



(12)发明专利

(10)授权公告号 CN 104378635 B

(45)授权公告日 2017. 12. 05

(21)申请号 201410589327.7

(22)申请日 2014.10.28

(65)同一申请的已公布的文献号
申请公布号 CN 104378635 A

(43)申请公布日 2015.02.25

(73)专利权人 西交利物浦大学
地址 215123 江苏省苏州市工业园区独墅湖高等教育区仁爱路111号

(72)发明人 罗天明 程飞

(74)专利代理机构 苏州创元专利商标事务所有
限公司 32103

代理人 范晴

(51)Int.Cl.

H04N 19/167(2014.01)

H04N 5/232(2006.01)

(56)对比文件

CN 101068308 A,2007.11.07,
CN 101511000 A,2009.08.19,
CN 101309410 A,2008.11.19,
CN 102160398 A,2011.08.17,
US 2010/0184011 A1,2010.07.22,

审查员 朱一雷

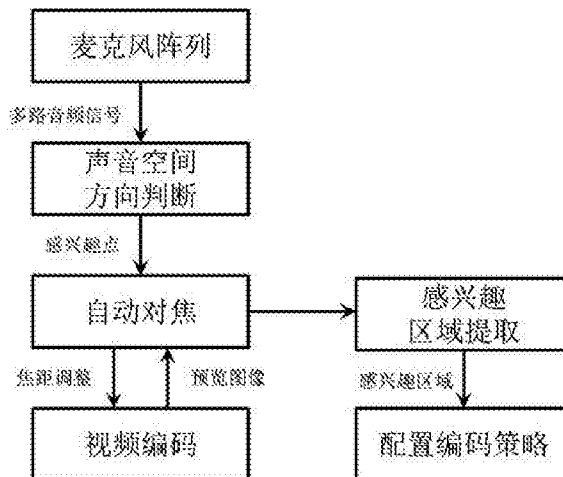
权利要求书1页 说明书4页 附图1页

(54)发明名称

基于麦克风阵列辅助的视频感兴趣区域的编码方法

(57)摘要

本发明公开了一种基于麦克风阵列辅助的视频感兴趣区域的编码方法,在拍摄视频中,通过麦克风阵列(两个或多个麦克风)获取视频拍摄中声音的空间方向,然后通过该方向确定视频拍摄中的自动对焦点,通过对焦点确定视频帧的感兴趣区域(Region of interest,ROI),并且最终通过不同的编码策略对视频帧进行编码的方法。该方法可以提升用户拍摄体验,动态的选取对焦点和感兴趣区域,并最终通过重新分配码流来提升视频主观观看体验。



1. 一种基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,包括以下步骤:

- (1) 采样麦克风阵列中各麦克风的音频信号;
- (2) 根据各音频信号,结合各麦克风的位置,计算出声音方向;
- (3) 根据声音方向结合相机参数,获取感兴趣点坐标 (x, y) , 并进行自动对焦;所述自动对焦方法为:

将原始视频帧的全部内容R,或者选取合焦点附近的一部分区域表示为灰度图P;

设计范围为f的低通滤波器对P进行滤波:

$$F(p_{x,y}) = \sum P \left(x - \frac{f}{2}, x + \frac{f}{2}, y - \frac{f}{2}, y + \frac{f}{2} \right) \cdot H ;$$

其中,H是 高斯低通滤波器,即: $H(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2}\right), (x, y \in [-f/2, f/2]), p_{x,y}$

为P中的一个像素点,x,y为像素的坐标,f是低通滤波器处理的范围,exp是以e的指数;

求原视频帧与滤波后的新视频帧的差距,使用差的绝对值或方差或者标准差来判断原视频帧和新视频帧之间的差距;

将差距最大的区域作为合焦区域的边界,以及分析合焦区域中的纹理细节;

获得合焦区域边界所围成的闭合区域C,按照P与R的截取关系将C映射到R中,记录为I,即为感兴趣区域;

- (4) 或者完成对焦后,判断对焦点周围的清晰区域,将该区域作为感兴趣区域;
- (5) 对感兴趣区域和非感兴趣区域设定不同编码参数,并对视频进行编码。

2. 根据权利要求1所述的基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,所述麦克风阵列的麦克风数量为2个或者2个以上。

3. 根据权利要求1所述的基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,所述声音方向的计算方法为:通过声音到达不同麦克风的时间差,求出双曲面渐进锥面,在距离麦克风一段距离之后,认定声源在锥面上或者通过声音到达不同麦克风之间的能量变化,结合麦克风之间的距离,计算出声源的方向。

