解码之时，视频流解除规范化，进入它们适当的位置并且被组合在一起产生原始的视频序列。

[0144] 在一个实施方案中，一个或多个对象是在视频流中探测到的，而与探测到的每个个别对象有关的像素随后被拆分，离开非对象像素。接下来，针对对象像素和非对象像素产生整体空间运动模型。这个整体模型用来完成对象像素和非对象像素的空间规范化。这样的规范化已经有效地把非平移的运动从视频流中除去并且已经提供一组影像，这组影像的相互遮挡已经被减到最少。这些是本发明的方法的两个有益的特征。

[0145] 像素已按空间规范化的对象和非对象的新影像是作为给传统的以区段为基础的压缩算法的输入提供的。在这些影像解码时，整体运动模型的参数被用来还原规范化的解码画面，对象像素一起合成到非对象像素之上，产生最初的视频流的近似。

[0146] 如图 6 所示，对于一个或多个对象 (630 和 650) 先前探测到的对象例证 (206 和 208) 每个都用传统视频压缩方法 (632) 的分开例证处理。此外，起因于对象的拆分 (230) 的非对象 (602) 也使用传统的视频压缩 (632) 压缩。这些分开的压缩编码 (632) 之中的每一个的结果是分开的传统编码流，每个编码流 (634) 分开地对应于每个视频流。在某个点，可能在传输之后，这些中间编码流 (234) 能被解压缩 (636) 成规范化的非对象 (610) 和许多对象 (638 和 658) 的合成物。这些合成后的像素能解除规范化 (640)，变成它们的已解除规范化的版本 (622、642 和 662)，把这些像素按空间相对于其它像素放置在正确的位置，以致合成程序 (670) 能把对象像素和非对象像素结合成完整的合成画面 (672)。

[0147] 混合编码解码的整合

[0148] 在把传统的基于区段的压缩算法和本发明描述的规范化一拆分方案结合起来时，有一些已经产生结果的本发明的方法。首先，有专门的数据结构和必要的通信协议。

[0149] 主要的数据结构包括整体空间变形参数和对象拆分规范掩模。主要的通信协议是包括传输整体空间变形参数和对象拆分规范掩模的各个层面。

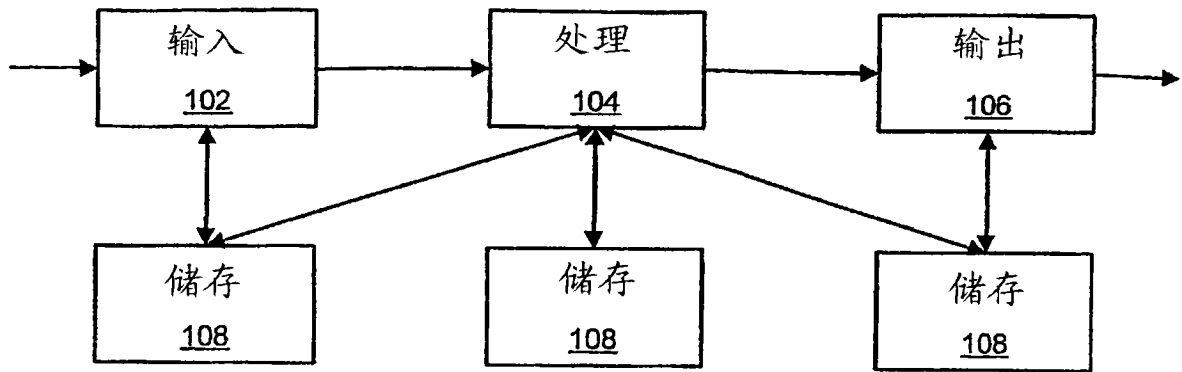


图 1

(现有技术)



