

基于 mean-shift 算法的人脸实时跟踪方法^{*}

蒋建国, 孙洪艳, 齐美彬

(合肥工业大学 计算机与信息学院, 合肥 230009)

摘 要: 复杂环境下实时人脸跟踪方法在视觉监控系统中具有很重要的意义, 但目前的跟踪算法普遍存在目标遮挡、尺寸变化等过于敏感的不足, 限制了其应用范围。提出了一种人脸检测、mean-shift 算法与卡尔曼滤波器相结合的实时全自动人脸跟踪算法。实验结果表明该算法实时性很强, 可以实现对运动人脸的快速跟踪, 同时对目标遮挡也有很好的鲁棒性。

关键词: 人脸跟踪; mean-shift 算法; 卡尔曼滤波器; 目标遮挡

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2008)07-2225-03

Real-time algorithm for face tracking based on mean-shift

JIANG Jian-guo, SUN Hong-yan, QI Mei-bin

(School of Computer & Information, Hefei University of Technology, Hefei 230009, China)

Abstract: A real-time face tracking under complex environment has significance in visual surveillance system, but one common problem to face tracking algorithms at present is their sensitivity to temporal occlusion and scale variations, which restrict them wide application. This paper proposed a real-time automatic face tracking algorithm which combined face detection and mean-shift algorithm with Kalman filter. In the experiments, the results show the proposed algorithm can track fast moving face successfully and have better robustness of occlusion.

Key words: face tracking; mean-shift algorithm; Kalman filter; temporal occlusion

在复杂环境下对目标进行实时有效的跟踪是计算机视觉中的关键技术, 它涉及到图像处理、模式识别、人工智能以及自适应控制等多领域问题的综合。人脸是人类在社会活动中相互辨别的首要依据, 提供了大量的视觉信息, 如人脸识别、表情识别等。近年来, 随着可视电话、视频会议、安全监控、人机交互等技术的发展, 迫切需要对环境中的人员进行跟踪, 所以获得清晰的、完备的人脸图像是视觉监控的一个重要任务, 根据人脸特征实现对特定人的连续跟踪已经成为新的研究热点。在过去的几年中, 已经有很多学者提出了不同的人脸跟踪算法。Mean-shift 算法是目前较为流行的无参数模式快速匹配算法, 此算法具有计算量小, 实时性好, 采用核函数直方图建模, 对边缘遮挡和背景运动均不敏感等优点^[1]。

本文通过基于肤色模型的快速人脸检测确定初始搜索窗口的位置与大小, 改善了 mean-shift 半自动跟踪的缺陷, 增加相似性测度相对阈值来更新模板, 解决尺度变化和人脸转动在人脸跟踪中的影响。但是在复杂背景中, 由于该算法只利用了颜色信息, 对运动物体不作任何预测, 抗干扰能力比较差, 容易导致跟踪失败, 需引入估计器预测运动参数。Kalman 滤波是一个在最小协方差准则下的最优估计方法, 能利用实际的运动参数不断修正未来运动状态的估计值, 提高估计的精度, 兼顾实时性和稳健性。本文结合 mean-shift 算法与 Kalman 滤波, 利用 Kalman 滤波预测 mean-shift 搜索窗口中心位置, 实验验证了此算法的有效性。

1 跟踪算法描述

Mean-shift 算法存在半自动跟踪的缺陷, 在起始跟踪帧, 需要通过手动确定搜索窗口来选择目标。本文通过基于肤色模型的人脸检测算法来确定起始帧搜索窗口的位置与大小实现自动跟踪。首先将起始帧图像转换成 HSV 图像, 再通过 H 分量阈值肤色区域分割, 建立肤色模型提取出人脸, 自动确定初始跟踪窗口的位置和大小, 然后利用 mean-shift 迭代计算跟踪窗口的新位置, 并将窗口的中心位置作为 Kalman 滤波的测量值, 校正运动目标位置, Kalman 滤波器的预测结果用于设置下一帧图像搜索窗口的中心位置。