4. 根据权利要求3所述的基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,所述麦克风为3个以上,通过声音到达不同麦克风的时间差,计算出声源所在的多个锥面,声源方向为锥面相交处。

5. 根据权利要求1所述的基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,所述对焦完成之后,再次通过拍摄设备中自动对焦系统,判断对焦点或对焦区域周围更大范围内的清晰图像区域,该区域为感兴趣区域。

6. 根据权利要求1-5任一项所述的基于麦克风阵列辅助的视频感兴趣区域的编码方法,其特征在于,对感兴趣区域采用容错策略以及保护策略进行保护。

基于麦克风阵列辅助的视频感兴趣区域的编码方法

技术领域

[0001] 本发明涉及一种基于感兴趣区域的视频编码方法,具体涉及一种基于麦克风阵列辅助的视频感兴趣区域的提取及编码的方法。

背景技术

[0002] 目前,高清晰度视频格式(High Definition,HD)越来越多的被用于各种视频录制与实时视频通信中。然而,存储和传输HD视频码流为存储设备和网络带宽带来了极大的挑战。尤其是对于便携式录像设备,例如智能手机和DV机,由于其存储空间受限,限制了HD视频的广泛使用。一种有效的解决方案是通过划分视频中感兴趣区域和非感兴趣区域,对不同区域使用不同编码策略,使用更多的码率编码感兴趣区域,反之亦然。

[0003] 在传统的基于感兴趣区域的视频编码方法中,大多数采用人脸识别、运动识别或者通过数学模型预测图像中的感兴趣区域,然后对不同区域采用不同的量化参数进行编码。然而,这些方案往往只是关注视频图像本身,而在录像的过程中,声音也是非常重要的一部分。通常,人们在观看视频的时候,会更关注发出声音的人或者物体,例如在多人视频会议场景中,观看者会更关注正在说话的人;在赛车比赛中,观看者会更加关注呼啸驶过的赛车。因此,发出声音的人和物体可以作为感兴趣点。

[0004] 大多数的视频录像场景都是同步采集声音的,而音频信号一般都是独立于视频信号进行编码。多数情况下,视频编码方案的研究中并不考虑音频信号本身提供的信息的。然而,对于同一个场景来说,音频和视频是以不同的形式描述同一个事物,中间必然有很多关联。尤其是双麦克风系统组成的立体声采集,不仅可以采集到立体声音频,同样可以用立体声信号来判断声音的方向。而麦克风阵列可以同时采样多个不同位置上的麦克风信号,然后通过幅值和相位等信息的差异,可以判断声音的空间方向。除此之外,通过声音方向的变换也可以作为视频编码中运动估计与运动补偿的辅助。

[0005] 智能手机是目前一种非常流行的视频拍摄设备,而相当多的智能手机已经具备了两个或更多的麦克风,通常一个用于受话器,另一个位于手机背面,用于通话降噪。由于这两个麦克风安装位置不同,因此可以用来在一定程度上判断声音的方向,用于辅助检测感兴趣区域。

发明内容

[0006] 本发明目的是:提供一种基于麦克风阵列辅助的视频感兴趣区域的编码方法,该方法可以提升用户拍摄体验,动态的选取对焦点和感兴趣区域,并最终通过重新分配码流来提升视频主观观看体验。

[0007] 本发明的技术方案是:一种基于麦克风阵列辅助的视频感兴趣区域的编码方法,包括以下步骤:

[0008] (1) 采样麦克风阵列中各麦克风的音频信号;

[0009] (2) 根据各音频信号,结合各麦克风的位置,计算出声音方向;

[0010] (3) 根据声音方向结合相机参数,获取感兴趣点坐标 (x, y) , 并进行自动对焦;

[0011] (4) 完成对焦后,判断对焦点周围的清晰区域,将该区域作为感兴趣区域;

[0012] (5) 对感兴趣区域和非感兴趣区域设定不同编码参数,并对视频进行编码。

[0013] 进一步的,所述麦克风阵列的麦克风数量为2个或者2个以上。

[0014] 进一步的,所述声音方向的计算方法为:通过声音到达不同麦克风的时间差,求出双曲面渐进锥面,在距离麦克风一段距离之后,近似认为声源在锥面上或者通过声音到达不同麦克风之间的能量变化,结合麦克风之间的距离,计算出声源的方向。

