

(12) 发明专利

(10) 授权公告号 CN 101482923 B

(45) 授权公告日 2012.05.23

(21) 申请号 200910013650.9

(22) 申请日 2009.01.19

(73) 专利权人 刘云

地址 266061 山东省青岛市高新区松岭路
69号青岛科技大学506号信箱

(72) 发明人 刘云 王传旭 张祥光

(74) 专利代理机构 青岛联智专利商标事务所有
限公司 37101

代理人 邵新华

(51) Int. Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

(56) 对比文件

CN 1700238 A, 2005.11.23, 全文.

CN 101251890 A, 2008.08.27, 全文.

EP 1300804 A2, 2003.04.09, 全文.

US 6678407 B1, 2004.01.13, 全文.

G. Gomez, etc.. On selecting an
Appropriate Colour Space for Skin
Detection. Lecture Notes in Computer
Science. Springer Berlin, 2002, 2313 第 69-78
页.

王传旭. 视频图像中人体目标的检测方法研
究. 中国博士学位论文全文数据库. 中国海洋大
学, 2008, 第 59、66 页.

审查员 李劲嫻

权利要求书 1 页 说明书 11 页 附图 1 页

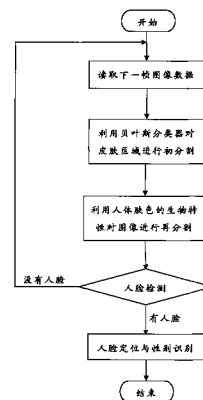
(54) 发明名称

视频监控中人体目标的检测与性别识别方法

(57) 摘要

本发明公开了一种视频监控中人体目标的检测与性别识别方法,包括:a、根据采集到的图片数据库在 HSV 彩色空间建立前景皮肤和背景的直方图;b、利用贝叶斯分类器对输入视频的每个像素点进行前景和背景的判决,进而分割出其中的人体皮肤区域;c、利用人体肤色的生物特征对图像进行再分割,以消除背景中肤色相近色物体的干扰,获得准确的人体目标;d、根据获得的人体目标中的人脸图像利用模式识别方法进行性别识别。本发明将人的面部皮肤特征作为人体的主要标识,对存在于视频图像中的人体目标进行检测,从而克服了光照变化和背景相近色干扰所产生的不利影响,提高了目标性别的正确识别。

CN 101482923 B



1. 一种视频监控中人体目标的检测与性别识别方法,包括以下步骤:

a、根据采集到的图片数据库在 HSV 彩色空间建立前景皮肤和背景的直方图;

b、利用贝叶斯分类器对输入视频的每个像素点进行前景和背景的判决,进而分割出其中的人体皮肤区域;

c、利用人体肤色的生物特征在复合彩色空间 (E, R/G, H) 下对前景皮肤和背景进行再次分割,以消除背景中肤色相近色物体的干扰,获得准确的人体目标;

d、根据获得的人体目标中的人脸图像利用模式识别方法进行性别识别;

其中,在所述步骤 a 中,当光照发生剧烈变化时,采用三维仿射变换对皮肤分布区域的帧间变化进行建模,通过三维仿射变换公式预测出当前帧中皮肤区域在 HSV 坐标系中的分布区域,进而统计出当前帧中前景皮肤和背景的直方图;所述三维仿射变换所需的 8 个参数 $X = [T_H, T_S, T_V, S_H, S_S, S_V, \theta, \phi]^T$ 采用线性组合预测模型进行预测;所述线性组合预测模型是采用基于平方和误差最小原理将维纳一步预测模型和二阶马尔科夫预测模型进行线性组合构建的;其中, T_i 表示平移; S_i 表示旋转; θ, ϕ 表示几何体旋转的角度。

2. 根据权利要求 1 所述的视频监控中人体目标的检测与性别识别方法,其特征在于:通过所述线性组合预测模型预测得到第 t+1 时刻输入帧的前景皮肤区域在 HSV 彩色空间分布的 8 个参数后,首先利用所述的三维仿射变换和直方图统计求出第 t+1 时刻输入帧的前景皮肤的预测直方图,然后利用 t 时刻前景皮肤的直方图和所述预测直方图的加权平均值,求出第 t+1 时刻输入帧的前景皮肤的直方图。

3. 根据权利要求 1 所述的视频监控中人体目标的检测与性别识别方法,其特征在于:采用过去两帧的背景直方图的加权平均得到当前帧的背景直方图分布。

4. 根据权利要求 1 至 3 中任一项所述的视频监控中人体目标的检测与性别识别方法,其特征在于:在所述步骤 c 中,利用人体肤色中的红色分量占据主导位置的这一生物特性,在复合彩色空间 (E, R/G, H) 下建立空间模型,利用参数阈值: $E > 13.4224$ 、 $R/G < 1.7602$ 、 $H < 23.89$ 对前景皮肤和背景进行再次分割,以消除背景中肤色相近色物体的干扰;其中,所述 H 分量来自 (H, S, V) 空间,表示色度;R/G 是 (R, G, B) 空间中 R 和 G 分量的比值;E 分量来自 (Y, E, S) 空间,表示色度,且 $E = 0.5 * (\text{red} - \text{green})$ 。

5. 根据权利要求 1 至 3 中任一项所述的视频监控中人体目标的检测与性别识别方法,其特征在于:在所述步骤 c 和 d 之间还包括步骤 c1:

c1、将经过再次分割后的图像输入形态学滤波器,进而利用邻域相关性特点将人脸修复成完整图像。

6. 根据权利要求 5 所述的视频监控中人体目标的检测与性别识别方法,其特征在于:将通过所述形态学滤波器产生的人脸图像进行尺度规范化,然后利用训练好的支持向量机对人脸图像的性别进行识别。

