



(12) 发明专利

(10) 授权公告号 CN 1711551 B

(45) 授权公告日 2010.04.28

(21) 申请号 200380103136.5

(22) 申请日 2003.11.07

(30) 优先权数据

10/293,976 2002.11.12 US

(85) PCT申请进入国家阶段日

2005.05.12

(86) PCT申请的申请数据

PCT/IB2003/005009 2003.11.07

(87) PCT申请的公布数据

W02004/044830 EN 2004.05.27

(73) 专利权人 诺基亚有限公司

地址 芬兰埃斯波

(72) 发明人 M·K·B·H·米勒德 F·策比尔

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 吴立明

(51) Int. Cl.

G06K 9/36 (2006.01)

G06K 9/00 (2006.01)

H04B 1/66 (2006.01)

H04N 5/225 (2006.01)

(56) 对比文件

US 6418166 B1, 2002.07.09, 全文.

审查员 杜轶

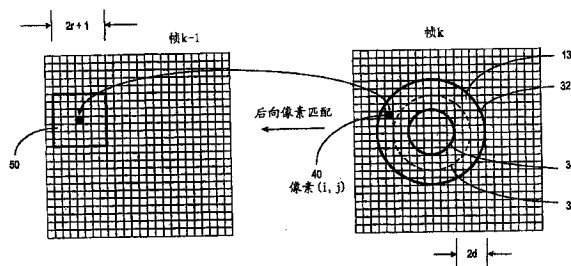
权利要求书 3 页 说明书 8 页 附图 5 页

(54) 发明名称

用于基于小波的视频编码的感兴趣区域跟踪方法和装置

(57) 摘要

一种用于跟踪图像帧序列中的感兴趣区域的方法和装置,其中将在先帧的目标区域的边界投影到当前帧上,以致可以建立当前帧中的搜索区。针对当前帧中的搜索区中的每个像素,在前一帧中建立搜索窗,以在该搜索窗中找出匹配像素。如果该匹配像素处于在前帧的 ROI 中,则将当前帧中的相应像素初步认定为当前帧的 ROI 内的像素。这一后向匹配是使用小波域中的低通子频带进行的。然后使用高频子频带中的边缘检测细化当前帧中的初步 ROI。



1. 一种在图像帧序列中的在先图像帧的目标区域的基础上跟踪图像帧中的目标区域的方法,所述图像帧序列中的每一个帧包括多个像素,所述方法的特征在于:

根据所述在先帧中的目标区域的至少一部分确定所述图像帧中的搜索区,所述搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

对于搜索区中的第一像素:

确定所述在先帧中的另一个搜索区,所述另一个搜索区包括在先帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

在第二像素值当中找出与所述第一像素的第一像素值之间的匹配,以定位参考第二像素;和

根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定所述图像帧中的目标区域。

2. 按照权利要求 1 所述的方法,其特征不在于,第二像素的区域状态代表所述第二像素是否位于所述在先帧中的目标区域内。

3. 按照权利要求 1 或 2 所述的方法,其特征不在于,所述在先帧中的目标区域的所述至少一部分具有用于定义所述图像帧中的相应轮廓的轮廓,并且其中第一像素包括与所述图像帧中的相应轮廓相邻的像素。

4. 按照权利要求 1 或 2 所述的方法,其中所述图像帧的目标区域包括边界,并且所述图像帧中的所述多个像素包括与该边界相邻的多个第三像素,每个第三像素具有至少一个相应的第三像素值,所述方法的特征此外还在于

确定第三像素的边缘类型,以根据第三像素的边缘类型修改所述图像帧中的目标区域。

5. 按照权利要求 4 所述的方法,其特征不在于,边缘类型是根据小波系数的亮度分量确定的。

6. 按照权利要求 1 或 2 所述的方法,其特征不在于,第一和第二像素值代表小波系数。

7. 按照权利要求 1 或 2 所述的方法,其特征不在于,第一和第二像素值代表低子频带中的小波系数。

8. 按照权利要求 1 或 2 所述的方法,其特征不在于,第一和第二像素值代表低子频带中的小波系数的色度分量的至少之一。

9. 按照权利要求 4 所述的方法,其特征不在于,第三像素值代表小波系数。

10. 按照权利要求 4 所述的方法,其特征不在于,第三像素值代表高子频带中的小波系数。

11. 按照权利要求 4 所述的方法,其特征不在于,第三像素值代表高子频带中的小波系数的亮度分量。

12. 一种跟踪设备,用于在图像帧序列中的在先图像帧的目标区域的基础上跟踪图像帧中的目标区域,所述图像帧序列中的每一个帧包括多个像素,所述跟踪设备包括:

用于根据所述在先帧中的目标区域的至少一部分确定所述图像帧中的搜索区的装置,所述搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

对于搜索区中的第一像素,还包括:

用于确定所述在先帧中的另一个搜索区的装置,所述另一个搜索区包括在先帧中的多

个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

用于在第二像素值当中找出与所述第一像素的第一像素值之间的匹配以定位参考第二像素的装置;和

用于根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态以根据所述至少一个第一像素的区域状态确定所述图像帧中的目标区域的装置。

13. 按照权利要求 12 所述的跟踪设备,其特征在于,第一和第二像素值代表低子频带中的小波系数的色度分量的至少之一。

14. 按照权利要求 12 或 13 所述的跟踪设备,其中所述图像帧的目标区域包括边界,并且所述图像帧中的所述多个像素包括与该边界相邻的多个第三像素,每个第三像素具有至少一个相应的第三像素值,所述跟踪设备还包括:

用于确定第三像素的边缘类型以根据第三像素的边缘类型修改所述图像帧中的目标区域的装置。

15. 按照权利要求 14 所述的跟踪设备,其特征在于,第三像素值代表高子频带中的小波系数的亮度分量。

16. 按照权利要求 12 或 13 所述的跟踪设备,其特征在于,第二像素的区域状态代表所述第二像素是否位于所述在先帧中的目标区域内。

17. 一种图像编码器,用于对包括至少一个第一图像帧和在先第二图像帧的图像帧序列进行编码,第一和第二图像帧中的每一个均包括多个像素,第二图像帧具有目标区域,所述图像编码器具有:

用于将各个图像帧分解成多个子频带分量的装置;和

用于将子频带分量编码成码流的装置,所述图像编码器还包括:

第一装置,响应于子频带分量,用于根据第二图像帧中的目标区域的至少一部分定义第一图像帧中的搜索区,搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

第二装置,响应于第一像素,用于:

