

基于视频引导模型更新的视频肤色增强方法

丁少博*, 张雅荔, 张坤

天津大学微电子学院, 天津 300072

摘要 在视频的采集和跨媒体再现过程中,相机的色域限制往往会导致获取的图像色彩失真。其中肤色作为人眼最敏感的色彩之一,肤色失真可能会降低观众的视觉体验。肤色增强是一种针对图像或视频中人类肤色的处理技术,通过对失真的肤色进行调整,达到提高显示质量的目的。尤其是在视频处理领域,在提高肤色模型自适应性的同时,还需要考虑算法的实时性与计算工作量,因此,提出一种面向实时视频处理的自适应肤色增强方法。利用镜头边界来引导肤色模型更新,以此降低模型更新的计算工作量;其次,建立一种根据镜头边界更新的动态肤色模型进行皮肤检测;最后,基于主观实验建立了偏好肤色模型,并且实现了针对不同种族的肤色增强。与现有方法相比,所提方法在主观评价实验中取得了更高的平均主观分值。实验结果表明,所提方法可以显著降低模型更新带来的计算工作量,同时实现有针对性的肤色增强。

关键词 图像处理; 图像增强; 镜头边界检测; 肤色建模

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP220685

Video Skin-Color Enhancement Method Based on Video-Guided Model Updates

Ding Shaobo*, Zhang Yali, Zhang Kun

School of Microelectronics, Tianjin University, Tianjin 300072, China

Abstract During the acquisition and cross-media reproduction of videos, colors can be distorted because the color gamut of the camera is limited and may differ from the color gamut of the display device. Skin color is among the most sensitive colors to the human eye. Therefore, skin-color distortion can deteriorate viewers' video experience. Skin-color enhancement is a processing technology that adjusts a distorted skin color to improve the display quality. Particularly in video processing, the self-adaptability of a skin-color model must be improved while considering the real-time performance and computational load of the algorithms. For these purposes, the present paper proposes an adaptive skin-color enhancement method for real-time video processing. The update of the skin-color model is guided by shot boundaries, which can reduce the computational load of updating. Second, a dynamic skin-color model updated with the shot boundary is built for skin detection. Finally, the preferred skin-color model and skin-color enhancements for different races are achieved through subjective experiments. The proposed method achieved higher mean opinion scores than the existing methods in subjective evaluation experiments. In addition to achieving the targeted skin-color enhancement, the proposed method significantly reduced the computational load of model update.

Key words image processing; image enhancement; shot boundary detection; skin color modeling

1 引言

随着光电信息技术的发展,图像与视频已成为人类生活中常见的信息载体,不同类型的显示设备在信息社会中成为人们生活与工作的必需品。通过电子设备获取的各种视觉信息往往是借助摄影设备从外界获取的图像与视频,最终通过显示设备对外呈现。而在拍摄过程中,由于部分摄影设备自身的色域限制,获得

图像的色彩会与事物原本的色彩出现偏差,导致最终在显示设备上呈现的画面出现色彩失真。随着画面尺寸的扩展,这种色彩失真也会愈发明显^[1-2],进而可能会影响观众的观看体验。研究表明,在长期的观察中,人们对皮肤、天空、草地等许多常见事物的颜色会形成后天性主观认知^[3-4]。肤色作为人类重要的外貌特征之一,在图像与视频中更容易吸引观众的注意力。因此,对图像与视频中的人类肤色进行针对性的色彩增

收稿日期: 2022-02-08; 修回日期: 2022-03-05; 录用日期: 2022-03-14; 网络首发日期: 2022-03-24

通信作者: *akat@tju.edu.cn

强,使其符合人们的主观认知,是一种改善画质的有效方法。

20 世纪末以来,学者们相继提出了不同的肤色增强技术。随着液晶显示器逐渐取代阴极射线管显示器,人们更加关注使用肤色增强算法来提高显示器的色彩质量的内容。目前研究领域中实现肤色增强的方法主要分为两类。第一类为基于目标偏好点实现肤色增强的方法。Kim 等^[5]在研究肤色增强的过程中设置了多个目标偏好点,以此对不同人种的肤色进行差异化处理,实现了对不同种族的肤色增强。王涛等^[6]在 YCbCr 色彩空间将亮度分量均分为三个等级,在三个亮度区间内分别统计肤色像素点的色度均值,并记为初始聚类中心,然后基于 K 均值聚类法计算得到目标偏好点坐标,一定程度提升了肤色增强在非均匀光照下的自适应性。该类方法计算工作量低,具备较好的实时性能,可应用于实时视频处理任务,但由于皮肤区域的检测存在误差,检测区域的边界可能出现在真实皮肤区域的平滑处。对于在检测区域边界出现的色彩突变问题,该类方法缺乏有效的解决方案。第二类为基于模型映射实现肤色增强的方法。2005 年, Kim 等^[7]提出了一种基于模型映射的肤色增强方法,该方法在 YCbCr 色彩空间建立了从皮肤椭圆模型到人眼偏好椭圆模型的仿射变换,有效地避免了区域边界色彩突变的问题。Park 等^[8]同样以 YCbCr 色彩空间的椭圆模型为基础,提出了与椭圆模型同圆心的一种偏好肤色椭圆模型,将分布于两椭圆之间环形区域内的像素点视为需要被色彩增强的肤色像素点,在 Cb-Cr 平面上沿偏好肤色椭圆模型的方向进行色度偏移。吴平等^[9]基于主观实验所得数据,在 Cb-Cr 平面建立了人类偏好肤色的分布模型,并与经典椭圆模型建立两者之间的映射关系,以此实现肤色增强。蒋崛等^[10]基于认知心理学,建立了人类肤色与偏好肤色在 HSV 色彩空间的分布模型,通过两者之间的映射关系实现肤色增强。该类方法通过建立更具体的模型来描述人类肤色与偏好肤色的分布特征,对肤色区域边界处色彩过渡的处理更自然。但该类

方法往往基于现有的静态肤色模型进行肤色检测,在复杂环境中自适应性较差。此外,近年来随着深度学习技术的发展,采用人工神经网络模型对图像中的皮肤区域进行分割,再通过肤色增强技术对皮肤区域进行色彩增强,也成为了一种有效的肤色增强手段^[11]。高性能的深度语义分割网络可显著提高皮肤检测精度,从而实现更高质量的肤色增强,如 DeepLabV3+ 模型^[12-13],在包括肤色检测在内的图像语义分割任务中,表现出显著优于自适应检测方法的性能;U-Net 模型,作为一种高性能经典语义分割网络模型,在医学图像处理领域中具有广泛应用^[14-15]。但值得注意的是,这一类基于深度学习的人工神经网络模型往往具有极高的模型复杂度,硬件友好性较低,虽然在图像处理领域中表现出色,但却难以满足视频处理的实时性要求。

为了同时兼顾肤色增强的自适应性及算法实时性,提出一种基于视频镜头引导模型更新的自适应肤色增强方法。该方法可随着视频场景的变化对皮肤检测模型进行更新,并对检测到的皮肤区域进行肤色分类与针对性肤色增强。由于视频片段中同一镜头内每一帧的肤色分布高度相似^[16],因此在同一镜头内,大量的模型更新操作可以省略。应用镜头边界检测技术引导皮肤检测模型仅在镜头边界处更新,以此取代现有方法中逐帧更新模型的操作,显著提高了算法的实时性能。在肤色检测与增强方面,构建了一种动态皮肤检测模型,该模型可根据视频镜头边界的变化进行实时更新;同时,基于主观实验建立了不同种族的偏好肤色模型,实现更有针对性的肤色增强,有效避免边界处剧烈的色彩突变。

