极低码率下的 2D-3D 人脸视频编解码

於 俊,汪增福
(中国科学技术大学自动化系,安徽合肥 230027)

摘 要: 针对动态变化背景下的人脸视频编解码问题,该文提出了一种 2D-3D 混合编解码系统;具体包括:(1)基于多种观测信息,在线外观模型和粒子滤波的人脸三维运动跟踪;(2)结合参数模型与肌肉模型的 3D 人脸动画合成;(3)基于头发检测和 3D 头发模型的头发合成;(4)无缝地拼接前景的三维编解码结果和背景的二维编解码结果.在极低码率下,客观实验表明,该系统在编码效率和解码质量上有较好的综合优势.主观实验表明,该系统的解码结果在脸部具有较高的辨识度.

关键词: 视频编解码; 人脸运动跟踪; 人脸动画; 头发分析与建模
中图分类号: TB391.9 **文献标识码:** A **文章编号:** 0372-2112 (2013)01-0185-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.01.032

2D-3D Facial Video Coding/Decoding at Ultra-low Bit-rate

YU Jun, WANG Zeng-fu
(*Department of Automation, University of Science & Technology of China, Hefei, Anhui* 230027, *China*)

Abstract: For facial video coding/decoding with dynamic background, a 2D/3D mixed system was proposed, it includes: (1) multi-measurements and online appearance models were applied to track 3D facial motion from video in framework of particle filter; (2) 3D facial animation was produced combining parameterized model and muscular model; (3) 3D hair was synthesized based on hair detection and 3D hair model; (4) 3D coding/decoding result of foreground and 2D coding/decoding result of background were stitched seamlessly. At ultra-low bit-rate, the object experiment confirmed the comprehensive advantage between coding efficiency and decoding quality of this system, the between subjects experiment indicated the suitability of subjective face identification by it.

Key words: video coding/decoding; facial motion tracking; facial animation; hair analysis and modeling

1 引言

传统的视频编码方法(H.264 等)通常把图像作为随机信号进行处理,并根据统计特性消除其空间冗余度与时间冗余度.与此不同,三维模型基视频编解码(MBC)[1~6]把图像看作具有知识的结构化模型.该知识可从输入视频中实时识别或从训练库中学习来得到.工作过程如下:假设编、解码端有描述处理对象的三维结构模型;首先,在编码端,通过对象分割,跟踪来获得形状,纹理和运动方面的特征参数;然后,将之编码传送至解码端;最后,在解码端,根据三维结构模型和解码后的特征参数综合出解码视频.由于仅需对特征参数进行编码传输,MBC 的效率极高且没有传统方法在码率较低时出现的块效应.理论上,MBC 适合于任何对象,但从三维结构模型和应用场合考虑,目前以三维模型基人脸视频编解码(MBC-Face)为主.

1.1 单视频下的人脸三维运动跟踪

它可以通过基于特征的方法或基于外观模型的方法来完成.前者[7]基于由光流或块匹配获得的二维运动信息来估计三维模型的运动.由于人脸没有足够稳定的特征以及当前帧的运动估计依赖于以前各帧的结果,从而随着估计误差的积累,存在运动漂移问题.为了克服该问题,需要引入关键帧等措施[8~10]或者融合多种特征信息来提高跟踪的鲁棒性[11].后者通过建立和更新外观模型来匹配当前帧,没有运动漂移问题,但对模型的健壮性和灵活性有较高要求.它也有确定性和统计性之分.前者[12,13]在视频开始捕捉参考纹理作为跟踪对象,然后根据估计得到的运动参数不断对参考纹理进行几何变形以更新跟踪对象.它对于光照和表情变化下的人脸对象效果有限.后者进一步有离线和在线之分.离线方法[14~16]利用目标的统计及先验知识来构造模型,

收稿日期:2012-03-10;修回日期:2012-08-02
基金项目:国家自然科学基金(No.60875026, No.60805019);安徽省科技攻关计划语音专项(No.11010202192);中国博士后科学基金(No.2012M521248)

鲁棒性强,但训练复杂且无法适应新情况.在线方法[17,18]通过在线不断学习目标的变化情况,能够学习到离线模型没有学习到的新情况,故较离线方法有更大优势,但如何更新模型以适应目标的变化以及应对光照和个体相关性的影响是难点.针对于此,一种途径[19,20]是挖掘在线外观模型的潜力;另一种途径[21]是融合形状等多种信息.近期,G-模型[22]将上述方法纳入到一个通用框架来解释.

1.2 运动滤波策略

人脸运动是非线性非高斯分布的,故需要滤波算法来提取出真实状态.在优化问题中,粒子滤波(PF)[20]因其具有的全局优化特性及对模型的弱依赖性,得到了广泛应用.然而,在人脸三维运动跟踪中采用 PF 面临着挑战:(1)计算盲目.在更新粒子过程中没有考虑状态的最新值,由此产生的粒子与真实后验概率产生的粒子偏差较大;(2)计算量大.为了达到较高的估计精度,需要大量的粒子来模拟状态的后验分布;(3)粒子退化.因为选择的观测似然度难以反映真实情况,所以根据它得到的粒子权值和重采样难以有效控制粒子退化.

1.3 人脸动画

参数模型[23]通过对表情参数和构造参数的控制,合成人脸动画.它简单实用,但真实感有待改进;物理模型[24]将人脸动画建模为动态质点-弹簧系统,真实感较强,但计算量太大且参数设置复杂;学习模型[25]运用统计分析方法从人脸图像库中合成期望的人脸,真实感较强,但合成结果受限于训练库的描述力且代价昂贵;肌肉模型[26,27]基于人脸生理知识来合成人脸动画.如果肌肉模型建立得足够好,可以足够表示人脸运动.但目前的建模复杂度和真实感有待改进.

