

基于颜色和边缘特征自适应融合的人脸跟踪算法

周平平¹, 万洪林¹, 刘慧², 李天平¹

(1. 山东师范大学物理与电子科学学院, 山东济南 250358; 2. 山东财经大学计算机科学与技术学院, 山东济南 250014)

摘要: 针对传统的基于颜色直方图跟踪算法不能精确跟踪的缺陷, 提出了一种基于粒子滤波的自适应融合多特征的人脸跟踪算法. 该方法首先在视频序列中提取人脸的肤色和边缘特征, 并以加权颜色直方图和边缘直方图描述人脸特征; 然后采用自适应融合方法计算粒子集权重, 这种自适应融合方法, 有效地增强了人脸跟踪的可靠性. 实验结果表明, 在视频人脸存在类肤色以及光照变化等复杂背景下, 该方法改善了跟踪效果并且具有较强的鲁棒性.

关键词: 人脸跟踪; 粒子滤波; 加权颜色直方图; 边缘方向直方图; 自适应融合方法

中图分类号: TP391 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2017.10.005

引用格式: 周平平, 万洪林, 刘慧, 等. 基于颜色和边缘特征自适应融合的人脸跟踪算法[J]. 中国科学技术大学学报, 2017, 47(10): 837-842.

ZHOU Pingping, WAN Honglin, LIU Hui, et al. Face tracking algorithm of color and edge features based adaptive fusion[J]. Journal of University of Science and Technology of China, 2017, 47(10): 837-842.

Face tracking algorithm of color and edge features based adaptive fusion

ZHOU Pingping¹, WAN Honglin¹, LIU Hui², LI Tianping¹

(1. School of Physics and Electronics, Shandong Normal University, Jinan 250358, China;

2. Department of Computer Science & Technology, Shandong University of Finance and Economics, Jinan 250014, China)

Abstract: In view of the imprecision of traditional tracking algorithms based on color histogram, a face tracking algorithm combining face multiple features based on adaptive fusion in the basic frame of particle filtering was presented. First, the color and edge features of the human face were extracted in the video sequence, while the weighted color histogram and edge orientation histogram (EOH) described facial features. Then, a self-adaptive features fusion strategy was employed to calculate particle set weight. The reliability of face tracking was enhanced by the self-adaptive features fusion strategy. Experimental results show that in the cases of complex backgrounds such as similar skin color, illumination change and so on, the proposed approach improves the tracking effect and has strong robustness.

Key words: face tracking; particle filtering; weighted color histogram; edge orientation histogram; self-adaptive features fusion

收稿日期: 2017-05-16; 修回日期: 2017-06-22

基金项目: 山东省科技发展计划项目(2014gsf116004)资助.

作者简介: 周平平, 女, 1991年, 硕士生, 研究方向: 视频图像处理. E-mail: 285048768@qq.com

通讯作者: 李天平, 硕士/副教授. E-mail: sdsdltp@163.com

0 引言

近年来,人脸跟踪是计算机视觉的一个重要领域.研究者一直在寻求能够提高视频跟踪结果的实时性、准确性、鲁棒性的方法.视频目标跟踪算法一般分为两类:确定性跟踪算法与随机性跟踪算法.最为常见的确定性跟踪算法属均值漂移(mean-shift)算法,其效率较高,但容易陷入局部最优,从而导致跟踪失败^[1].目前最常用的贝叶斯滤波方法,属于随机性跟踪算法,它通过最优估计后验概率 $p(x_k | y_{1:k})$,进而确定目标的位置.卡尔曼滤波^[2]是用状态方程描述系统动态模型,用观测方程描述系统观测模型,它是一种适用于平稳和非平稳随机过程的线性最小方差估计方法,适用于线性高斯系统.考虑到大多数动态系统都是非线性的,于是 Anderson 等^[3]提出了扩展卡尔曼滤波,将非线性系统模型线性化近似后再进行卡尔曼滤波.Julier^[4]提出了无迹卡尔曼滤波.上述两种方法均改善了跟踪效果,但在原理上它们都是通过高斯函数逼近后验概率.由于人脸是非刚性的,具有非线性、非高斯、多模特性,Isard^[5]首次提出粒子滤波算法(particle filter,PF),它是一种基于非参数化的 Monte Carlo 模拟方法实现递推贝叶斯估计的算法.