2 研究方法

所提方法包括三个功能模块,整体系统流程如图 1 所示。镜头边界检测模块将视频中每个镜头的起始帧标记为一个节点;当检测到镜头边界时,皮肤分割模块在该节点处更新皮肤检测模型,该模型将用于提取同一镜头内所有帧的皮肤区域,直至检出下一镜头

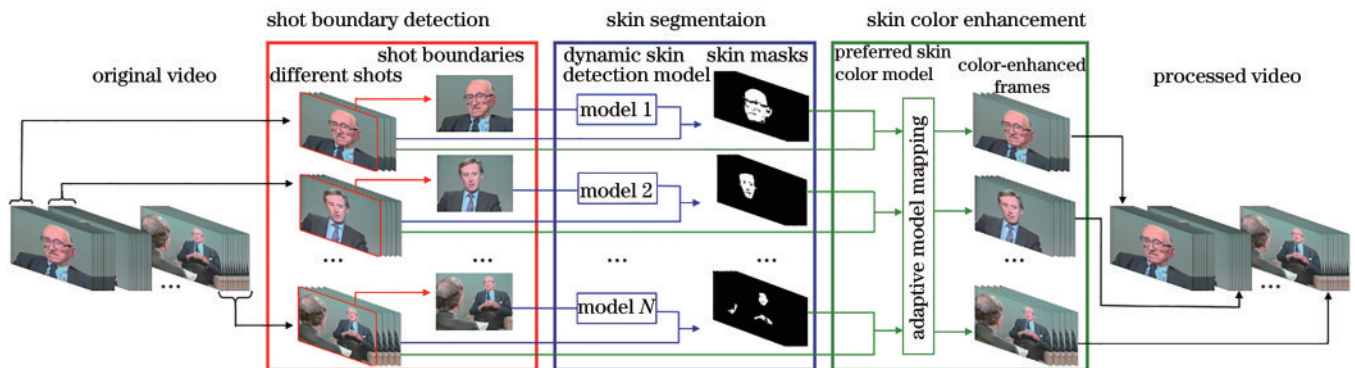


图 1 所提方法的系统结构

Fig. 1 System structure of the proposed method

边界;肤色增强模块对检测区域内的肤色进行分类,通过自适应模型映射实现肤色增强。

2.1 镜头边界检测

镜头边界检测模块的作用在于检测视频中的视觉不连续性,以镜头为单位将视频序列分割成片段,引导皮肤检测模型仅在每一个镜头片段的起始节点更新参数,可以有效降低模型更新所需的计算工作量,提高算法整体的实时性。



图 2 常见镜头边界类型。(a)突变;(b)渐变

Fig. 2 Common shot boundary types. (a) Hard transition; (b) soft transition

以 F_1 与 F_2 表示视频中不同的两帧,引入感知哈希算法^[17]分别计算 F_1 与 F_2 的图像指纹,记为 P_1 与 P_2 ,计算 P_1 与 P_2 之间的汉明距离,记为 $D_H(P_1, P_2)$,以此作为量化帧间结构差异的参数;分别计算 F_1 与 F_2 在 HSV 色彩空间 Hue 分量的直方图,记为 H_1 与 H_2 ,计算 H_1 与 H_2 之间的巴氏距离,记为 $D_B(H_1, H_2)$,以此作为量化帧间色彩差异的参数。其中, D_H 取值范围为 $[0, 64]$ 且与结构差异呈正相关, D_B 取值范围为 $[0, 1]$ 且与色彩差异呈负相关。将两者均归一化至区间 $[0, 1]$, 且使其均与对应的特征差异呈正相关,计算融合特征差异参数:

$$\begin{cases} D_{H-N}(P_1, P_2) = \frac{D_H(P_1, P_2)}{64} \\ D_{B-N}(H_1, H_2) = 1 - D_B(H_1, H_2) \end{cases}, \quad (1)$$

$$R_{\text{fusion}}(F_1, F_2) = \alpha D_{H-N}(P_1, P_2) + (1 - \alpha) D_{B-N}(H_1, H_2), \quad (2)$$

式中: $D_{H-N}(P_1, P_2)$ 为归一化后的帧间结构差异参数; $D_{B-N}(H_1, H_2)$ 为与帧间色彩差异呈正相关的色彩差异参数; R_{fusion} 为融合特征差异参数; α 为融合权重。当原始视频输入时,镜头边界检测模块将逐帧计算帧间的融合特征差异参数 R_{fusion} 并与阈值 T_1 与 T_2 ($T_1 > T_2$) 进行比较,以此确定帧间的相关性。为便于表述,将当前帧记为 F_C ,前一相邻帧记为 F_P ,前一镜头边界记为 F_B 。从首帧起计算 $R_{\text{fusion}}(F_P, F_C)$,若 $R_{\text{fusion}}(F_P, F_C) > T_1$,判定两帧属于不同镜头,帧间存在突变,当前帧被标记为新的镜头边界,如图 3 所示。

若当前位置不存在突变,算法将在当前位置继续检测渐变。在渐变发生的期间,当前帧与上一镜头边界之间的特征差异会逐渐变大,因此若 $R_{\text{fusion}}(F_P, F_C) < T_1$ 且 $R_{\text{fusion}}(F_B, F_C) > T_1$,则判定当前帧正处于渐变过程中。由于渐变过程中相邻帧间的差异较小,为了确定

常见的镜头边界主要有突变(HT)与渐变(ST),如图 2 所示。突变表现为帧间画面的突然切换,发生在未经编辑直接衔接的两帧画面之间;渐变表现为视觉上缓慢的画面变化,通常是借助编辑手法跨越多帧实现的画面变化过程。镜头边界检测的主要思路是量化图像特征并比较帧间差异,如结构特征与色彩特征。为了能准确检测这两种镜头边界类型,提出一种基于特征融合的镜头边界检测方法。

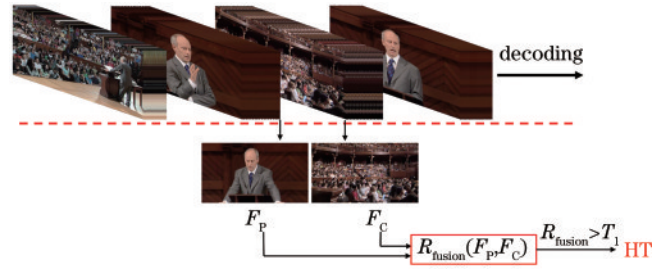


图 3 镜头突变的确认过程

Fig. 3 Confirmation process of HT

渐变结束的位置,算法将继续比较相邻帧间的特征差异,若 $R_{\text{fusion}}(F_P, F_C) < T_2$,则判定渐变过程已结束,视频画面已切换至新的镜头,当前帧被标记为新的镜头边界,如图 4 所示。

由于各类视频内容上的不同,视觉上存在明显差异的两帧的融合特征差异参数 R_{fusion} 可能较大也可能较小。为了保证该方法在不同视频场景中的自适应性,根据动态范围内帧间差异的变化动态更新阈值 T_1 与 T_2 :

$$\begin{cases} T_1 = \frac{6}{L} \sum_{i=1}^{L-1} R_{\text{fusion}}(F_{i-1}, F_i) \\ T_2 = \frac{3}{L} \sum_{i=1}^{L-1} R_{\text{fusion}}(F_{i-1}, F_i) \end{cases}, \quad (3)$$