1.4 头发与模型基视频编解码领域的契合

目前 MBC-Face 的相关研究均忽略了头发的作用,这是因为对人脸的运动跟踪和动画合成没有得到较好解决的缘故.由于头发对于人脸的真实感表现具有不可替代的作用[28],如何在 MBC-Face 中加入头发是亟待解决的问题.

1.5 背景和前景的拼接

如何实现对背景的编解码是左右 MBC-Face 系统能否在实际中获得广泛应用的一个重要因素[29,30].因为背景的复杂性,无法像人脸那样用结构化模型来描述,所以目前对它的处理多采用传统编解码方法,但如何将之与前景的 MBC-Face 无缝地结合以构成混合系统是难点.

2 系统框架和创新点

图 1 是系统框架.(1)编码端:在首帧,首先进行人脸模型特定化,接着根据特定化的结果检测头发区域.在后续帧,首先进行人脸运动跟踪,接着在姿态变化较大的时候更新人脸/头发的纹理,然后采用 H.264 对背景(人脸和头发以外的区域)进行编码,并获取区分前景(人脸和头发)和背景的判别信息.(2)信道:在首帧只传输人脸模型特定化和头发区域检测的结果.在后续帧传输人脸运动参数,人脸/头发更新纹理,判别信息,背景编码信息.(3)解码端:在首帧,首先根据特定化结果特定化人脸模型,接着根据头发区域检测结果特定化头发模型.在后续帧,首先根据人脸运动参数,人脸/头发更新纹理进行人脸动画,接着根据 H.264,判别信息和背景编码信息对背景进行解码.

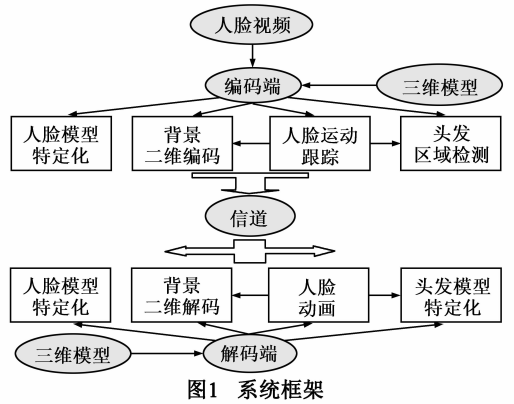
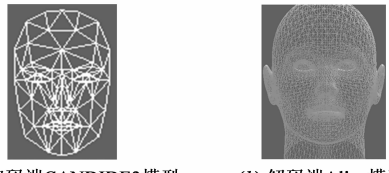


图1 系统框架

较当前 MBC-Face 系统[29,30]和传统方法(H.264),该系统具有以下特点:(1)对象包含 3D 人脸,3D 头发,2D 背景;(2)在极低码率下,具有较高的率失真,且解码结果在脸部具有较高的辨识度.

3 编、解码端:人脸三维模型

在编码端,基于实时性,采用简单有效的 CANDIDE3[31]三维模型(图 2(a)).对于该模型中所有顶点坐标的级联向量 $\boldsymbol{g}$,可由式(1)来表示:



(a) 编码端CANDIDE3模型 (b) 解码端Alice模型

图2

$$\boldsymbol{g} = \bar{g} + S\boldsymbol{\sigma} + A\boldsymbol{\alpha} \tag{1}$$

$\bar{g}$ 是模型的标准形状,S 和 A 的列是形状单元和表情单元,$\boldsymbol{\sigma}$ 和 $\boldsymbol{\alpha}$ 分别表示形状和局部运动参数.

接着,根据文献[14],定义全局运动参数为:

$$\boldsymbol{z}_g = [\theta_x, \theta_y, \theta_z, s, t_x, t_y]^{\mathrm{T}} \tag{2}$$

最后,综合以上两类参数定义人脸运动参数为:

$$\boldsymbol{b} = [\boldsymbol{z}_g^{\mathrm{T}}, \boldsymbol{\sigma}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}]^{\mathrm{T}} = [\theta_x, \theta_y, \theta_z, s, t_x, t_y, \boldsymbol{\sigma}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}]^{\mathrm{T}} \tag{3}$$

进一步,在后续的人脸运动跟踪中,b 可简化为:

$$\boldsymbol{b}=[\theta_x,\theta_y,\theta_z,s,t_x,t_y,\boldsymbol{\alpha}^{\mathrm{T}}]^{\mathrm{T}} \tag{4}$$

在解码端,基于可控性和表达力,采用网格点数目较多的 Alice[32] 人脸三维模型(图 2(b))。该模型的表面细腻平滑,且包括皮肤,眼睛,牙齿,舌头等器官,从而对于人脸运动的描述力较强.

4　编、解码端:人脸模型特定化

在编码端.对于包含正面人脸的第一帧图像:(1)由 Adaboost[33] 来检测人脸,AAM[15,34] 来定位人脸特征点;(2)由数个图像特征点来确定全局运动参数[35];(3)在(2)的基础上对所有图像特征点进行最小二乘拟合得到局部非刚体运动参数以及特征网格点的位移;(4)根据特征网格点的位移,进行径向基插值[36]得到其它网格点的位移.