视频监控中人体目标的检测与性别识别方法

技术领域

[0001] 本发明属于图像信息处理技术领域,具体地说,是涉及一种在视频监控图像中对人脸进行自动检测并利用模式识别方法对检查到的人脸进行性别识别的技术。

背景技术

[0002] 人脸是人类的重要生物特征,可以反映了一个人的身份、年龄、情感和性别等重要信息。而性别分类在心理学领域和机器视觉领域得到了广泛关注,早期的计算机视觉性别分类大都是基于神经网络的机器学习方法,包括 Gollomb 等的两层神经元网络 SEXNET,在 30×30 的人脸图片上进行性别判别,在 90 张图片(男女各 45 张)中性别正确识别率达到 91.9%;Edelman 等用人脸的 3 个不同部分(整个人脸、人脸上半部分和下半部分)分别训练,采用线性神经网络进行分类,并对 3 种情况的分类性能做了分析比较;Tamura 等利用了与 Gollomb 相近的技术,采用多层神经网络对不同分辨率的人脸进行了分类分析,在 8×8 的超低分辨率下可以得到平均 7% 的错误率。以上基于神经网络的性别识别方法,需要的人脸训练样本数量比较大,且泛化能力差;另外,神经网络的训练速度比较慢,难以在视频监控中实时实现人脸的性别识别。近年,BabackMoghaddam 和 M. H. Yang 首次将支持向量机方法引入性别判别领域,采用像素特征并与 RBF 网络、FLD、最邻近分类器等经典分类方法做了比较,用 FERET 人脸库做了分类实验,证明了支持向量机(SVM)分类的有效性。

[0003] 但是,上述这些方法大都只能对静态人脸图像进行性别识别,而视频监控场景受到光照变化的影响以及背景相近色的干扰,因而对人脸的检测和性别的识别带来很大挑战。

发明内容

[0004] 本发明针对现有人体目标检测方法中普遍存在的“光照变化”和“背景相近色干扰”两个棘手问题,提供了一种全新的人体目标检测方法,利用人的面部皮肤特征作为人体的主要标识,来对存在于视频图像中的人体目标进行检测,从而实现对人脸的可靠分割;进而针对获得的人脸图像进行性别分类,从而实现了在视频监控中人体目标的性别识别。

[0005] 为解决上述技术问题,本发明采用以下技术方案予以实现:

[0006] 一种视频监控中人体目标的检测与性别识别方法,包括以下步骤:

[0007] a、根据采集到的图片数据库在 HSV 彩色空间建立前景皮肤和背景的直方图;

[0008] b、利用贝叶斯分类器对输入视频的每个像素点进行前景和背景的判决,进而分割出其中的人体皮肤区域;

[0009] c、利用人体肤色的生物特征在复合彩色空间(E, R/G, H)下对前景皮肤和背景进行再次分割,以消除背景中肤色相近色物体的干扰,获得准确的人体目标;

[0010] d、根据获得的人体目标中的人脸图像利用模式识别方法进行性别识别。

[0011] 其中,在所述步骤 a 中,当光照发生剧烈变化时,采用三维仿射变换对皮肤分布区域的帧间变化进行建模,通过三维仿射变换公式预测出当前帧中皮肤区域在 HSV 坐标系中

的分布区域,进而统计出当前帧中前景皮肤和背景的直方图。

[0012] 进一步的,所述三维仿射变换所需的 8 个参量采用线性组合预测模型进行预测;其中,所述线性组合预测模型采用对两个或者多个现有的预测模型进行线性组合的方法来构建。

[0013] 优选的,所述线性组合预测模型采用维纳一步预测模型和二阶马尔科夫预测模型组成。

[0014] 其中,在所述线性组合预测模型中,采用平方和误差为精度综合评价指标来确定所述维纳一步预测模型和二阶马尔科夫预测模型的最优组合权重,即加权系数,以构建出所述的线性组合预测模型。

[0015] 又进一步的,在对前景皮肤直方图进行更新时,通过所述线性组合预测模型预测得到第 $t+1$ 时刻输入帧的前景皮肤区域在 HSV 彩色空间分布的 8 个参数向量后,首先利用所述的三维仿射变换和直方图统计求出第 $t+1$ 时刻输入帧的前景皮肤的预测直方图,然后利用 t 时刻前景皮肤的直方图和所述预测直方图的加权平均值,求出第 $t+1$ 时刻输入帧的前景皮肤的直方图。

[0016] 而对于背景直方图的更新,则采用过去两帧的背景直方图的加权平均得到当前帧的背景直方图分布。

[0017] 再进一步的,在所述步骤 c 中,利用人体肤色中的红色分量占据主导位置的这一生物特性,在复合彩色空间 (E, R/G, H) 下建立空间模型,利用参数阈值: $E > 13.4224$ 、 $R/G < 1.7602$ 、 $H < 23.89$ 对前景皮肤和背景进行再次分割,以消除背景中肤色相近色物体的干扰;其中,所述 H 分量来自 (H, S, V) 空间,表示色度;R/G 是 (R, G, B) 空间中 R 和 G 分量的比值;E 分量来自 (Y, E, S) 空间,表示色度,且 $E = 0.5 * (\text{red} - \text{green})$ 。

[0018] 更进一步的,在所述步骤 c 和 d 之间还包括步骤 c1:

[0019] c1、将经过再次分割后的图像输入形态学滤波器,进而利用邻域相关性特点将人脸修复成完整图像。

[0020] 然后,将通过所述形态学滤波器产生的人脸图像进行尺度规范化,进而利用训练好的支持向量机对人脸图像的性别进行识别。

[0021] 与现有技术相比,本发明的优点和积极效果是:本发明将人的面部皮肤特征作为人体的主要标识,对存在于视频图像中的人体目标进行检测,进而克服了光照变化和背景相近色干扰对人体识别所造成的不利影响,提高了在视频监控系统中人体目标的性别识别。