[0015] 进一步的,所述麦克风为3个以上,通过声音到达不同麦克风的时间差,计算出声源所在的多个锥面,声源方向为锥面相交处。

[0016] 进一步的,所述对焦完成之后,再次通过拍摄设备中自动对焦系统,判断对焦点或对焦区域周围更大范围内的清晰图像区域,该区域为感兴趣区域。

[0017] 进一步的,所述自动对焦方法为:将原始视频帧的全部内容,或者选取合焦点附近的一部分区域表示为灰度图P;

[0018] 设计范围为f的低通滤波器对P进行滤波:

$$[0019] \quad F(p_{x,y}) = \sum P \left(x - \frac{f}{2}, x + \frac{f}{2}, y - \frac{f}{2}, y + \frac{f}{2} \right) \cdot H$$

[0020] 其中,H是高斯低通滤波器,即: $H(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2}\right), (x, y \in [-f/2, f/2])$,

$p_{x,y}$ 为P中的一个像素点,x,y为像素的坐标,f是低通滤波器处理的范围,exp是以e的指数;

[0021] 求原视频帧与滤波后的新视频帧的差距 $E = |P - F(P)|$,使用差的绝对值或方差或者标准差来判断原视频帧和新视频帧之间的差距;

[0022] 差距最大的区域为合焦区域的边界,以及合焦区域中的纹理细节;

[0023] 获得合焦区域边界所围成的闭合区域C,按照P与R的截取关系将C映射到R中,记录为I,即为感兴趣区域。

[0024] 进一步的,对感兴趣区域采用容错策略以及保护策略进行保护。

[0025] 本发明的优点是:

[0026] 1.通过麦克风阵列(两个或多个麦克风)获取视频拍摄中声音的空间方向,然后通过该方向确定视频拍摄中的自动对焦点,完成自动对焦。可以使拍摄者更轻松地完成拍摄;

[0027] 2.通过对焦点来确定视频帧的感兴趣点,进而通过感兴趣点查找视频帧的感兴趣区域;

[0028] 3.对于感兴趣区域和非感兴趣区域使用不同的编码策略进行编码压缩,对于感兴趣区域,使用更好的码率来压缩,以实现更好的效果,相反,对于经常被观看者忽视的非感兴趣区域,使用低码率来压缩。对于整帧视频来说,可以在总码率不变的基础上,实现更好的主观观看效果;

[0029] 4.该感兴趣区域不仅可以用于视频编码压缩,也可以用于其他应用,例如图像分割与模式识别等领域。

附图说明

[0030] 下面结合附图及实施例对本发明作进一步描述:

[0031] 图1为本发明基于麦克风阵列辅助的视频感兴趣区域的编码方法的流程图；

[0032] 图2为本发明基于麦克风阵列辅助的视频感兴趣区域的编码方法的智能手机背面麦克风阵列示意图。

[0033] 其中,1、通话麦克风,2、第一降噪麦克风,3、摄像头,4、第二降噪麦克风。

具体实施方式

[0034] 为使本发明的目的、技术方案和优点更加清楚明了,下面结合具体实施方式并参照附图,对本发明进一步详细说明。应该理解,这些描述只是示例性的,而并非要限制本发明的范围。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本发明的概念。

[0035] 实施例:

[0036] 一种基于麦克风阵列辅助的视频感兴趣区域提取及编码的方法,其特征在于所述方法中,硬件方面:在传统的拍摄设备中需要麦克风阵列,即两个及以上的麦克风的支持;在软件方面:需要通过声音方向检测算法获取声音的空间方向,然后通过自动对焦系统来获取感兴趣区域,或者采用相关算法获取感兴趣区域,然后通过不同的编码策略来编码感兴趣区域和非感兴趣区域。

[0037] 硬件方面:该方法在智能手机中的应用,如图2所示。在该手机硬件上,需要三个麦克风。其中通话麦克风1设置在壳体的下端,第一降噪麦克风2和第二降噪麦克风4设置在摄像头3的两端用于通话降噪,这是一般智能手机的三个麦克风的。在拍摄视频中,三个麦克风同时采样音频信号,一方面用于立体声采样,另一方面用于对声源方向的定位。

[0038] 如图1所示,该方法的具体实施步骤如下:

[0039] (1) 采样麦克风阵列中各个麦克风的音频信号,需要确保同步采样各麦克风的音频信号;

