

文章编号: 1004-4353(2017)02-0144-06

基于改进 Camshift 的人脸跟踪算法

骆绍烨¹, 刘丽桑²

(1. 莆田学院 信息工程学院, 福建 莆田 351100; 2. 福建工程学院 信息科学与工程学院, 福建 福州 350118)

摘要: 研究了一种基于改进 Camshift 的人脸跟踪算法. 该算法采用分块加权的直方图匹配方法增强人脸的辨识度, 通过人脸形态约束筛选不合理的人脸形态变化, 并结合 Kalman 滤波器预测修正跟踪人脸位置. 实验表明, 改进后的算法比经典 Camshift 算法有更强的抗肤色干扰能力和跟踪准确性.

关键词: 人脸跟踪; Camshift; 分块颜色直方图; Kalman 滤波

中图分类号: TP391.41 **文献标识码:** A

Face tracking algorithm based on improved Camshift

LUO Shaoye¹, LIU Lisang²

(1. College of Information Engineering, Putian University, Putian 351100, China;
2. College of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China)

Abstract: The improved Camshift algorithm for face tracking is proposed, which enhances the identification of facial structure by using the block weighted color histogram in template matching, filtrates the unreasonable facial morphological change according to the face shape detection, predicts and amends the accuracy of face tracking through Kalman filter. Experimental results prove the improved algorithm become stronger interference ability, higher tracking accuracy than the classical Camshift algorithm.

Keywords: face tracking; Camshift; block color histogram; Kalman filter

人脸跟踪是在已知人脸窗口的基础上,通过一定的方法在视频图像中标示出该人脸的位置和大小变化的过程. 1993 年, A. Azarbayejani 等^[1]首次提出人脸跟踪问题的解决方案之后, 许多学者针对人脸跟踪问题进行了进一步的研究, 例如: A. Eleftheriadis 等通过人脸头部的轮廓形状特征和图像的帧差实现人脸跟踪^[2], K. Fukunaga 等将均值偏移(MeanShift)算法^[3]应用于人脸跟踪问题, 梁路宏等利用人脸检测来关联视频序列上的人脸运动关系^[4], 等等. 在这些后续研究中, MeanShift 算法因算法简单和效果良好而受到关注. 1998 年, G. R. Bradski 对 MeanShift 算法进行了改进, 提出了连续自适应的均值偏移(Continuously Adaptive Mean Shift)算法^[5], 即 Camshift 算法, 该算法因具有较好的抗遮挡性、鲁棒性和实时性而成为当前使用最广泛的人脸跟踪算法之一.

1 Camshift 算法

Camshift 算法是一种高效的非参数估计算法^[6], 该算法通过视频图像中运动物体颜色信息的计算来达到跟踪的目的. 因此, Camshift 算法既无需计算整个区间的概率密度, 也不用事先假定区间中的样本分布模型, 完全依靠训练数据(特征空间中足够多的样本点)进行估计. Camshift 算法主要包含以下

步骤:

1) 计算被跟踪目标的色彩直方图. 首先给定一个区域作为被跟踪人脸的目标区域, 为了减少光照等外界因素对人脸肤色的影响, 一般将图像原有的 RGB 空间模型转换到具有颜色直观特性的 HSV 颜色空间模型中, 并对其中的 H 分量构造人脸颜色概率查找表. 假设目标区域 B 有 N 个像素点 $\{x_i\}$ ($i = 1, 2, \dots, N$), 式中的 $x_i = (X_i, Y_i)$ 为像素坐标, 则目标的直方图模型为

$$q_u = C_h \sum_{i=1}^N k \left(\left\| \frac{x_i}{h} \right\|^2 \right) \delta [b(x_i) - u], \quad (1)$$

其中归一化系数 C_h 是用以保证 $\sum_{u=1}^m q_u = 1$, δ 为 Kronecker delta 函数, h 为限定候选目标的像素个数的窗口带宽矩阵; $b(x_i)$ 为像素 x_i 所对应的特征值映射到相应的 bin 值时的量化函数.