[0022] 结合附图阅读本发明实施方式的详细描述后,本发明的其他特点和优点将变得更加清楚。

附图说明

[0023] 图 1 是发明所提出的视频监控中人体目标的检测与性别识别方法的一种实施例的结构示意图。

具体实施方式

[0024] 下面结合附图和具体实施方式对本发明作进一步详细地说明。

[0025] 在连续的视频图像序列中,人体目标的运动情况是随机的,即有时运动有时静止,基于运动特征(如高斯混合模型方法、背景差分法等)分割人体目标的方法,当人体运动很缓慢或者静止的时候,就会变得无效。而皮肤颜色是人体表面最为显著的特征之一,不依赖于细节特征,具有较高的稳定性且与大多数背景物体的颜色相区别。因此,采用基于人体肤色特征的人体目标分割方法,即使在分割、预测和跟踪过程中光照变化较大时,这种方法也能对皮肤进行很可靠的分割。

[0026] 本发明所提出的人体目标检测方法,其设计思路是:借助采集到的图片数据库在 HSV 彩色空间建立初始的前景皮肤和背景的直方图,进而求出各类概率值。在光照基本不发生变化时,利用 Bayes 分类器对输入视频的每个像素点进行前景和背景的判决,进而分割出其中的人体皮肤区域。当光照发生剧烈变化时,前景皮肤和背景的直方图分布会随之变化,因此,分割之前应该对当前帧的直方图分布进行重新估计。为保证光照变化条件下的分割精度,假设每帧肤色区域像素在 HSV 彩色空间的分布构成相对集中的“点云”三维几何体,光照变化时每帧“点云”三维几何体在 HSV 空间位置的变化都可以通过平移、缩放和旋转来参数化,因而可用三维仿射变换对皮肤分布区域的帧间变化进行建模(假设皮肤区域的变化是整体变化,不考虑局部的变形)。在一定长度的观察窗内,首先,采用维纳一步预测模型和二阶马尔科夫预测模型组成的线性组合预测模型来对三维仿射变换所需的 8 个参量进行预测;其次,通过三维仿射变换公式预测出当前帧中皮肤区域在 HSV 坐标系中的分布区域,进而可统计出当前帧中前景皮肤和背景的直方图分布。然后,再利用 Bayes 分类器对当前帧进行皮肤区域的分割,以保证光照变化时人体目标分割算法的鲁棒性;最后,为了克服复杂背景中的肤色相近色物体的干扰,利用人体肤色生物特征并结合形态学滤波器,可有效克服初分割结果中大片相近色的干扰,从而得到理想的人体目标。

[0027] 在获得了准确的人体目标后,根据检测到的人脸图像利用模式识别方法便可以进行性别识别。

[0028] 下面以一个具体的实施例来详细阐述所述视频监控中人体目标的检测与性别识别方法的具体实现过程。

[0029] 实施例一,在本实施例中,利用视频监控系统获取连续的视频图像序列,以建立起图像数据库,进而调用如图 1 所示的人体目标检测与性别识别方法来获得待检测目标的性别。下面详细描述所述人体目标检测与性别识别方法的具体执行步骤。

[0030] 第一步,在所述视频图像序列中读取一帧图像数据。在本实施例中,为了保证检测程序能够循环运行,这里定义为读取下一帧图像数据。

[0031] 第二步,利用贝叶斯分类器进行肤色检测。

[0032] 基于像素级的人体肤色贝叶斯分类检测原理,是根据当前帧的肤色区域与背景区域(非肤色区域)的颜色直方图分布,分别计算某一像素属于肤色区域和非肤色区域的条件概率来进行分类的。该统计分类器的关键是计算出当前待分割帧的肤色区域与背景区域的颜色分布直方图。

[0033] 为了获得所述的颜色分布直方图,本实施例采用三维仿射变换对皮肤分布区域的帧间变化进行建模,以预测当前待分割帧的颜色直方图分布。

[0034] 通过视频监控系统采集到的视频图像一般在 RGB 颜色空间,颜色分辨率为 $256 \times 256 \times 256$ 。为了减少运算量,在保证精度的条件下,可以采用 $32 \times 32 \times 32$ 颜色分辨

率来计算直方图的分布。得到的结果显示：该分辨率在分割中不仅能保证分辨率的要求，而且还会降低运算量。另外，低维直方图的优势主要是需要较少的训练像素，倘若估计 $256 \times 256 \times 256$ 的直方图，则需要多出 512 倍的训练数据。

[0035] 统计每个直方图条中某一颜色 (rgb) 像素的数目以及前景皮肤和背景的像素的总数，可以计算得到各类概率密度，记为： $p(\text{fg})$ 、 $p(\text{bg})$ 、 $P(\text{rgb}|\text{fg})$ 和 $P(\text{rgb}|\text{bg})$ 。其中，fg 代表前景，bg 代表背景，rgb 为像素颜色， $\text{rgb} \in \mathbb{R}^3$ 。考查皮肤颜色分布的不同表现，可以推断出皮肤颜色分布的一些结构特点，即皮肤颜色在彩色空间分布很集中地占据一个紧凑的区域。近年来，在皮肤物理外观上的研究表明，皮肤反射区直接紧密地与光照相关，而且试验证明：当光照强度没有剧烈变化时，皮肤颜色分布变化是平滑的，每一帧皮肤颜色的分布在彩色空间中，都仅占某一小部分区域，因此，在背景光照变化的环境下，可以通过跟踪那一小部分区域的轨迹变化来表征相应的皮肤颜色变化。

[0036] 为了模型化和预测这种变化，需要假设在彩色空间皮肤颜色的分布所经历的运动类型。这里假设肤色分布作为整体变化，即不能有任何局部的变形，这类似于假设全局的光照在图像的所有皮肤片上有均匀的表现。皮肤颜色的分布区域上 HSV (hue、saturation、value，即色相、饱和度和亮度) 参数随每帧时间间隔非线性平滑改变。为了进一步简化预测模型，限定该区域仅发生最重要的三种仿射转变：平移、旋转和缩放。

