

## 面向互联网视频的人体皮肤区域检测

郭俊波<sup>1,2</sup>, 张冬明<sup>1</sup>, 张勇东<sup>1</sup>, 徐杰<sup>1,2</sup>, 汪文英<sup>1,2</sup>

(1. 中国科学院 计算技术研究所, 北京 100190; 2. 中国科学院 研究生院, 北京 100190)

**摘要:** 提出了一种改进的皮肤区域检测算法, 首先以改进的皮肤颜色检测模型为基础检测可能的肤色点, 进一步对可能的肤色区域进行腐蚀、膨胀处理以去除零散肤色点, 获得连通性较好的可能皮肤区域, 之后通过灰度共生矩阵度量纹理特性, 去除非肤色区域。实验数据表明该算法能够有效提高视频图像的皮肤区域检测的准确率。

**关键词:** 皮肤区域检测; 肤色模型; 灰度共生矩阵; 线段编码; 连通区域

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-436X(2009)04-0006-05

## Human skin region detection in Web video

GUO Jun-bo<sup>1,2</sup>, ZHANG Dong-ming<sup>1</sup>, ZHANG Yong-dong<sup>1</sup>, XU Jie<sup>1,2</sup>, WANG Wen-ying<sup>1,2</sup>

(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

2. Graduate School of Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** An improved detection algorithm was presented. Firstly, an improved skin color model, which adopted many preprocessing including optical offset, auto white balance, and auto gamma correction, was built to detect candidate skin pixels. Secondly, the image was filtered with erosion and dilation to exclude unconnected candidate pixels. Lastly, gray co-matrix was used to exclude those candidate human skin regions with complex texture. Experimental results show that this algorithm improves the skin region detection precision.

**Key words:** skin region detection; skin color model; gray co-matrix; line segment coding; connected region

### 1 引言

人是视频图像活动对象的重要的主体之一, 皮肤是人体最显著的生理特征之一, 皮肤区域检测已经成为提取视频中人体语义概念的重要技术基础。皮肤检测一方面是人脸检测和跟踪、人体姿态识别等应用的基础和关键步骤; 另一方面, 由于互联网上大量包含敏感色情内容的图片和视频, 对人们特别是青少年产生了不可忽视的毒害作用, 而采用图

像内容分析技术来过滤可能的色情图片是有效的解决方案, 其中, 皮肤检测是至关重要的一个步骤。本文研究面向互联网视频的皮肤区域检测技术, 为各类视频中人体语义概念提取应用提供技术基础。

图像和视频中的皮肤区域检测具有许多共同的技术基础, 就其本质来说, 视频中皮肤区域检测是对时间轴上连续的多帧图像进行皮肤区域检测, 通常可利用多帧图像的检测结果, 综合判定皮肤区域, 有可能降低误判率。然而, 视频图像的分辨率

收稿日期: 2008-11-12; 修回日期: 2009-01-15

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2007CB311100); 国家高技术研究发展计划(“863”计划)基金资助项目(2007AA01Z416); 国家自然科学基金资助项目(6077356); 北京市科技新星计划基金资助项目(2007B071)

**Foundation Items:** The National Basic Research Program of China (973 Program) (2007CB311100); The National High Technology Research and Development Program of China (863 Program) (2007AA01Z416); The National Natural Science Foundation of China(6077356); Beijing New Star Project on Science & Technology(2007B071)

较低,特别是面向在线播放业务的视频内容,其中80%以上分辨率仅在QVGA(320×240)左右,而且压缩比高,图像质量较差,因此,就单帧视频图像的皮肤区域检测来说,其面临比图像检测更大的挑战,传统的皮肤区域检测技术对网络视频图像往往会有较高的误判率或漏判率。

## 2 相关工作介绍

皮肤区域检测技术属于基于内容的图像检测技术(CBIR, content-based image retrieval)范畴,跟其他CBIR一样,都是对某个对象或概念的抽象描述,这种抽象一般很难有一个完整标准的定义,因此要将概念转换成算法,就不可能只从单一的特征着手。因此,皮肤区域检测,通常都会用到多项特征,例如颜色、形状、面积、纹理等。

