

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200810132831.9

[51] Int. Cl.

H04N 7/64 (2006.01)

H04N 7/14 (2006.01)

G06K 9/00 (2006.01)

G06T 5/00 (2006.01)

[43] 公开日 2008年11月19日

[11] 公开号 CN 101309426A

[22] 申请日 2008.7.10

[21] 申请号 200810132831.9

[71] 申请人 北京邮电大学

地址 100876 北京市海淀区西土城路10号

[72] 发明人 郑中亮 郭彦东

[74] 专利代理机构 北京纽乐康知识产权代理事务所

代理人 田磊

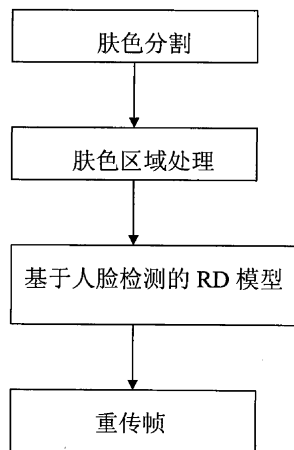
权利要求书2页 说明书6页 附图2页

[54] 发明名称

基于人脸检测的可视电话视频传输抗误码方法

[57] 摘要

本发明涉及一种基于人脸检测的可视电话视频传输抗误码方法，选择基于肤色的人脸检测，该方法包括以下步骤：肤色分割：选择 YCbCr 空间作为肤色分布统计的映射空间，其中 Y 分量表示亮度、Cb、Cr 表示色度，肤色点能够形成较好的肤色聚类；肤色区域处理：首先对分割出来的皮肤区域进行腐蚀与膨胀操作，除去噪声对肤色分隔的影响；然后通过人脸肤色区域的统计特征分析，对分割出来的肤色区域作以下预处理，得到最为可能的人脸区域；基于人脸检测的 RD 模型：将人脸检测模型引入 RD 模型中；重传帧。本发明的有益效果：提高对人脸检测精度、速度性能；以及对帧内，帧间选择子模式的选择模型进行优化。



1、一种基于人脸检测的可视电话视频传输抗误码方法，选择基于肤色的人脸检测，其特征在于：该方法包括以下步骤：

肤色分割：选择 YCbCr 空间作为肤色分布统计的映射空间，其中 Y 分量表示亮度、Cb、Cr 表示色度，肤色点能够形成较好的肤色聚类；利用椭圆模型对肤色建模，经过非线性分段色彩变换得到的色彩空间用 YCb'Cr'来表示，肤色聚类在 YCb'Cr'空间中的分布呈明显的椭圆分布，椭圆内区域为肤色区域，其他区域为非肤色区域，从而得到比较理想的二值化分割图像；

肤色区域处理：首先对分割出来的皮肤区域进行腐蚀与膨胀操作，除去噪声对肤色分隔的影响，然后通过人脸肤色区域的统计特征分析，对分割出来的肤色区域作预处理，得到最为可能的人脸区域；

基于人脸检测的 RD 模型：在基于人脸的 RD 模型 FDRD 中，将人脸检测的信息用作引导模式的选择，在端到端 RD 模型的基础上，改进的基于人脸的 RD 模型 FDRD 的代价方程为：

$$J(o) = \alpha(F, o, R)(D(o) + D_{ep_ref}(o) + \lambda R(o))$$

加权因子 $\alpha(F, o, R)$ 根据视频块是否在人脸区域内，以及是否属于重传帧来确定；

重传帧：将帧内编码的加权因子设置为 1，而将帧间编码的加权因子设置为 $+\infty$ ，这样，感兴趣的人脸区域将会被完整的进行帧内刷新，其中加权因子只在帧内、帧间编码的选择层面产生作用，并不影响帧内以及帧间的子模式，子模式的选择仍然由端到端模型来决定。

2、根据权利要求 1 所述的基于人脸检测的可视电话视频传输抗误码方法，其特征在于：在肤色区域处理步骤的对分割出来的肤色区域作预处理中，对人脸来说，其长宽比为 1，计算肤色区域的长宽比，取人脸的长宽比为 0.4~0.5，长宽比不满足此范围的区域，被认为是非人脸区域而被剔除。