式中: L 表示从上一镜头边界 F_B 到当前帧 F_C 的总帧数; F_i 表示 F_B 之后的第 i 帧。

为了确定式(2)中融合权重参数 α 的取值,以达到最优的检测精度,对镜头边界检测精确率随融合权重参数 α 变化的情况进行分析。随机从不同题材、分辨率为 3840×2160 的视频中选取若干片段,通过剪辑将这些片段拼接为一段超过 2000 帧的视频片段。采用

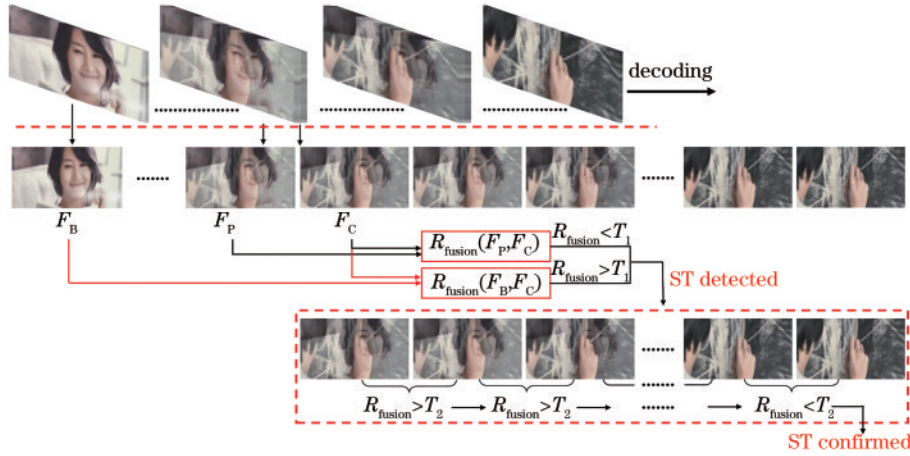


图 4 镜头渐变的确认过程

Fig. 4 Confirmation process of ST

该方法对这一视频片段进行镜头边界检测,记录检测结果中的真正例(TP)与假正例(FP),并计算检测精确率。从 $\alpha=0$ 开始,以0.01为步长逐渐增加权重值,直至 $\alpha=1$,得到精确率随 α 的变化曲线,如图5所示。由图5可知:当 α 明显趋于0或趋于1,即主要依据色彩特征或主要依据结构特征进行镜头边界检测时,检测精确率均不理想;当 $\alpha=0.58$ 时,检测精确率达到最高。因此,设定 $\alpha=0.58$ 为融合权重参数。

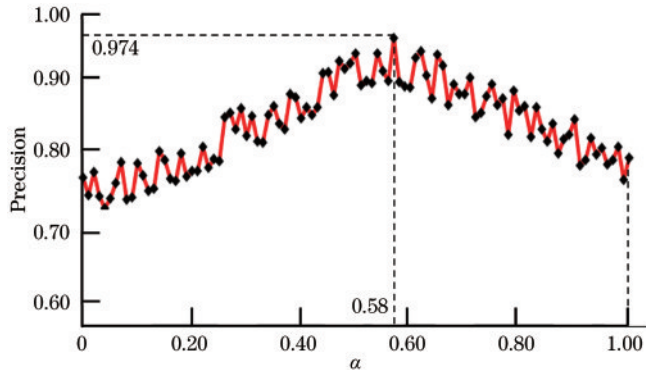


图 5 镜头边界检测精确率随融合权重的变化

Fig. 5 Variation of shot boundary detection precision with fusion weight

2.2 皮肤区域分割

当镜头边界检测模块对新的节点进行标记后,为保证在不同场景下均能较为准确地检出皮肤区域,皮肤区域分割模块需要针对新的镜头场景更新模型参数。为了同时兼顾肤色检测的精确性与实时性,提出一种通过人脸采样获取肤色参考点建立肤色阈值模型的方法。

首先需要对标记为镜头边界的帧图像进行人脸检测,以此确定获取特征点的采样区域。较为经典的Viola-Jones人脸检测器^[18]是一种基于Haar-like特征的人脸检测方式,对正脸检测表现出较好的精确性与实时性。但是考虑到视频中的人脸往往存在不同的角

度,该方法对视频图像进行人脸检测的效果并不理想,因此引入Shi等^[19]提出的一种级联递进校正网络(PCN)实现快速且准确的多角度人脸检测。为了避免非皮肤区域对人脸肤色采样的干扰,采样的肤色参考点应尽可能从人脸皮肤的平滑区域选取,因此需要进一步对面部区域各类特征进行定位。Kazemi等^[20]提出的基于级联回归树生成68个人脸特征点的方法在确定人脸区域后进行人脸对齐。虽然这些特征点多数位于人脸的边缘处,但特征点之间两两连线的中点大多位于人脸皮肤的平滑区域,因此该方法通过计算人脸特征点之间连线的中点坐标可以确定若干肤色参考点。

图6(a)中展示了沿人脸边缘轮廓分布的68个人脸特征点。为了使特征点之间连线的中点尽可能位于皮肤平滑区域,将面部由下至上分为3个区域,如图6(b)所示,在同一区域内选择特征点进行连线。如图6(c)所示,1号区域中,沿下颌轮廓分布的特征点可与沿唇部轮廓分布的特征点分别连线;2号区域中,沿面部轮廓分布的特征点及沿双眼轮廓分布的特征点可与中央区域分布的特征点分别连线;3号区域中,沿不同眼眉分布的特征点间可分别连线。以此类推,共可确定210组特征点之间的连线关系,从而确定210个肤色参考点,如图6(d)所示。

由于人类肤色在YCbCr色彩空间呈现出明显的聚类特征^[21],通过对大量视频中的皮肤区域像素信息进行统计分析,得到的皮肤像素分布如图7所示。值得注意的是,这些像素点在Cb与Cr分量上的分布与高斯分布类似。

基于这种分布特征,将所获取的肤色参考点的色度分量视为服从高斯分布的序列,根据置信区间理论得到一组阈值作为肤色判别条件。其中,舍弃偏差较大的样本可以提高模型的可靠性。如在某些视频中,肤色受到灯光、滤镜或特殊妆容的影响,其色度会严重偏离正常肤色。为减小这类不可靠肤色对建模的影