[0037] 每帧皮肤区域各像素点的色彩坐标 (h, s, v) 在 HSV 彩色空间组成一个紧凑的三维几何体 (离散的)，相邻两帧皮肤区域的这种三维几何体的演变 (由于光照变化和运动引起) 可以借助仿射变换的旋转、平移、缩放过程来描述，该仿射变换需要 8 个参数 $X = [T_H, T_S, T_V, S_H, S_S, S_V, \theta, \phi]^T$ 来刻画，其中， T_i 表示平移； S_i 表示旋转； θ, ϕ 表示几何体旋转的角度。

[0038] 为了获得所述三维仿射变换所需要的 8 个参数，通常采用的方法是通过选择一种现有的预测模型来对该仿射变换所需的 8 个参量进行预测。由于没有任何一种预测方法或预测模型会在各种情况下都比其他方法或模型表现得更好。因此，为了实时地估计和预测肤色分布状态，需要对其建立一个全新的预测模型。对复杂系统的预测仅仅应用一种预测方法，往往很难进行精确而可靠的预测。如果对同一预测问题采用各种不同的预测方法并加以适当的组合，则可以较充分地利用各种信息，达到提高预测精度的目的。因此，本实施例采用对现有的两个或者多个预测模型进行线性组合的方法来预测当前待预测帧的仿射变换所需要的 8 个参数。

[0039] 本实施例以采用维纳 (Weiner) 一步预测模型和二阶马尔科夫 (Markov) 预测模型为例来组成线性组合预测模型，以预测下一帧的 8 个参数。为了达到更好的预测效果，基于平方和误差最小原理将这两种预测算法进行线性组合，以构造一种新的线性组合预测模型。

[0040] 下面对维纳一步预测模型和二阶马尔科夫预测模型以及采用这两种预测模型所构造出的线性组合预测模型分别进行详细描述。

[0041] 1、维纳一步预测模型

[0042] 视频图像皮肤区域在 HSV 彩色空间随光照的帧间变化是符合布朗运动规则的，而布朗漫步运动可用稳态高斯过程描述。因此，本实施例选用维纳一步预测方法来预测当前待预测帧的仿射变换所需要的 8 个参数。将仿射变换中 8 维列向量 X_n 的任一分量表示为

$x(n)$, 可通过维纳一步预测模型预测 $x(n)$ 随光照的变化。维纳一步预测的原理是由过去 p 个观测数据 $x(n-1), x(n-2), \dots, x(n-p)$ 的某种线性组合, 来预测当前数据 $x(n)$, 用公式表示为:

$$[0043] \quad \hat{x}(n) = -\sum_{k=1}^p a_{pk} x(n-k) \quad (1)$$

[0044] 如果令 $a_{pk} = -h(k)$, 则 $\hat{x}(n) = \sum_{k=1}^p h(k)x(n-k) = x(n) * h(n)$ 由此可见, 维纳一步预测器对应一个长度为 p 的 FIR 滤波器, 其单位脉冲响应即为 $h(n)$ 。

[0045] 误差函数 $e(n)$ 表达式为:

$$[0046] \quad e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^p a_{pk} x(n-k) = \sum_{k=0}^p a_{pk} x(n-k) \quad a_{p0} = 1 \quad (2)$$

[0047] 均方误差 $E[e^2]$ 为:

$$[0048] \quad E[e^2] = E[x(n) - \hat{x}(n)]^2 = E[x(n) + \sum_{k=1}^p a_{pk} x(n-k)]^2 \quad (3)$$

[0049] 为求出使 $E[e^2(n)]$ 最小的 $\{a_{pk}\}$, $k = 1, 2, \dots, p$, 将上式对各 a_{pk} 求偏导数, 并令其等于 0, 得

$$[0050] \quad \frac{\partial E[e^2]}{\partial a_{pl}} = 2E \left[e \frac{\partial e}{\partial a_{pl}} \right] = 2E[ex(n-l)] = 0 \quad l = 1, 2, \dots, p \quad (4)$$

[0051] 整理可得:

[0052]

$$\begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) & \cdots & \phi_{xx}(p) \\ \phi_{xx}(1) & \phi_{xx}(0) & \cdots & \phi_{xx}(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(p) & \phi_{xx}(p-1) & \cdots & \phi_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{p1} \\ \vdots \\ a_{pp} \end{bmatrix} = \begin{bmatrix} E[e^2(n)]_{\min} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (5)$$

[0053] 此式称为 Yule-walker 方程, 它反映了相关函数与最佳单位脉冲响应之间的关系。

[0054] Yule-walker 方程的解就是维纳预测器的系数, 也就是 FIR 数字滤波器的单位脉冲响应 $h(n)$, 此时维纳预测器的输出是信号的最佳一步预测。其中, Yule-walker 方程可通过 Levinson 算法求解。

[0055] 2、二阶 Markov 预测模型

[0056] 记 X_n 为已知的第 n 帧的 8 维列向量 $X = [T_H, T_S, T_V, S_H, S_S, S_V, \theta, \phi]^T$, \bar{X} 为 X_n 的期望值, X_{n+1} 为要预测的第 $n+1$ 帧的 8 维列向量, A_0, A_1 是 8×8 矩阵, B 为 8×1 矩阵, ω_n 为随机部分。二阶马尔科夫预测模型用公式表示为:

$$[0057] \quad X_{n+1} = A_0 X_{n-1} + A_1 X_n + (1 - A_0 - A_1) \bar{X} + B \omega_n \quad (6)$$