基于统计的肤色检测是皮肤区域检测的基础,主要步骤是颜色空间变换和肤色建模。肤色的统计特征在不同颜色空间会有不同的呈现,选择颜色空间本身就是选择肤色检测的最基本特征表示,研究分析结果表明颜色空间的变换并不能改善肤色紧致性、肤色和非肤色可分辨性以及分类等性能,但RGB及线性变换空间却具有较好的类可分辨性和分类性能<sup>[1]</sup>。

肤色检测的相关理论,文献[2,3]中做了较好的综述,这里再做简要总结。通过训练样本集建立肤色模型是肤色检测的关键,根据不同应用可以将肤色建模分为静态和动态2类。目前常用的静态肤色建模基本方法有3种:肤色范围、高斯密度函数估计和直方图统计,3者分别对应阈值化、参数化和非参数化方法。

用数学表达式明确规定肤色范围是一种简单的肤色建模方法,例如Chai采用 $YCbCr$ 颜色空间的 $C_bC_r$ 平面,如果输入像素的颜色落入 $RC_r=[133, 173]$ 和 $RC_b=[77, 127]$ 限定的矩形区域,则判定属于肤色像素<sup>[4]</sup>。其优点是算法简单,非识别准确率不高。

高斯密度函数估计是一种参数化建模,可以用单峰高斯模型SGM<sup>[5]</sup>或高斯混合模型GMM<sup>[6]</sup>,并依靠模型计算像素点为肤色点和非肤色点的概率,根据概率大小判定是否属于肤色点。

直方图肤色模型是一种非参数模型。利用肤色样本的直方图统计可以构造肤色概率图,它为离散化的颜色空间中的不同区域赋予不同的肤色概率值,将输入像素点转换到与之相同的颜色空间,然

后在肤色概率图中查找它所处区域属于肤色点的概率,以判决它是否为肤色点。与此不同,贝叶斯分类器则同时利用肤色样本和非肤色样本的概率图,通过这2个先验概率,利用贝叶斯来计算获得像素点为肤色点的概率<sup>[7]</sup>,由于该方法同时使用了正反样本,理论上可获得更好的分类效果。然而,样本的选择是影响直方图肤色模型性能的决定因素,必须要选择典型的样本,否则样本会呈指数式增长。

肤色的动态建模方法可以分成2类:第一类方法是可以将肤色模型参数调节到适应某幅静态图像;第二类方法是针对序列图像,能适应成像条件随时间的变化。这些自适应方法所建立的动态肤色模型的通用性可能较差,但是在特定条件下的性能更优,能获得高检测率和低误检率。

单纯以肤色进行皮肤区域的检测,有可能产生误判情形:类似皮肤颜色的对象或背景,比如沙漠、黄色的墙壁、沙发、地板、卡通场景对象等。因此,有必要引入更多的皮肤区域的特征,比如,利用Sobel边缘检测<sup>[8]</sup>和Gabor滤波获得皮肤纹理信息<sup>[9]</sup>来确定可能的皮肤区域;提取皮肤区域的颜色向量集中程度以滤除误检测的皮肤区域。利用这些综合特征信息,借助Adaboost算法自动学习皮肤区域的颜色、形状和纹理等特征,以建立更好的皮肤区域分类器,从而获得更精准的检测结果。然而,这些肤色检测方法在提高肤色检测准确率的同时也导致了检测速度严重下降,不适合实时性要求高的应用。

本文充分利用互联网视频的特点,从改进皮肤区域检测入手,综合应用人脸检测和关键部位检测,有效提高了视频中皮肤区域检测的准确率和召回率,并尽可能保证检测算法的速度,以满足实时检测的需求。

## 3 本文工作

为适应低分辨率和低质量视频中皮肤区域检测任务的需求,首先对现有肤色像素点检测算法改进,之后利用区域分布特性、基于灰度共生矩阵的纹理特性等特征进一步去除可能的误检区域,细化实际的皮肤区域。

### 3.1 改进的肤色检测

互联网上视频图像受分辨率、码率限制,图像一般较差,尤其是视频分享网站中自拍视频,光线变化范围非常大。受这些因素影响,原肤色检测算

法性能大打折扣，本文设计了一种改进的肤色检测算法。

首先，选择合适的颜色空间。虽然，颜色空间的变换并不能从本质上改善肤色检测的效率，为了尽可能提高肤色检测的效率，选择在 YUV 空间进行肤色检测，其原因在于几乎所有视频编码标准都以 YUV 作为其标准的输入输出，无需进行额外的空间转换。