1.1 人脸检测定位

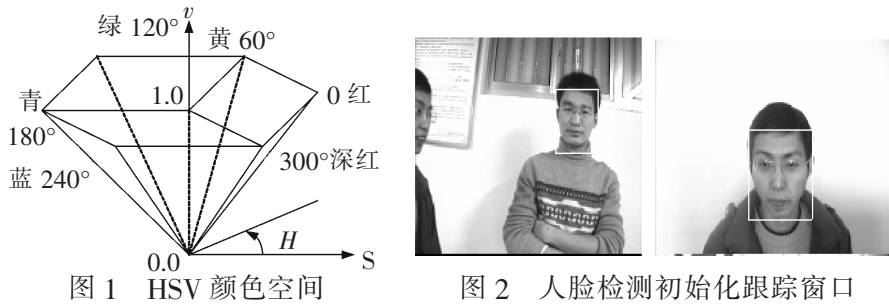
人脸检测是人脸跟踪的前提。肤色是人脸恒定特征的重要信息, 不依赖于面部的细节特征, 对于旋转、表情等变化情况均适用, 具有相对的稳定性, 因此肤色特征建模在人脸检测中是最常用的一种方法^[2]。在通常的光照条件下, 人脸肤色会聚集在色彩空间中某个特定的区域内, 通过对肤色采样寻找肤色分布的合适阈值, 就能够将肤色区域提取出来。

HSV 颜色模型就是一种与人眼的视觉特征相匹配的颜色模型。其中, H 代表色调, 用角度表示, 其变化范围为 $0 \sim 360^\circ$; S 表示色饱和度, 变化范围为 $0 \sim 1$; V 为亮度, 变化范围为 $0 \sim 1$ 。它通过对 RGB 颜色空间进行变换, 使信息结构更加紧凑各个分量的独立性增加, 颜色信息丢失减少。图 1 即为 RGB 与 HSV 颜色空间变换关系。

收稿日期: 2007-07-19; 修回日期: 2007-10-24 基金项目: 国家科技部创新基金资助项目(04c26213401216)

作者简介: 蒋建国(1955-), 男, 安徽黄山人, 教授, 博导, 主要研究方向为信号与信息处理、传感与智能控制; 孙洪艳(1980-), 女, 硕士研究生, 主要研究方向为数字图像处理、DSP 技术及应用(sunhongyan325@163.com); 齐美彬(1969-), 男, 副教授, 博士研究生, 主要研究方向为分布式人工智能、数字图像处理、DSP 技术及应用。

通常情况下,人脸肤色区域 H 值变化范围比较集中。因此只要选择合适的 H 阈值就可以迅速地检测人脸区域。根据大量肤色 H 直方图统计得出经验值: H 值分布在 $0.015 \sim 0.136$ 的绝大部分肤色图像代表了通常情况下的人体肤色。考虑实际情况,并使选择的阈值具有一定的抗噪能力,选取用作肤色分割的 H 阈值为 $0.03 \sim 0.128$ 。首先将起始帧图像色彩空间的值转换到 HSV 空间,然后提取其中的 H 分量做一维直方图,根据经验阈值进行肤色区域分割。提取出人脸区域,确定了初始目标位置和搜索窗口的初始大小,完成了目标选择,图 2 即为自拍的 QCIF 实验序列通过人脸检测初始化跟踪窗口的结果。



在这里需要考虑 H 分量的取值范围的问题, H 分量的取值范围是 $[0, 360]$, 为了能用一个 Byte 表示这个取值范围, 需要将 H 值作适当的量化处理, 这里将 H 分量的范围归一化到 $[0, 255]$ 。

1.2 Mean-shift 算法及改进

Mean-shift 的核心思想是: 在起始跟踪帧计算核函数加权下的搜索窗口的直方图分布, 用同样的方法计算当前帧候选目标的直方图分布, 用 Bhattacharyya 系数来描述两个分布的相似程度, 以两个分布的相似性最大为原则, 通过 mean-shift 向量的迭代运算, 使搜索窗口沿密度增加最大的方向运动, 收敛于最佳位置^[3~5]。

假设通过人脸检测获得的跟踪窗口中心位于 x_0 , 用直方图分布来描述目标模板则可以表示为

q_u = C \sum_{i=1}^m k((x_i - x_0)/h)^2 [b(x_i) - u]

