彩色视频序列图像中的人脸跟踪方法

夏思宇 夏良正 金立左
(东南大学自动控制系,南京 210096)

摘 要 针对彩色视频序列图像的人脸检测,提出了一种基于肤色的人脸跟踪方法。该方法首先在 Hsu 提出的肤色模型基础上,采样一种自肤色分割算法来提取复杂背景下人脸的肤色特征,与传统的采用固定肤色模型的检测算法相比,该方法具有更好的检测效果;然后,在人脸跟踪过程中采用 Condensation 滤波跟踪算法,并对算法做了两点改进,即在跟踪过程中采用基于 Metropolis 算法的重采样方法以及自适应的动态模型,实现了复杂背景下的人脸自由运动的跟踪,并从各种影片中截取了彩色视频序列图像进行了测试实验。实验结果表明,该方法有效地解决了复杂背景下人脸自由运动、光照变化及部分遮挡的问题,且精度较高。

关键词 彩色视频序列 人脸跟踪 Condensation 滤波 自肤色分割
中图法分类号:TP391.4 **文献标识码**:A **文章编号**:1006-8961(2006)09-1249-06

Face Tracking in Color Image Sequences

XIA Si-yu, XIA Liang-zheng, JIN Li-zuo
(*Department of Automatic Control Engineering, Southeast University, Nanjing* 210096)

Abstract Human face tracking has received extensive attention because of its potential applications in many fields, such as video coding, surveillance and human-computer interface. The aim of this paper is to track human face in color image sequences effectively. A method using human skin feature integrated with Condensation (conditional density propagation) algorithm is proposed. At first, this paper presents a method of self-skin segmentation based on Hsu R L's skin color model, to detect human skin in color image correctly and effectively. Compared with conventional skin color model methods, this method has more precise detection performance in varying lighting condition. Additionally, this paper presents two improvements for Condensation algorithm: a simple and efficient sampling method based on the Metropolis algorithm and an adaptive dynamical model are used in Condensation algorithm for robust face tracking. Experimental results show that in complex backgrounds and occlusion case, this approach could track human face's free movement.

Keywords color image sequence, face tracking, Condensation algorithm, self-skin segmentation

1 引 言

近年来,随着计算机技术的高速发展与商业应用需求的推动,人脸检测与跟踪技术得到了很大的发展,不仅在电视电话会议、远程教学、监视与监控等场合,都需要对特定人脸目标进行实时跟踪、分析和传递,而且可视电话、视频会议、基于内容的压缩与检索、身份鉴别、人机智能交互等许多应用都与好的人脸跟踪算法紧密相关。

目标跟踪技术包括基于运动的方法及基于模型的方法。前者是采用运动分割、光流、立体视觉等方法,利用时空梯度、卡尔曼滤波器等来跟踪人脸运动;后者是首先通过获取目标的先验知识来构造目标模型,然后对输入的每一帧图像通过滑动窗口进行模型匹配。在人脸跟踪中,往往将这两种方法结合使用[1]。目前的人脸跟踪系统中普遍存在对人脸姿态变化、光线变化等比较敏感,且不能正确处理

基金项目:浙江省长三角联合攻关项目(2005E60007)
收稿日期:2005-12-30;改回日期:2006-04-03
第一作者简介:夏思宇(1978~),男,2000 年、2003 年先后获南京航空航天大学学士、硕士学位,现为东南大学博士研究生。研究方向为人脸检测、跟踪与识别。E-mail:xia081@gmail.com

多目标背景下目标遮挡时的跟踪等问题。

众所周知,肤色是人脸的重要特征,其不仅不依赖于面部的细节信息,而且肤色提取具有速度快、姿态不变性等特点。文献[2]提出通过人脸肤色来跟踪人脸的思想,并利用统一的高斯模型来描述所有人的肤色,以用于进行人脸检测与跟踪,但它忽略了不同光照环境下的肤色差异,这往往使得由于检测精度低而破坏了整个系统的性能。

针对彩色视频序列图像,本文提出一种基于自肤色分割的方法,用以提取人脸的肤色特征,并对 Condensation(conditional density propagation)滤波跟踪方法进行改进,并将其用于人脸的跟踪,从而不但有效地保证了复杂背景下多目标跟踪的准确性,而且实现了人脸姿态变化与遮挡情况下的人脸跟踪。