在解码端.(1)根据 Alice 模型和编码端传来的人脸运动参数,得到特征网格点的位移;(2)根据(1)的结果,进行径向基插值得到其他网格点的位移.

5　编码端

5.1　人脸三维运动跟踪

5.1.1　提取多种观测量

观测量从对应于输入人脸图像的几何归一化人脸图像(GNI)[14]中提取.GNI 由以下过程(图 3)获得:(1)三维模型根据人脸运动参数投影到像平面上得到投影坐标;(2)令(1)的坐标为纹理坐标,将输入人脸图像纹理映射到三维模型上,三维模型投影到像平面上得到投影图像;(3)标准形状以垂直正面姿态投影到像平面上得到二维三角网格,将(2)的投影图像分段仿射变换到该二维三角网格的内部.另外,把它的额头部分去掉来避免头发的干扰;对颜色值去均值/归一化来减少光照的影响;为了检测眨眼幅度,使它的眼睛处于闭合状态.

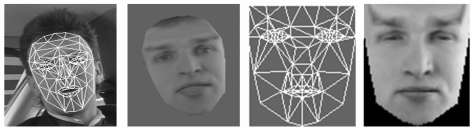


图3　GNI的获取过程

本文从 GNI 中提取两部分外观观测量:颜色值和对光照比图像进行 Gabor 小波变换得到的系数.

第二部分观测量的获取依据和过程如下:光照比图像[21]与表面反射性质无关,故它刻画的是人脸的共性部分,从而消除了个体相关性,而且 WEN Z[21]证明了它的高频分量受光照影响较小;另外,Gabor 小波能够捕捉对应于空间尺度,位置及方向的局部结构信息.鉴于此,本文将当前帧的 GNI 和首帧的 GNI 之间的光照比图像的 Gabor 小波变换系数(8 个方向,5 个尺度)作为第二部分的观测量.为了克服噪声的影响,只采用系数幅值.

5.1.2　构建在线外观模型与融合多种观测量

(1)构建第一个在线外观模型

设大小为 d 的 $y(b_t)$,缩写为 y_t,为第一部分观测量的级联向量表示形式.如文献[20]中的 W,S,L 分量,我们将 y_t 建模为 3 个分量的混合高斯模型(观测模型),且令 $\{\mu_{i,t};\ i=s,w,l\}$ 为高斯均值向量,$\{\Sigma_{i,t};i=s,w,l\}$为方差矩阵,$\{m_{i,t};\ i=s,w,l\}$为权值,$\{\sigma_{i,t};i=s,w,l\}$为$\{\Sigma_{i,t};i=s,w,l\}$对角元素的平方根所组成的向量.观测似然度为 $p(y_t/b_t)=\prod_{j=1}^{d}\{\sum_{i=s,w,l}m_{i,t}(j)N(y_t(j);\mu_{i,t}(j),\sigma_{i,t}^2(j))\}$.

定义在时刻 t 的在线外观模型 A_t 代表了到时刻 $t-1$为止 $y_{1:t-1}$的变化过程.为了使它能够跟踪目标的变化,定义$\{m_{i,t};i=s,w,l\}$和稳定分量(S 分量)$\mu_{s,t}$,$\sigma_{s,t}$的更新方程,$\beta=0.2$ 是遗忘因子:

$$\begin{aligned}
m_{i,t}&=(1-\beta)m_{i,t-1}+\beta m_{i,t-1}N(y_{t-1};\mu_{i,t-1},\sigma_{i,t-1}^2)\\
\mu_{s,t}&=(1-\beta)\mu_{s,t-1}/m_{s,t}+\beta y_{t-1}m_{s,t-1}/m_{s,t}\\
\sigma_{s,t}^2&=(1-\beta)\sigma_{s,t-1}^2/m_{s,t}+\beta y_{t-1}^2m_{s,t-1}/m_{s,t}-\mu_{s,t-1}^2
\end{aligned} \tag{5}$$

(2)构建第二个在线外观模型

设 $G(b_t)$,缩写为 G_t,为第二部分观测量的级联向量形式.与 y_t 相似,同样将它建模为混合高斯模型,$p(G_t|b_t)$,$\{\mu_{i,t}',\Sigma_{i,t}',m_{i,t}',\sigma_{i,t}';i=w,s,l\}$,$A_t'$ 和更新方程的定义与上节类似.

(3)基于外观模型融合多种观测量

采用乘法方式进行融合:$p(y_t/b_t)\cdot p(G_t/b_t)$,该融合结果将在 5.1.4 节中被设定为粒子的权重.

5.1.3　应对遮挡的措施

首先给人脸三维模型中每个三角面片赋予与其它三角面片不同的颜色,接着根据人脸运动参数投影到屏幕上,然后在三角面片的屏幕重心坐标处读取屏幕缓存的颜色值,获取的值与赋给的值相同则该三角面片没有被遮挡,否则被遮挡.