[0040] (2) 根据各个音频信号,结合各麦克风的位置,计算出声音方向;

[0041] (3) 根据声音方向结合相机参数,获取感兴趣点坐标(x,y),并进行自动对焦;

[0042] (4) 完成对焦后,判断对焦点周围的清晰区域,将该区域作为感兴趣区域;

[0043] (5) 对感兴趣区域和非感兴趣区域设定不同编码参数,并对视频进行编码。

[0044] 麦克风阵列需要由两个及以上的麦克风组成。如果麦克风阵列由三个及以上的麦克风组成,则可以检测声音的空间方向;如果麦克风阵列由两个麦克风组成,则只可以检测声音的方向。

[0045] 步骤(2)中,可以有多种方法来计算声音方向:

[0046] a) 到达时间差(TDOA)方法:可以通过声音到达不同麦克风的时间差,求出双曲面渐进锥面,在距离麦克风一段距离之后,可以近似认为声源在上述锥面之上;

[0047] b) 双耳音级差(Interaural Level Difference)方法:通过声音到达不同麦克风之间的能量变化,结合麦克风之间的距离,可以计算出声源的方向。

[0048] 如果同时存在3个及以上的麦克风,则可以计算出声源所在的多个锥面,锥面相交出则可以认为是声源方向。

[0049] 步骤(3)中,主要的相机参数为传感器尺寸,传感器分辨率。通过上述参数,则可以计算出声源方向在传感器上对应的像素坐标,从而进行对焦。

[0050] 自动对焦方案有：

[0051] a) 直方图对焦(软件对焦)：根据对焦区域图像的直方图更加均匀分布的特点，调整镜头马达，进行自动对焦。

[0052] b) 相位检测对焦(硬件对焦)：通过在图像传感器或者对焦器安装用于对焦的传感器硬件，常用于单反相机中，对焦速度快于软件对焦方法，但是支持的对焦点一般比软件对焦少。

[0053] 步骤(4)中，对焦完成之后，再次通过拍摄设备中自动对焦系统，判断对焦点或对焦区域周围更大范围内的清晰图像区域，该区域即可以作为感兴趣区域ROI使用。

[0054] 所得到的ROI区域可以是不规则形状，ROI的最大尺寸可以通过参数限定，以免将ROI区域范围识别的过大。

[0055] 如果所使用拍摄设备的自动对焦系统不包括硬件对焦系统，可以考虑使用如下软件算法实现：

[0056] a) 将原始视频帧的全部内容，或者选取合焦点附近的一部分区域表示为P(灰度图)；

[0057] b) 设计范围为f的低通滤波器对P进行滤波：

$$[0058] \quad F(p_{x,y}) = \sum_{\left(x-\frac{f}{2}, x+\frac{f}{2}, y-\frac{f}{2}, y+\frac{f}{2}\right)} P \cdot H ;$$

[0059] 其中，H可以是高斯低通滤波器，即： $H(x,y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2+y^2}{2}\right), (x,y \in [-f/2, f/2])$ ；

$p_{x,y}$ 为P中的一个像素点。

[0060] c) 求原视频帧与滤波后的新视频帧的差距 $E = |P - F(P)|$ ，使用差的绝对值或方差或者标准差来判断原视频帧和新视频帧之间的差距；

[0061] d) 差距最大的区域为合焦区域的边界，以及分析合焦区域中的纹理细节；

[0062] e) 获得合焦区域边界所围成的闭合区域C，按照矩阵P和矩阵R的关系将C映射到矩阵R中，记录为I，即所述感兴趣区域。

[0063] 可以影响到视频编码质量的参数主要是QP(Quantization Parameter, 量化参数)，但不局限于今后新的视频编码算法中会使用其他参数控制视频编码质量。

[0064] 除了可以对ROI和非ROI区域使用不同的编码质量之外，也可以将ROI适用于网络传输中的容错策略以及保护策略，重点对ROI区域进行保护。

[0065] 应当理解的是，本发明的上述具体实施方式仅仅用于示例性说明或解释本发明的原理，而不构成对本发明的限制。因此，在不偏离本发明的精神和范围的情况下所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。此外，本发明所附权利要求旨在涵盖落入所附权利要求范围和边界、或者这种范围和边界的等同形式内的全部变化和修改例。