2 人脸肤色提取

众所周知,肤色特征在人脸检测中是最常用的一种特征。传统的肤色检测算法都是采用一种或两种固定的肤色模型去处理所有的彩色人脸图像,但是由于每种模型都有自身的局限性,因此不可能在各种光照与背景环境下都有效。实验表明,一方面光线的变化会影响到肤色的统计特性;另一方面,肤色模型是根据大量的肤色像素点统计得到的,但不一定“十分适合”每张输入图像。对单幅图像来说,由于肤色点的聚类分布范围很小,肤色模型很容易将非肤色点也包含进来,从而导致肤色点的“过检测”。为此,根据同一图片中人脸颜色与非人体背景的色彩具有相对差异性以及具体图片中人脸的颜色具有一致性的事实,本文提出了一种自肤色检测算法。

自肤色人脸检测算法的主要思路是针对单幅图像中人脸的肤色分布,在色度空间中进行区域分割,同时结合肤色统计信息来自动选取最类似的区域作为肤色区域,以克服传统的事先通过大量肤色样本统计得到的肤色模型,在用肤色检测时容易导致的过检测问题。

2.1 肤色空间变换

本文的检测算法选择在 YC_rC_b 色彩空间进行的,理由如下:(1) YC_rC_b 颜色空间受到亮度影响较小,能更好地利用色彩空间的颜色信息进行建模;(2)采用 YC_rC_b 颜色空间在视频中可以避免额外色彩空间转换的计算;(3)在亮度与色度的分离和肤色聚类上,YC_rC_b 颜色空间类似于 TSL 色度空间。

将图像的色度空间由 RGB 空间转换为 YC_rC_b 空间之后,再统计每个像素的 C_rC_b 值,并取其对数值,即形成了一个 C_rC_b 值的 2 维直方图分布矩阵。通过大量实验发现,由于正常光照下的某个特定人脸的绝大部分区域的颜色具有较强的一致性,因此肤色点 C_rC_b 分量的 2 维直方图呈一定的峰状(如图版Ⅰ图 1(b)所示)。

2.2 标记提取

为了消除干扰,并提取 2 维直方图中的峰值,本文引入了一个参考值 K,即

$$K = C/H \tag{1}$$

其中,C 表示从峰值到邻近峰谷的垂直最短距离,H 表示峰值高度(如图 1 所示)。

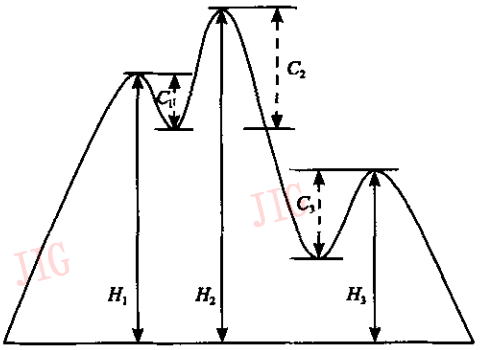


图 1 标记提取示意图

Fig. 1 Markers extraction

标记提取时,首先,检测 2 维直方图中的各个峰值区域,然后将参考值大于阈值的峰值作为标记(这里阈值取为 0.1),以便于进行下一步的肤色检测。

2.3 肤色检测

类似于带标记的分水岭算法,本文将 2 维直方图看成一个位于水平面以下的各个标记过的山峰,然后通过调整水位,使之逐步下降,这样各个山峰便依次显露出来。可见算法的过程实际上就是一个在 2 维直方图上不断“降低水位”的过程,即首先设定水平面高度 H,然后下调水位 d,以检测是否有高于水平面($H-d$)的标记峰值,再接着调整水位,并检测位于 $H-2d$ 与 $H-d$ 之间的像素值,若该像素是新的标记峰值,则将它作为一个新的分割区域中心,若不是,则查找周围最近的已经标记过的点,并分配相同标记。如此反复,直到所有点都分配到一个标记。

接着，通过比较各个峰值点与一个标准中心点距离来得到一个最近的峰值点。在这里，标准中心点取自文献[3]中椭圆模型的中心点坐标，即

$$\frac{(x-x_c)^2}{a^2}+\frac{(y-y_c)^2}{b^2}=1$$
$$\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}C_b-c_x\\C_r-c_y\end{bmatrix} \tag{2}$$