[0058] 二阶马尔科夫预测模型共有三个参数 A_0, A_1 和 B 。矩阵 A_0, A_1 表示帧间的相关性, 决定了运动模型的确定部分; 而 B 决定随机部分, 随机部分通过标准高斯分布来模型化, 它通过加权系数 B 加入到 X_{n+1} 中。所以, 可将公式 (6) 变形为:

$$[0059] \quad (X_{n+1} - A_0 X_{n-1} - A_1 X_n) \sim N(0, |B|^2) \quad (7)$$

[0060] 采用 MLE 算法估计这些参数,并且在估计过程中并不直接求解 B,而是估计 $C = BB^T$ 。在长度为 m 帧的观察窗内,前 m-2 帧用来估计参数 A_0 、 A_1 和 B,第 m-1 和第 m 帧与要预测的第 m+1 帧相关性强,用来预测第 m+1 帧的 8 维参数向量 X_{m+1} 。前 m-2 帧的仿射变换参数向量 $(X_{n+1}-A_0X_{n-1}-A_1X_n)$ 的联合概率密度函数为:

$$[0061] \quad L(X_1, \dots, X_n | A_0, A_1, B) = -\frac{1}{2} \sum_{n=1}^{m-2} |B^{-1}(X_{n+1} - A_0 X_{n-1} - A_1 X_n)|^2 - (m-2) \log |B| \quad (8)$$

[0062] 将上式分别对 A_0 、 A_1 求偏导,并令偏导为零,得出下面的等式:

[0063]

$$S_{20} - \hat{A}_0 S_{00} - \hat{A}_1 S_{10} = 0$$

[0064]

$$S_{21} - \hat{A}_0 S_{01} - \hat{A}_1 S_{11} = 0 \quad (9)$$

[0065] 其中,

$$[0066] \quad S_{ij} = \sum_{n=1}^{m-2} X_{(n-1)+i} X_{(n-1)+j}^T, \quad i, j = 0, 1, 2$$

$$[0067] \quad (10)$$

[0068] 从而求出 A_0 、 A_1 。

[0069] 将公式 (9) 对 B 求偏导,令偏导为零,并将求出的 A_0 、 A_1 代入得:

[0070]

$$\hat{C} = \frac{1}{m-2} Z(\hat{A}_0, \hat{A}_1) \quad (11)$$

[0071] 其中,

$$[0072] \quad Z(A_0, A_1) = S_{22} + A_1 S_{11} A_1^T + A_0 S_{00} A_0^T - S_{21} A_1^T$$

$$[0073] \quad -S_{20} A_0^T + A_1 S_{10} A_0^T - A_1 S_{12} - A_0 S_{02} + A_0 S_{01} A_1 \quad (12)$$

[0074] 如果将三维仿射变换的 8 个参数视为独立的变量,则估计动态模型的参数仅需要较少的观察帧 (m)。本系统至少需要 4 帧,但是,帧数越多估计得越好。在本实施例中,优选取 8 到 30 帧效果较好。最后将前 m-2 帧估计得到的参数 A_0 、 A_1 、B 和第 m-1、m 帧的两个向量 X_{m-1} 、 X_m 代入公式 (6),即可计算出要预测的第 m+1 帧的 8 维参数向量 X_{m+1} 。

[0075] 3、线性组合预测模型

[0076] 建立线性组合预测模型的一个重要步骤是对组合权重进行优化,其目的是通过选择组合权重向量,使线性组合预测模型的精度满足预测对象的要求。常见的预测模型精度综合评价指标有:平方和误差 (SSE)、平均误差 (MAE)、均方误差 (MSE)、平均百分比误差 (MAPE)、均方百分比误差 (MSPE) 等。本实施例从线性组合预测模型绝对误差序列出发,根据统计学思想对线性组合预测模型进行精度分析,以选择合适的组合权重。

[0077] 用 $f_{t,i}$ $i = 1, 2$ 表示组合中的 t 时刻第 i 个预测模型,则线性组合预测模型可以表示为 $f_t = k_1 f_{t,1} + k_2 f_{t,2}$, 其误差为 $e_t = \sum_{i=1}^2 k_i e_{t,i}$, 其中, k_i 为第 i 种预测模型的组合权重。设

t 时刻第 i 个预测模型的预测绝对误差为 $e_{t,i} = f_{t,i} - y_t$ ($i = 1, 2$)。线性组合预测模型 f 的精度信息包含在绝对误差序列 e_1, e_2 中。

[0078] 为使线性组合预测模型绝对误差序列分布为优化模型,本实施例以平方和误差

(SSE) 为精度综合评价指标, 观察窗长度为 N, 即

$$\begin{aligned}
 [0079] \quad SSE &= \frac{1}{N} \sum_{i=1}^N e_i^2 = \frac{1}{N} \sum_{i=1}^N \left[\sum_{j=1}^2 k_j e_{i,j} \right]^2 = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} e_{i,1} \\ e_{i,2} \end{bmatrix} \begin{bmatrix} e_{i,1} & e_{i,2} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \\
 [0080] \quad &= \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} e_{i,1}^2 & e_{i,1}e_{i,2} \\ e_{i,1}e_{i,2} & e_{i,2}^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} k_1 & k_2 \end{bmatrix} A_m \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \frac{1}{N} K^T A_m K
 \end{aligned} \tag{13}$$

[0081] 满足平方和误差 (SSE) 的 $K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ 即为线性组合预测模型的最优组合权重, 即通过求解下式:

$$\begin{aligned}
 [0082] \quad \min(SSE) &= \frac{1}{N} \min(K^T A_m K) \\
 [0083] \quad st \quad &\begin{cases} \sum_{i=1}^2 k_i = 1 \\ k_i \geq 0 \quad i=1,2 \end{cases}
 \end{aligned} \tag{14}$$