其次，对视频帧图像进行预处理。对图像进行光线补偿、自动白平衡、自动伽玛矫正等预处理。预处理尤其对于互联网上自拍类视频的皮区域检测有重要作用。这类视频受拍摄条件限制，光线、曝光度明显区别于专业摄制的视频，画面要么严重偏暗，要么过亮，或者受光线影响，皮肤区域严重偏色。预处理提高了这类视频中皮肤区域的检测的准确率和召回率，是关键步骤之一。

接下来，建立 YUV 空间上的肤色模型。在 YUV 空间中，定义肤色色调，根据大量图像的彩色分析，可以确定人的肤色色调的变化范围。但是该模型也会把一些与人的肤色比较接近的其他区域当作肤色区域。例如该模型难以把肤色与棕色头发或灰色背景分开。为此，需要在 YIQ 颜色空间中滤掉其他成分。如图 1 所示，在 YIQ 空间中，I 分量代表从桔黄到蓝绿的色调，I 值越小，包含的黄色越多，蓝绿色越小，所以肤色在 YIQ 空间内 I 值在一个范围里变化，通过实验，确定其合适的参数范围，该参数可用于在自适应的肤色检测控制。

应用该肤色模型，遍历图像中像素点，判定其为肤色点或非肤色点。对所有肤色点用线段编码的方式进行记录。该算法能够将连续的肤色点识别为同一个对象。同时将非肤色点像素重置为黑色，以便于后续处理。图 2 是处理前后的 2 帧图像实例。

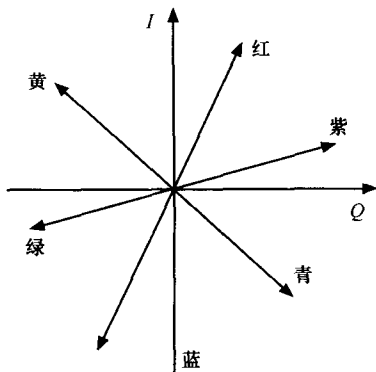


图 1 YIQ 颜色空间中色度信号矢量图



图 2 肤色点检测前后的图像比较

### 3.2 连通区域特性

记录获得的皮肤区域是进一步提取视频中人体语义概念的基础。在肤色检测过程中，以线段编码算法记录可能的皮肤区域。

获得可能皮肤区域的线段编码之后，进一步通过腐蚀和膨胀的方式，对所获得的区域进行裁剪，去除非皮肤区域，裁剪原则包括：

- 1) 皮肤区域的宽度大于阈值  $T_1$ ， $T_1$  与视频图像宽度成正比；
- 2) 皮肤区域的高度大于阈值  $T_2$ ， $T_2$  与视频图像高度成正比；
- 3) 皮肤区域的面积应大于阈值  $T_3$ ， $T_3$  与视频图像面接成正比。

制定以上原则的理由在于，在以人的活动为主的视频中，人绝大部分情况下为内容的主体，相应地，皮肤区域也占据较大的区域比例。如果区域尺度过小，通常为颜色上比较接近肤色的其他对象区域。

### 3.3 基于灰度共生矩阵的纹理度量

为避免颜色类似肤色的对象被认为是皮肤，引入纹理特性来进一步分辨可能的皮肤区域。人眼对亮度分量的感知明显高于色度分量，因此只考虑亮度分量的纹理，具体使用其共生矩阵 (co-occurrence matrix) 予以衡量。灰度共生矩阵<sup>[10]</sup>用 2 个位置的像素的联合概率密度来定义，它不仅反映亮度的分布特性，也反映具有同样亮度或接近亮度的像素之间的位置分布特性，是有关图像亮度变化的二阶统计特征。它是定义一组纹理特征的基础。

一幅图像的灰度共生矩阵能反映出图像灰度关于方向、相邻间隔、变化幅度的综合信息，它是

分析图像的局部模式和它们排列规则的基础。

设  $f(x, y)$  为一幅二维数字图像，其大小为  $M \times N$ ，灰度级别为  $L$ ，则满足一定空间关系的灰度共生矩阵为

$$P(i, j) = \text{num}\{f(x_1, y_1) = i \cap f(x_2, y_2) = j, (x_1, x_2 \in (1, M), y_1, y_2 \in (1, N))\}$$