其中: x_i 是第 i 个像素的坐标; $k(x)$ 是核函数; h 是核函数的带宽; 函数 b 是像素 x_i 在直方图中的颜色索引; C 是一个标准化常量系数, 使得所有特征值的概率和为 1。

核函数 $k(x)$ 是一个凸的单调下降函数, 作用是给目标模板区域的像素设置权值。由于跟踪目标受到背景和一些遮挡的影响, 周边的像素可靠性低, 离中心越远的像素分配的权值越小, 此处理可以提高估计值的鲁棒性, 提高搜索跟踪能力。核函数的选择可以参考文献[3]。

候选目标在当前帧中以 y 为中心, 选用相同的核函数和核半径, 则候选目标模板可以描述为

q_u(y) = C \sum_{i=1}^m k((x_i - y)/h)^2 [b(x_i) - u]

其中: x_i , $k(x)$, h , b , C 的含义与目标模板相同。

通常情况下用 Bhattacharyya^[6] 系数 $p(y)$ 度量候选目标分布 $q_u(y)$ 与目标分布 p_u 的相似性: $(y) [p(y), q] = \frac{1}{\sqrt{p_u(y) q_u}}$

因此物体跟踪可以简化为通过 mean-shift 向量反复迭代, 目标由当前帧 y_0 位置不断向 y_1 移动, 寻找使得候选目标特征 $p_u(y)$ 与目标特征 q_u 最相似的位置。

y_1 = \sum_{i=1}^m x_i w_i g((y_0 - x_i)/h)^2 / \sum_{i=1}^m w_i g((y_0 - x_i)/h)^2

其中: $g(x) = -k'(x)$ 是核函数的导数, $w_i = \frac{1}{\sum_{i=1}^m q_u p_u(y_0) (b(x_i) - u)}$ 是像素权值。

传统的 mean-shift 算法目标模板统计始终选用起始帧来计算目标直方图分布 q_u , 没有对目标模板进行实时更新, 一段时间间隔过后, 候选目标模板 $p_u(y)$ 与初始目标模板 q_n 之间就会产生偏差, 容易导致目标跟踪失败。同时, 由于核函数半径 h 始终不变, 无法自适应跟踪渐变的目标。本文通过增加相似性测度 (y) 相对变化量 方法来更新模板, 解决目标尺度变化的问题。整个跟踪过程中, 当 Bhattacharyya 系数的相对变化量 超过一定阈值, 说明当前候选目标与目标模板已有很大的偏差, 这时重新启动人脸检测, 根据新的跟踪窗口的大小与位置, 对初始目标分布 q_n 进行更新, 以提高跟踪的准确性, 同时更新的跟踪窗口也可以更好地适应跟踪尺寸变化的目标。如图 3 就是通过判断相似性测度相对改变量 来更新目标模板。

1.3 Kalman 滤波

在复杂背景中当运动目标运动过快或出现遮挡, 由于只利用颜色信息, 对运动物体不作任何预测, 抗干扰能力比较差, 容易导致跟踪失败, 引入二维 Kalman 滤波器预测窗口中心来解决这个问题。本文将第 k 帧 mean-shift 算法迭代的中心位置作为 Kalman 滤波器的观测值, 首先用 Kalman 滤波器预测 $k+1$ 帧目标窗口的中心位置, 作为 mean-shift 算法中搜索窗口的起始位置, mean-shift 会在这个位置的邻域内找到目标最优的位置。再以这个窗口中心位置作为 Kalman 滤波器的观测值, 进行下一帧的预测运算。

设系统状态方程: $X(K) = AX(k-1) + W(k)$

观测方程: $Z(k) = HX(k) + V(k)$

其中: $X(k)$ 是 k 时刻的系统状态; A 是系统状态转移矩阵; $W(k)$ 表示过程噪声, 假定其服从零均值的高斯分布, 且具有协方差 Q ; $Z(k)$ 是 k 时刻的测量值, H 为观测矩阵; $V(k)$ 表示测量噪声。假定其服从零均值的高斯分布, 且具有协方差 R (这里假设 Q, R 不随系统状态变化而变化)^[7]。