被遮挡三维三角面片对应的 GNI 中的二维三角面片(图 4)是处理的对象.对于在每一次迭代搜索 b_t 后得到的 GNI 中第 k 个三角面片内的第 j 个像素位置,令 $Tr_t^{(k,j)}$ 为 GNI 在此处的颜色值或 Gabor 小波系数,$Tr\mu_{s,t-1}^{(k,j)}$,$Tr\mu_{s,t-1}^{'(k,j)}$,$Tr\sigma_{s,t-1}^{(k,j)}$,$Tr\sigma_{s,t-1}^{'(k,j)}$ 为 $\mu_{s,t-1}$,$u_{s,t-1}'$,$\sigma_{s,t-1}$,$\sigma_{s,t-1}'$在此处的值.当第 k 个三角面片被遮挡时,$Tr_t^{(k,j)}$ 由前一时刻的值和在线外观模型来估计:当 $Tr_t^{(k,j)}$ 为颜色值时,

$$\begin{cases} Tr_t^{(k,j)} = Tr_{t-1}^{(k,j)} + Tr\mu_{s,t-1}^{(k,j)} + Tr\sigma_{s,t-1}^{(k,j)}, & Tr_{t-1}^{(k,j)} \geq Tr\mu_{s,t-1}^{(k,j)} \\ Tr_t^{(k,j)} = Tr_{t-1}^{(k,j)} + Tr\mu_{s,t-1}^{(k,j)} - Tr\sigma_{s,t-1}^{(k,j)}, & Tr_{t-1}^{(k,j)} < Tr\mu_{s,t-1}^{(k,j)} \end{cases} \tag{6}$$

当 $Tr_t^{(k,j)}$ 为 Gabor 小波系数时,与上述方程类似.

5.1.4 改进的粒子滤波算法

(1)结合局部优化降低计算盲目性和计算量

局部优化可以结合当前时刻的最新观测值来生成与真实后验概率产生的粒子偏差较小的粒子,从而降低了计算盲目性,而且粒子数目可以根据局部优化的结果进行调整来降低计算量.鉴于此,本文在更新粒子权重之前增加局部优化:

$$\min_{b_t} e(b_t) = \min_{b_t} D(y_t, \mu_t) = \sum_{i=1}^{d} ((y_{i(t)} - \mu_{i(t)})/\sigma_{i(t)})^2 \tag{7}$$

从 b 的初始值(b_{t-1})出发,计算 b 的更新值:

$$\Delta b = -(GD_t^T GD_t)^{-1} GD_t^T r, r(b) = (y_{t-1} - \mu_t) \tag{8}$$

$GD_t = \partial r/\partial b$ 是梯度矩阵.进而得到新的 b, e:

$$b' = b + \lambda \Delta b, e' = e(b'), \lambda = 0.5 \tag{9}$$

如果 $e' < e$,则 b 如上更新直至本次匹配误差与上次匹配误差的差值在小阈值之内即收敛,否则换更小的 λ 直至收敛,如果超时不收敛则终止更新.

在得到 b_t 后,对当前帧摄动计算得到 GD_t,接着设置粒子数目正比于局部优化后得到的误差 e.

(2)改进的重采样

首先对粒子进行 PERM 采样[37]:权值 π_t^j 小于 π^- 的粒子按照概率 a 被接受,若被接受,新权值为 π_t^j/a;权值 π_t^j 大于 π^+ 的粒子 K 次重采样,新权值为 π_t^j/K;处于 π^- 和 π^+ 之间的粒子直接被保留,权值不变.然后对粒子进行标准重采样.这样做的原因是,少数权值很大而有可能不是真实人脸运动的粒子也被复制多次,但权值被自适应地减小,少数权值较小而有可能是真实人脸运动的粒子以一定的概率保留,且权值被自适应地增大.从而弥补了只单独使用标准重采样的不足.

另外,PERM 采样的参数进行在线调整:首先根据当前帧的人脸运动参数和人脸三维模型合成虚拟图像,然后计算它与当前帧像素值差的绝对值和.如果该值小则表示当前保留的粒子是好的,就弱化 PERM 采样:增大,减小;否则强化 PERM 采样:减小,增大.

5.1.5 眨眼幅度检测

在 GNI 对应的二维三角网格中,令上眼皮下端顶点的坐标和离它最近的下眼皮上端顶点的坐标等于这两个坐标的平均值,从而使得 GNI 中的眼睛处于闭合状态,那么学习到的外观模型的均值也是闭眼状态的.如果当前帧中的眼睛是张开的,那么根据从当前帧中提取的观测量与外观模型之间在眼部的差异,匹配检测出眨眼幅度.

5.2 头发区域的检测

(1)根据人脸运动跟踪的结果,将额头和脸颊区域确定为人脸肤色区域,将眉毛上方人脸轮廓附近的区域确定为头发的初始搜索区域[38].(2)首先由肤色区域中像素来建立肤色模型,接着由该模型得到头发初始搜索区域中头发像素的分布,然后以这些头发像素作为种子在当前搜索区域的上方寻找新的头发像素,同时根据新的头发像素来更新种子.(3)首先由(2)的结果确定头发周围物体的大致位置;然后将这些位置和(2)的结果输入到 Matting 算法[39]得到精确检测结果.

5.3 人脸和头发纹理的更新

首先由当前帧的人脸运动跟踪结果中绕 Y 轴的旋转角是否大于一定阈值来判断是否需要更新纹理.其次由人脸三维模型根据人脸运动跟踪结果投影到像平面上的投影来确定人脸外轮廓.再次对头发区域进行检测来确定头发外轮廓.然后根据人脸外轮廓和头发外轮廓得到它们的闭合区域.最后将闭合区域中的图像传输给解码端.

5.4 背景(2D)-前景(3D)的无缝拼接

在当前帧中:(1)首先进行人脸运动跟踪;(2)接着按照 H.264 的方式进行分块;(3)然后对于人脸和头发模型在图像平面上的投影区域以外的块,根据 H.264 进行编码;(4)最后对于处在背景和以上投影区域的边界上的块,仍然由 H.264 进行编码,并且对于其中像素获得如下判别信息:如果像素位置处于以上投影区域以外,该值为真,否则为假.