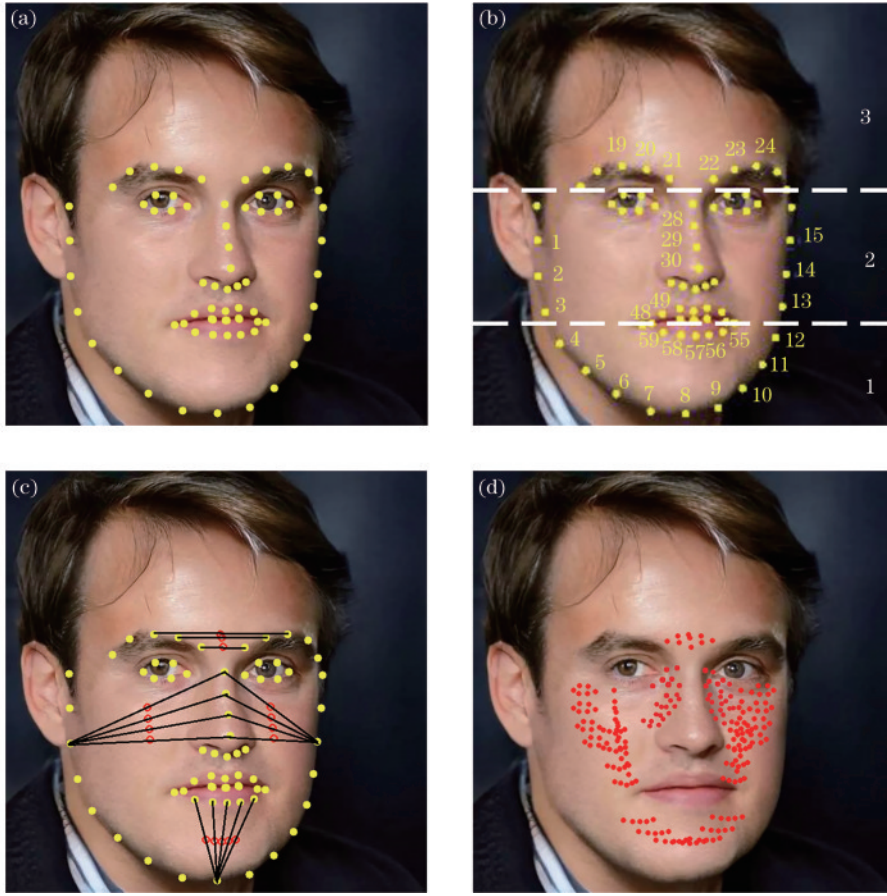


图 6 肤色采样过程。(a)68个人脸特征点;(b)特征点编号与区域划分;(c)计算中点坐标确定参考点;(d)210个肤色参考点
Fig. 6 Skin color sampling process. (a) 68 feature points; (b) feature point number and area division; (c) calculating the midpoint coordinates to determine the reference point; (d) 210 skin-color reference points

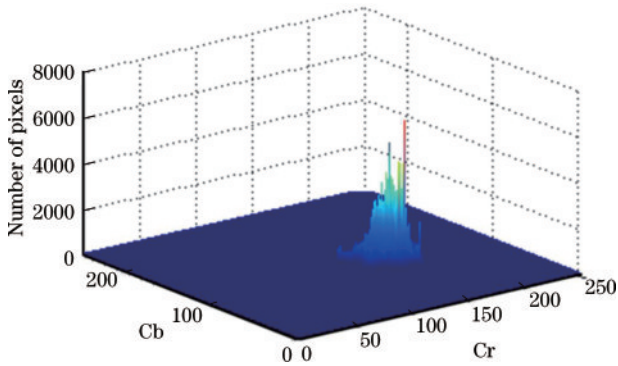


图 7 人类肤色在 YCbCr 色彩空间的分布
Fig. 7 Distribution of human skin color in the YCbCr color space

响,首先采用基于经验的固定阈值 $C_b[89, 142]$ 、 $C_r[126, 176]$ 对肤色参考点进行筛选。若该序列中超过 50% 的参考点不可靠,则丢弃本次人脸采样结果。对符合固定阈值的所有参考点,计算 95% 置信区间建立肤色模型,并采用该阈值对当前帧内所有像素进行条件判定,得到当前帧对应的肤色二值掩模,具体的公式为

$$\begin{cases} T_{Cb}: \mu_{Cb} - 1.96\sigma_{Cb} < P_{Cb} < \mu_{Cb} + 1.96\sigma_{Cb} \\ T_{Cr}: \mu_{Cr} - 1.96\sigma_{Cr} < P_{Cr} < \mu_{Cr} + 1.96\sigma_{Cr} \end{cases}, \quad (4)$$

$$M(m, n) = \begin{cases} 1, & P_{Cb} \in T_{Cb} \& \& P_{Cr} \in T_{Cr} \\ 0, & \text{else} \end{cases}, \quad (5)$$

式中: μ 与 σ 分别表示肤色参考点序列在两种色度分量上的平均值与标准差; P_{Cb} 和 P_{Cr} 分别表示像素点在 YCbCr 空间中的两种色度分量; M 表示基于当前阈值条件所得当前帧的肤色掩模; (m, n) 表示图像中任意像素点的坐标。

若当前帧存在复数个人脸,为减少人脸肤色采样所需的计算量,所提方法在同一帧内最多选择 4 个尺寸较大的人脸区域进行采样,并计算对应的肤色掩模,再将当前帧内所有掩模合并,如图 8 所示。掩模合并通过逻辑“或”实现,表达式为

$$M_{final} = M_1 \parallel M_2 \parallel \dots \parallel M_N, \quad (6)$$

式中: N 表示当前帧内被算法检测并采样的人脸个数,且 $N \leq 4$ 。

若从当前帧内未检测到可靠人脸,或人脸角度偏转过大导致无法生成人脸特征点时,则直接采用固定阈值 $C_b[89, 142]$ 、 $C_r[126, 176]$ 限定采样区域,并从符合阈值条件的像素点中随机选取 210 个参考点进行肤色建模。此外,为了避免无肤色区域或肤色区域不明显的帧对计算资源的浪费,若固定阈值限定的采样区域小于图像全局尺寸的 5%,则停止对当前帧

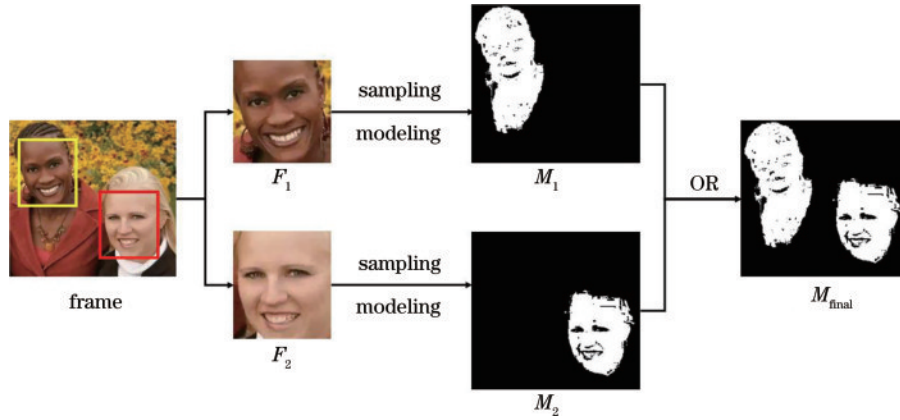


图 8 肤色掩模合并

Fig. 8 Skin mask merging

的处理,并跳过与当前帧属于同一镜头内的全部连续帧。

2.3 肤色增强

由于世界上不同人种之间的肤色存在较大差异,需要对不同人种的肤色匹配更具有针对性的色彩增强方案。因此,从当前帧内获取肤色区域后,肤色增强模块需要对区域内的肤色进行人种分类,并采用对应的色彩映射模型再演出更符合观众主观偏好的肤色。