2) 反向投影. 将当前搜索帧图像搜索窗口中每个像素的值换算成为其颜色出现的概率值, 并将该值映射到 $0 \sim 255$ 的范围内, 映射后的值构成了一个颜色概率分布图.

3) 利用 MeanShift 算法计算当前帧人脸的位置和大小. 选择搜索窗口并计算窗口的质心, 将窗口的中心设置在计算出的质心处, 完成算法的一次偏移. 如果中心偏移的距离大于预设的固定阈值则重复该步骤, 直至算法收敛, 由此得到新窗口的位置, 窗口的大小可以利用窗口的二阶矩计算得到. 在 MeanShift 算法的每一次偏移过程中, 位于位置 y 的候选目标的模型为

$$p_u(y) = C_h \sum_{i=1}^{N_h} k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta [b(x_i) - u]. \quad (2)$$

通常选择用于评价边缘检测的 Bhattacharyya 系数^[7]来衡量目标和候选区域间的相似性程度, 相似性程度的计算公式为

$$\rho[p(y), q] = \sum_{b=1}^m \sqrt{p_b(y) q_b}, \quad (3)$$

$\rho[p(y), q]$ 值越大说明两个颜色模型的相似度越高. 本步骤即是在当前的视频图像中寻找能使 $\rho[p(y), q]$ 值最大的图像区域. 将式(3)在 y_0 处 Taylor 展开并近似可以得到

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u(y_0) q_u} + \frac{C_h}{2} \sum_{i=1}^{N_h} \omega_i k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right), \quad (4)$$

其中加权系数 $\omega_i = \sqrt{\frac{q_u}{p_u(y_0)}} \delta [b(x_i) - u]$. 式(4)右端的第一项与候选目标像素的位置无关, 第二项即是该候选区域在 y 处的概率密度估计, 第二项的极大值可以通过均值偏移向量计算得到, 由此获得候选区域的新位置 y_1 , 并有

$$y_1 = \sum_{i=1}^{N_h} x_i \omega_i g \left(\left\| \frac{y_0 - x_i}{h} \right\|^2 \right) / \sum_{i=1}^{N_h} \omega_i g \left(\left\| \frac{y_0 - x_i}{h} \right\|^2 \right). \quad (5)$$

判断 $\rho[q(y_0), p] < \rho[q(y_1), p]$ 是否成立, 若不成立则令 $y_1 = (y_1 + y_0)/2$, 重复该判断直至不等式成立. 然后将此次中心的位移 $\|y_1 - y_0\|$ 和给定的阈值相比较, 若大于阈值则设定此时的 y_1 为 y_0 , 重复以上过程进行下一次的偏移, 否则输出当前的中心位置.

4) 对视频逐帧重复以上过程直至结束. 且在重复的过程中将上一帧搜索窗的大小和中心作为下一帧算法搜索窗的初始值, 实现目标人脸在连续视频序列中的跟踪.

Camshift 算法能有效解决被跟踪人脸目标变形的问题, 在简单背景下能够取得良好的跟踪效果, 但由于 Camshift 算法单纯地考虑颜色直方图, 而忽略了目标的空间分布特性, 所以当视频中具有较大面积类人脸肤色的背景干扰时, 往往会造成跟踪失败或跟踪不准确. 针对这一问题, 张丽媛等^[8]通过优化目标选取和动态阈值来提高了算法效果, 但该算法较适用于跟踪较小的目标, 不完全适合人脸的跟

踪. 黄亚勤等^[9]在 Camshift 算法中加入人眼模板并以此作为算法收敛的条件, 但当人眼部分出现遮挡时, 跟踪效果并不理想. 林建华等^[10]采用在肤色之外融合纹理和边缘特征的方法来匹配目标, 但该方法对遮挡的抵抗性仍然不强.