确定第二图像帧中的另一个搜索区,所述另一个搜索区包括第二图像帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

根据第二像素值当中与所述第一像素的第一像素值之间的匹配确定所述另一个搜索区中的参考第二像素;和

根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定第一图像帧中的目标区域。

18. 按照权利要求 17 所述的图像编码器,其特征在于,第一和第二像素值代表低子频带中的小波系数的色度分量的至少之一。

19. 按照权利要求 17 或 18 所述的图像编码器,其中第一图像帧的目标区域包括边界,并且第一图像帧中的所述多个像素包括与该边界相邻的多个第三像素,每个第三像素具有至少一个相应的第三像素值,所述图像编码器的特征此外还在于

用于确定第三像素的边缘类型以便根据第三像素的边缘类型修改所述图像帧中的目标区域的装置。

20. 按照权利要求 19 所述的图像编码器,其特征在于,第三像素值代表高子频带中的

小波系数的亮度分量。

21. 按照权利要求 17 或 18 所述的图像编码器,其特征在于,图像编码器适用于以高于所述第一图像帧中的另一个区域的视觉质量对第一图像帧中的所述目标区域进行编码。

22. 一种编解码系统,具有用于将图像帧序列编码为码流的编码器和用于根据该码流重构图像帧序列的解码器,其中图像帧序列包括至少一个第一图像帧和在先第二图像帧,第一和第二图像帧中的每一个均包括多个像素,第二图像帧具有目标区域,所述编码器具有:

用于将各个图像帧分解成多个子频带分量的装置;和

用于将子频带分量编码成码流的装置,所述编码器还包括:

第一装置,响应于子频带分量,用于根据第二图像帧中的目标区域的至少一部分定义第一图像帧中的搜索区,搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

第二装置,响应于第一像素,用于:

确定第二图像帧中的另一个搜索区,所述另一个搜索区包括第二图像帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

根据第二像素值当中与所述第一像素的第一像素值之间的匹配确定所述另一个搜索区中的参考第二像素;和

根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定第一图像帧中的目标区域。

用于基于小波的视频编码的感兴趣区域跟踪方法和装置

技术领域

[0001] 本发明总地来讲涉及基于子频带分解的图像编解码器,并且更加具体地讲,涉及贯穿一系列图像的运动物体的编码。

背景技术

[0002] 我们一般都知道,图像压缩在减小用于传送或存储的图像数据量方面是很有效的。具体来说,通过引入象 JPEG2000 这样的可伸缩图像编码格式,使得接收图像文件的一小段并且还要在接收端重构高质量图像成为可能。从图像中略去的部分通常包含描述存在于图像中的高频分量的信息,相当于人类视觉系统 (HVS) 不是很敏感的细节。

[0003] JPEG 指的是针对图像压缩的联合图像专家组。在 1988 年,这个委员会通过了它的第一个标准,称为 JPEG 基准,它基于离散余弦变换 (DCT) 和霍夫曼 (Huffman) 编码。在 2001 年, JPEG 委员会开发了一种新的压缩标准,名为 JPEG2000。这种新标准实现了低位速率操作,速率失真和主观图像质量性能优于现有的标准,而没有牺牲速率失真频谱中其它点上的性能。更重要的是, JPEG2000 能够实现由同样的码流表示提取压缩图像不同的分辨率和像素逼真度。它还提供了诸如感兴趣区域 (ROI) 编码和随机访问图像区之类的功能。这使得用户能够由图像的 JPEG2000 位流表示针对任何目标装置仅仅操作、存储或传送该图像的基本信息。

[0004] JPEG2000 是基于子频带分解的位平面编码器。它在变换阶段使用小波。将图像分解为多个拆分结果。每个拆分结果由代表低和 / 或高频率分量的子频带构成。然后从最重要的位平面开始,在位平面内对这些子频带中的采样进行编码。小波变换和位平面编码方案的采用实现了 JPEG2000 的可伸缩性特色。

[0005] 运动 JPEG 2000 是一种视频压缩方法,它基于使用 JPEG 2000 的帧内编码。按照运动 JPEG 2000,将视频序列中的帧编码为独立的图像,即,不进行图像之间的预测。这种编码方案提供了重要的功能,比如质量和分辨率的可伸缩性、对于误码的强健性和帧编辑。不过,与其它的标准相比,比如 MPEG-4 (使用运动向量估计来编码中间帧),在压缩能力方面,这种编码方案效果较差。另一方面,可以利用 JPEG2000 的特点来提高运动 JPEG 2000 的压缩能力,比如 ROI 编码。

[0006] ROI 编码是 JPEG 2000 中一种有用的功能。在对图像进行编码的时候,它能够比同一图像中的其它区域更多地在 ROI 中分配位。通过不均等地对图像的各部分进行编码,以致与不太重要的对象相比,重要的对象可以每像素得以赋予更多的位,获得了更好的视频序列视觉感受,这使得这一特点非常有用,尤其是在低数据速率应用中更为有益。要将视频序列中的重要对象编码为 ROI,对这些对象进行跟踪对于产生高质量的视频流非常重要。

[0007] ROI 跟踪是很多与视觉相关的应用中的重要要素,比如基于视觉的控制、人机交互、监视、农业自动化、医学成像和图像重构。视频序列中的 ROI 跟踪的主要困难是由于帧序列内的目标区域中的可能变化造成的。这些变化通常是由于姿势、形状变形、照度和目标的局部或全部遮挡而造成的。在设计跟踪算法的时候,应当将这些变化的缘由考虑进来,以

便保证强健性和可靠性。

[0008] 在现有技术中,大多数 ROI 跟踪方法依赖于多级处理策略,这种策略由三个连续的步骤组成。在第一步中,进行基于颜色的前景分段,并且将与目标区域颜色相近的区域屏蔽起来。第二步包括,利用分段图像的水平垂直求和投影,通过求所估计向量的最小值定位 ROI,或者采用基于目标区域的大小、形状、位置、纵横比和颜色一致性的试探法定位 ROI。最后一个阶段是细化和平滑区域边界。