式中，$c_x=109.38$，$c_y=152.02$，$\theta=2.53$，$x_c=1.60$，$y_c=2.41$，$a=25.39$，$b=14.03$。这个模型是 Hsu 从 heinrich-hertz-institute(HHI)图像库中选取的 137 张图像，共计 853 571 个肤色点进行统计得来的[3]。文中的标准中心点取为这个椭圆模型的中心点，即

$$\begin{cases}x=x_c\\y=y_c\end{cases} \tag{3}$$

此时，可求得 C_b，C_r 分量平均值为

$$\begin{cases}\bar{C}_b\approx 108\\ \bar{C}_r\approx 150\end{cases} \tag{4}$$

本文将(108，150)这点定义为肤色区域在 C_rC_b 平面上的标准中心点。

最后，用与这个峰值点的标记相同的点来对整个图像进行肤色检测，即对图像的每个像素而言，若它在 YC_rC_b 空间中的 C_rC_b 分量属于这个标记，则定义为肤色值，否则为非肤色值。实验结果如图版 I 图 2(c)所示。

3　人脸跟踪

人脸跟踪通常可以看成是在给定观测量下，求解系统隐含状态量的估计问题。贝叶斯估计方法是将未知(隐含)状态量的先验知识和描述观测量与状态量关系的似然函数结合起来，利用贝叶斯公式来得到未知量的后验概率，但是迭代贝叶斯估计只在特定的模型和假设下才有解析解，包括线性高斯状态空间模型(卡尔曼滤波)和有限状态空间隐含马尔可夫模型(隐马尔可夫滤波)。然而，由于在许多实际问题中，状态空间模型含有非线性和非高斯成分，因此没有闭合的最优解。近 30 年来，人们提出了许多次优方案，如扩展卡尔曼滤波(extended Kalman filter，EKF)及确定性数值积分方法等，但当状态维数增加时，近似误差的收敛率下降。

3.1　Condensation 滤波跟踪方法

1996 年，Isard 与 Blake 将基于贝叶斯规则的概率模型引入到计算机视觉中，用于跟踪非刚体的、多关节表示的手掌运动，称之为 Condensation 滤波跟踪方法[4]。与 EKF 不同的是，Condensation 方法不是近似模型，而是利用加权的采样值来逼近真实后验概率值。

通常，通过建立一个状态空间模型，可以将视频序列的跟踪问题表示为一个动态系统的状态估计问题。在状态空间模型中，状态向量 X_t 用来描述时刻 t 系统的信息，观测向量 Z_t 为对应于时刻 t 状态的观测值。状态方程与观测方程分别为

$$X_t=f(X_{t-1},W_{t-1}) \tag{5}$$
$$Z_t=g(X_t,V_t) \tag{6}$$

其中，W、V 分别为对应的过程噪声与观测噪声。

以贝叶斯学派的观点，跟踪问题就是在给定时刻 t 观测向量 Z_t 的条件下，估算状态向量 X_t 的值，即估计后验概率密度函数 $p(X_t/Z_t)$。假设在时刻 $t-1$ 的概率密度函数 $p(X_{t-1}/Z_{t-1})$ 已知，则利用系统模型(式(5))就可以预测以下时刻 t 的先验概率密度：

$$p(X_t/Z_{t-1})=\int p(X_t/X_{t-1})p(X_{t-1}/Z_{t-1})\,\mathrm{d}X_{t-1} \tag{7}$$

注意，在式(7)中，利用了式(5)所描述的一阶马尔可夫过程，即

$$p(X_t/X_{t-1},Z_{t-1})=p(X_t/X_{t-1}) \tag{8}$$

而利用贝叶斯规则则可得

$$p(X_t\mid Z_t)=\frac{p(Z_t\mid X_t)p(X_t\mid Z_{t-1})}{p(Z_t\mid Z_{t-1})} \tag{9}$$

由于式(8)与式(9)是最优贝叶斯估计的一般概念表达式，因此通常不可能对它进行精确的分析。

Condensation 滤波跟踪方法的主要思想是利用一组带有相关权值的随机样本，以及基于这些样本的估算来表示后验概率密度 $p(X_t\mid Z_t)$。当样本数非常大时，这种概率估算将等同于后验概率密度。假设在时刻 t，系统包含 N 个样本，每个样本 S_t 都由加权样本对表示(s_t,ω_t)，其中 s_t 为一个样本的状态向量，ω_t 为样本的权值。其过程是，首先，对这 N 个样本，以其权值 ω_t 为概率进行采样，并通过抑制权值较小的样本与增强权值较大的样本来得到 N 个新的样本 $\tilde{S}_t$；然后，对每个样本 $\tilde{S}_t$ 进行预测，从 $p(X_{t+1}\mid X_t=s_t)$ 中生成 s_{t+1}，其中 $p(X_{t+1}\mid X_t)$ 表示系统状态的演化，由系统动态模型决定；最后，对权值进行更新，即 $\omega_{t+1}=p(Z_{t+1}\mid X_{t+1}=s_{t+1})$，其中，$p(Z_{t+1}\mid X_{t+1})$ 由系统观测模型决定，这样在时刻