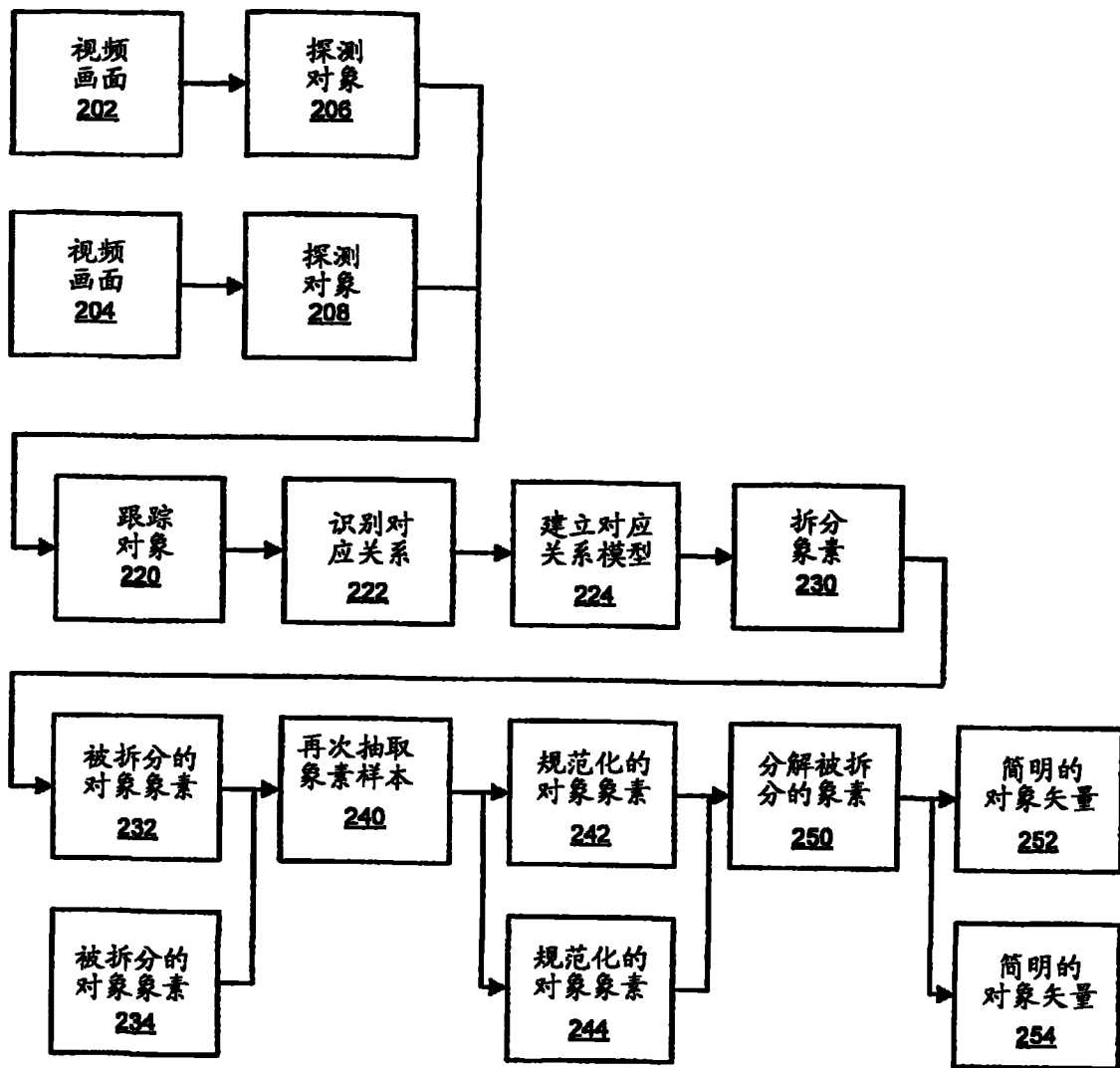


图 2