[0009] Hager 等人(《采用几何和照度的参数模型的有效区域跟踪 (Efficient Region Tracking with Parametric Models of Geometry and Illumination)》, IEEE 模式分析和机器智能会报 (IEEE Transactions on Pattern Analysis and Machine Intelligence), 第 20 卷, 第 10 期, 1998 年 10 月)利用了线性模型来简化光流公式,并且利用了一组基本向量通过对目标的训练序列进行奇值分解来模拟照度的变化。Kruger 等人(《用于实时面部跟踪的小波子空间方法 (Wavelet Subspace Method for Real-time Face Tracking)》,图形识别会议纪要,第 23 届 DAGM 论坛 (Proc. Pattern Recognition, 23th DAGM Symposium), 德国慕尼黑, 2002 年)使用大量的小波函数来构建代表目标特征的小波网络,然后根据小波系数之间的欧几里得距离跟踪 ROI。

[0010] Kundu(美国专利 US5974192)公开了一种根据纹理、起支配作用的边缘像素以及强度和颜色匹配图像序列中的块的方法。Ponticos(美国专利 US6035067)公开了一种使用“势阱”利用预先定义的用于对图像区域进行分类的标准来防止前景对象的大运动的方法。Poggio 等人(美国专利 US6421463)公开了一种在图像中搜索对象的方法,其中必须使用小波系数模板对搜索系统进行训练。

[0011] 现有方法的主要缺点是运算复杂。而且,它们中的一部分需要针对同样类型、颜色或形状的 ROI 对跟踪算法进行训练。

[0012] 这样,提供具有低运算复杂性和适于运动 JPEG 2000 编解码器的存储成本的用于任意形状和类型的 ROI 跟踪的方法和装置是颇有益处且众望所归的。

发明内容

[0013] 本发明的主要目的是提供一种用于跟踪图像序列中的对象的至少一部分的方法和装置,其中将这些图像编码为独立的图像,并且跟踪不需要模板和预先训练过的对象。这一目的可以通过将通过色度空间中的像素匹配进行的低频带中的边缘检测与通过使用亮度分量进行的边缘分析在高频带中进行的检测到的边缘的细化相结合而实现的。

[0014] 这样,按照本发明的第一个方面,提供了一种在图像帧序列中的在先图像帧的目标区域的基础上跟踪图像帧中的目标区域的方法,所述图像帧序列中的每一个帧包括多个像素。所述方法的特征在于

[0015] 根据所述在先帧中的目标区域的至少一部分确定所述图像帧中的搜索区,所述搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

[0016] 针对搜索区中的第一像素:

[0017] 确定所述在先帧中的另一个搜索区,所述另一个搜索区包括在前帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

[0018] 在第二像素值当中找出与所述第一像素的第一像素值之间的匹配,以定位参考第

二像素 ;和

[0019] 根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定所述图像帧中的目标区域。

[0020] 第二像素的区域状态代表所述第二像素是否位于所述在前帧中的目标区域内。

[0021] 由此,在前帧中的目标区域的至少一部分具有用于定义所述图像帧中的相应轮廓的轮廓,并且第一像素包括与所述图像帧中的相应轮廓相邻的像素。

[0022] 所述图像帧的目标区域包括边界,并且所述图像帧中的所述多个像素包括与该边界相邻的多个第三像素,每个第三像素具有至少一个相应的第三像素值。所述方法的特征此外还在于

[0023] 确定第三像素的边缘类型,以根据第三像素的边缘类型修改所述图像帧中的目标区域。

[0024] 第三像素值代表高子频带中的小波系数的亮度分量。

[0025] 按照本发明的第二个方面,提供了一种计算机程序,用在具有用于将图像帧序列编码为码流的装置的图像编码器中,所述图像帧序列具有至少一个第一图像帧和在前第二图像帧,第一和第二图像帧中的每一个均具有多个像素,其中第二图像帧具有目标区域,所述计算机程序的特征在于:

[0026] 用于根据第二图像帧中的目标区域的至少一部分定义第一图像帧中的搜索区的代码,所述搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值 ;和

[0027] 用于针对搜索区中的第一像素进行下述操作的代码 :

[0028] 确定第二图像帧中的另一个搜索区,所述另一个搜索区包括第二图像帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态 ;

[0029] 根据第二像素值当中与所述第一像素的第一像素值之间的匹配确定所述另一个搜索区中的参考第二像素 ;和

[0030] 根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定第一图像帧中的目标区域。

[0031] 第二像素的区域状态代表所述第二像素是否位于所述在前帧中的目标区域内。

[0032] 由此,第一和第二像素值代表低子频带中的小波系数的色度分量的至少之一。

[0033] 有利地,第一图像帧的目标区域包括边界,并且第一图像帧中的所述多个像素包括与该边界相邻的多个第三像素,每个第三像素具有至少一个相应的第三像素值。所述计算机程序的特征此外还在于

[0034] 用于确定第三像素的边缘类型以便根据第三像素的边缘类型修改第一图像帧中的目标区域的代码。

[0035] 由此,第三像素值代表高子频带中的小波系数的亮度分量。

[0036] 按照本发明的第三个方面,提供了一种图像编码器,用于对包括至少一个第一图像帧和在先第二图像帧的图像帧序列进行编码,第一和第二图像帧中的每一个包括多个像素,第二图像帧具有目标区域,所述图像编码器具有 :

[0037] 用于将各个图像帧分解成多个子频带分量的装置 ;和

[0038] 用于将子频带分量编码成码流的装置。所述图像编码器的特征在于

[0039] 第一算法,响应于子频带分量,用于根据第二图像帧中的目标区域的至少一部分

定义第一图像帧中的搜索区,搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

[0040] 第二算法,响应于第一像素,用于:

[0041] 确定第二图像帧中的另一个搜索区,所述另一个搜索区包括第二图像帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

[0042] 根据第二像素值当中与所述第一像素的第一像素值之间的匹配确定所述另一个搜索区中的参考第二像素;和

[0043] 根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定第一图像帧中的目标区域。

[0044] 该图像编码器适用于以高于所述第一图像帧中的另一个区域的视觉质量对第一图像帧中的所述目标区域进行编码。

[0045] 按照本发明的第四个方面,提供了一种图像编码系统,具有用于将图像帧序列编码为码流的编码器和用于根据该码流重构图像帧序列的解码器,其中图像帧序列包括至少一个第一图像帧和在先第二图像帧,第一和第二图像帧中的每一个均包括多个像素,第二图像帧具有目标区域,所述编码器具有:

[0046] 用于将各个图像帧分解成多个子频带分量的装置;和

[0047] 用于将子频带分量编码成码流的装置。所述编码器的特征在于

[0048] 第一算法,响应于子频带分量,用于根据第二图像帧中的目标区域的至少一部分定义第一图像帧中的搜索区,搜索区包括多个第一像素,每个像素具有至少一个相应第一像素值;和