2 改进的 Camshift 算法

针对 Camshift 算法在人脸跟踪时出现的上述不足, 本文提出一种改进的 Camshift 算法. 为了加强人脸的空间结构描述, 该算法在进行 MeanShift 算法模型匹配的过程中对人脸目标模型进行分块加权, 每一分块的权重不仅受到距离的影响, 还受到分块中人脸结构细节量的影响, 即具有更多细节结构的分块将在目标模型对比时拥有更大的权重. 之后, 根据人脸区域的高宽比不能过于悬殊以及短时间内变化不会过大的特点, 新算法对候选区域的大小变化及高宽比进行限定约束, 并在跟踪过程中融入 Kalman 滤波预测^[11], 以此进一步增强了算法的抗干扰能力.

2.1 算法的改进

1) 模型的分块与加权. 目标区域分块有助于增强跟踪目标的空间结构描述^[12]. 为了保证分块在跟踪过程中能够保持较好的区分性和稳定性, 分块的数量不能太多, 也不能过少. 分块过多将造成抗形变能力变差, 导致分块不稳定; 而分块过少则无法体现面部细节特征. 通常将目标区域划分为 9 块或 16 块, 即在目标区域横向和纵向上各做 3 或 4 等分.

假设将目标矩形区域 B 分为 S 个大小相同的矩形分块 $B_k (k=1, 2, \dots, S)$, 对于每个分块 B_j 由 M 个像素点 $\{x_i\} (i=1, 2, \dots, M)$ 组成, 且 $N=M \times S$, 则目标区域每一分块 B_j 的直方图模型为

$$q_{uj} = C_h \sum_{x_i \in B_j} k \left(\left\| \frac{x_i}{h} \right\|^2 \right) \delta [b(x_i) - u]. \quad (6)$$

类似地, 候选目标对应分块 B'_j 的模型定义为

$$p_u(y) = C_h \sum_{x_i \in B'_j} k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta [b(x_i) - u]. \quad (7)$$

由于各个分块大小一致且相互独立, 于是整个目标区域与候选区域之间的相似度由各个对应分块间的相似度 $\rho [p_j(y), q_j]$ 和对应分块的权重 w_j 决定, 即

$$\rho [q(y), p] = \sum_{j=1}^S \rho [q_j(y), p_j] w_j. \quad (8)$$

目标分块 B_j 和候选分块 B'_j 间的相似度 $\rho [p_j(y), q_j]$ 的计算方法与原有的区域相似度计算方法相同, 只是将颜色直方图参考模型由整个区域变为各个分块, 所以由式(4)可得

$$\rho [q_j(y), p_j] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u(y_0) q_u} + \frac{C_h}{2} \sum_{x_i \in B_j} \omega_i k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right). \quad (9)$$

式(8)中的 w_j 为对应分块 B_j 的权重, 该权重由分块的目标人脸直方图计算得到. 在目标人脸中, 由于外围的分块容易受到背景等因素的干扰, 所以分块的权重与分块到中心的距离成反比; 而人脸的细节结构主要体现在眼睛、嘴唇等五官特征上, 拥有这些结构的分块将获得更大的权重, 这些特征由于颜色与肤色不同而在肤色直方图的反向投影中主要显示为暗色, 具有相对较小的平均颜色概率值. 因此, 为了提高这些分块的权重值, 将其与整个区域的平均颜色概率做比较, 由此可得

$$w_j = K_l \frac{G_0}{G_j}. \quad (10)$$

式(10)中的 K_l 为分块到目标中心的距离系数, G_0 为完整直方图模型下的整个目标的平均颜色概率, G_j 是完整直方图模型下的分块 B_j 的平均颜色概率值. 由此可得候选区域的新位置 y_1 满足下式:

$$y_1 = \sum_{j=1}^M \sum_{x_i \in B_j} x_i \omega_i w_i g \left(\left\| \frac{y_0 - x_i}{h} \right\|^2 \right) / \sum_{j=1}^M \sum_{x_i \in B_j} \omega_i w_i g \left(\left\| \frac{y_0 - x_i}{h} \right\|^2 \right). \quad (11)$$