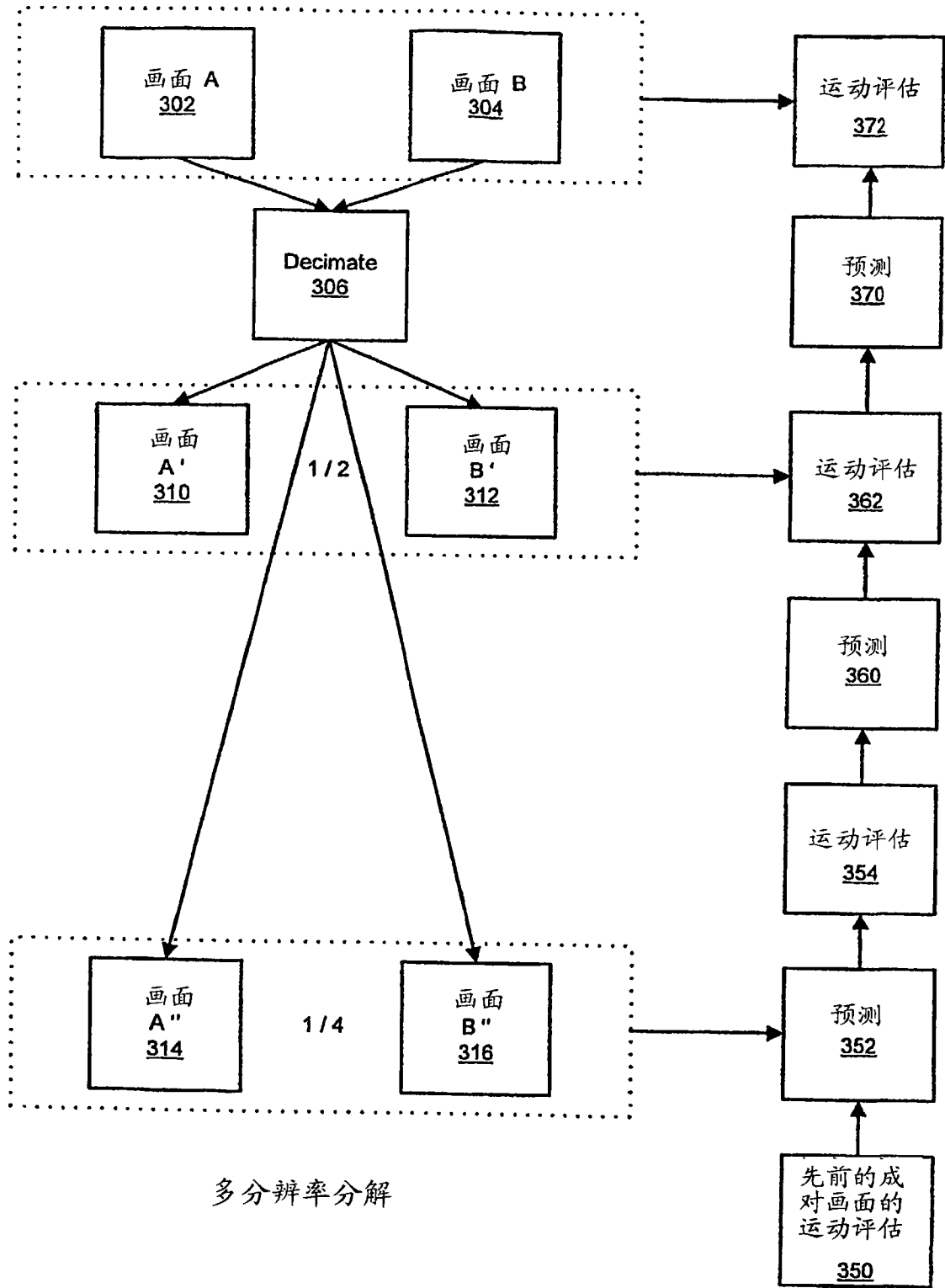


图 3