[0049] 第二算法,响应于第一像素,用于:

[0050] 确定第二图像帧中的另一个搜索区,所述另一个搜索区包括第二图像帧中的多个像素当中的多个第二像素,每个第二像素具有至少一个相应第二像素值和区域状态;

[0051] 根据第二像素值当中与所述第一像素的第一像素值之间的匹配确定所述另一个搜索区中的参考第二像素;和

[0052] 根据参考第二像素的区域状态确定所述第一像素中的至少一个的区域状态,以根据所述至少一个第一像素的区域状态确定第一图像帧中的目标区域。

[0053] 在结合附图 1 到 5 阅读说明书的情况下,本发明将会变得清晰明了。

附图说明

[0054] 附图 1 是说明将在先图像帧中的 ROI 的边界投影到当前图像帧中的示意性表示。

[0055] 附图 2 是说明当前图像帧中用于像素匹配的包围着投影得到的 ROI 边界的区域中的像素的示意性表示。

[0056] 附图 3a 是说明估计的当前帧中的 ROI 的示意性表示。

[0057] 附图 3b 是说明水平和垂直边缘检测的示意性表示。

[0058] 附图 4 是说明按照本发明的 ROI 搜索过程的流程图。

[0059] 附图 5 是说明能够进行按照本发明的 ROI 跟踪的图像编码系统的框图。

具体实施方式

[0060] 按照本发明,将 ROI 跟踪方法设计成针对序列通过运动 JPEG 2000 进行压缩。该方法由两个步骤构成,这两个步骤在小波域中的低分辨率子频带中进行。在第一个步骤中,通过根据在先帧中的相应目标区域进行后向像素匹配,进行当前帧中的目标区域的估计。使用低频子频带降低了运算复杂程度并且还在跟踪之中引进了稳定性。后向像素匹配估计针对运动采用模型的需要。优选地,第一步是使用用于跟踪的色度或色品 (U, V) 分量来进行的。与亮度不同,色度分量通常在整个视频序列内是稳定的,并且具有很小的变化。这使得跟踪算法对照度的变化更加强健,从而消除了对入射照度估计的训练序列的需要。在第二步中,使用亮度分量 Y 的高频子频段 (即,HL 和 LH) 来进行边缘检测,以细化所估计的当前帧中的目标区域。使用 Y 分量能够实现边缘的强健检测,即使当目标区域与同一颜色类型的另一个区域相接触时也是如此。

[0061] 附图 1 表示按照本发明的 ROI 搜索中的第一个步骤,该第一个步骤是为了在当前图像帧 k 中获得估计的 ROI 边界。如附图 1 中所示,将 ROI 20 在图像帧 (k-1) 中的轮廓投影到图像帧 k 上。帧 (k-1) 是在先帧。投影到帧 k 上的轮廓 30 表示为虚线圆。由外边界 32 和内边界 34 界定出来的像素带用作当前帧 k 中的搜索区 130。像素带 130 的宽度是 2d。如附图 2 所示,搜索区 130 表示 ROI 轮廓的运动空间,而 d 是任何位移向量的上边界。只有搜索区 130 内的像素或者“相关”像素才根据色度分量的低通小波系数进行后向匹配。应当注意,由圆表示的 ROI 20 的边界仅仅用于图解说明的目的。一般来说,ROI 的边界或轮廓的形状是不规则的,并且搜索区 130 由一组包围着投影得到的 ROI 轮廓的“相关”像素占据。

[0062] 后向像素匹配包括将每个“相关”像素映射为在先帧 (k-1) 中的另一个像素。映射判据是色度分量 (U, V) 的低通小波系数之间的欧几里得距离。通过小窗口内的局部搜索使得这一度量标准最小。如附图 2 所示,使用帧 k 中的像素 40 来找出帧 (k-1) 中围绕着相应像素的搜索窗 50 内的匹配。搜索窗的面积为 $(2r+1)^2$,其中 r 是大于或等于 d 的整数。如果将像素 40 表示为 $x_k(i, j)$,则搜索窗 50 包括一组像素 $x_{k-1}(p, q)$,其中 $(i-r) \leq p \leq (i+r)$, $(j-r) \leq q \leq (j+r)$ 。对于帧 k 中的搜索区 130 内的每个像素,进行后向搜索是为了在搜索窗 50 内找到帧 (k-1) 内的相应最佳像素 x_{opt} 。具体来说,如果 w_k 帧 k 的低通频带,而 w_{k-1} 是相同的帧 (k-1) 的低通频道,那么将最佳像素 x_{opt} 定义为

$$[0063] \quad x_{opt} = \min_{p, q} |w_k(i, j) - w_{k-1}(p, q)| \quad (1)$$

[0064] 同样,搜索窗 50 是一组可能的匹配,并且匹配是通过依据色度分量求 $w_k(i, j)$ 和 $w_{k-1}(p, q)$ 之间的欧几里得距离的最小值而实现的。忽略亮度分量 Y 使得该算法对照度的变化更加强健。应当注意,分量 $w_k(i, j)$ 和 $w_{k-1}(p, q)$ 与色度空间中的小波系数有关。如果我们将 U 空间和 V 空间中的小波系数记为 u 和 v,则 w 既可以是 u 也可以是 v,或者优选地, w 是具有值 (u, v) 的 2D 向量。

[0065] ROI 搜索中的下一步骤是得到各个像素 $x_k(i, j)$ 的区域状态-属于图像的背景 (非 ROI) 还是属于 ROI。像素 $x_k(i, j)$ 的区域状态取决于相应像素 x_{opt} 的区域状态,后者是已知的。这样,如果帧 (k-1) 中的最佳像素在 ROI 中,则将相应像素 $x_k(i, j)$ 也设定为在 ROI 中。类似地,如果帧 (k-1) 的最佳像素在背景中,则将相应像素 $x_k(i, j)$ 也设定为是背景像素。应当注意,必须将序列中的第一个帧的 ROI 提供给搜索算法,从而可以据此标注第一帧 ROI 内和外的每个像素的区域状态。

[0066] 搜索区 130 内具有背景状态的那些像素 $x_k(i, j)$ 将用来形成用于定义估计的帧 k 中的 ROI 的屏蔽。附图 3a 是在帧 k 中经过屏蔽之后估计得到的 ROI 的示意性表示。

[0067] 优选地,使用针对最小欧几里得距离规定的上边界来判定匹配是否成功。所规定的上边界例如可以根据面积密度的平均值来设定。如果公式 1 中的 x_{opt} 大于所规定的上边界,则将该 x_{opt} 相应的像素 $x_k(i, j)$ 的区域状态设定为背景。