2) 人脸的形态约束. Camshift 算法在跟踪过程中能够根据跟踪目标形态的变化而不断改变其跟踪目标区域的大小,因此增强了算法的抗人脸形变的能力;但当有类人脸肤色物体在人脸周围干扰时,算法则很容易将该物体识别为人脸的一部分. 本文中的改进算法虽然通过分块加权增强了抗干扰能力,但并不能完全解决该问题. 当类肤色物体干扰到跟踪人脸时,算法计算出的跟踪人脸区域往往会发生高度和宽度比例严重失调或面积突然增大的情况. 为进一步增强算法的抗干扰性能,本文通过对跟踪区域的高宽比和面积大小两个方面进行约束来增强算法的抗干扰性能.

首先,在算法计算得出的人脸跟踪区域执行高宽比约束. 考虑到人脸在活动中的转向同样也会造成一定程度的比例变化,所以将窗口高宽比的范围限制为 $0.8 \sim 2.5$. 当计算出的人脸窗口超过这一范围时,认为人脸因受到周边干扰而导致过度变形,所以此时计算出的窗口的中心是不准确的,应将质心还原为上一帧的窗口中心位置,并将超过约束的高度或宽度设定为边界值. 设当前帧由算法计算出的人脸窗口的高度和宽度分别为 H_e 和 W_e , 约束控制后的高度和宽度分别为 H_c 、 W_c , 则满足以下约束条件:

$$\begin{cases} H_c = 2.5W_c, W_c = W_e, H_e > 2.5W_e; \\ H_c = H_e, W_c = W_e, 0.8W_e < H_e < 2.5W_e; \\ H_c = H_e, W_c = 0.8H_e, H_e < 0.8W_e. \end{cases} \quad (12)$$

其次,对候选窗口的面积变化率进行验证. 窗口的高宽比验证能够解决单方向上的干扰,但如果干扰的物体相对人脸区域较大,如果单一的进行高宽比的验证就无法识别. 若候选窗口面积相对前一帧窗口增大超过一定比例(本文设定为 25%),则认为人脸受到了大面积的类肤色干扰,此时同样应将质心还原为上一帧的质心,且窗口大小与前一帧相同,即跟踪窗口与上一帧相同.

3) 融入 Kalman 滤波预测. 为了进一步弥补 Camshift 算法在受到干扰和遮挡时跟踪位置不准确的问题,本文的改进算法融入了 Kalman 滤波器动态预测人脸的运动状态. Kalman 滤波的基本思想是利用线性系统状态方程得到状态变量和输出信号的估计值,通过系统输出数据校正状态变量的估计值,从而得到状态变量估计误差的均方差最小的最优估计^[13]. 将 Kalman 应用于人脸位置跟踪,即是算法的匹配结果作为系统观测值,并运用运动预测进行修正,从而获得更接近实际的最优人脸位置. Kalman 滤波器主要使用系统状态方程(式(13))和系统观测方程(式(14))来估计目标的状态:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{W}_t, \quad (13)$$

$$\mathbf{Z}_t = \mathbf{H}\mathbf{X}_t + \mathbf{V}_t. \quad (14)$$

式中的 \mathbf{X}_t 和 \mathbf{X}_{t-1} 分别表示对应时刻的状态向量,分别由中心所在的 x 和 y 轴坐标和对应方向上的速度构成,由于视频图像相邻帧的时间间隔很短,因此可以认为目标在做匀速直线运动; \mathbf{Z}_t 是系统观测向量,只由观测所得的中心位置决定; \mathbf{A} 和 \mathbf{H} 分别为系统状态转移矩阵和系统观测矩阵; \mathbf{W}_t 和 \mathbf{V}_t 是互不相关且均值都为 0 的系统噪声和观测噪声.