该方法依据肤色差异将世界人类分为以亚洲人为

代表的黄色人种、以欧洲人为代表的白色人种、以非洲人为代表的黑色人种。该方法分析了来源于不同类型视频的 10793616 个皮肤像素点在 YCbCr 色彩空间中的色彩信息,其中包括黄色人种 3589544 个,白色人种 3611707 个,黑色人种 3592365 个。将色彩空间中 Y 分量由低到高平均划分为 5 个亮度等级,并统计归纳了三类人种在不同亮度等级下的肤色分布情况,如表 1 所示。表 1 列出了三类肤色在每个亮度等级下的分布频率与平均色度(\overline{D}_{Cb} 和 \overline{D}_{Cr})。

表 1 三类人种在每个亮度等级下的肤色分布与平均色度

Table 1 Distribution of skin color and average chrominance of three races at each luminance level

Luminance level	Asian			Caucasian			African		
	Frequency / %	\overline{D}_{Cb}	\overline{D}_{Cr}	Frequency / %	\overline{D}_{Cb}	\overline{D}_{Cr}	Frequency / %	\overline{D}_{Cb}	\overline{D}_{Cr}
1	3.4	117	140	9.1	123	140	16.6	125	132
2	12.6	110	147	21.5	109	145	38.5	120	143
3	34.9	107	145	33.4	101	155	36.7	111	151
4	38.4	105	149	28.1	106	153	5.6	118	140
5	10.7	117	140	7.9	116	140	2.6	123	131

当同一帧内存在多处不连续的肤色区域时,需要对每一个连通域进行独立分析。基于以上统计数据,通过将当前帧内肤色区域每一个连通域的亮度与色度信息与表 1 数据对比,确定与连通域内肤色更接近的人种,如图 9 所示。首先计算每个连通域的平均亮度,确定该连通域在表 1 内对应的亮度等级;然后计算连通域内全部像素在两个色度分量上的平均值,由此在 Cb-Cr 平面确定一个点 P_{Domain} ;通过查阅表 1,可以确定当前亮度等级下三类种族肤色在 Cb-Cr 平面对应的点 $P_{Asian}, P_{Caucasian}, P_{African}$ 。分别计算点 P_{Domain} 与其他三点之间的二维距离,并以对应的分布频率作为调节因子,计算得到量化当前连通域内肤色与每一类种族肤色间差异的参数,取差异参数最小者对应的种族作为肤色分类结果。

$$R_{diff(Race)} = \frac{\sqrt{(R_{Cb(mean)} - R_{Cb(Race)})^2 + (R_{Cr(mean)} - R_{Cr(Race)})^2}}{R_{freq(Race)}} \quad (7)$$

式中: Race 表示三类人种中任意一类; $R_{Cb(mean)}$ 和 $R_{Cr(mean)}$ 表示当前连通域的平均色度; $R_{Cb(Race)}$ 和 $R_{Cr(Race)}$ 表示某一种族在当前亮度等级下的平均色度; $R_{freq(Race)}$ 表示某一种族肤色在当前亮度等级下的分布频率; $R_{diff(Race)}$ 为计算所得量化肤色差异的参数。

另一方面,为了保证调整后的肤色符合人们的主观印象,该方法基于主观实验建立了针对三类人种的偏好肤色映射模型。选择 90 幅含有人体皮肤区域的测试图像,每类人种各 30 张。通过调整测试图像肤色区域在 YCbCr 色彩空间中的色度,可以基于每幅测试图像生成一组候选图像集。皮肤区域 Cb 和 Cr

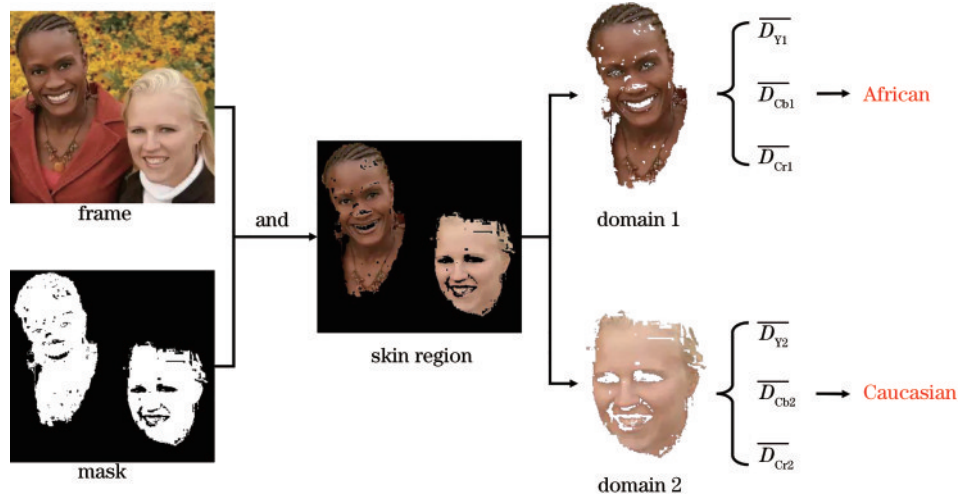


图 9 对当前帧内不同连通域进行肤色分类的结果

Fig. 9 Skin-color classification results for different connected domains in the current frame

分量的调整范围为 $[-2, 2]$, 步长为 1。因此每个候选图像集共包括 25 幅图像。邀请 82 位不同年龄、职业的志愿者参与到肤色主观评价实验中。每一位志愿者需在同一显示设备上依次观看 30 组候选图像集, 从每组候选图像集中标记出肤色更符合他们主观

偏好的 3 幅图像。实验结束后, 选出每组候选图像集中被标记次数最多的图像, 并记录该图像中皮肤区域的平均亮度与平均色度, 在 YCbCr 色彩空间确定对应的点。最终共确定 90 个点, 每类人种对应 30 个点, 如图 10 所示。

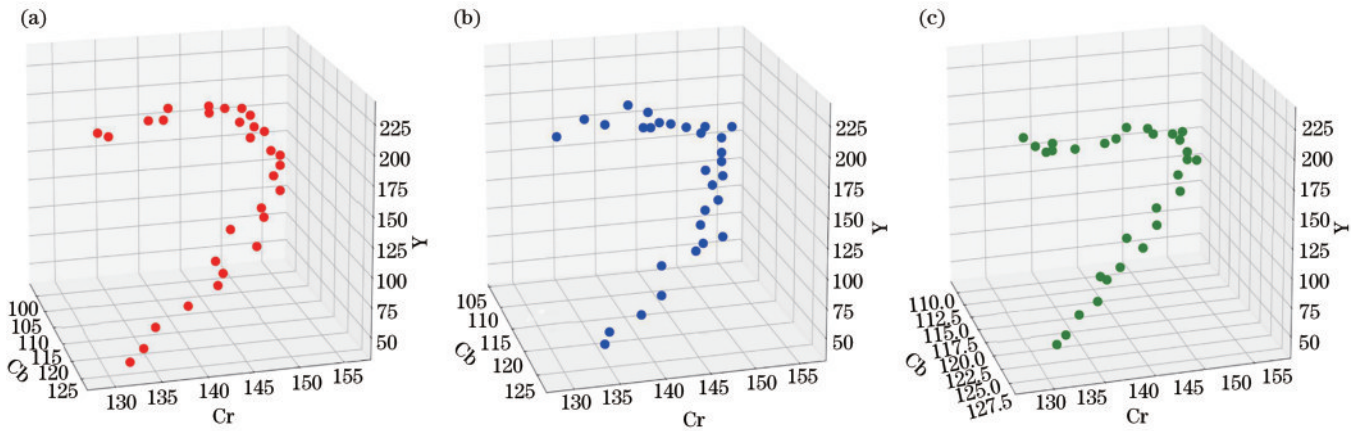


图 10 三类人种的偏好肤色分布。(a)黄色人种;(b)白色人种;(c)黑色人种

Fig. 10 Distribution of preferred skin color of three races. (a) Asian; (b) Caucasian; (c) African