文献[6]提出了非均匀的空间直方图的颜色跟踪线索以及一种新的重采样方法,保证了粒子多样性.文献[7-8]将 Adaboost 人脸检测算法与粒子滤波相结合实现人脸跟踪,但是每一帧都要用两种算法,计算复杂度较高.文献[9]在粒子滤波的框架下,针对人脸肤色和纹理两个特征,建立了肤色直方图模型和 Gabor 纹理特征模型,并使用融合策略在视频中实现对人脸区域的准确跟踪.

本文在粒子滤波算法的框架下,利用加权颜色直方图和边缘方向直方图描述人脸特征,并以一种自适应融合方式实现上述两个特征在人脸观测上的使用.

1 粒子滤波

粒子滤波跟踪算法是一种贝叶斯递推估计算法,目的是构造目标状态的后验概率分布 $p(x_t | y_{1:t})$.通常假设动态模型具有一阶马尔可夫状态转移性质,并且观测值关于状态是条件独立的,且目标系统的初始状态概率密度函数 $p(x_0 | y_0) = p(x_0)$ 是假定为已知的,其中目标系统的初始状态向量用 x_0 表示,系统量测值的最初信息用 y_0 表示.那么粒子滤

波步骤分为目标预测和状态更新两个过程.

(I) 预测

根据 Chapman-Komogorov 方程得

$$p(x_t | y_{1:t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} \quad (1)$$

式中, $p(x_{t-1} | y_{1:t-1})$ 为上一时刻目标状态的后验分布, $p(x_t | x_{t-1})$ 为目标先验概率.

(II) 更新

$$p(x_t | y_{1:t}) = \frac{p(y_t | x_t) p(x_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \quad (2)$$

式中, $p(y_t | x_t)$ 为观测似然函数, $p(y_t | y_{1:t-1})$ 为归一化常数.

从式(1)和式(2)可以看出,基于粒子滤波器的跟踪算法需要建立状态空间模型(state space model,SSD),即

$$\left. \begin{aligned} x_t &= f_t(x_{t-1}, \xi_t) \\ y_t &= h_t(x_t, \omega_t) \end{aligned} \right\} \quad (3)$$

式中, $f_t(\cdot)$, $h_t(\cdot)$ 分别为目标状态转移函数和观测函数, ξ_t , ω_t 分别是系统噪声和观测噪声.

粒子滤波是基于序列蒙特卡罗和递归贝叶斯的估计方法,则后验概率分布 $p(x_t | y_{1:t})$ 可近似表述为

$$p(x_t | y_{1:t}) \approx \sum_{i=1}^N w_t^i \delta(x_t - x_t^i) \quad (4)$$

式中, N 为粒子数; w_t 为权值,满足 $\sum_{i=1}^N w_t^i = 1$,且权值更新为

$$w_t^i \propto \frac{p(x_{0:t}^i | y_{1:t}^i)}{q(x_{0:t}^i | y_{1:t}^i)} = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | x_{t-1}^i)}{q(x_t^i | x_{t-1}^i, y_t)} \quad (5)$$

式中, $q(x_t^i | x_{t-1}^i, y_t)$ 为重要性密度函数.

2 人脸特征描述

若实际应用中的重要性密度函数取为 $q(x_t^i | x_{t-1}^i, y_t) = p(x_t^i | x_{t-1}^i)$,则式(5)中的粒子权值满足 $w_t^i \propto w_{t-1}^i p(y_t | x_t^i)$.在粒子滤波跟踪算法中,观测是基于目标特征的.在人脸视频中,由于类肤色、光照变化等干扰因素,会直接影响我们对人脸的观测效果.本文在结合粒子滤波实现人脸跟踪框架下,提出利用加权颜色直方图和边缘方向直方图描述人脸特征.

2.1 颜色特征

颜色直方图^[10]模型因对同一目标的尺寸变化、旋转、部分遮掩具有不敏感性而被广泛使用,但在跟

踪过程中,这一特征易受光照及背景变化的影响,人脸肤色的差异主要体现在亮度上,而皮肤的色调是均匀一致的,因此本文在 HSV 颜色空间中,只计算 H-S 直方图,一定程度上减轻了光照变化的影响,加权颜色直方图不仅仅计算颜色特征的频率,而且解决了直方图缺乏空间特性的问题.由于 Epanechnikov 核函数权值减少比较缓慢,本文利用该核函数突出不同位置的像素对颜色直方图的不同贡献.

$$K_E(x) = \begin{cases} c(1 - \|x\|^2), & \|x\| < 1 \\ 0, & \|x\| \geq 1 \end{cases} \quad (6)$$

式(6)表明,在目标中心区域的像素对颜色统计贡献最大,在目标边缘处像素对颜色统计贡献最小.