6 信道

(1)传输的编码信息有:只在首帧中传输的人脸模型特定化结果(人脸纹理和形状参数)和头发区域检测结果(头发纹理和轮廓信息);在后续帧中传输的人脸运动参数,人脸和头发的更新纹理,判别信息,背景编码信息;(2)人脸的特定化形状参数和头发的轮廓信息由熵编码进行压缩;(3)人脸和头发纹理根据 JPEG 方式进行压缩;(4)人脸运动参数按照 MPEG-4 中 FAP 的预测编码方式进行压缩;(5)采用均匀量化方式进行量化.

7 解码端

7.1 人脸动画

本文结合参数模型和肌肉模型来合成动画,前者采用径向基插值,后者采用 Waters 模型[26].

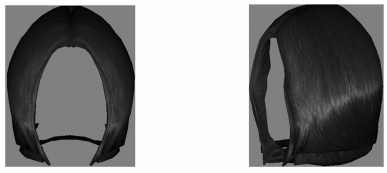
因为人脸中不同器官具有不同的运动特征,所以本文划分了人脸功能区,把网格点和人脸运动参数划分到各功能区,由各功能区去驱动它们.其中,网格点

分为以下三类:主特征点,次特征点和非特征点.主特征点是模型中与 MPEG-4 FDP 相对应的网格点;次特征点是在主特征点附近的肌肉带上选取的网格点;其他网格点统称为非特征点.在某一功能区,首先,主特征点的位移等于编码端传过来的人脸运动参数值;然后,根据主特征点的位移,利用 Waters 肌肉模型求出次特征点的位移,最后,根据主特征点的位移,利用径向基插值求出非特征点的位移.

Waters 肌肉模型对其影响范围内的所有网格点产生作用,导致计算量较大;而本算法的肌肉模型只对影响范围内的次特征点产生作用,非特征点的位移由参数模型求出,故大大减少了计算量.

7.2　三维头发模型特定化

首先在头发区域的检测结果上选取若干二维特征点,接着在三维头发模型(图 4)[32]上选取对应的三维特征点,然后采用径向基插值算法对三维头发模型进行调整,使得它在图像平面的投影与头发区域的检测结果相吻合.



(*a*) 三维头发模型的正面　(*b*) 三维头发模型的侧面

图4

7.3　人脸动画纹理映射图的更新

首先,求出当前帧与首帧的人脸全局运动参数的差值;接着,传输过来的更新纹理根据该差值变换到首帧中人脸纹理所在的坐标系下;然后,对于变换后的更新纹理和首帧中的纹理的交界区域利用均值滤波进行平滑;最后,将平滑后得到的人脸和头发纹理全景图作为人脸动画的纹理映射图.

7.4　背景(2D)-前景(3D)的无缝拼接

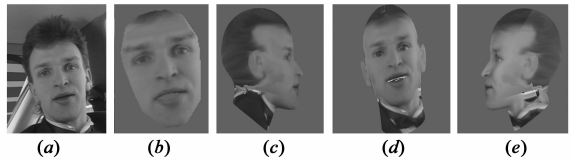
在当前帧中:(1)首先进行人脸动画合成;(2)接着按照 H.264 的分块方式,对当前帧分块;(3)然后对于三维人脸和头发模型在图像平面上的投影区域以外的块,根据 H.264 进行解码;(4)最后对于处在背景和以上投影区域的边界上的块,仍由 H.264 解码,但根据判别信息,只显示属于背景的像素.

8　实验结果与分析

实验配置为:CPU 3.01G,内存 2G,GT200.

8.1　编、解码端:人脸模型特定化

图 5 是在编、解码端对 MPEG-4 序列 Carphone 的人脸模型特定化结果,其中后者的结果是人脸和头发纹理更新后的结果.由此可见,它们与输入图像之间具有较强的相似性.



(*a*)　(*b*)　(*c*)　(*d*)　(*e*)

图5　(*a*)输入视频Carphone的首帧;(*b*)编码端CANDIDE3模型的特定化结果;(*c*)解码端Alice模型特定化结果的右侧面;(*d*)(*c*)的正面;(*e*)(*c*)的左侧面.

8.2　编码端

8.2.1　人脸运动跟踪结果

从图 6 可见,即使在大姿态,夸张表情,强光照下,人脸运动也被准确地跟踪到.特别是处于闭眼状态的最左边图像,眨眼幅度被准确地检测出来了.

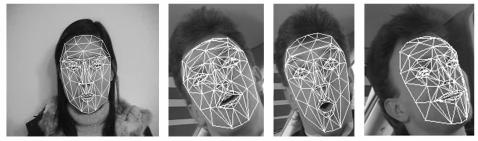


图6　人脸运动跟踪结果

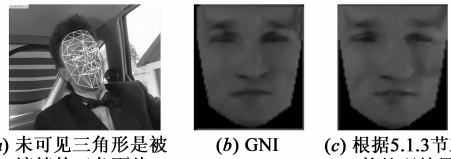
8.2.2　单个观测量 Vs 多个观测量

对于光照变化较大的 Foreman,由图 7 可见,采用单个观测量的结果不好,而采用多个观测量的结果会更好.8.2.4 节将给出它们的量化比较.

图7　左列是Forman;中/右列是多/单个观测量的结果.

8.2.3　应对遮挡

从图 8 可见,(*c*)中右眉毛的右边比(*b*)中的更加合理,从而可从中提取出更加合理的观测量.