[0068] 下一个步骤是调谐过程,这一步骤用于在所估计的 ROI 的边界像素的基础上细化帧 k 中的 ROI。调谐过程的目的是通过这些边界像素周围的局部搜索检测任何边缘,以确定帧 k 中最终的 ROI。按照本发明,用于调谐过程的搜索域由亮度空间中的 HL 和 LH 子频带的系数构成,这分别实现了垂直和水平边缘的提取(附图 3b)。使用亮度分量 Y 能够实现边缘的强健检测,即使当目标区域同一颜色类型的另一个区域相接触时也是如此。具有较大量级的像素有可能是边缘像素。如果边缘像素恰巧是在背景中,那么在将其包含在 ROI 中之前应当核对其颜色系数。必须当心不要将不同颜色的像素包括到 ROI 的轮廓区域中。这样的像素可能会在对整个图像序列进行跟踪的时候导致误差的蔓延。

[0069] 算法

[0070] 在将帧 (k-1) 的 ROI 轮廓投影到帧 k 上之后,对帧 k 中的 ROI 的搜索是按照下述算法进行的:

[0071] 遍及帧 k 中的搜索区域内的所有“相关”像素 $x_k(i, j)$ 进行循环;

[0072] 为各个 $x_k(i, j)$ 在帧 (k-1) 中定义一个搜索窗;

[0073] 遍及所有窗口像素 $x_{k-1}(p, q)$ 进行循环;

[0074] 计算欧几里得距离 $|w_k(i, j) - w_{k-1}(p, q)|$;

[0075] 选择最小欧几里得距离,以确定 x_{opt} ;

[0076] 按照 x_{opt} 判定 $x_k(i, j)$ 的区域状态;

[0077] 根据 $x_k(i, j)$ 的区域状态建立估计的 ROI 边界;

[0078] 遍及估计得到的 ROI 的边界像素进行循环;

[0079] 区分边界类型;

[0080] 比较高频系数并且核对颜色系数,以选定边缘像素。

[0081] 应当注意,上述算法中使用的数学运算是减法和比较。不需要使用乘法。假设编码器已经求出了小波系数,该算法中的最高复杂性在于搜索循环。因为跟踪是以低分辨率进行的,并且目标对象一般来说仅占据图像的一小部分(压缩效率所要求的),所以循环的平均总数也是很小的。

[0082] 按照本发明,跟踪一系列图像帧中的一个图像帧内的目标区域的方法在流程图 500 中进行了概括,如附图 4 所示。

[0083] 如附图 4 所示,在步骤 510 中,将帧 (k-1) 中的 ROI 轮廓投影到帧 k 上。在步骤 512 中,建立包含帧 k 中的包围投影得到的 ROI 轮廓的多个像素的搜索区。在步骤 514 中,对于搜索区中的每个像素 $x_k(i, j)$,根据欧几里得距离确定帧 (k-1) 中的搜索窗内的最佳像素 $x_{k-1}(p, q)$ 。在步骤 516 中,根据相应的最佳像素确定 $x_k(i, j)$ 的区域状态。如果在步骤 518 中将搜索区中每个 $x_k(i, j)$ 的区域状态都确定为已判定的,那么在步骤 520 中根据 $x_k(i, j)$ 的区域状态建立帧 k 中估计的 ROI。在步骤 522 中,使用亮度空间中的高频子频带来检测估计得到的 ROI 边界周围的边缘像素,以便细化帧 k 中的 ROI。如果在步骤 524 中判

定当前帧 k 是序列中的最后一帧, 则处理继续进行到下一序列。否则, 该处理循环回到步骤 510, 以确定下一帧中的 ROI。

[0084] 附图 5 中给出了能够执行按照本发明的目标区域跟踪方法的图像编码系统。该图像编码系统 600 包括编码器 610 和解码器 630。如附图 5 中所示, 将由多个各自可分成三个颜色平面 (R, G, B) 的数字图像帧组成的图像序列 700 提供给转换器模块 612。转换器 612 将数字图像帧转换成 YUV 格式的一系列数字图像帧 710。使用前向小波变换模块 614 将各个图像帧 710 中的像素变换为各种不同子频带中的小波系数。将经过变换的图像数据 720 提供给多分辨率分配模块 616, 在这个模块 616 中将每像素更多的位分配给帧中的重要图像区域。最好, 将第一个帧中的各个重要图像区域的边界提供给 ROI 跟踪模块 618, 以使后续帧中的重要图像区域能够得以自动跟踪。ROI 跟踪模块 618 使用按照本发明的 ROI 跟踪算法来判定后续帧中的重要图像区域。将多分辨率图像数据 730 传送给用于编码和压缩的编码模块 620。可以对码流 740 进行实时传送, 或者对其进行存储, 以备将来使用。在接收端, 解码器 630 在码流 740 的基础上对输入图像序列 700 进行重构。如图所示, 使用解码模块 632 来根据码流 740 对经过变换的数据进行重构。由逆小波变换模块 634 将重构后的经变换数据 750 变换为 YUV 空间内的一系列数字图像帧 760。在由转换模块 636 进行了转换之后, 图像数据变成了三个颜色平面 (R, G, B) 内的一系列数字图像 770。

[0085] 仿真结果

[0086] 对三个不同的图像序列“MOBILE”、“FOREMAN”和“IRENE”使用了按照本发明的 ROI 跟踪方法。所有序列具有 CIF 大小 (公用中间格式) 和每秒 30 帧的速率。为了以较低的速率工作, 在“MOBILE”和“FOREMAN”序列中每个跟踪重复过程跳过了 3 个帧, 从而跟踪是以每秒 7.5 帧的速率进行的。在“IRENE”序列中, 保持了原始的帧频。

[0087] 在“MOBILE”序列中, ROI 是具有小的周围白色区域的红球。该球进行来来回回的旋转和平移运动, 并且与其它相近颜色的区域进行接触。

[0088] 在“FOREMAN”序列中, 所跟踪的人脸改变姿势和位置。在序列的末端, 出现场景的急剧变化并且目标区域得以完全关闭。

[0089] 在“IRENE”序列中, 选取了两个 ROI, 它们各自位于演讲者的两只手上。这两只手表现出了相对快速的位置、姿势和颜色变化。这两只手还与其它肤色区域相重叠, 比如脖子区域, 这使得跟踪更加困难。