3、根据权利要求 1 所述的基于人脸检测的可视电话视频传输抗误码方法，

其特征在于：在肤色区域处理步骤的对分割出来的肤色区域作预处理中，肤色区域小于 400 像素被认为是噪声干扰而被除去。

基于人脸检测的可视电话视频传输抗误码方法

技术领域

本发明涉及一种基于人脸检测的可视电话视频传输抗误码方法。

背景技术

近年来，抗误码视频编码技术被广泛的研究，特别是在无线信道环境下，前向差错更正，多描述编码，分层视频编码技术，联合编码以及信道编码等技术被广泛研究，此外还有鲁棒小波编码等。在最近的国际视频编码标准 H.264/AVC 中，又有一些新的算法用来提高编码的容错能力，比如复杂宏块排序，参数设置，SP 变换等。虽然有这么多容错技术，帧内刷新（Intra refresh）技术，即将某些选定的视频数据进行帧内编码来阻止差错扩散，仍然是易错信道中视频传输的最简单和有效的方法。

现有的基于块的混合视频编码结构中有两种编码模式：帧内编码与帧间编码，其中帧内编码只用当前帧的信息，而帧间编码模式用相邻帧的信息作为参考帧来进一步减少时间冗余。当一个编码后的视频在倾向于出错的网络中传输时，无码会在某一帧中出现，然后传播到后面的几帧：因为帧间编码的预测。甚至在解码端采用一些错误检测方法，也效果十分有限。相反地，帧内编码模式可以抑制差错传播因为他并不参考相邻帧的信息。因此，通过设定某些规则，帧内编码，或者说帧内刷新，是在抗误码视频编码中修复视频主观质量损失的一个基本的，有效的方法。I-帧是帧内刷新的一个特殊情况，在 I-帧情况下，整帧图像用帧内模式编码。但是，编码这么多 I 帧将会产生大量的比特流，并且产生一个剧烈的比特率波动。因此，有许多帧内刷新技术的改进：包括随机重传，周期性重传，基于运动信息的重传，基于端到端的 RD 模型的重传，基于反馈的重传，基于丢包率的重传，以及这些方法的结合。

在上述方法中，基于端到端模型的帧内刷新技术将有高的，或者复杂的运动信息的块设置为帧内刷新块，因为如果将它们用帧间模式编码的话，与用帧内编码模式相比，由于是误码扩散，会有更大的端到端的失真。虽然将高运动

信息的块用帧内模式编码是看起来很合理，但是，从感知的角度看，这个解决方案仍然不够完美，尤其是在视频电话的应用中。在可视电话或者视频会议中，人们更倾向于关注可视电话中的“人脸”，而不是其他区域。而往往人脸区域并没有很大的运动向量信息。因此，一个适用于可视电话的容错视频编码算法应该尽可能保护在人脸的区域的信息，从而提高视频的主管质量。

在 2002 年，M.H.Yang 在 IEEE Trans. PAMI 上发表的“Detecting Faces in Images: A survey”，的综述文章，将人脸检测技术分为几类进行了详细介绍。主要包括基于几何特征的人脸检测方法（包括基于先验知识的方法，基于特征不变性的方法，基于模板的方法），基于肤色模型的人脸检测方法，以及基于统计理论的人脸检测方法（包括子空间方法，神经网络方法，支持向量方法，隐马尔可夫模型方法，以及 Boosting 方法）。在这些方法中，基于肤色模型的人脸检测方法应用广泛，具有易于应用，对人脸姿态不敏感，在背景单调的环境下检测率高的优点。此外，对于色度空间的选择非常重要，很大程度上影响整个算法的效果。基于肤色模型的方法在光照和背景复杂的情况下，效果不是很理想，现在往往和其他方法相结合使用。近期出现的使用红外线作为光源的方法可以部分解决光照复杂的问题。在基于几何特征的人脸检测方法中基于先验知识的方法可以直观的寻找到简单的判定准则，适合简单背景下的正面单人脸检测。因为针对人脸面部器官制定一系列精确恰当的判定准则非常困难，而且该方法在复杂背景下的检测效果不很理想，所以在实际系统中应用并不广泛。与以上两种相似，基于特征不变性的方法主要的问题是在面部存在遮挡和复杂背景的情况下很难获得理想的效果。基于模板方法的优点在于构造简单，但是面对人脸的多样性，很难设计出精确匹配的标准模板。此外对于应用于人脸配准上的变形模板，模板初始位置必须要在待检人脸四周一定的范围内，否则不能收敛。