[0084] 便可以得到 $K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ 。

[0085] 由于 A_m 为对称正定矩阵, 可进行 Cholesky 分解 $A_m = BB^T$, 其中, B 为下三角矩阵, 公式 (13) 可写为 $\min(SSE) = \frac{1}{N} \min(K^T B^T BK) = \frac{1}{N} \min(\|BK\|_2^2)$ 可见最优解 K 即为满足一

定条件的方程 $BK = 0$ 的最小二乘解, 即 $K \in B\{1,4\}$, $B\{1,4\}$ 表示 B 的广义 $\{1,4\}$ -逆。

[0086] 根据以上的分析, 如果对每帧的单项预测模型进行无偏性调整, 并不影响组合权重向量的计算。为了叙述方便, 用 est_x_LPC 表示维纳一步预测模型得到的预测信号, est_x_MLE 表示二阶马尔可夫预测模型得到的预测信号。无偏性调整的过程可以用信号的值减去信号的均值, 利用 $mean()$ 表示求解信号均值的函数, 则

$$\begin{aligned}
 [0087] \quad e_LPC &= x - est_x_LPC; \\
 [0088] \quad e_MLE &= x - est_x_MLE; \\
 [0089] \quad f1 &= e_LPC - mean(e_LPC); \\
 [0090] \quad f2 &= e_MLE - mean(e_MLE)
 \end{aligned}$$

[0091] 显然, f_1 和 f_2 代表无偏性调整后的误差。

[0092] 根据式 (14) 研究如何求解组合权重向量。 R_m 为元素全为 1 的 m 维行向量, E 可根据式 (13) 得到。具体到组合预测模型只用到了两种基本的预测模型, 则

$$\begin{aligned}
 [0093] \quad E &= \begin{pmatrix} f_1 & f_2 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \\
 [0094] \quad R_m = R_2 &= (1 \quad 1)
 \end{aligned} \tag{15}$$

[0095] 组合权值的计算表示为:

$$[0096] \quad W_0 = (inv(E) * R_m') / (R_m * inv(E) * R_m'); \tag{16}$$

[0097] 其中, $inv()$ 返回一个矩阵的逆矩阵:

$$[0098] \quad E^{-1} = inv(E) \tag{17}$$

[0099] 此时, 已经求出了加权系数 W_0 , 线性组合预测模型的预测信号只需要由两种基本

模型的预测信号线性加权求和即可得到：

$$[0100] \quad \text{est_x_LIN} = W_0(1) * \text{est_x_LPC} + W_0(2) * \text{est_x_MLE}。 \quad (18)$$

[0101] 第 t+1 时刻输入帧的前景皮肤区域在 HSV 彩色空间分布的 8 维参数向量 X_{n+1} 通过上述线性组合预测模型 (18) 预测得到后, 利用三维仿射变换和直方图统计可求出该帧的预测直方图 $\text{Prdct_H}_{h, s, v}(t+1)$ 。为考虑直方图的关联性, 第 t+1 时刻输入帧分割的直方图应为当前帧和预测结果 $\text{Prdct_H}_{h, s, v}(t+1)$ 的加权平均值：

$$[0102] \quad H_{h, s, v}(t+1) = (1-a_{fg})H_{h, s, v}(t) + a_{fg} \cdot \text{prdct_H}_{h, s, v}(t+1) \quad (19)$$

[0103] a_{fg} 在 0 到 1 之间取值, 表示更新的速度。当 $a_{fg} = 0$ 说明无适应, $a_{fg} = 1$ 说明直方图无记忆地全更新。严格来说, 背景和皮肤颜色分布都要更新适应, 但当摄像机固定不动或缓慢移动时, 可以认为背景分布变化缓慢, 因而认为背景分布的改变远小于前景分布的改变。这里只对前景颜色的变化预测, 对背景直方图分布的更新是用过去两帧的背景直方图的加权平均得到：

$$[0104] \quad H_{h, s, v_bg}(t+1) = (1-a_{bg})H_{h, s, v_bg}(t-1) + a_{bg}H_{h, s, v_bg}(t) \quad (20)$$

[0105] 最佳更新常数的选择一般通过经验得到, 通过实验表明: 对于半适应的系统, 取 $a_{fg} = 0.2$ 、 $a_{bg} = 0.10$ 效果最好。

[0106] 在获得了当前待分割帧的直方图分布后, 统计每个直方图条中某一颜色 (rgb) 像素的数目以及前景皮肤和背景的像素的总数, 可以计算得到各类概率密度: $p(fg)$ 、 $p(bg)$ 、 $P(rgb|fg)$ 和 $P(rgb|bg)$, 进而利用 Bayes 分类器对当前帧进行皮肤区域的分割。

[0107] 下面具体阐述一下如何通过 Bayes 分类器来分割当前帧中的皮肤区域。

[0108] 针对当前帧图像中任意一点像素的颜色 rgb, 通过上述颜色分布直方图得到的是条件概率 $P(rgb|fg)$ 和 $P(rgb|bg)$, 而在皮肤分割时更关心的是 $P(fg|rgb)$ 和 $P(bg|rgb)$, 利用贝叶斯公式得：

$$[0109] \quad P(fg|rgb) = \frac{P(rgb|fg) \cdot P(fg)}{P(rgb|fg) \cdot P(fg) + P(rgb|bg) \cdot P(bg)} \quad (21)$$

$$[0110] \quad P(bg|rgb) = \frac{P(rgb|bg) \cdot P(bg)}{P(rgb|fg) \cdot P(fg) + P(rgb|bg) \cdot P(bg)} \quad (22)$$

[0111] 令 $P(fg|rgb)$ 和 $P(bg|rgb)$ 的比率称为分类界限 K, 是误分类的极限值, 则

$$[0112] \quad K < \frac{P(fg|rgb)}{P(bg|rgb)} = \frac{P(rgb|fg)P(fg)}{P(rgb|bg)P(bg)} \quad (23)$$