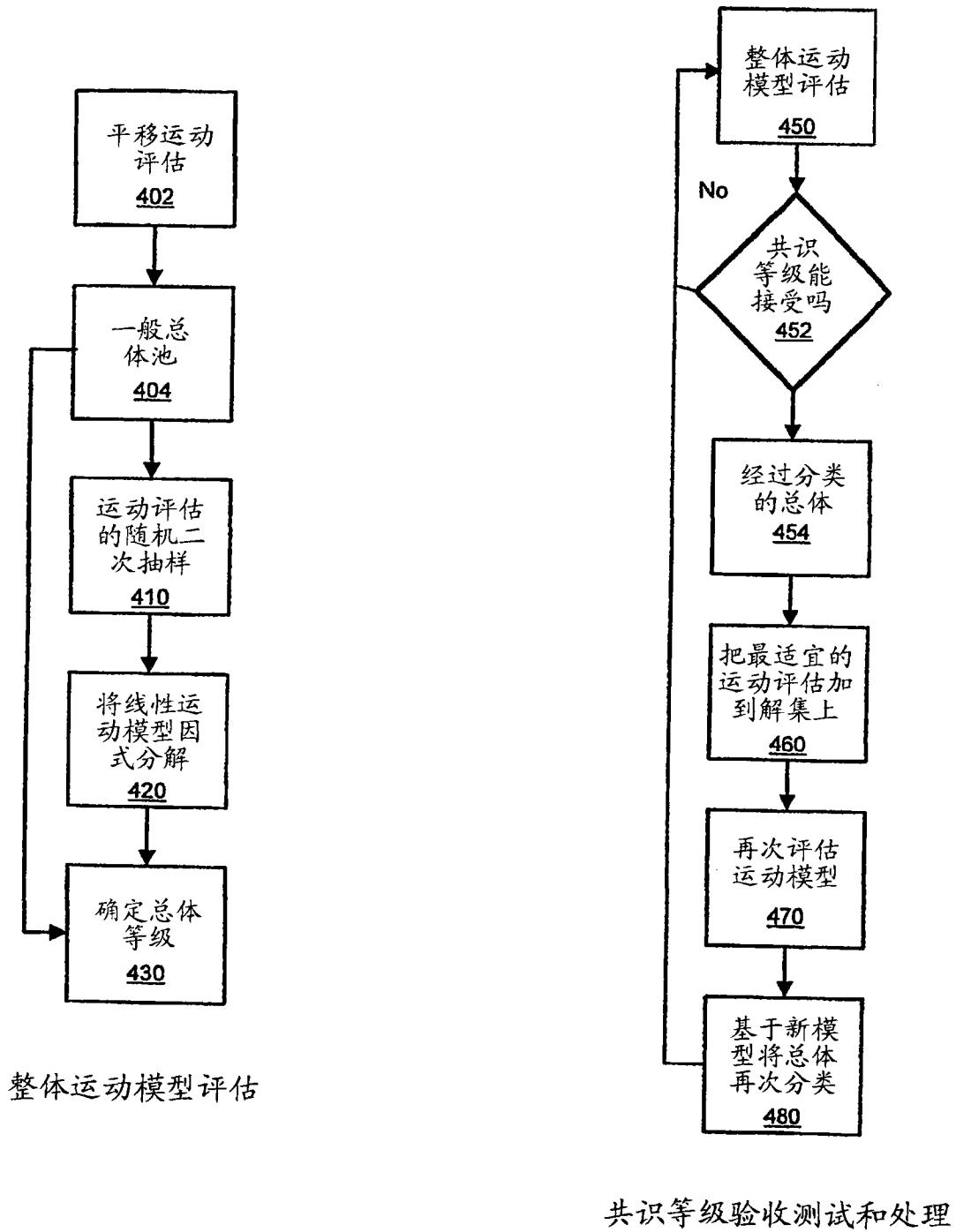


图 4