(*a*) 未可见三角形是被遮挡的三角面片　(*b*) GNI　(*c*) 根据5.1.3节对(*b*)的处理结果

图8

8.2.4　人脸运动跟踪算法的评测与比较

鉴于文献[19]是在线外观模型中最重要的方法之一,我们以它为基准来比较.尽管本实验所在的环境下人脸的真实运动值是不知道的,但我们可以用计算机图形学的技术来间接获得,即在给定人脸运动(真实值)和光照下绘制人脸图像,然后对绘制的人脸图像进行人脸运动跟踪,最后将估计出来的运动值与真实值进行比较.量化指标为:估计出来的运动值与真实值之间误差的平均值.

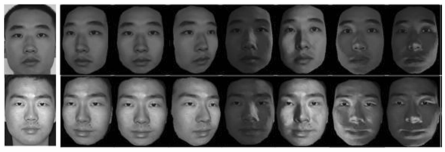


图9 在不同姿态和光照下绘制的人脸图像

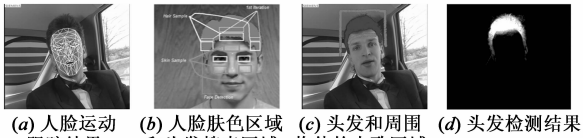
根据在前续工作[40,41]中合成的在不同运动和光照下的人脸图像(图 9),实验结果如表 1 所示.从中可以看到,本文算法在跟踪精度上优于文献[19],且采用多个观测量要优于采用单个观测量.

表 1 本文算法和文献[19]的量化指标

人脸的旋转角度	X 轴	Y 轴	Z 轴
单个观测量	2.38	1.75	1.69
多观测量	2.73	1.94	1.86
文献[19]	2.71	1.98	1.85

8.2.5 头发区域的检测

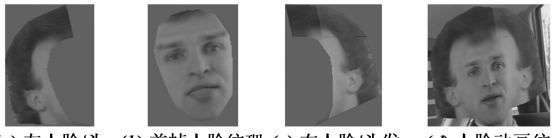
由图 10 可见,基于人脸运动跟踪的结果,头发区域被较好地检测出来了.



(*a*) 人脸运动跟踪结果 (*b*) 人脸肤色区域和头发搜索区域 (*c*) 头发和周围物体的大致区域 (*d*) 头发检测结果

图10

8.2.6 更新人脸头发纹理和人脸动画纹理映射图



(*a*) 左人脸/头发纹理更新图像 (*b*) 首帧人脸纹理 (*c*) 右人脸/头发纹理更新图像 (*d*) 人脸动画纹理映射图

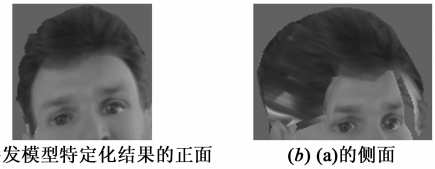
图11

由图 11 可见,基于人脸运动跟踪和头发检测,人脸和头发的侧面纹理被较好地提取出来,且在解码端形成了无缝的人脸和头发纹理的全景图.

8.3 解码端

8.3.1 头发模型特定化

由图 12 可见,根据编码端的头发检测结果,在解码端可得到一个针对特定人的三维头发模型.



(*a*) 头发模型特定化结果的正面 (*b*) (a)的侧面

图12

8.3.2 验证是否满足模型基视频编解码的要求

该要求是:根据从编码端输入视频得到的编码信息,解码端合成出与原始视频差异很小的动画.该要求也可以说是对编解码质量的要求.

对于 Carphone 视频的解码结果如图 13 所示.

图13 上行为输入视频,下行为解码视频

定义评价 MBC-Face 质量的量化指标:$Q_Y = \sum_{i=1}^{N}\sum_{j=1}^{M} abs(y_{ani}^{(i,j)} - y_{org}^{(i,j)})/(N\cdot M)$.

N 为输入视频的帧数,M 为该视频第 i 帧的像素个数,$y_{org}^{(i,j)}$ 为该帧第 j 个像素的颜色值,$y_{ani}^{(i,j)}$ 为解码视频第 i 帧第 j 个像素的颜色值.

将 8.2 中提取出来的人脸运动参数,输入到本文的人脸动画模型和 Waters 模型中来.从表 2 可见,本系统满足了上述要求,并且与后者相比,以牺牲较少的真实感为代价,大大减少了计算量.

表 2 本文算法和 Waters 模型的性能比较

	平均每帧耗时	Q_Y 的均值
本文算法	0.023s	3.9
Waters 模型	0.07s	3.8

8.4 客观评测与比较

对于 8.2 中用到的人脸视频,分别利用本系统,文献[30]和 H.264 进行编解码.由图 14 和表 3 可见,一方面,本系统在极低码率下(低于 7.2Kbit/s 时)的率失真优于 H.264.另一方面,相比于 2D/3D 方案[30],本系统的率失真是优于它的.

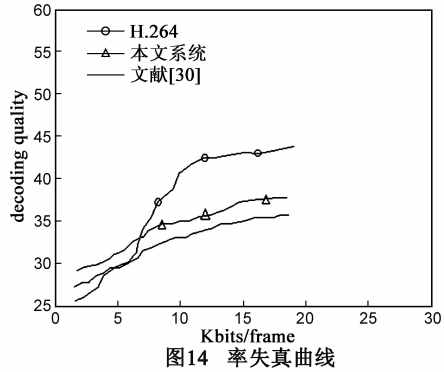


图14 率失真曲线

表 3　本文系统,文献[30]和 H.264 的性能比较

	Q_Y 的均值	平均信道传输码率	平均帧率
本文算法	3.9	5.9k/s	24frame/s
文献[30]	5.1	7.2k/s	24frame/s
H.264	2.1	10.4k/s	24frame/s

8.5　在极低码率下的主观评测与比较

该主观评测的调查对象是 34 个人,他们的年龄,性别以及出生地的情况如表 4 所示.从中可以看到,调查对象的分布具有广泛的代表性.