人脸区域的颜色加权分布模型^[11]为

$$p_c(n) = C \sum_{i=1}^N K_E\left(\frac{x_i - x_0}{a}\right) \delta[h(x_i) - n] \quad (7)$$

式中, $n \in [1, M]$, 观测区域内共有 N 个像素, a 是核带宽, x_i 是观测区域内像素的位置, x_0 是观测区中心位置, $\delta(\cdot)$ 是 Kronecker 函数, $h(x_i)$ 是 x_i 位置上像素在颜色加权直方图中的 bin, M 是颜色加权直方图总 bin 数. C 是归一化系数,且满足

$$\sum_{n=1}^m p_c(n) = 1 \text{ 和 } C = \frac{1}{\sum_{i=1}^N K_E\left(\frac{x_i - x_0}{a}\right)}.$$

2.2 边缘特征

边缘特征是人脸的另一关键特征,且对光照变化不敏感,本文使用边缘方向直方图 (edge orientation histogram)^[12]来描述目标的边缘特征,能够有效应对光照变化和背景颜色变化的混淆.本文采用 Sobel 算子进行边缘检测,首先将视频帧转换为灰度图像,再计算每个像素 $I(x, y)$ 的梯度幅值 G 与方向 τ ,即

$$\left. \begin{aligned} G &= \sqrt{G_x^2 + G_y^2} \\ \tau &= \arctan\left(\frac{G_y}{G_x}\right) \end{aligned} \right\} \quad (8)$$

式中, G_x, G_y 分别为图像的水平核与垂直核, $0 \leq \tau \leq \pi$. 统计对应角度的像素梯度幅值,最后进行归一化处理,得到人脸的边缘方向直方图 $p_e(n)$.

2.3 相似性度量

颜色和边缘的相似性度量均选择常用的 Bhattacharyya 系数^[13],表达式为

$$\rho[p, q] = \sum_{n=1}^M \sqrt{p(n)q(n)} \quad (9)$$

Bhattacharyya 距离为

$$d = \sqrt{1 - \rho[p, q]} \quad (10)$$

这里, $p(n)$ 和 $q(n)$ 分别表示候选区直方图和模板直方图, ρ 的变化区间在 $[0, 1]$ 内, ρ 越大,两个直方图就越接近;当 $d = 0$ 时,则意味着完全匹配.

3 基于颜色和边缘特征的人脸跟踪算法

3.1 状态转移模型

本文动态模型基于自回归模型 (AR), 假设一阶运动系统描述为

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\xi_t \quad (11)$$

式中, \mathbf{A} 为确定性漂移系数矩阵; \mathbf{B} 为随机扩散系数矩阵, 且 $\mathbf{A} = \begin{bmatrix} 1 + \Delta T & 0 \\ 0 & 1 + \Delta T \end{bmatrix}$ 和 $\mathbf{B} =$

$$\begin{bmatrix} \frac{\Delta T^2}{2} & \frac{\Delta T^2}{2} \end{bmatrix}, \text{ 其中 } \Delta T \text{ 为采样周期.}$$

3.2 观测模型

目标区域颜色和边缘特征单独观测结果如下:

$$p(y_c | x) = \frac{1}{\sqrt{2\pi\sigma_c}} \exp\left(-\frac{1 - \rho_c}{2\sigma_c^2}\right) \quad (12)$$

$$p(y_e | x) = \frac{1}{\sqrt{2\pi\sigma_e}} \exp\left(-\frac{1 - \rho_e}{2\sigma_e^2}\right) \quad (13)$$

式(12)和(13)中, σ_c 和 σ_e 为高斯方差, ρ_c 和 ρ_e 为 Bhattacharyya 系数. 当采用这两个特征进行联合观测时, 所得的观测结果为

$$p(y | x) = \theta_c p(y_c | x) + \theta_e p(y_e | x) \quad (14)$$

式中, θ_c 和 θ_e 分别代表基于颜色特征权重和基于边缘特征权重, 特征权重值的大小表明某特征的似然观测效果好坏, 且满足 $\theta_c + \theta_e = 1$.