$t+1$,系统状态的期望就可表示为

$$E[f(X) \mid Z_{t+1}] \approx \sum_{i=1}^{N} \omega_{t+1}^{(i)} f(s_{t+1}^{(i)}) \tag{10}$$

Condensation 方法适合跟踪那些非线性非高斯的运动,即并不要求概率密度函数为高斯分布,而这是卡尔曼滤波器所无法胜任的。另外,Condensation 方法的鲁棒性很强,能适应复杂的跟踪环境。在实际跟踪过程中,由于光照变化、遮挡或其他干扰物的出现,因而常常会出现一些不确定的观测数据。Condensation 的鲁棒性源于对这些不确定性的描述,其在跟踪过程中并不是只保留权值最大的样本,那些权值较小的样本仍然有可能保留下来,在后续的跟踪中起作用。

3.2 基于肤色特征的 Condensation 滤波跟踪改进算法

这里采用第 1 节提出的自肤色分割方法来提取人脸的肤色特征,并结合 Condensation 算法对人脸进行跟踪。具体步骤如下:

(1)初始化 为了标记人脸在图像中的位置,本文取系统的状态向量为 $X=[x,y,w,h]^{\mathrm{T}}$,其中 x,y 为人脸区域中心点的坐标,w,h 分别为人脸区域的宽与高。初始化时,首先在给定的人脸区域附近随机地选取 $N=100$ 个样本(这里假定初始帧的人脸区域位置已得到);然后,分别计算每个样本的观测值

$$G_i = \frac{N_i}{w_i \times h_i} \tag{11}$$

其中,N_i 为第 i 个样本中肤色像素点的数量;最后,将其归一化后即得到各个样本的权值

$$\omega_i = \frac{G_i}{\sum_{k=1}^{N} G_k} \tag{12}$$

这里的状态向量与权值用来预测下一帧中样本的概率分布。

(2)采样 本文采用一种基于 Metropolis 算法的采样方法[5,6],依据每个样本所对应的权值来进行采样,采样后的样本将用来做下一步的预测。Metropolis 是一种有效的重点抽样法,其算法为:系统从能量一个状态变化到另一个状态时,相应的能量从 E_1 变化到 E_2,其概率为

$$p = \exp[-(E_2 - E_1)/kT] \tag{13}$$

式中,k 为权值参数,T 为控制参数。如果 $E_2 < E_1$,则系统接收此状态,否则,以一个随机的概率接收此状态或丢弃此状态。这样经过一定次数的迭代后,系统会逐渐趋于一个稳定的分布状态。本文将这种抽样方法应用到 Condensation 滤波跟踪过程中,以便能更加有效地保留高权值的样本及更好地反映样本分布。从下面的实验结果(图 2)中可以看到,基于 Metropolis 算法采样的跟踪结果更为准确。

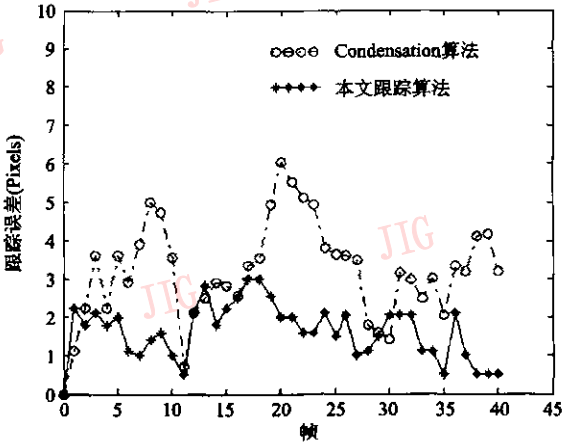


图 2 跟踪误差

Fig. 2 The error of face tracking

(3)预测 本文采用一个自适应的动态模型来预测下一帧中样本的各个状态量,其定义如下:

$$\begin{cases} x_{t+1} = x_t + N(\mu_x, k_1\sigma_x) \\ y_{t+1} = y_t + N(\mu_y, k_1\sigma_y) \end{cases} \tag{14}$$

$$\begin{cases} w_{t+1} = w_t \cdot N(\mu_{\mathrm{width}}, k_2\sigma_{\mathrm{width}}) \\ h_{t+1} = h_t \cdot N(\mu_{\mathrm{height}}, k_2\sigma_{\mathrm{height}}) \end{cases} \tag{15}$$