跟踪过程中由于两帧图像时间间隔较短, 运动目标状态变化较小, 所以采用匀速运动模型用 Kalman 滤波器预测窗口中心位置。定义状态向量 $X(k) = \{x(k), v_x(k), y(k), v_y(k)\}^T$ 。其中, $x(k)$, $v_x(k)$, $y(k)$, $v_y(k)$ 分别表示目标中心 x, y 轴上的位置和速度。根据系统状态方程和观测方程得到状态转移矩阵 A 和观测矩阵 H 为:

A = [1 T 0 0; 0 1 0 0; 0 0 1 T; 0 0 0 1]; H = [1 0 0 0; 0 0 1 0]

其中: T 表示两帧之间的时间间隔。

Kalman 预测迭代过程如下:

- a) $G(k) = AP(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1}$
- b) $X(k+1|k) = AX(k|k-1) + G(k)(Z(k) - HX(k|k-1))$
- c) $P(k) = (I - A^{-1}G(k)H)P(k|k-1)$
- d) $P(k+1|k) = AP(k)A^T + Q$

其中: $P(k)$, $Q(k)$ 分别表示系统协方差矩阵、Kalman 增益。

Kalman 迭代初始条件为 $X(1|0)$, $P(1|0)$ 如下:

a) $X(1|0) = [x(2) [x(2) - x(1)]/T \ y(2) [y(2) - y(1)]/T]^T$

b) $P(1|0) = AP(0)A^T + Q$

2 跟踪算法步骤

- a) 将整个图像设为搜索区域, 采用人脸检测自动确定跟踪目标, 初始化搜索窗口的大小和位置。
- b) 计算搜索窗口内目标特征的概率分布, 运行 mean-shift 算法迭代获得最佳搜索窗口新的位置。在这个最佳位置上通过判断相似性测度相对改变量是否超出阈值决定是否重新启动人脸检测对模板进行更新。
- c) 将搜索窗口新位置作为观测值传递给 Kalman 滤波器, 运行 Kalman 滤波器预测下一帧目标搜索窗口的位置。
- d) 在下一帧视频图像中, 用 c) 获得的位置初始化搜索窗口的大小和位置, 跳转到 b) 继续运行。

3 实验结果分析

对自拍的 QCIF 测试序列进行实验, 图 4 是跟踪过程中存在目标局部遮挡或完全遮挡的两组视频序列跟踪结果。



图 3 人脸跟踪过程中模板更新图 4 部分遮挡的人脸跟踪

第一帧用人脸检测确定跟踪窗口起始位置, 运行 mean-shift 搜索候选目标, 通过 Kalman 预测降低搜索范围, 实验表明在发生目标跟踪区域较大部分被遮挡的情况下, 利用 mean-shift 算法与 Kalman 滤波器相结合, 通过预测仍然对人脸运动

进行了正确的跟踪, 增强了跟踪的鲁棒性。证明本文提出的算法是成功的。

4 结束语

通过人脸检测实现运动目标全自动跟踪, 以 mean-shift 算法输出的窗口中心位置为测量信号, 采用 Kalman 滤波算法校正窗口中心位置, 克服了 mean-shift 算法的缺陷。实验表明, 利用 mean-shift 与 Kalman 滤波结合的方式可以实时准确地跟踪运动目标, 同时可以有效地解决部分遮挡问题。此系统便于移植到 DSP, 结合控制云台, 易于扩展为全自动动态人脸跟踪系统。这也是本文下一步研究的重点。

参考文献:

[1] ORIN C, PETER M. Robust analysis of feature spaces: color image segmentation[C] //Proc of IEEE Conf on CVPR. Puerto Rico: [s. n.], 1997: 750-755.

[2] FIEGUTH P, TERZOPOULOS D. Color-based tracking of heads and other mobile objects at video frame rates [C] //Proc of IEEE Conference on CVPR. Puerto Rico: [s. n.] 1997: 21-27.

[3] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift [C] //Proc of IEEE Conference on CVPR. Hilton Head Island: South Carolina, 2000: 142-149.