由图 10 可知, 三类人种的偏好肤色分布呈现出相似的特征。以黄色人种为例, 其偏好肤色分布如图 11(a) 所示, 通过观察散点分布情况提出假设, 即这些散点分布轨迹近似一条空间曲线。为验证这一假设, 对这些散点分别进行一维平面拟合与二维曲面拟合, 如图 11(b) 与图 11(c) 所示。实验证明, 这些散点的分布轨迹可以视为平面与曲面相交确定的一条曲线, 如图 11(d) 所示。

上述偏好肤色参考点的分布特征对白色人种与黑色人种同样成立, 因此通过对三类偏好肤色参考点分别进行数据拟合, 确定对应不同种族的偏好肤色映射模型。将当前帧内肤色连通域的平均亮度值代入

$$\begin{cases} P_{\text{Race}}: P(R_{Y(\text{mean})}, R_{\text{Cb}(\text{pre})}, R_{\text{Cr}(\text{pre})}) = 0 \\ C_{\text{Race}}: C^2(R_{Y(\text{mean})}, R_{\text{Cb}(\text{pre})}, R_{\text{Cr}(\text{pre})}) = 0 \end{cases}, \quad (8)$$

式中: P_{Race} 与 C_{Race} 分别表示通过数据拟合得到的一维平面与二维曲面的方程; $R_{Y(\text{mean})}$ 表示当前连通域的平均亮度; $R_{\text{Cb}(\text{pre})}$ 和 $R_{\text{Cr}(\text{pre})}$ 分别表示由方程解得目标偏好肤色的两种色度分量。这样可计算对应的偏好肤色色度值。

为了避免肤色增强时在肤色连通域边界处出现剧烈的色彩突变, 根据肤色连通域内每一像素色度与肤色阈值模型中值的距离, 为不同位置的像素赋予对应的映射权重, 越接近区域边界的位置所得映射权重越小, 进而其色彩变化越小, 表达式为

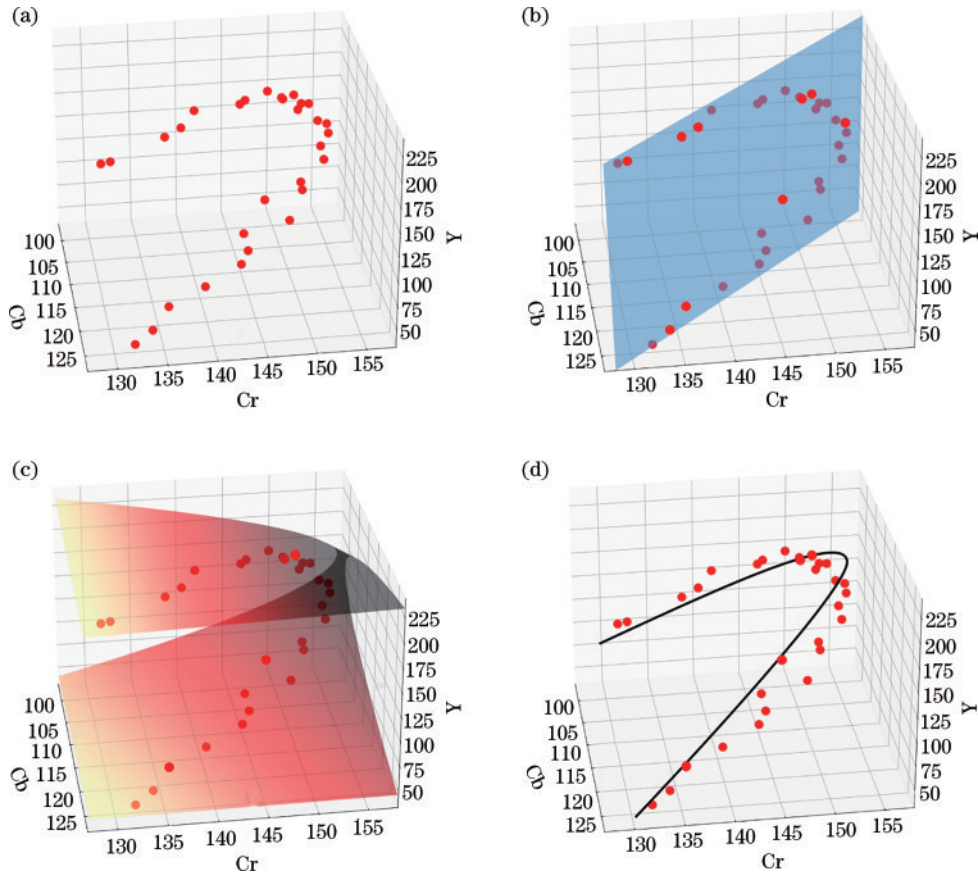


图 11 黄色人种的偏好肤色分布。(a) 30 个点在 YCbCr 色彩空间的分布;(b) 一维平面拟合;(c) 二维曲面拟合;(d) 一维平面与二维曲面相交确定的曲线

Fig. 11 Distribution of preferred skin color of Asian. (a) Distribution of 30 points in the YCbCr color space; (b) 1D plane fitting; (c) 2D quadric surface fitting; (d) curve determined by intersection of the 1D plane and the 2D quadric surface

$$\begin{cases} \omega_{Cb}(m, n) = c \exp \left[- \left(D_{Cb}(m, n) - \frac{T_{Cb(max)} + T_{Cb(min)}}{2} \right)^2 \right] \\ \omega_{Cr}(m, n) = c \exp \left[- \left(D_{Cr}(m, n) - \frac{T_{Cr(max)} + T_{Cr(min)}}{2} \right)^2 \right] \end{cases} \quad (9)$$

最后,对肤色连通域内所有像素进行色彩映射,实现肤色增强,表达式为

$$\begin{cases} D'_{Cb}(m, n) = D_{Cb}(m, n) + \omega_{Cb}(m, n)(R_{Cb(pre)} - R_{Cb(mean)}) \\ D'_{Cr}(m, n) = D_{Cr}(m, n) + \omega_{Cr}(m, n)(R_{Cr(pre)} - R_{Cr(mean)}) \end{cases}, \quad (10)$$

式中: c 为常量且取值 $c=0.31$; D_{Cb} 与 D_{Cr} 分别表示当前肤色连通域的Cb分量与Cr分量; $T_{Cb(max)}$, $T_{Cb(min)}$, $T_{Cr(max)}$, $T_{Cr(min)}$ 分别表示式(4)所得阈值模型的边界值; D'_{Cb} 与 D'_{Cr} 分别表示经过肤色增强后肤色连通域的Cb分量与Cr分量。

3 实验与分析

对所提肤色增强方法进行三方面的实验分析。首先,面向实时视频,保证算法执行保持良好的实时性能,并使输出视频保持良好的视觉连续性。其次,所提肤色增强方法基于主观实验建立偏好肤色模型,通过主观评价来验证肤色增强的有效性。实验均在配备 Intel i7-8750H 2.2 GHz CPU 的计算机上进

行,实验图像与视频均选自 Pratheepan^[22]、ECU Face and Skin Detection^[23]、Feeval Skin video DB^[24]三个数据集。