表 4　调查对象的分布情况

组	可能的回答	频率
年龄	1.小于 20;2.20 和 30 之间;3.大于 30.	8/18/8
性别	1.男性;2;女性.	20/14
籍贯	1.东部地区;2.中部地区;3.西部地区.	7/14/13

第一步　建立问卷.　表 5 展示了其中的问题.对于它们的回答从"绝对不同意"到"完全同意"分为 5 级.并且采用 Cronbach's alpha 测试来验证了它的内在可信度(测试结果在 0.7 以上).

表 5　问卷的 Cronbach's alpha 测试结果

分组	问题	Cronbach 结果
运动	1.我识别出解码视频中虚拟人头的表情 2.解码视频中虚拟人头的运动是自然的	0.743
外貌	1.我喜欢解码视频中虚拟人头的外貌 2.解码视频中虚拟人头看起来像个人	0.811

第二步　比较极低码率下解码结果在脸部的辨识度.　首先本文系统,文献[30]和 H.264 在 1Kbit/s 的信道下传输数个人脸视频,然后让调查者观察输入视频和解码视频的差异,最后让他们填写问卷,评测的最高分是 10 分,最低分是 0 分.由表 6 可见,在极低码率下,本系统的得分最高,且在 7.5 以上,表明它对于脸部的解码结果具有较高的辨识度.

表 6　本文系统,文献[30]和 H.264 的平均得分

分组	问题	本文系统平均得分	文献[30]平均得分	H.264 平均得分
运动	1.我识别出解码视频中虚拟人头的表情; 2.解码视频中虚拟人头的运动是自然的.	7.87 8.06	6.73 7.01	4.12 5.54
外貌	1.我喜欢解码视频中虚拟人头的外貌; 2.解码视频中虚拟人头看起来像个人.	7.83 8.12	6.48 7.47	4.36 5.59

9　结论与展望

本文从系统的角度研究 MBC-Face 问题,提出了背景(2D)-前景(3D)混合编解码的 MBC-Face 系统.该系统可以极低的传输码率在接收端再现真实感的人脸视频.下步计划是:建立集骨骼,肌肉,皮肤为一体的人脸动画计算模型来提高人脸动画的真实感.利用人脸运动参数进行表情识别.

参考文献

[1] Liu YC, et al. A virtual teleconferencing system based on face detection and 3D animation in a low bandwidth environment [J]. JIST, 2010, 20(4): 323 – 332.

[2] 普波.基于视频的三维人脸动画驱动的设计与实现[D].电子科技大学,2010.
Pu B. Design of 3D Facial Animation Based on Video[D]. University of Electronic Science and Technology, 2010. (in Chinese)

[3] Forchheimer R. Image coding: from waveforms to animation [J]. TASSP, 1989, 37(12): 2008 – 2023.

[4] B Moghaddam, et al. An automatic system for model based coding of faces[A]. DSC[C]. UT, Snowbird: IEEE Explore, 1995.568 – 573.

[5] Chai J X, Xiao J, Hodgins J. Vision-based control of 3D facial animation[A]. Eurographics[C]. Switzerland, Aire-la-Ville: ACM DL, 2003.193 – 206.

[6] Aizawa K, Harashima H, Saito T. Model-based analysis synthesis image coding(MBASIC) system for a person's face[J]. Image Communication, 1989, 1(2): 139 – 152.

[7] S. Gokturk et al. A data-driven model for monocular face tracking[A]. ICCV[C]. Vancouver, BC: IEEE Xplore, 2001.701 – 708.

[8] Wei Zhang, et al. Real Time feature based 3-D deformable face tracking[A]. ECCV[C]. Berlin, Heidelberg: Springer-Verlag, 2008.720 – 732.

[9] Q Wang, et al. Real-time bayesian 3-D pose tracking[J]. TCSVT, 2006, 16(12): 1533 – 1541.

[10] L Vacchetti, et al. Stable real-time 3D tracking using online and offline information[J]. PAMI, 2004, 26(10): 1385 – 1391.

[11] Wei-Kai Liao, et al. Integrating multiple visual cues for robust 3d face tracking[A]. AMFG[C]. Berlin, Heidelberg: Springer-Verlag, 2007.109 – 123.

[12] J Strom. Model-Based Head Tracking and Coding[D]. Sweden: Linköping University, 2002.

[13] M L Cascia, et al. Fast, reliable head tracking under varying illumination: An approach based on registration of texture mapped 3D models[J]. TPAMI, 2000, 22(4): 322 – 336.

[14] J Ahlberg. Model-Based Coding: Extraction, Coding, and Evaluation of Face Model Parameters[D]. Sweden: Linköping University, 2002.

[15] I Matthews, et al. 2D vs 3D deformable face models: Representational power, construction, and real-time fitting[J]. IJCV, 2007, 75(1): 93 – 113.