式中,$N(\cdot)$ 表示正态分布函数,μ_x,μ_y,μ_{width},μ_{height} 与 σ_x,σ_y,σ_{width},σ_{height} 分别为相应的均值与方差。其中

$$\begin{cases} \mu_x = x_t - x_{t-1} \\ \mu_y = y_t - y_{t-1} \end{cases} \tag{16}$$

$$\begin{cases} \mu_{\mathrm{width}} = w_t / w_{t-1} \\ \mu_{\mathrm{height}} = h_t / h_{t-1} \end{cases} \tag{17}$$

k_1,k_2 为方差系数(实验中,k_1 取为 2,k_2 取为 0.5),σ 取值为其相应均值的绝对值,这是考虑到目标移动的速度越快,其机动性可能越高的缘故。式(16)与式(17)为实验中所采用的估计公式。样本下一帧状态向量中的坐标 x_{t+1},y_{t+1} 为当前值加上一个变化量,这个变化量变化的范围由目标位置移动与形状改变的快慢来决定,这样就能解决由于人脸运动突然停止或改变方向而导致跟踪失败的问题。方差系数的选取是根据实验经验得到,因为这样可保证

模型既能适应缓慢的人脸运动，又能及时跟踪目标的快速移动。同时，由于考虑到目标的平移变化的速度要大于前后变化的速度，所以反映跟踪区域位置的动态模型的方差系数值要比反映跟踪区域大小的动态模型的大。

(4) 更新　计算预测得到的每个样本 S_{t+1} 的观测量

$$\begin{cases} G_{t+1} = \dfrac{N_{t+1}}{w_{t+1} \times h_{t+1}} \cdot \theta \\ \theta_{t+1} = \dfrac{N_{t+1}}{N_0} \end{cases} \tag{18}$$

其中，N_{t+1} 是由上一节的自肤色分割算法检测得到的肤色像素点数目，N_0 为初始帧中肤色像素点的数目，θ_{t+1} 表示样本中肤色点所反应出的人脸的信息量大小。

更新每个样本的权值：

$$\omega_{i,t+1} = \frac{G_{i,t+1}}{\sum_{k=1}^{N} G_{k,t+1}} \tag{19}$$

在时刻 $t+1$，系统状态向量均值与方差分别为

$$\bar{X} = \sum_{i=1}^{N} \omega_{i,t+1} X_i \tag{20}$$

$$\sigma_X = \sum_{i=1}^{N} \omega_{i,t+1} X_i^2 - \bar{X}^2 \tag{21}$$

3.3　人脸区域的消失与新的人脸区域的出现

在对人脸进行跟踪的过程中，可每隔 N 帧就对整幅图像进行人脸检测。如果检测到新的区域，且离现在所跟踪的区域较远，则标记为新的人脸区域，并进行跟踪。这种策略同时也能解决两个人脸区域重合后又分开的情况，因为两个跟踪系统可能会在分开之后集中到同一个人脸区域，采用这样的方法后，那个漏掉的人脸区域就会被重新检测到，并归为新的目标区域。

如果每个样本的观测值都很小，甚至为零，且一直持续 N 帧，在这种情况下，就认为人脸已消失，不再进行跟踪。如果在 N 帧之内，样本的观测值又恢复到原先的范围，则继续跟踪，因为有时由于光线变化剧烈而致使人脸区域检测到的肤色点过少，故会影响跟踪效果，而当光线恢复正常后，肤色点的数目又会增加。

4　实验结果

为验证本文算法效果，采用各种典型视频序列图像，如 American Miss、Carphone、IIT-NRC facial video database[7,8] 等对本文算法进行了实验，另外也从各种影片中截取了视频序列图像进行测试。这里以 3 组实验为例来进行说明，其一是来自人脸视频数据库中的序列图像，另外两组是来自影片中的部分片断，图像大小均调整为 160 × 120Pixels。实验平台是 P4-2. 4G 的微机，处理速度在 10fps 左右。

图版 I 图 3 是对来自 IIT-NRC 人脸视频数据库中的视频图像序列进行跟踪的部分结果。由该图可以看出，在人脸大小不断变化的情况下，算法能很好地进行跟踪。本文将跟踪得到的人脸区域的中心点位置与人工提取的相应帧的中心点位置进行了比较，结果如图 2 所示，由图 2 可以看出，改进后算法的平均跟踪误差在 1 ~ 2Pixels 之间。