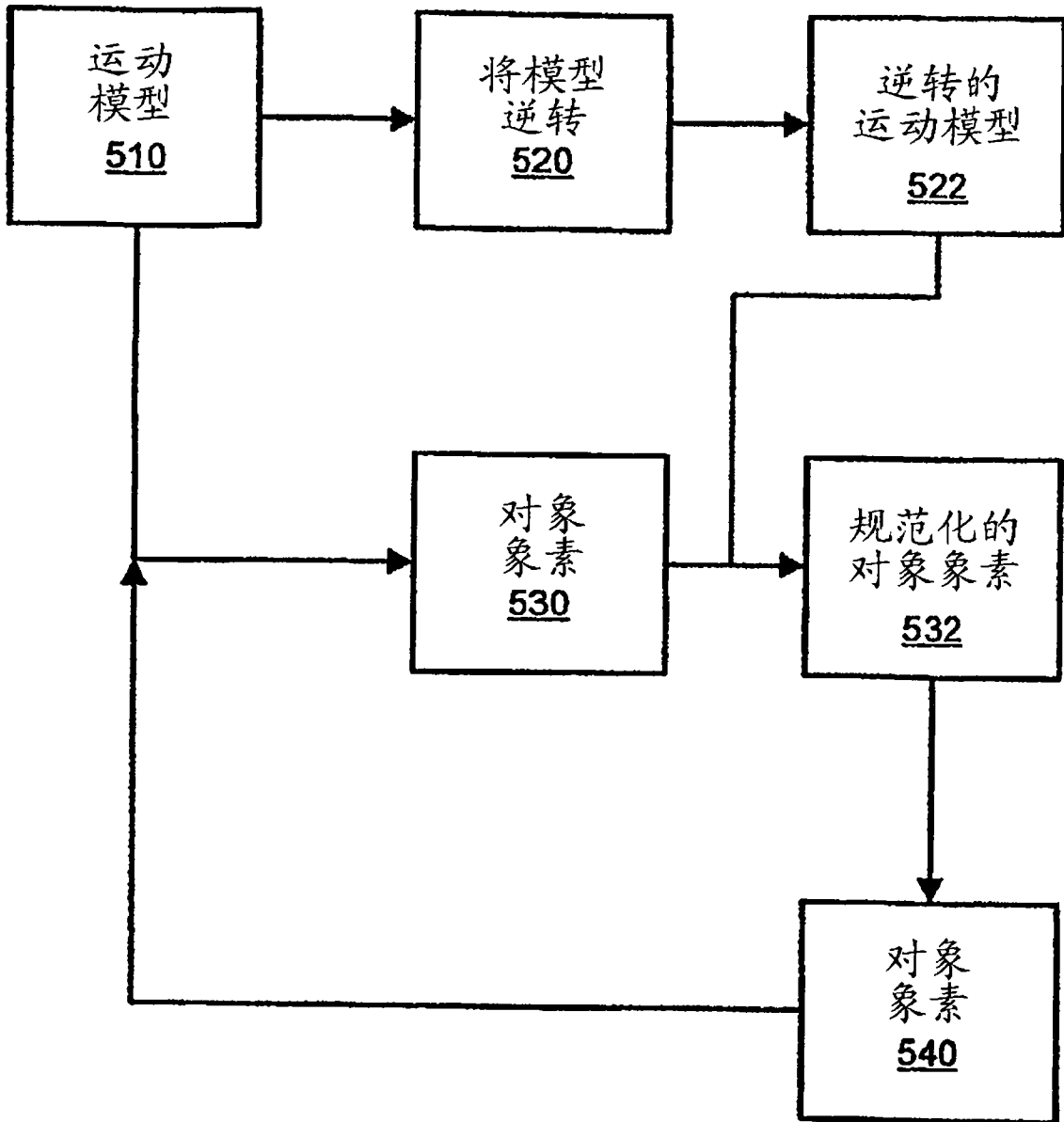