[16] J Sung, T Kanade, D Kim. Pose robust face tracking by combining active appearance models and cylinder head models

[J]. IJCV, 2008, 80(2): 260 – 274.
[17] A D Jepson, et al. Robust online appearance models for visual tracking[J]. TPAMI, 2003, 25(10): 1296 – 1311.
[18] Lui YM, et al. Adaptive appearance model and condensation algorithm for robust face tracking[J]. TSMC Part A, 2010, 40(3): 437 – 448.
[19] F Dornaika, F Davoine. Simultaneous facial action tracking and expression recognition in the presence of head motion [J]. IJCV, 2008, 76(3): 257 – 281.
[20] S Zhou, et al. Visual tracking and recognition using appearance-adaptive models in particle filters[J]. TIP, 2004, 13(11): 1491 – 1506.
[21] Z Wen, T Huang. Capturing subtle facial motions in 3d face tracking[A]. ICCV[C]. France, Nice: IEEE Xplore, 2003. 1343 – 1350.
[22] Tim K Marks, et al. Tracking motion, deformation, and texture using conditionally gaussian processes[J]. TPAMI, 2010, 32(2): 348 – 363.
[23] F I Parke, K Waters. Computer Facial Animation[M]. Wellesley, MA: A K Peters, 1996. 1 – 365.
[24] Wang W M, et al. A physically-based modeling and simulation framework for facial animation[A]. ICIG[C]. Xi'an, Shanxi: IEEE Explore, 2009. 521 – 526.
[25] Volker Blanz, et al. Reanimating faces in images and video [J]. Computer Graph, 2003, 22(3): 641 – 650.
[26] K. Waters. A muscle model for animating three dimensional facial expression[J]. Computer Graphics, 1987, 22(4): 17 – 24.
[27] Marcos S, Bermejo JGG, Zalama E. A realistic facial animation suitable for human-robot interfacing[A]. ICIRS[C]. Nice: IEEE Explore, 2008. 1-3. 3810 – 3815.
[28] 张黎. 真实感三维头发的建模及动态模拟算法研究[D]. 浙江大学, 2010.
Zhang L. Research on 3D Hair Modeling and Dynamic Simulation[D]. University of Zhejiang, 2010. (in Chinese)
[29] 杨晓辉. 一种采用模型基辅助的混合视频编码方法[J]. 电路与系统学报, 2002, 7(2): 35 – 38.
Yang X H. A novel approach of model-aided hybrid coding [J]. Journal Circuits and Systems. 2002, 7(2): 35 – 38. (in Chinese)
[30] Peter Eisert, et al. Model-aided coding: a new approach to incorporate facial animation into motion-compensated video coding[J]. TCSVT, 2000, 10(3): 344 – 358.
[31] J. Ahlberg. CANDIDE3-an Updated Parameterized Face[R]. LiTH-ISY-R-2326, Department of Electrical Engineering, Linköping University, 2001. 1 – 16.
[32] Koray Balcı, et al. Xface open source project and SMIL-agent scripting language for creating and animating embodied conversational agents[A]. ICM[C]. Augsburg, Germany: ACM DL, 2007. 1013 – 1016.
[33] 吴暾华, 周昌乐. 平面旋转人脸检测与特征定位方法研究[J]. 电子学报, 2007, 35(9): 1714 – 1718.
Wu T H, Zhou C L. Study on face detection under rotation in image plane and facial features localization[J]. Acta Electronica Sinica, 2007, 35(9): 1714 – 1718. (in Chinese)
[34] 吴证, 周越, 杜春华等. 彩色图像人脸特征点定位算法研究[J]. 电子学报, 2008, 36(2): 309 – 313.
Wu Z, Zhou Y, et al. Research on facial feature points extraction in color images[J]. Acta Electronica Sinica, 2008, 36(2): 309 – 313. (in Chinese)
[35] M Kampmann. Automatic 3-D face mode adaption for model-based coding of videophone sequences[J]. TCSVT, 2002, 12(3): 172 – 182.
[36] 薛峰, 丁晓青. 基于形状匹配变形模型的三维人脸重构[J]. 电子学报, 2006, 34(10): 1896 – 1899.
Xue F, Ding X Q. 3D reconstruction of human face based on shape match morphing model[J]. Acta Electronica Sinica, 2006, 34(10): 1896 – 1899. (in Chinese)
[37] P Grassberger. The pruned-enriched rosenbluth method: simulations of theta polymers of chain length up to 1,000,000[J]. Physical Review E, 1997, 56: 3682 – 3693.
[38] Yaser Yacoob, Larry S. Davis. Detection and analysis of hair [J]. TPAMI, 2006, 28(7): 1164 – 1169.
[39] Levin A, Rav-Acha A, Lischinski D. Spectral matting[A]. CVPR[C]. Minneapolis, Minnesota, USA: IEEE Xplore, 2007. 1 – 8.
[40] Hu Yuankui, et al. Reconstruction of 3D face from a single 2D image for face recognition[A]. International Workshop on VSPETS[C]. San Diego, California USA: IEEE Xplore, 2005. 217 – 222.
[41] Hu Yuankui, et al. A Low-dimensional illumination space representation of human faces for arbitrary lighting conditions [A]. ICPR[C]. Hong Kong, China: IEEE, 2006. 1147 – 1150.

作者简介

於 俊 男, 1983 年生于安徽滁州. 中国科学技术大学自动化系博士后. 研究方向为人机情感接口, 智能机器人.
E-mail: harryjun@ustc.edu.cn

汪增福 男, 1960 年生于安徽合肥. 教授, 研究方向为计算机视听觉, 智能机器人.
E-mail: zfwang@ustc.edu.cn