图版 I 图 4 与图 5 分别是对来自两段影片中的视频序列图像进行跟踪的部分结果。从图版 I 图 4 中可以看出，在复杂背景及人脸姿态发生变化的情况下，算法仍能进行成功跟踪。对该序列的实验表明，文中提出的自肤色分割算法能在复杂背景下很好地检测到人脸的肤色区域，且跟踪算法对人脸姿态变化具有较好的适应能力。图版 I 图 5 是人脸快速运动，且出现部分遮挡情况下的跟踪结果。因为本文算法是基于肤色特征的跟踪，所以即使在人脸的面部特征(如眼睛、鼻子等) 被遮挡的情况下仍能取得很好的跟踪结果。

5　结　论

本文提出了一种自肤色分割的方法用于人脸的肤色检测，并对 Condensation 滤波跟踪算法做了改进，即在跟踪过程中采用了基于 Metropolis 算法的采样方法以及自适应的动态模型，不仅实现了复杂背景下的对人脸自由运动的跟踪，且精度较高。基于自肤色分割的检测方法不仅提高了系统对环境的适应能力，而且 Condensation 滤波鲁棒性强，适合跟踪人脸这种非线性非高斯的运动。实验证明，基于自肤色检测与 Condensation 滤波的人脸跟踪方法能有效解决人脸姿态变化与遮挡情况的跟踪问题，但目前这种方法存在的问题是肤色特征过于单一，还需加入其他人脸特征，以便对人脸运动进行准确跟踪。

参考文献(References)

1　Gong S，Psarrou A. Tracking and recognition of face sequences[A].

In：Proceedings of European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production[C]，Hamburg，Germany，1994：97～112.

2 Yang J，Waibel A. Tracking human faces in real-time[R]. Technical Report CMU-CS-95-210，Carnegie Mellon University，Pittsburgh，PA，USA：1995.

3 Hsu R L，Abdel-Mottaleb M，Jain A K. Face detection in color images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2002，**24**(5)：696～706.

4 Isard M，Blake A. Condensation—Conditional density propagation for visual tracking[J]. International Journal of Computer Vision，1998，**29**(1)：5～28.

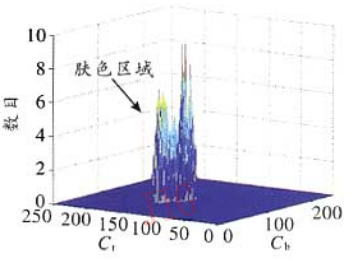
5 Metropolis N，Rosenbluth A W，Rosenbluth M N，*et al*. Equation of state calculations by fast computing machines[J]. The Journal of Chemical Physics，1953，**21**(6)：1087～1092.

6 Guo Hong，Hou Wen-chi，Yan Feng，*et al*. A monte carlo sampling method for drawing representative samples from large databases[A]. In：Proceedings of the 16th International Conference on Scientific and Statistical Database Management[C]，Santorini，Greece，2004：1～2.

7 Gorodnichy D O. IIT-NRC facial video database[DB/OL]. http://synapse.vit.iit.nrc.ca/db/video/faces/cvglab，2005-12-06

8 Gorodnichy D O. Video-based framework for face recognition in video[A]. In：Proceedings of Second Canadian Conference on Computer and Robot Vision（CRV'05）[C]，Victoria，BC，Canada，2005：330～338.

(a) 彩色图像



(b) 图1(a)的C_rC_b2维直方图分布

图1 肤色点C_rC_b分量的2维直方图分布

Fig.1 2D histogram of C_rC_b Component of the pixels

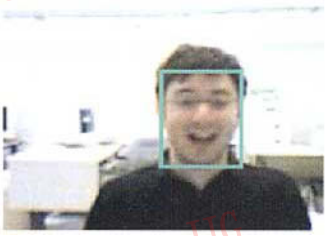
(a) 原图

(b) 肤色检测方法分割结果

(c) 本文方法分割结果

图2 肤色点分割结果

Fig.2 The result of skin color segmentation



(a) 初始帧

(b) 第20帧

(c) 第40帧

图3 IIt-NRC人脸视频数据库的部分跟踪结果

Fig.3 The result of face tracking in IIt-NRC facial video database

图4 来自影片1 中的视频序列图像的部分跟踪结果

Fig.4 The result of face tracking in video 1

图5 来自影片2 中的视频序列图像的部分跟踪结果

Fig.5 The result of face tracking in video 2