RD 模型几乎是当前所有视频编码技术通过在特定比特率下，得到最小失真的，决定编码模式的基本策略。这个受限的最优化问题可以被拉格朗日优化方法解决。对于每一个编码模式 o 来说，代价函数 $J(o)$ 用下式计算。最优的编码模式 o ，就是使得代价方程值最小的编码模式：

$$J(o) = D(o) + \lambda R(o)$$

与传统的 RD 模型不同的是，端到端的 RD 模型进一步考虑了视频传输与解码的影响，因此他使用端到端的失真代替了原有的只考虑在编码端的失真。端到端的失真定义“失真”为传输，差错检测后的解码数据与编码前的原始视频数据的差异。相应地，从解码端进行考虑的端到端的 RD 模型的失真就由三部分组成：由量化产生的失真，由错误扩散产生的失真，由丢失当前帧产生的失真。当当前数据块收到的时候，只有第一个和第二个部分有效果。数据块是否收到可以由之前假设的 PLR 模型设定。一个简单有效的端到端 RD 模型在中提出，错误扩散被大致用基于块的失真图模拟。在对每个块进行编码之前，错误扩散失真可以由前一帧/参考帧的失真图得到。端到端的 RD 模型的代价方程为：

$$J(o) = D(o) + D_{ep_ref}(o) + \lambda R(o)$$

端到端的 RD 模型的主要问题在于，人脸不一定是运动信息值大的区域，运动信息小的区域未必在可视电话中不引人注目。在并不引人注目的区域进行帧内刷新，会导致一些不必要的码率提升。

发明内容

本发明的目的涉及一种基于人脸检测的可视电话视频传输抗误码方法，以进一步去除肤色分割后包括的非人脸区域，实现更精准的人脸定位，从而尽可能多地检测到人脸的同时降低误检率。

本发明的目的是通过以下技术方案来实现：

一种基于人脸检测的可视电话视频传输抗误码方法，选择基于肤色的人脸检测，该方法包括以下步骤：

肤色分割：选择 YCbCr 空间作为肤色分布统计的映射空间，其中 Y 分量表示亮度、Cb、Cr 表示色度，肤色点能够形成较好的肤色聚类；对人脸进行肤色检测首先要对肤色建模，人脸建模的方法诸如高斯模型、椭圆模型，其中对于椭圆模型，经过非线性分段色彩变换得到的色彩空间用 YCb'Cr'来表示，肤色聚类在 YCb'Cr'空间中的分布呈明显的椭圆分布，椭圆内区域为肤色区域，其他区域为非肤色区域，从而得到比较理想的二值化分割图像；

肤色区域处理：首先对分割出来的皮肤区域进行腐蚀与膨胀操作，除去噪声对肤色分隔的影响；然后通过人脸肤色区域的统计特征分析，对分割出来的肤色区域作预处理，得到最为可能的人脸区域；

基于人脸检测的 RD 模型：将人脸检测模型引入 RD 模型中，在基于人脸的 RD 模型 FDRD 中，将人脸检测的信息用作引导模式的选择，在端到端 RD 模型的基础上，改进 FDRD 的代价方程为：

$$J(o) = \alpha(F, o, R)(D(o) + D_{ep_ref}(o) + \lambda R(o))$$

加权因子 $\alpha(F, o, R)$ 根据视频块是否在人脸区域内，以及是否属于重传帧来确定；

重传帧：将帧内编码的加权因子设置为 1，而将帧间编码的加权因子设置为 $+\infty$ ，这样，感兴趣的人脸区域将会被完整的进行帧内刷新，提升了视频的主观质量，加权因子只在帧内、帧间编码的选择层面产生作用，并不影响帧内以及帧间的子模式，子模式的选择仍然由端到端模型来决定。

在对实时性要求不高的场合，可以用 Adaboost 级联分类器对上面的方法进行改进。首先用肤色模型检测肤色区域，然后用膨胀，腐蚀去噪；再将这些区域作为输入图像用训练好的 Adaboost 级联分类器进行检测，进一步去除肤色分割后包括的非人脸区域，实现更精准的人脸定位，从而尽可能多地检测到人脸的同时降低误检率。同时，也避免了只用级联算法需要扫描整幅图像得到所有子窗口的过程，大大加快了算法的运行速度，也避免了 Adaboost 算法检测多人脸图像时效果不理想的情况

本发明的有益效果：提高对人脸检测精度、速度性能；以及对帧内，帧间选择子模式的选择模型进行优化。

附图说明

图 1 是本发明实施例所述的基于人脸检测的可视电话视频传输抗误码方法的流程图；

图 2A-2D 是几种差错控制方式的比较图。

具体实施方式

如图 1 所示，本发明实施例所述的基于人脸检测的可视电话视频传输抗误码方法，选择基于肤色的人脸检测，该方法包括以下步骤：

在步骤 10 中，进行肤色分割：选择 YCbCr 空间作为肤色分布统计的映射空间，其中 Y 分量表示亮度、Cb、Cr 表示色度，该空间的优点是可以将亮度和色度分开单独处理，实现亮度和色度分量比较彻底的分离，Cb、Cr 是两维独立分布，能较好地限制肤色分布区域，肤色点能够形成较好的肤色聚类；利用椭圆模型对肤色建模，经过非线性分段色彩变换得到的色彩空间用 YCb'Cr'来表示，肤色聚类在 YCb'Cr'空间中的分布呈明显的椭圆分布，可用以下的公式来匹配两个色度分量的距离：

$$(x - ec_x)^2 / a^2 + (y - ec_y)^2 / b^2 = 1$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} C'_b & -c_x \\ C'_r & -c_y \end{bmatrix}$$