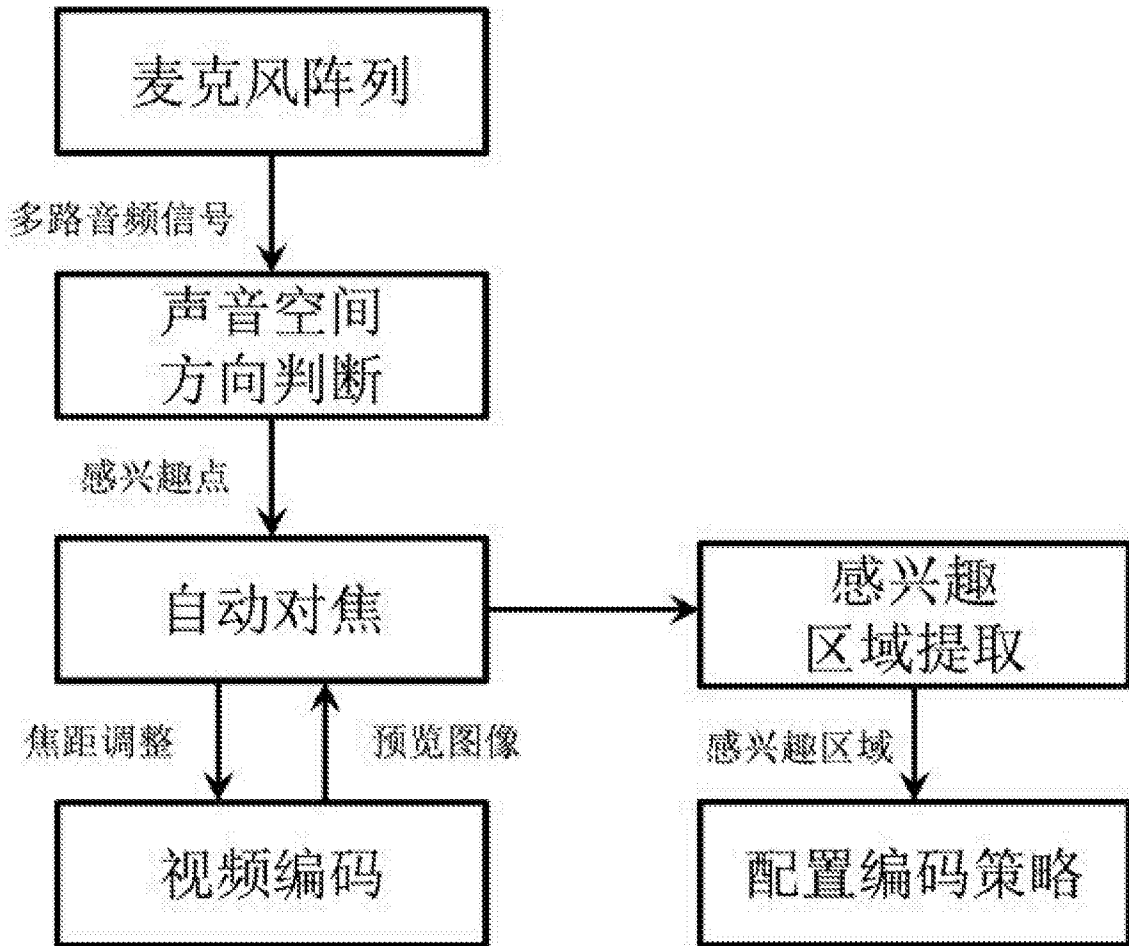


图1

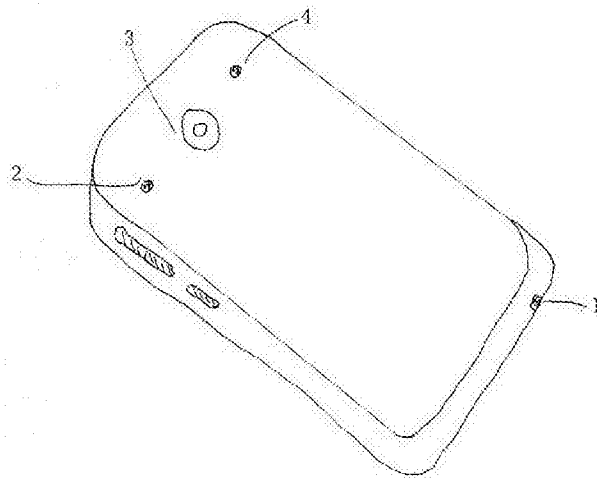


图2