[0090] 在仿真过程中, 将第一帧中的 ROI 提供给该算法。在每次重复过程中, 仅使用在先帧的小波变换。搜索是以第二分辨率 (LL 子频段) 进行的, 并且将上边界 d 选取为等于 6 个像素, 从而搜索窗的大小为 16×16 像素 ($r = 8$)。在“MOBILE”和“FOREMAN”序列中, 最小欧几里得距离的上边界为 3。在“IRENE”序列中, 没有设定上边界, 以便补偿 ROI 中颜色的变化。

[0091] 在“MOBILE”序列中, 主要的困难是存在很多颜色与目标相近的背景区域。这些区域中的一部分一开始并没有出现在场景中 (全部或部分包含)。在第一个帧中, 目标区域不包含整个球, 因为仅仅出现了球的一部分。仿真结果表明, 本跟踪算法能够在视频序列中自始至终保持对球的跟踪。不过, 随着重复过程的进行, 将背景区域的一部分包含到了 ROI 中。所包含的区域在颜色和位置上非常接近于目标区域, 并且它们一开始并不在该场景中。

[0092] 在“FOREMAN”序列中, 本算法表现出了跟踪面部的能力, 尽管在序列的某些周期期

间面部姿态以相对较快的运动发生改变。随着重复过程的进行,本算法将脖子的一部分包含在了 ROI 中。所包含的区域没有分隔边缘地与脸部连在一起。由于颜色分量的相近造成墙壁的一部分也包含在 ROI 中。仿真结果还表现出了本算法检测目标区域从场景中完全遮挡的能力。

[0093] 在“IRENE”序列中,以每秒 30 帧的速率对演讲者的两只手都进行了跟踪。由于位置、姿势(翻面和不同的手指位置)和颜色的明显变化,造成这样的目标区域很难跟踪。为了克服颜色变化的问题,放宽了最小欧几里得距离的上边界。在一些帧中,两只手都会翻面。不过,本算法依然能够跟踪 ROI,错过的仅仅是右手的一部分。在一些帧中,尽管与脖子和面部重叠,也实现了良好的跟踪性能。当手从其它肤色区域移开时,错失了与背景相关的右手的主要部分。最终,对右手的整个 ROI 完全消失了。这主要是因为右手移动相对较快,并且因此很难进行跟踪。应当注意,能够在本算法中加入反馈功能来防止 ROI 消失。例如,如果在经过了一定数量的帧时 ROI 逐渐变小,则可以增大搜索区和搜索窗。

[0094] 按照本发明的目标区域跟踪方法的主要优点是,不依赖于关于感兴趣区域的先验信息。例如,在跟踪人脸过程中,本算法不依赖于诸如颜色值之类的特定特征,和眼睛及嘴的特殊性。通常,离线进行目标训练是需要特性信息的。按照本发明的跟踪算法适用于用户指定的非特定的感兴趣区域。

[0095] ROI 跟踪在图像编解码(比如 JPEG 2000)中是一项非常有用的功能。它能够实现,在编码图像帧的时候,将比图像帧中其它区域多的位分配给目标区域。这样,可以实现目标区域周围的高视觉质量。

[0096] 虽然本发明是针对其优选实施方式加以介绍的,但是本领域的技术人员应当理解,可以对前述内容和各种不同的其它变化进行其形式上和细节的省略和转义,而不会超出本发明的范围。