[4] COMANICIU D, RAMESH V. Mean shift and optimal prediction for efficient object tracking[C] //Proc of IEEE International Conference on Image Processing. Vancouver: [s. n.], 2000: 70-73.

[5] CHENG Y. Mean-shift, mode seeking and clustering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1995, 17 (8) : 790-799.

[6] DJOUADI A, SNORRASON O, GARBER F D. The quality of training-sample estimates of the Bhattacharyya coefficient[J]. IEEE Trans on Pattern Analysis and Machine Intelligent, 1990, 12: 92-97.

[7] KALMAN R E. A new approach to linear filtering and prediction problems [J]. Trans of the ASME-Journal of Basic Engineering, 1960: 35-45.

(上接第 2224 页) 不能够准确地追踪到眼睛的外形, 以及虹膜、瞳孔等位置。这也是下一步研究的重点。部分实验图片如图 3 所示。

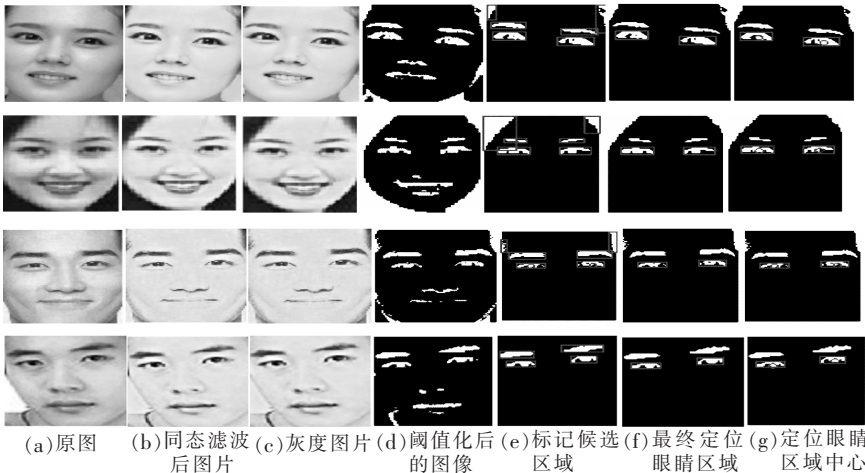


图 3 部分实验图片

5 结束语

本文所提算法在 MATLAB 7.0 下演示成功, 提出了一种基于最小化外接矩形的自定义眼睛模板实施人眼的检测, 利用了同态滤波算法来补偿亮度信息的影响, 同时利用了迭代式阈值选择算法对图像进行二值化处理, 增强了算法的鲁棒性以及通用性。利用质心原理标定出了眼睛的中心位置, 为下一步研究打下了良好的基础。实验表明在脸部倾斜不是很大的图片中,

检测精度达到 95% 以上。但是本文还有明显的不足, 对于脸部倾斜较大的图像容易产生误检, 另外对于配戴眼镜的人脸图像, 由于眼镜边缘以及光亮的影响造成检测率不高, 而且还无法准确跟踪到眼睛的轮廓, 这是下一步研究解决的重点。最后, 希望本文能够对相关研究提供一定的参考和帮助。

参考文献:

[1] 张敏, 陶亮. 人脸图像中人眼的检测与定位 [J]. 光电工程, 2006, 33(8) : 33-36.

[2] TIAN Ying-li, KANADE T, JEFFREY F C. Eye-state action unit detection by Gabor wavelets[C] // Proc of ICMI. 2000: 143-150.

[3] KUMAR R T, RAJA S K, RAMAKRISHNAN A G. Eye detection using color cues and projection functions[C] // Proc of IEEE ICIP . 2002: 337-340.

[4] YUILLE A L, HALLINAN P W, COHEN D S. Feature extraction from faces using deformable template [J]. International Journal of Computer Vision, 1992, 8(2) : 99-111.

[5] XIE X, SUDHAKAR R, ZHUANG H. Corner detection by a cost minimization approach [J]. Pattern Recognition, 1993, 26 (8) : 1235-1243.

[6] 黄万军, 尹宝才, 陈通波, 等. 基于三维可变形模板的眼睛特征提取 [J]. 计算机研究与发展, 2002, 39(4) : 495-501.