粒子权重更新为

$$\begin{aligned} w_t^i &\propto w_{t-1}^i p(y_t | x_t^i) = \\ &w_{t-1}^i [\theta_c p(y_c | x) + \theta_e p(y_e | x)] \end{aligned} \quad (15)$$

3.3 多特征自适应融合

在实际视频跟踪中, 同一目标不同帧的特征权重值并不一定是相等的, 因此有必要提出一种能够具有动态自适应能力的似然观测方式达到较好的观测效果, 本文采用了多特征自适应融合方法计算粒子集权重^[14].

设预测后得到的粒子集为 $\{x^i, \omega^i\}$, 共有 M 个特征进行联合. 对于第 $s (s \in [1, M])$ 个特征, 其似然观测结果为 p_s^i .

$$\delta_s = \frac{1}{N} \sum_{i=1}^N |p_s^i - \overline{p_s}| \quad (16)$$

式中, $\overline{p_s} = \frac{1}{N} \sum_{i=1}^N p_s^i$, δ_s 表达第 S 个特征似然度值

形成观测峰的情况, δ_s 值越大越好.

第 S 个特征的权重为

$$\theta_s = \frac{\delta_s}{|x_s^{\text{peak}} - \bar{x}^{\text{peak}}|} \quad (17)$$

式中, $\bar{x}^{\text{peak}} = \frac{1}{M} \sum_{s=1}^M x_s^{\text{peak}}$, x_s^{peak} 是第 S 个特征的似然观测峰值的空间位置. 似然观测峰与目标真实状态越接近, 分母的值越小.

3.4 算法步骤

在粒子滤波的框架下, 本文算法步骤如下:

步骤 1 初始化, $t=0$. 用鼠标在第一帧视频图像中框选出人脸区域, 分别计算人脸模板颜色直方图、边缘直方图 p_c^0, p_e^0 , 粒子初始状态为 $\{x_0^i, \omega_0^i\}_{i=1}^N$.

步骤 2 由式(11)预测人脸当前状态 x_t .

步骤 3 权值更新:

1) 由式(12)计算每个粒子状态对应的颜色观测似然函数值 $p(y_c | x)$;

2) 由式(13)计算每个粒子状态对应的边缘观测似然函数值 $p(y_e | x)$;

3) 分别将 $p(y_c | x)$ 和 $p(y_e | x)$ 的值代入式(16)得到 δ_c 和 δ_e , 再代入式(17)求出 θ_c 和 θ_e , 并由式(14)计算每个粒子状态对应的总的观测似然函数值 $p(y | x)$;

4) 由式(15)计算权重 ω_t^i , 并归一化权重

$$\tilde{\omega}_t^i = \frac{\omega_t^i}{\sum_{i=1}^N \omega_t^i}$$

步骤 4 输出 t 时刻人脸状态:

$$E(x_t) = \sum_{i=1}^N \tilde{\omega}_t^i x_t^i$$

步骤 5 对粒子重采样, 得到 100 个新的粒子的集合 $\{\tilde{x}_t^i | i=1, 2, \dots, N\}$.

步骤 6 判断跟踪是否结束. 若未结束转步骤 2.

4 实验结果与分析

为了验证本文方法的跟踪效果, 在不同视频序列上将文献[11]的跟踪结果与本文算法进行比较. 实验采用手动初始化人脸模板, 粒子数 $N=100$. 本实验平台选择基于 opencv2.4.8 的 Visual Studio 2010 软件环境.

本文使用欧式距离计算人脸跟踪结果的误差, 欧式距离的计算公式表示为

$$d = \sqrt{(x_t - x_0)^2 + (y_t - y_0)^2} \quad (18)$$

式中, (x_t, y_t) 为 t 时刻目标位置估计值, (x_0, y_0) 为 t 时刻目标真实位置.

实验一是 David 视频(图像大小为 480×360 , 帧长为 119). 图 1(a) 是文献[11]跟踪的部分结果, 从误差曲线(图 2)可以明显看出, 该跟踪算法对人脸表情、姿态变化不敏感, 但容易受到肤色背景干扰, 如第 82 帧人手区域属于类肤色区域, 对脸部跟踪产生干扰, 跟踪结果偏离了真实人脸位置. 图 1(b) 是本文算法的跟踪结果比较理想, 解决了类肤色的影响.