其中， $\text{num}\{x\}$  表示集合  $x$  中的元素个数，显然  $P$  为  $L \times L$  的矩阵，若  $(x_1, y_1)$  与  $(x_2, y_2)$  间距离为  $d$ ，两者与坐标横轴的夹角为  $\theta$ ，则可以得到各种间距及角度的灰度共生矩阵  $P(i, j, d, \theta)$ 。

纹理特征提取的一种有效方法是以灰度级的空间相关矩阵即共生矩阵为基础的，因为图像中相距  $(\Delta x, \Delta y)$  的 2 个灰度像素同时出现的联合频率分布可以用灰度共生矩阵来表示。若将图像的灰度级定为  $L$  级，那么共生矩阵为  $L \times L$  矩阵，可表示为  $M(\Delta x, \Delta y)(h, k)$ ，其中位于  $(h, k)$  的元素  $m(h, k)$  的值表示一个灰度为  $h$  而另一个灰度为  $k$  的 2 个相距为  $(\Delta x, \Delta y)$  的像素对出现的次数。

对粗纹理的区域，其灰度共生矩阵的  $m(h, k)$  值较集中于主对角线附近。因为对于粗纹理，像素对趋于具有相同的灰度。而对于细纹理的区域，其灰度共生矩阵中的  $m(h, k)$  值则散布在各处。为了能更直观地以共生矩阵描述纹理状况，从共生矩阵导出 2 种反映矩阵状况的参数，进行区域纹理复杂性度量。

1) 能量：灰度共生矩阵元素值的平方和，所以也称能量，反映了图像灰度分布均匀程度和纹理粗细度。如果共生矩阵的所有值均相等，则值小；相反，如果其中一些值大而其他值小，则值大。当共生矩阵中元素集中分布时，该值大。该值大时表明一种较均一和规则变化的纹理模式。

2) 逆差距：反映图像纹理的同质性，度量图像纹理局部变化的多少。其值大则说明图像纹理的不同区域间变化少，局部非常均匀。

为降低纹理特性度量开销，仅对上节检测出的可能皮肤区域进行灰度共生矩阵的计算。

#### 4 实验结果

为了验证算法在皮肤区域检测上的查准率、查全率和检测速度等性能指标，设计了如下系统。

系统从特定网络路由由网关上捕捉视频解码数据流量，从中选取 1 000 个访问频率较高的节目，

分辨率主要包括  $320 \times 240$ 、 $352 \times 240$  等。

将这些节目数据解码，每个节目截取有代表性的 10 帧图像，对共计 10 000 幅图像按照包含皮肤区域所占百分比进行分类、标注，获得如表 1 所示的结果。

表 1 测试数据分类结果

皮肤区域百分比	数目
>50%	874
(20%, 50%)	1 736
(5%, 20%)	2 823
(0.1%, 5%)	2 341
≤1%	2 226

用设计的皮肤区域检测算法进行检测，记录每幅图像所检测出的皮肤区域所占百分比，并与标注百分比进行比对，并计算相应的准确率和召回率。

表 2 中的准确率和召回率分别按照式 (1) 和式 (2) 计算。

$$\text{准确率} = \frac{\text{检测正确的图像数}}{\text{检测出的图像总数}} \quad (1)$$

$$\text{召回率} = \frac{\text{检测正确的图像数}}{\text{图像总数}} \quad (2)$$

式 (1) 和式 (2) 中，检测正确的图像数是指检测出的皮肤区域与实际百分比之绝对差小于段区间 10% 的图像数目，比如对区间  $(20\%, 50\%]$  而言，检测结果与实际百分比绝对差小于，则认为检测结果正确。

表 2 皮肤区域检测结果

皮肤区域百分比	准确率	召回率
>50%	800/830=96.4%	800/874=91.5%
(20%, 50%)	1 313/1 367=96.0%	1 313/1 736=75.6%
(5%, 20%)	1 687/1 776=95.0%	1 687/2 823=60.0%
(1%, 5%)	1 425/1 507=94.6%	1 425/2 341=60.9%
≤1%	2 016/2 139=94.2%	2 016/2 226=90.6%
合计	7 241/7 619=95.0%	7 241/10 000=72.4%