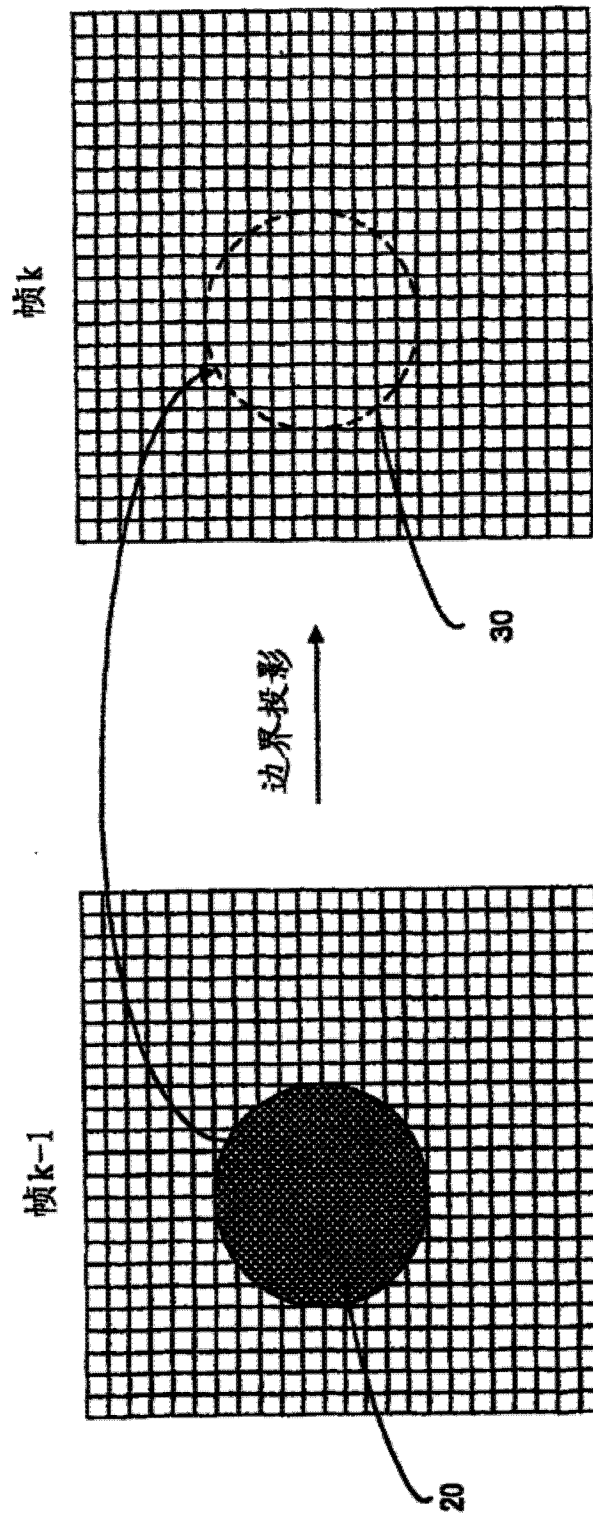


图 1

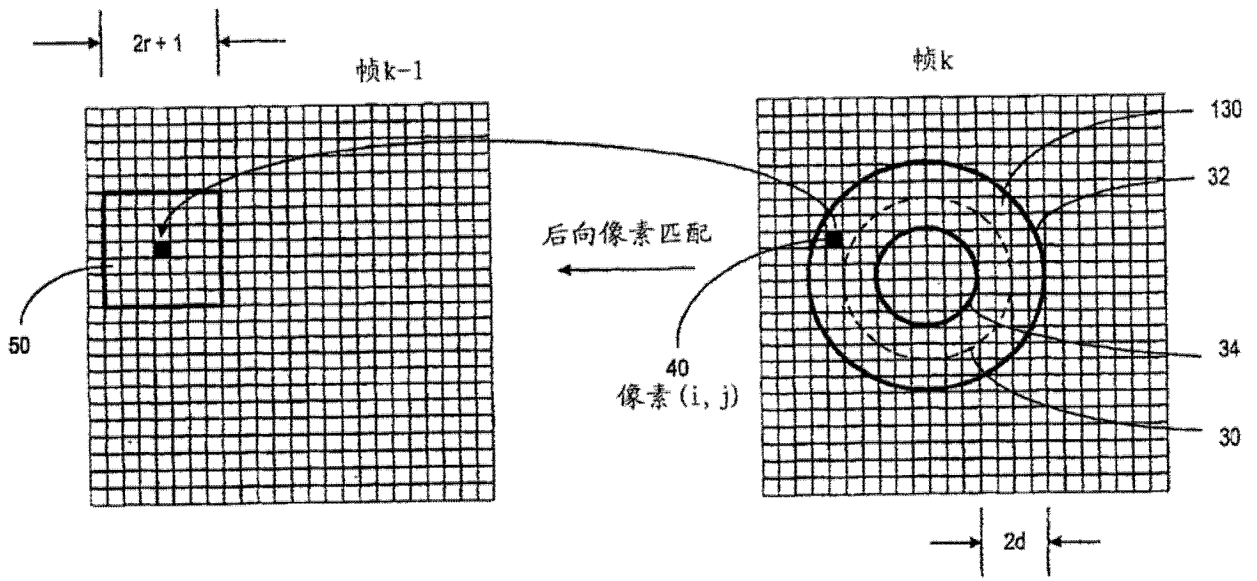


图 2

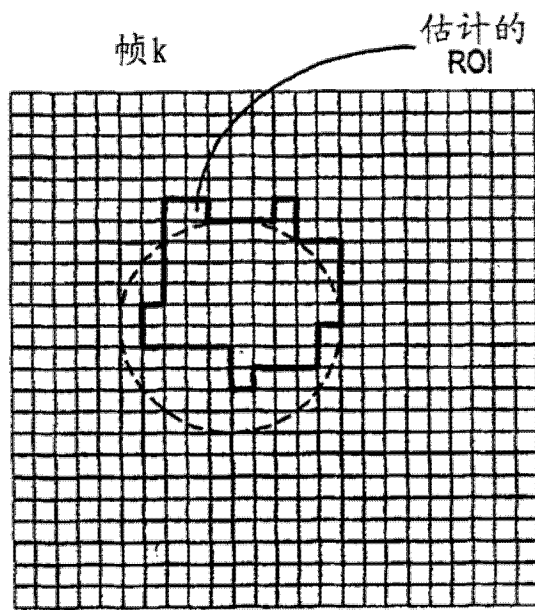


图 3a