以上表达式中的常量分别为

$$c_x = 109.38, \quad c_y = 152.02,$$

$$ec_x = 1.60, \quad ec_y = 2.41,$$

$$\theta = 2.53, \quad a = 25.39, \quad b = 14.03$$

椭圆内区域为肤色区域，其他区域为非肤色区域，从而得到比较理想的二值化分割图像；

在步骤 20 中，进行肤色区域处理：经过肤色分割得到一系列的连通区域，这些区域包括皮肤区域及与皮肤颜色相近的背景区域，在复杂背景下，有可能存在很多与人脸肤色相近的非人脸区域被分割出来，如手、胳膊等，这就需对这些肤色区域作进一步的处理，首先对分割出来的皮肤区域进行腐蚀与膨胀操

作，除去噪声对肤色分隔的影响；然后通过人脸肤色区域的统计特征分析，对分割出来的肤色区域作以下预处理，得到最为可能的人脸区域：

a)对人脸来说，其长宽比一般为 1 左右，计算肤色区域的长宽比，取人脸的长宽比为 0.4~0.5，长宽比不满足此范围的区域，被认为是非人脸区域而被剔除；

b)由于人脸区域有一定的大小，将肤色区域小于 400 像素（小于 2×2 个宏块单位）被认为是噪声干扰而被除去；

在步骤 30 中，基于人脸检测的 RD 模型：将人脸检测模型引入 RD 模型中，在基于人脸的 RD 模型（FDRD）中，将人脸检测的信息用作引导模式的选择，在端到端 RD 模型的基础上，改进 FDRD 的代价方程为：

$$J(o) = \alpha(F, o, R)(D(o) + D_{ep_ref}(o) + \lambda R(o))$$

加权因子 $\alpha(F, o, R)$ 根据视频块是否在人脸区域内，以及是否属于重传帧来确定；

在步骤 40 中，进行重传帧：将帧内编码的加权因子设置为 1，而将帧间编码的加权因子设置为 $+\infty$ ，这样，感兴趣的人脸区域将会被完整的进行帧内刷新，提升了视频的主观质量，加权因子只在帧内、帧间编码的选择层面产生作用，并不影响帧内以及帧间的子模式，子模式的选择仍然由端到端模型来决定。

如图 2A-2D 所示，比较了 4 种情况：图 2A、不进行差错控制；图 2B、随机选择的帧内刷新技术；图 2C、端到端模型；图 2D、FDRD 模型。设置丢包率为 10%，20%对多种格式的序列进行测定。图 2A-2D 是在丢包率为 20%的情况下，CIF 格式的序列“Foreman”的一个截图。可以看出，FDRD 模型对人脸区域保护非常好，与其他三种方法相比，得到了最好的主观质量。

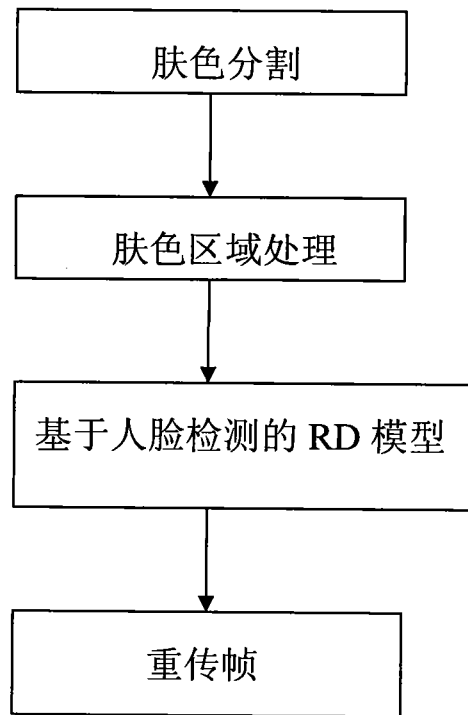


图 1



图 2A



图 2B



图 2C



图 2D