(a) 文献[11]的结果 (b) 本文算法

图 1 类肤色背景下的跟踪(由上至下分别为第 6 帧、第 38 帧、第 82 帧、第 117 帧)

Fig.1 Tracking under skin like color background

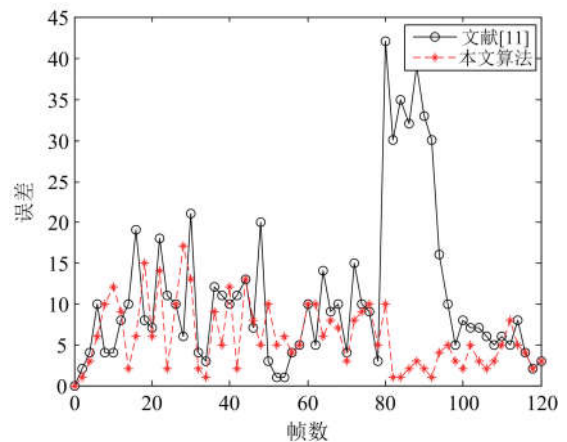


图 2 David 跟踪误差

Fig.2 David tracking error

实验二是 Trellis 视频(图像大小为 480×360 ,

帧长为 182)。图 3 是在光照变化下的实验结果,图 3 (a)在跟踪过程中,对于光照变化不明显的情况,跟踪效果比较理想;当光照剧烈变化时,文献[11]的跟踪结果偏离了人脸位置.而图 3(b)能自适应融合人脸肤色与边缘特征的跟踪算法,这得益于边缘特征不受光照影响,能够准确跟随人脸的运动.图 4 对比了文献[11]和本文算法跟踪结果的误差,结果显示自适应融合两种特征能充分发挥各自的优势,减少误差,从而获取更准确、更具鲁棒性的跟踪结果.

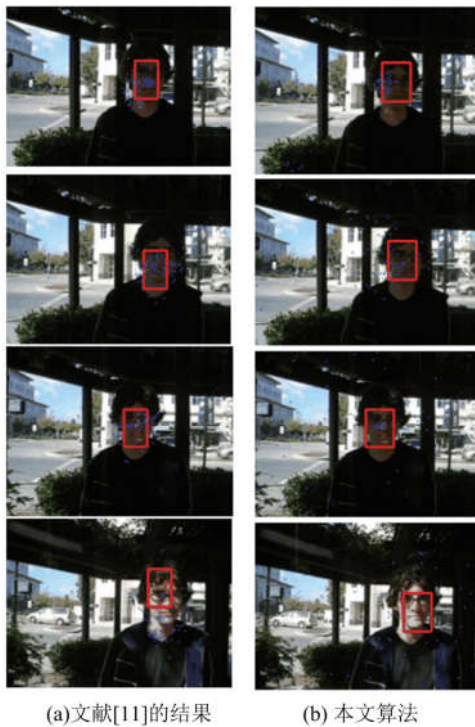


图 3 光照变化下的跟踪(由上至下分别为第 24 帧、第 36 帧、第 52 帧、第 138 帧)

Fig.3 Tracking under illumination change

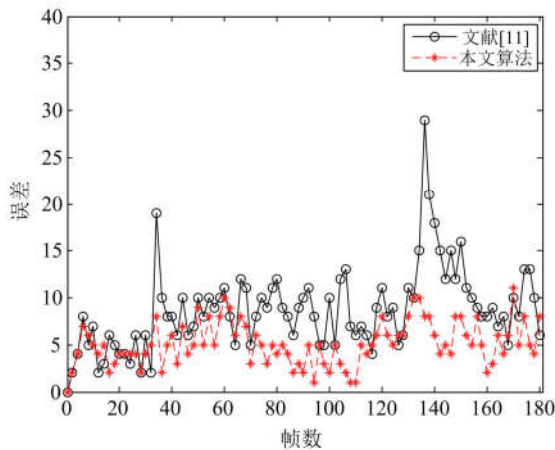


图 4 Trellis 跟踪误差

Fig.4 Trellis tracking error

实验三是 David2 视频(图像大小为 480×360 , 帧长为 45)。从图 5、6 明显看出,文献[11]的跟踪结果不理想,这是因为背景颜色与人脸肤色相近,从第 11 帧开始跟踪结果均偏离了正确人脸位置.本文跟踪算法很好地解决了背景与人脸肤色相似带来的干扰.