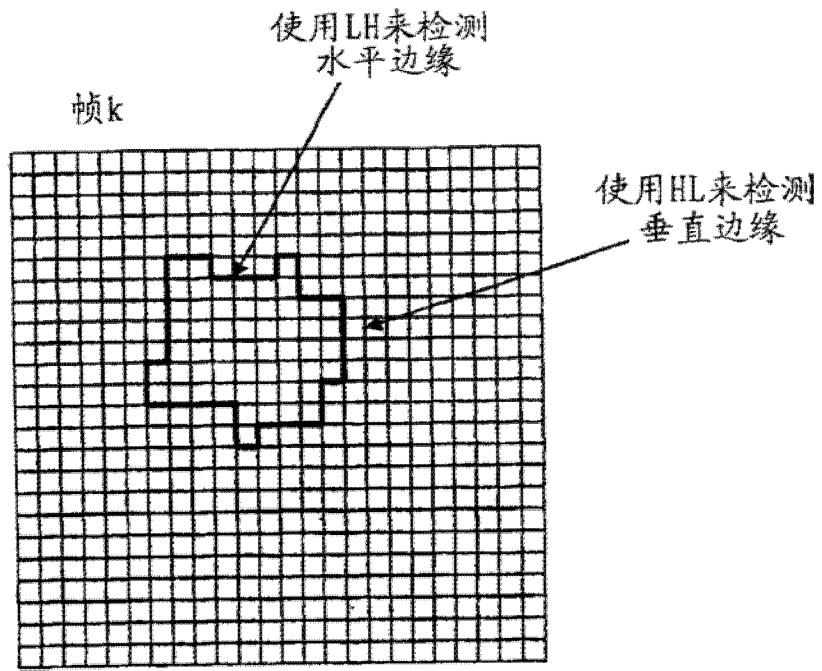


图 3b

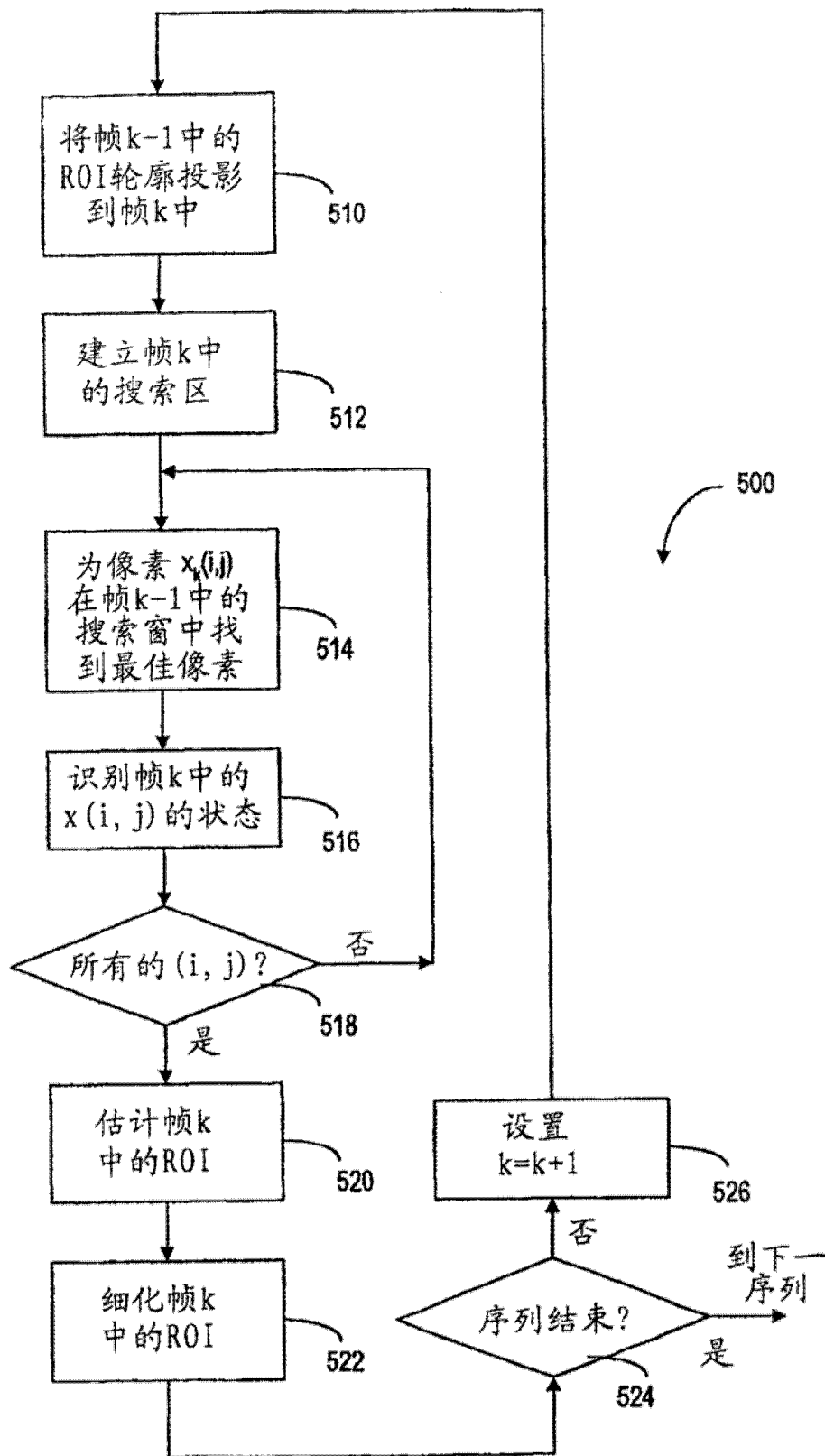


图 4

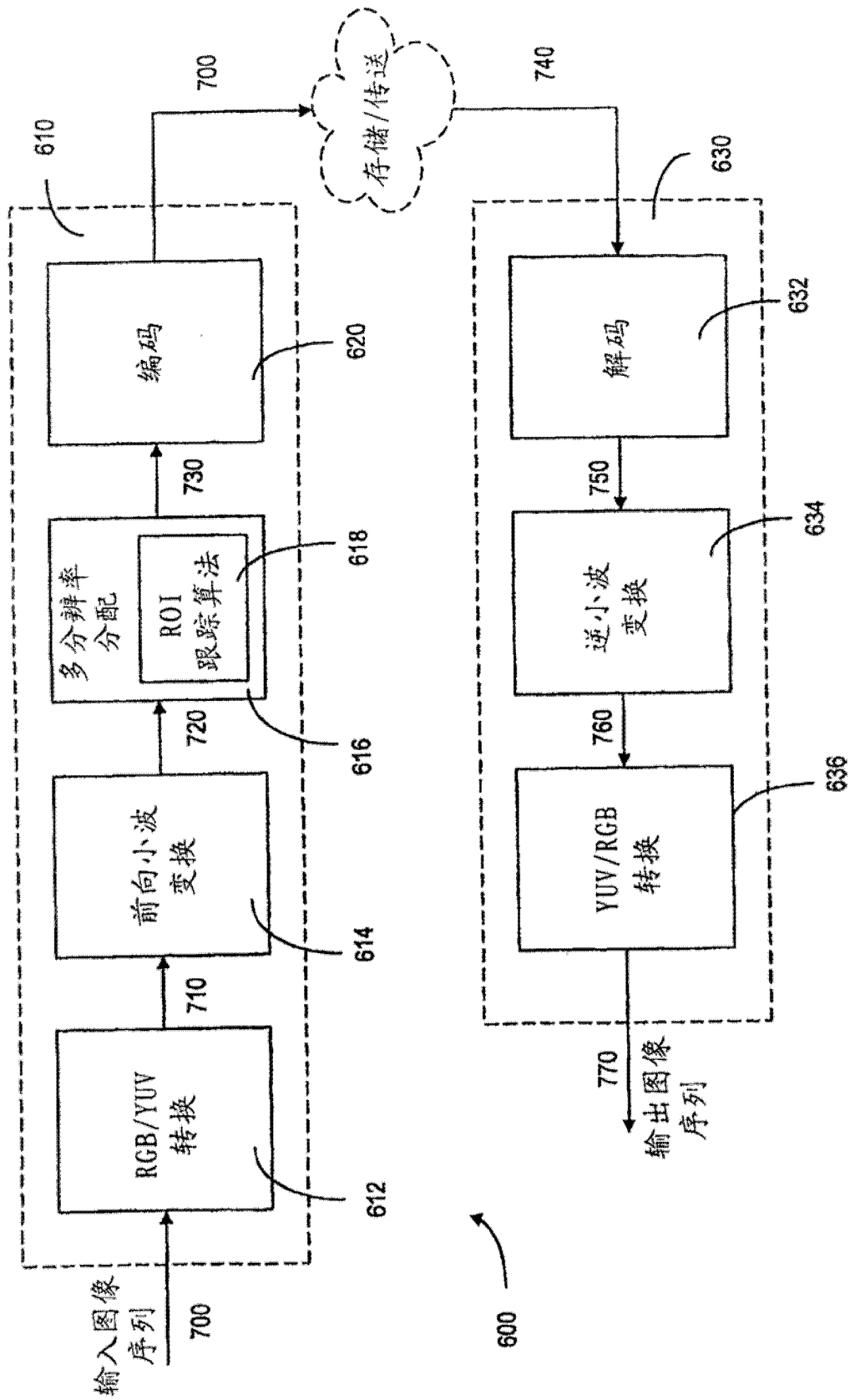


图 5