[0113] 计算得：

$$[0114] \quad K \times \frac{1 - P(fg)}{P(fg)} < \frac{P(rgb|fg)}{P(rgb|bg)} \quad (24)$$

[0115] 其中, $P(fg)$ 是图像中任意像素可能为皮肤的概率。很明显, 这个概率随图像不同而改变, 但利用一个足够大的数据集就可得到一个最佳的概率。用整个图像数据库作为数据集, 则 $P(fg)$ 可以表示为：

$$[0116] \quad P(fg) = \frac{N_{foreground}}{N_{foreground} + N_{background}} \quad (25)$$

[0117] 其中, $N_{foreground}$ 是前景直方图中像素的总数, $N_{background}$ 是背景直方图中像素的总数。

在基于人脸的视频监控系统中,一般要求 $P(fg)$ 不小于 0.08。极限 K 的选择应使正确分类达到至少 85%,并且错误警报少于 25%,根据大量实验获得:极限取 $K = 0.067$ 。

[0118] 上面所介绍的像素分类的结果是一个二进制的图像模板,当 (24) 式成立时,该像素位置判为 1,与前景皮肤像素相对应;否则,判为 0,与背景像素相对应。为了将噪声影响最小化,在二进制模板进入学习阶段前对其采用形态学滤波器进行滤波处理。

[0119] 第三步,利用人体肤色的生物特性对图像进行再分割,并结合形态学滤波器消除相近色干扰,修复人脸图像。

[0120] 在文献 G.Gomez, M.Sanchez, and L.Enrique Sucar, "On Selecting an Appropriate Colour Space for Skin Detection", MICAI 2005, LNAI 2313, pp.69-78, 2002. 中,对 Compaq 数据库中 12000 个具有肤色和非肤色图像进行统计分析发现:人体肤色在彩色空间 (R, G, B) 中,具有更高成分的红色 (Red) 分量,尤其是 R/G 的比值更为突出,它明显地区别于非肤色区域。而且,该比值随光照不会发生明显的变化。文献中所揭示的图形数据显示肤色 (R, G, B) 分量之间的比值稳定在: $1.1 \leq R/G \leq 3, 1.0 \leq G/B \leq 1.3$; 而非肤色区域三个分量的比值关系一般为: $0.8 \leq R/B \leq 12, 0.8 \leq R/G \leq 1.4, 0.5 \leq G/B \leq 1.7$ 。可见,肤色区域的分布相对较集中,而非肤色区域的分布更广泛,具有更大的随机性。

[0121] 通常情况下分割是在单独的肤色空间,如 RGB, YCBCR, HSV 等空间中,根据肤色聚类性设定合适的阈值来分割肤色。但是,单独在每个彩色空间下的分割都有其特定的局限性,上述在 RGB 空间虽然肤色与非肤色大致占据不同的区域,但它们仍有一部分重合,并且该空间对光照变化比较敏感。文献对常用的彩色空间 HSV、YIQ、RGB-Y、YES、YUV、CMY、CIE、XYZ、GB 等进行了分析比较后认为,没有一个空间能很好地将肤色和非肤色区域彻底分开,因而提出了一种组合彩色空间的概念,该三维空间的三个分量是互补的,它们的组合能够最大限度地 将肤色和非肤色区域分开,即尽可能地减少两者的重叠区域,以便更好的分割肤色。

[0122] 该文献提出的组合彩色空间为 (E, R/G, H), 其中, E 分量来自彩色空间 (Y, E, S); R/G 是 (R, G, B) 空间中 R 和 G 分量的比值; H 分量来自 (H, S, V) 空间。彩色空间 (Y, E, S) 各分量的含义如下: Y 是亮度, E 和 S 是色度空间, 其中, $E = 0.5 * (\text{red} - \text{green})$; 彩色空间 (H, S, V) 各分量的含义如下: 分量 H 和 S 是色度, V 是亮度。文献中给出了它们的取值范围, 分别为: $E > 13.4224, R/G < 1.7602, H < 23.89$ 。其中, E 和 R/G 两个分量充分考虑了人体肤色中的红色分量占据主导位置的这一生物特性, 两个分量联合起来表明: $20 * \text{green} < \text{red} < 1.7602 * \text{green}$ 。文献后面对 H 分量的取值做了进一步调整: $-17.4545 < H < 26.6666$, 这样得到的彩色空间对室内和室外采集到的图像进行实验, 肤色分割识别率分别为 97.95% 和 93.74%。

[0123] 总之, 该文献所提出的组合彩色空间 (E, R/G, H) 比常见的任何单独彩色空间相比, 具有将肤色和非肤色分离的特性。因此, 在上述贝叶斯初分割的基础上, 本实施例采用该文献所提出的方法, 利用人体肤色的生物特征在复合彩色空间 (E, R/G, H) 下进行再分割, 即建立 E, R/G, H 空间模型, 利用前面提到的给定的参数阈值: $E > 13.4224, R/G < 1.7602, -17.4545 < H < 26.6666$, 来进一步消除背景中肤色相近色物体的干扰。

[0124] 利用文献所提出的方法对图像进行再分割后, 可有效克服初分割结果中大片相近

色的干扰,但是同时也会将原本属于皮肤区域部分的一些像素当作背景去除掉,使输出的人脸图像产生空洞。基于此,本实施例将经过再次分割后产生的图像输入形态学滤波器,利用邻域相关性特点将人脸修复成完整图像,并去除掉预留的背景像素。

[0125] 第四步,检测通过所述形态学滤波器输出的皮肤图像中是否有人脸,若没有,则返回第一步;否则,执行下一步。

[0126] 第五步,人脸定位与性别识别。

[0127] 在获得了人脸的数据信息后,可以通过目前的多种模式识别方法来实现对性别的分类,比如基于神经网络的机器学习方法、最邻近分类器或者支持向量机 SVM 等。在本实施例中,优选采用 SVM 对人脸进行性别识别。