3.1 实时性能

所提肤色增强方法引入镜头边界检测技术引导皮肤检测模型以镜头为单位进行更新,显著降低了模型更新频率。如表 2 所示,在 4 类不同内容的视频中,所

表 2 不同皮肤模型的更新次数

Table 2 Update times of different skin models

Model	Movies	TV Dramas	Sports programs	Reality shows	Total
Guided by video	332	284	905	133	1654
Frame by frame	9547	9411	15462	9164	43584

提肤色增强方法的更新次数为逐帧更新模型的 3.79% 左右。

表 3 比较了所提肤色增强方法与肤色中心点法、椭圆映射法^[7]在相同设备上对分辨率为 3840×2160 的相同视频片段进行肤色增强的逐帧用时,表中数据均

为经过大量测试计算所得平均时间。实验结果证明,所提肤色增强方法虽然增加了镜头边界检测步骤,但显著降低了模型更新所需的计算工作量。总体而言,所提肤色增强方法在实时视频处理中表现出更好的实时性能。

表 3 各模块的平均执行用时
Table 3 Execution time of each module

Method	Execution time /ms			Total time /ms	Speed / (frames·s ⁻¹)
	Shot boundary detection	Skin segmentation	Color enhancement		
Skin color center	None	14.03	12.06	26.09	38.32
Ellipse mapping	None	13.45	11.99	25.44	39.31
Proposed method	1.53	9.12	12.14	22.79	43.87

3.2 视觉连续性

采用所提肤色增强方法与椭圆映射方法对属于同一镜头的视频片段进行肤色区域检测,部分实验结果如图 12 所示。由实验结果可知:由于椭圆映射法采用逐帧更新肤色模型,模型的频繁更新导致相邻帧间的

皮肤区域分割结果可能存在显著差异,导致色彩增强后的视频在视觉上出现帧间闪烁现象;所提肤色增强方法在降低了模型更新计算量的同时,在相同镜头内也能保持更好的视觉连续性。



图 12 视觉连续性比较。(a)原始帧;(b)椭圆映射法;(c)所提方法

Fig. 12 Comparison of visual continuity. (a) Original frames; (b) Ellipse mapping method; (c) proposed method

3.3 肤色主观评价

图 13 展示了所提肤色增强方法对图像中的人类皮肤区域进行肤色增强前后的部分效果对比。为了更清晰地展示皮肤色彩的变化,将帧图像中面部皮肤区域单独放大展示。每一组展示结果中左侧为原始帧图像及其面部区域,右侧为经过肤色增强后的帧图像及其面部区域。经所提肤色增强方法进行色彩增强后,肤色显得更为“红润”,更接近人们主观印象中的健康肤色。

根据国际电信联盟提出的电视图像质量主观评价建议,邀请另外 20 名志愿者参与到对肤色图像的主观评价实验中。主观评价实验采用单刺激连续质量评价

方法(SCQE),选择 10 段分辨率为 3840×2160 、帧速为 60 frame/s 的样本视频,分别采用肤色中心点法、椭圆映射法与所提肤色增强方法对每段视频片段进行肤色增强处理。将包括原始片段在内的所有视频片段重新随机排序并显示在同一设备上。参与主观评价的观众在无原始参考的情况下,分别观看不同的样本视频,并在 [1, 10] 区间内对所看到视频图像中的肤色进行主观评分。最后,计算不同方法对所有视频片段获得的平均主观分值(MOS),如表 4 所示。实验结果证明,所提肤色增强方法在每一组内均获得了更高的平均主观分值,实现了更有效的肤色增强。

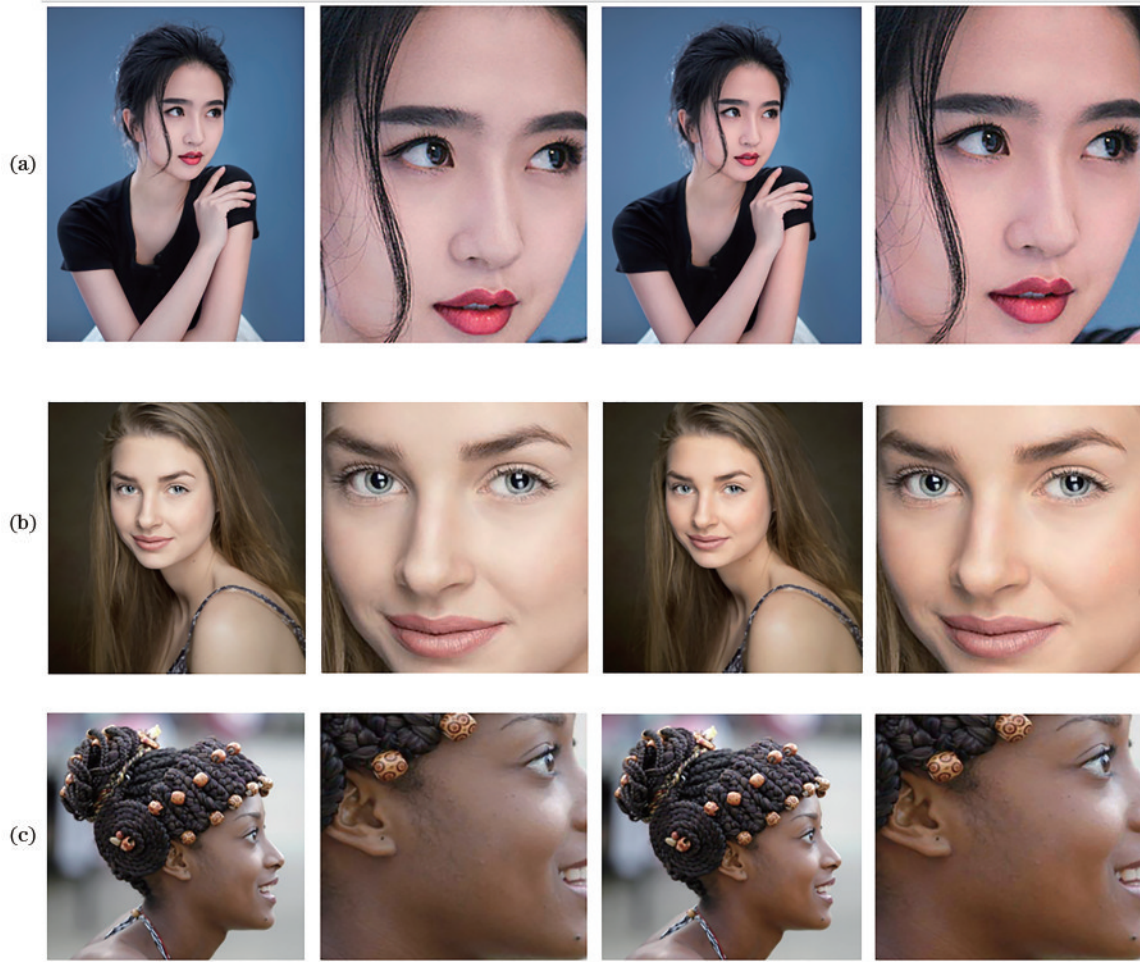


图 13 肤色增强。(a)黄色人种;(b)白色人种;(c)黑色人种
Fig. 13 Skin color enhancement. (a) Asian; (b) Caucasian; (c) African