从表 2 结果可以看出，本文算法在检测皮肤区域百分比不同的图像中，均有较高的准确率，而召回率在不同图像则有较大波动，在皮肤区域大于 50% 或小于 1% 时，召回率较高，而在其他情况下，召回率有较大降低，这主要受区域检测中的区域裁剪阈值设定的影响，需要设计自适应的区域裁剪算法以改善裁剪效果。

### 5 结束语

从网络随机选择的 1 000 个网络视频的实验结果表明: 该算法能够自适应地准确检测视频图像中的皮肤区域, 具有 95%左右的准确率和 72%左右的召回率。

算法在 YUV 空间进行皮肤区域的检测, 减少了色彩空间转换导致的计算量增加, 同时采用的检测技术进行优选、结构优化组合, 保证了算法有较好的运算速度。同时, 线段编码记录的皮肤区域可用于进一步的姿态识别、人脸跟踪等视频语义概念的提取。

### 参考文献:

[1] 徐战武, 朱森良. 肤色检测最优空间[J]. 计算机辅助设计与图形学学报, 2006, 18(9): 1350-1356.  
XU Z W, ZHU M L. Optimum colorspace for skin-tone detection[J]. Journal of Computer-Aided Design & Computer Graphics, 2006, 18(9): 1350-1356.

[2] 徐战武, 朱森良. 基于颜色的皮肤检测综述[J]. 中国图像图形学报, 2007, 12(3): 377-388.  
XU Z W, ZHU M L. Color-based skin detection: a survey[J]. Journal of Computer Graphics, 2007, 12(3): 377-388.

[3] 陈锻生, 刘政凯. 肤色检测技术综述[J]. 计算机学报, 2006, 29(2): 194-207.  
CHEN D S, LIU Z K. A survey of skin color detection[J]. Journal of Computers, 2006, 29(2): 194-207.

[4] CHAI D, NGAN K N. Locating facial region of a head- and-shoulders color image[A]. Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition[C]. Nara, Japan, 1998. 124-129.

[5] HSU R L, ABDEL MOTTALIB M, JOIN A. Face detection in color images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 696-706.

[6] YANG M H, AHUJA N. Gaussian mixture model for human skin color and its application in image and video databases[A]. Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases(SPIE 99) , 1999. 458-466.

[7] CHAI D, BOUZERDOUM A. A Bayesian approach to skin color classification in ycbcr color space[A]. Proceedings of IEEE Region Yen Conference(TENCON'2000)[C]. 2000.421-424.

[8] GUO Y C, HOU Z X, YU M. Automatic localization of pupils in color huma images[A]. Proceedings of the Third International Conference on Machine Learning and Cybernetics[C]. 2004. 4067-4070.

[9] 杨金锋, 傅周宇, 谭铁牛等. 一种新型的基于内容的图像识别与过滤方法[J]. 通信学报, 2004, 25(7): 93-106.

YANG J F, FU Z Y, TAN T N, et al. A novel algorithm for content-based image recognition and filtering[J]. Journal on Communications, 2004, 25(7): 93-106.

[10] KENNETH R, CASTLEMA N. Digital Image Processing[M]. Prentice-Hall, 1999.

### 作者简介:



郭俊波 (1975-), 男, 湖北公安人, 中国科学院计算技术研究所博士生, 主要研究方向为互联网视频搜索与分析。



张冬明 (1977-), 男, 江苏盐城人, 博士, 中国科学院计算技术研究所助理研究员, 主要研究方向为视频编解码、视频图像内容分析与理解。



张勇东 (1973-), 男, 山东掖县人, 博士, 中国科学院计算技术研究所副研究员、硕士生导师, 主要研究方向为视频编解码, 视频分析和检索。



徐杰 (1982-), 男, 山西忻州人, 中国科学院计算技术研究所博士生, 主要研究方向为视频编解码、图像内容分析。



汪文英 (1982-), 男, 安徽桐城人, 中国科学院计算技术研究所博士生, 主要研究方向为视频图像内容分析。