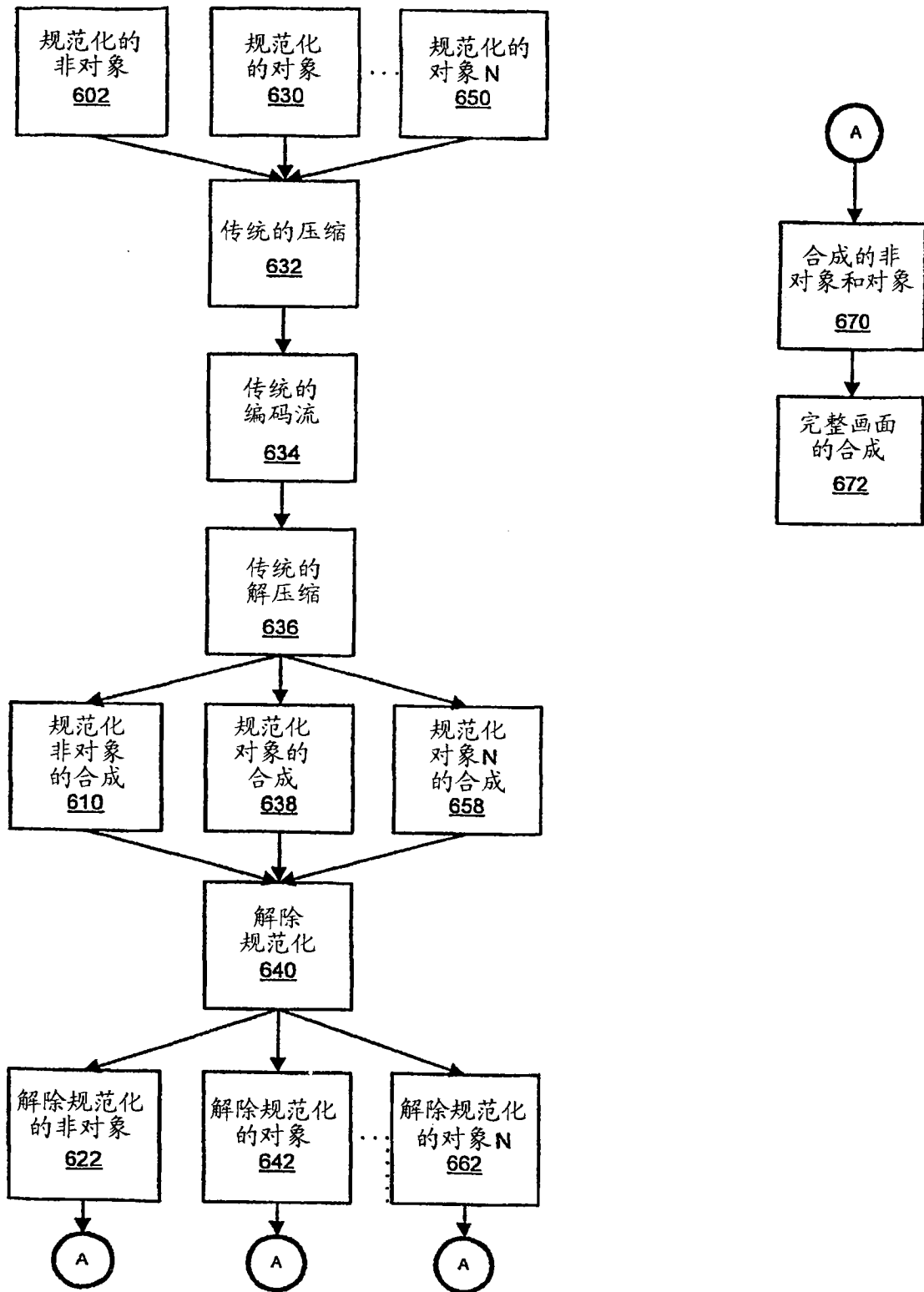


图 5



混合的

图 6