表 4 测试视频主观评价的平均主观分数
Table 4 Mean opinion score for subjective assessment of test video

Index	MOS			
	Original video	Skin color center	Ellipse mapping	Proposed method
1	5.1	5.9	5.2	6.4
2	5.2	5.1	5.6	5.8
3	5.6	5.6	5.4	5.7
4	4.8	4.4	4.2	4.9
5	4.9	4.2	4.3	5.4
6	4.5	4.0	4.3	4.7
7	4.6	4.9	4.7	5.2
8	6.9	6.8	6.4	7.2
9	5.8	6.3	6.1	6.6
10	6.8	7.0	7.4	7.8

4 结 论

提出了一种基于视频镜头引导模型更新的自适应肤色增强方法。首先,通过引入镜头边界检测来引导

皮肤检测模型以镜头为单元进行更新,有效降低了模型更新的计算工作量,提高视频处理的实时性;其次,可动态更新的皮肤检测阈值模型提高了算法在不同视频中的皮肤分割自适应性;最后,基于主观实验,建立了针对三类人种的偏好肤色模型,实现了对不同种族更有针对性的肤色增强,同时通过调整映射权重避免了皮肤区域边界色彩突变的问题。实验证明,所提方法可以满足视频处理任务的实时性能要求,并且可有效实现对不同人种的肤色增强,与现有方法相比,具备更优的综合性能。

参 考 文 献

- [1] Francke H, Ruiz-del-Solar J, Verschae R. Real-time hand gesture detection and recognition using boosted classifiers and active learning[M]//Mery D, Rueda L. Advances in image and video technology. Lecture notes in computer science. Heidelberg: Springer, 2007, 4872: 533-547.
- [2] Jones M J, Rehg J M. Statistical color models with application to skin detection[C]//Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 23-25, 1999, Fort Collins, CO, USA. New York: IEEE Press, 1999:

- 274-280.
- [3] Zafarifar B, Bellers E B, de With P H N. Application and evaluation of texture-adaptive skin detection in TV image enhancement[C]//2013 IEEE International Conference on Consumer Electronics, January 11-14, 2013, Las Vegas, NV, USA. New York: IEEE Press, 2013: 88-91.
- [4] Chen H S, Wang T M, Chen S H, et al. Skin-color correction method based on hue template mapping for wide color gamut liquid crystal display devices[J]. *Color Research & Application*, 2011, 36(5): 335-348.
- [5] Kim D C, Kyung W J, Ha Y H. Preferred racial skin color reproduction in an image based on race classification [J]. *Journal of Imaging Science and Technology*, 2015, 59(2): 020504.
- [6] 王涛, 王佰玲. 非均匀光照条件下的彩色人脸图像增强 [J]. *计算机工程与应用*, 2017, 53(5): 164-169.
Wang T, Wang B L. Enhancement method for color face image with non-uniform illumination[J]. *Computer Engineering and Applications*, 2017, 53(5): 164-169.
- [7] Kim D H, Do H C, Chien S I. Preferred skin color reproduction based on adaptive affine transform[J]. *IEEE Transactions on Consumer Electronics*, 2005, 51(1): 191-197.
- [8] Park D S, Kwak Y, Ok H, et al. Preferred skin color reproduction on the display[J]. *Journal of Electronic Imaging*, 2006, 15: 041203.
- [9] 吴平, 张霞. 偏好肤色的获取与再现研究[J]. *包装工程*, 2016, 37(3): 146-152.
Wu P, Zhang X. Acquisition and reproduction of preferred skin colors[J]. *Packaging Engineering*, 2016, 37(3): 146-152.
- [10] 蒋崛, 张小宁, 梁志虎, 等. HSV 空间中具有调整因子的肤色增强方法[J]. *真空科学与技术学报*, 2011, 31(5): 515-520.
Jiang J, Zhang X N, Liang Z H, et al. Skin-color enhancement with adjustment factor in color-space of hue-saturation-value[J]. *Chinese Journal of Vacuum Science and Technology*, 2011, 31(5): 515-520.
- [11] 陈芳芳, 计忠平. 人脸肤色检测与编辑[J]. *系统仿真学报*, 2019, 31(7): 1377-1386.
Chen F F, Ji Z P. Face skin detection and color transferring[J]. *Journal of System Simulation*, 2019, 31(7): 1377-1386.
- [12] 邱云飞, 温金燕. 基于 DeepLabV3+ 与注意力机制相结合的图像语义分割[J]. *激光与光电子学进展*, 2022, 59(4): 0410008.
Qiu Y F, Wen J Y. Image semantic segmentation based on combination of DeepLabV3+ and attention mechanism [J]. *Laser & Optoelectronics Progress*, 2022, 59(4): 0410008.
- [13] 姚燕, 胡立坤, 郭军. 基于改进 DeepLabv3+ 网络的轻量级语义分割算法[J]. *激光与光电子学进展*, 2022, 59(4): 0410015.
Yao Y, Hu L K, Guo J. Improved lightweight semantic segmentation algorithm based on DeepLabv3+ network[J]. *Laser & Optoelectronics Progress*, 2022, 59(4): 0410015.
- [14] 张欢, 仇大伟, 冯毅博, 等. U-Net 模型改进及其在医学图像分割上的研究综述[J]. *激光与光电子学进展*, 2022, 59(2): 0200005.
Zhang H, Qiu D W, Feng Y B, et al. Improved U-Net models and its applications in medical image segmentation: a review[J]. *Laser & Optoelectronics Progress*, 2022, 59(2): 0200005.
- [15] 易三莉, 王天伟, 杨雪莲, 等. 基于改进 U-Net 的肺野分割算法[J]. *激光与光电子学进展*, 2022, 59(2): 0210010.
Yi S L, Wang T W, Yang X L, et al. Lung field segmentation algorithm based on improved U-Net[J]. *Laser & Optoelectronics Progress*, 2022, 59(2): 0210010.
- [16] Jain V, Learned-Miller E. FDDB: a benchmark for face detection in unconstrained settings[D]. Amherst: University of Massachusetts Amherst, 2010.
- [17] Korsunov N I, Toropchin D A. Recognition method of near-duplicate images based on the perceptual hash and image key points using[C]//2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, September 24-26, 2015, Warsaw, Poland. New York: IEEE Press, 2015: 261-264.
- [18] Viola P, Jones M J. Robust real-time face detection[J]. *International Journal of Computer Vision*, 2004, 57(2): 137-154.
- [19] Shi X P, Shan S G, Kan M N, et al. Real-time rotation-invariant face detection with progressive calibration networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2295-2303.
- [20] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1867-1874.
- [21] Garcia C, Tziritas G. Face detection using quantized skin color regions merging and wavelet packet analysis[J]. *IEEE Transactions on Multimedia*, 1999, 1(3): 264-277.
- [22] Tan W R, Chan C S, Yogarajah P, et al. A fusion approach for efficient human skin detection[J]. *IEEE Transactions on Industrial Informatics*, 2012, 8(1): 138-147.
- [23] Phung S L, Bouzerdoum A, Chai D. Skin segmentation using color pixel classification: analysis and comparison [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(1): 148-154.
- [24] Stöttinger J, Hanbury A, Liensberger C, et al. Skin paths for contextual flagging adult videos[M]//Bebis G, Boyle R, Parvin B, et al. *Advances in visual computing. Lecture notes in computer science*. Heidelberg: Springer, 2009, 5876: 303-314.