(a) 文献[11]的结果 (b) 本文算法

图 5 类似背景下的跟踪(由上至下分别为第 6 帧、第 11 帧、第 21 帧、第 44 帧)

Fig.5 Tracking under similar background

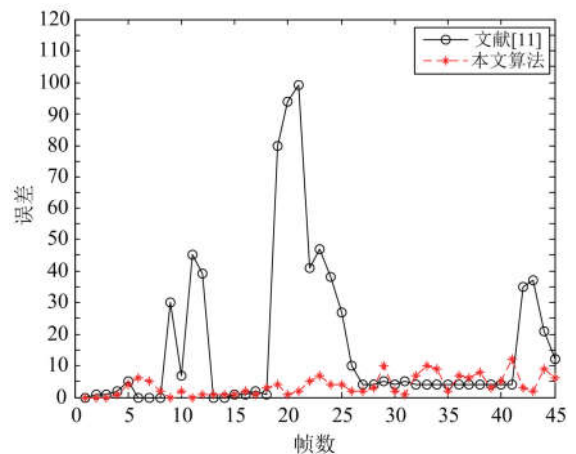


图 6 David2 跟踪误差

Fig.6 David2 tracking error

5 结论

本文针对人脸非线性和非高斯的特点,采用了

粒子滤波算法.为了克服相似背景和光照变化的干扰,本文添加了边缘方向直方图作为跟踪的另一有效线索,并结合加权颜色直方图描述人脸肤色特征,同时以一种自适应融合方式实现人脸跟踪,实验充分验证其较好的跟踪性能,在背景与人脸相似的情况下也能较好跟踪.

参考文献(References)

- [1] 王新红,王晶,田敏,等.基于空间边缘方向直方图的 Mean Shift 跟踪算法[J].中国图象图形学报,2008, 13(3): 586-592.
WANG Xinhong WANG Jing, TIAN Min, et al. Mean shift tracking algorithm based on spatial edge orientation histograms [J]. Journal of Image and Graphics, 2008, 13(3): 586-592.
- [2] WELCH G, BISHOP G. An Introduction to the Kalman Filter[M]// University of North Carolina at Chapel Hill, 2001, 8(7): 127-132.
- [3] ANDERSON B D, MOORE J B. Optimal Filtering [M]. New Jersey: Prentice-Hall, 1979.
- [4] JULIER S J, UHLMANN J K A. New extension of the Kalman filter to non linear systems [C]// Proceedings of Aero Sense; The 11th International Symposium on Aero space/Defense Sensing, Simulation and Controls. SPIE, 1997: 182-193.
- [5] ISARD M, BLAKE A. Condensation-conditional density propagation for visual tracking [J]. International Journal of Computer Vision, 1998, 29(1): 5-28.
- [6] YANG D L, ZHANG Y L, JI R D, et al. An improved spatial histogram and particle filter face tracking[C]// Proceedings of the 8th International Conference on Genetic and Evolutionary Computing. Nanchang, China: Springer, 2014:257-267.
- [7] DOU J F, LI J X, ZHANG Z, et al. Face tracking with an adaptive Adaboost-based particle filter [C]// 24th Chinese Control and Decision Conference. Taiyuan, China: IEEE, 2012, 23(1): 3626-3631.
- [8] CHEN W M, LIN Y L, HSIEH Y H. An adaptive particle filter based method for real time face tracking [J]. Journal of Software Engineering and Applications, 2016, 6(5): 1-5.
- [9] 田天,陈刚.基于肤色和 Gabor 纹理的粒子滤波人脸跟踪[J].计算机工程,2014, 40(7): 123-127.
TIAN Tian, CHEN Gang. Face tracking using particle filtering based on skin color and Gabor texture [J]. Computer Engineering, 2014, 40(7): 123-127.
- [10] SWAIN M J, BALLARD D H. Color indexing [J]. International Journal of Computer Vision, 1991, 7(1): 11-32.
- [11] VALISKA J, MARCHEVSKY S, KOKOSKA R. Object tracking by color-based particle filter techniques in video sequences[C]// 24th International Conference Radioelektronika. Bratislava, Slovakia: IEEE, 2014: 1-4.
- [12] ZHU W, LEVINSON S. Edge orientation-based multi-view object recognition[J]. International Conference on Pattern Recognition, 2000, 1(1): 1936-1939.
- [13] KAILATH T. The divergence and Bhattacharyya distance measures in signal selection [J]. IEEE Transaction on Communication Technology, 1967, 15(1): 52-60.
- [14] ZHOU S K, MOGHADDAM B. Appearance tracking using adaptive models in a particle filter [R]. Proceedings of Asian Conference Computer Vision, 2004.