[0128] SVM 是建立在 SLT 的 VC 维理论和结构风险最小化原理的基础上,根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和泛化能力(即无错误地识别任意样本的能力)之间寻找最佳折衷,以期获得最好的推广能力。SVM 能够有效地避免经典学习方法中出现的过学习、欠学习以及陷入局部极小点等诸多问题。SVM 将输入空间线性不可分问题,通过非线性变换转化为某个高维空间中的线性可分的问题,在高维的变换空间中寻找最优分类面,通过核函数巧妙的在输入空间实现高维空间中的内积运算实现分类,而且计算复杂度并没有增加,SVM 是针对二分类问题的强分类器,而人脸性别分类问题是典型的二分类问题,因此,本实施例采用 SVM 作为性别判别器。

[0129] 将检测到的人脸图像的尺寸进行规范化,定位出 21×18 的人脸图像,然后转变为每个向量的维数为 378×1 的向量。将这些特征向量送入训练好的支持向量机 SVM,调用识别函数 svm-predict,进而实现对人脸图像的性别测试。

[0130] 所述识别函数 svm-predict 的输入为检测出的人脸区域经过插值下采样变为规格大小的人脸图像,即 21×18 的人脸图像,并将其展开为像素值的向量作为测试样本,即 378×1 的向量。此外,还需要将训练生成的支持向量机模型送入 svm-predict 函数。函数的输出为类别标号,如: +1 代表男性, -1 代表女性,此处的标号应当是与训练时规定的一致。对各种不同的视频图像进行人脸检测与性别识别,识别率比较如表 1 所示。

[0131]

视频序列名称		正确率	错误率
1	有少量相近色干扰时的背景下 300 帧图像序列	90%	10%
2	有典型背景干扰色干扰时单人正面 376 帧视频	100%	0%
3	单人 159 帧无相近色干扰时人脸转动的视频	96%	4%
4	单人简单背景人脸正面表情丰富时的视频	100%	0%
5	单人表情丰富且多处肤色 170 帧视频	95%	5%
6	双人有背景干扰色干扰时的识别率	83%与 70%	17%和 30%
7	双人无干扰色干扰时的识别率	93%与 85%	7%和 15%

[0132] 表 1

[0133] 本发明的人体目标检测方法将人的面部皮肤特征作为人体的主要标识,对存在于视频图像中的人体目标进行检测,从而克服了光照变化和背景相近色干扰所产生的不利影

响,实现了在视频监控系统中人体目标的准确检测以及性别的正确识别。当然,上述说明并非是对本发明的限制,本发明也并不仅限于上述举例,本技术领域的普通技术人员在本发明的实质范围内所做出的变化、改型、添加或替换,也应属于本发明的保护范围。

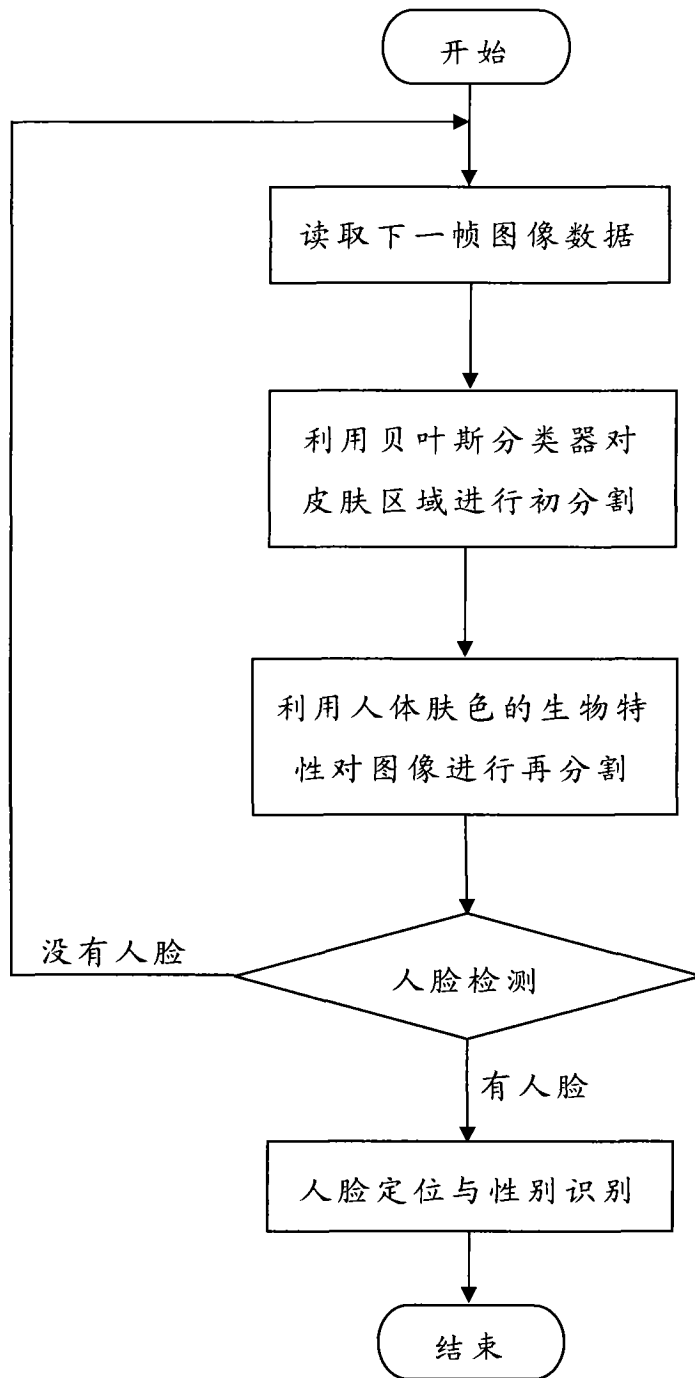


图 1