使用 Kalman 滤波预测人脸运动位置是一个预测和修正结合的循环过程^[12]. 系统的初始状态由人脸检测或手动选择决定,并假设初始速度为 0. Kalman 预测器根据前一帧中人脸的位置和运动状态来预测其在当前帧中的位置及速度,即系统先验估计值. 然后,利用该值来修正系统观测值(由改进 Camshift 跟踪算法计算获得)得到该帧的最终人脸位置和速度,并以此结果作为后一帧的先验估计值. 重复该过程直至跟踪完成.

2.2 算法流程

本文改进后的算法流程如下:

1) 在视频图像中设定初始目标人脸区域.

2) 初始化目标人脸模板和 Kalman 滤波器. 统计目标人脸窗口的整个窗口以及各个分块的颜色直方图,并依此计算各个分块的权重.

3) 根据上一步骤得到的人脸分块颜色直方图和各分块权重,利用改进后的相似度匹配算法计算出候

- 选窗口的大小和位置,并对候选窗口进行人脸形态约束检查修正,得到人脸目标的新的跟踪位置和尺寸。
- 4) 将上一步得到的人脸位置坐标作为系统观测值,使用 Kalman 滤波预测修正获得人脸位置的最优估计值。
- 5) 将跟踪结果设置为目标人脸模板,读入新的一帧,并回到步骤 2)继续执行,直至最后视频的一帧。

3 实验结果及分析

为了验证本文中的改进算法的效果,设计相关的实验进行验证。验证实验在 Windows 7 系统下开展,采用 Microsoft Visual Studio 2010 作为实验平台,并借助开源跨平台计算机视觉库 OpenCV(open source computer vision library, 2.4.10 版本)中的相关函数。实验所用视频均为摄像头本地录制,干扰源为网络皮肤颜色图片。为了验证该肤色图片的干扰作用,实验将人脸与其一同进行基于人脸直方图的反向投影。从图 1(b)可以看到,人脸区域与肤色图片区域都呈现出亮色,说明两者的颜色直方图类似。



图 1 肤色图片验证效果

针对测试者从肤色图片前方水平移动通过的实验视频,分别利用经典 Camshift 算法和本文算法跟踪测试者脸部区域。为了保证两次跟踪时所采用的初始人脸区域一致,实验的初始人脸窗口由 OpenCV 中的人脸检测函数在第 1 帧图像获得。在两次跟踪过程中分别截取部分对应帧进行比较,跟踪效果分别如图 2 和图 3 所示。



图 2 经典 Camshift 算法的跟踪结果



图 3 本文改进算法的跟踪结果

在使用经典 Camshift 算法跟踪时,当人脸靠近肤色图片并与面部部分重叠时,跟踪窗口出现了偏差,并逐渐将部分肤色图片一起包含到跟踪窗口中。随着人脸离开肤色图片,由于肤色图片面积大于人脸,算法错误地将肤色图片认作人脸,说明跟踪失败。而在使用本文算法跟踪时,跟踪窗口未随干扰而明

显改变.

为了进一步验证算法在不同场景下的效果,分别选取肢体(实验者的手掌)、类肤色图片、类肤色大物体(家具)等不同干扰源的 3 段视频作为实验素材. 为了便于检测验证,实验者在各视频中均保持正脸. 实验首先使用人脸检测函数逐帧获取检测人脸窗口并记录,再依照上述流程分别使用经典 Camshift 算法和本文算法对各视频片段进行人脸跟踪得到各帧的人脸跟踪窗口,然后逐帧对比检测结果和跟踪结果. 同一帧中,若检测窗口与跟踪窗口的重复像素占两者所含像素之和的比例达到或超过 40%,则认为该帧跟踪成功,否则判定跟踪失败. 各视频在不同算法下的跟踪成功率(跟踪成功率=跟踪成功的帧数/视频总帧数)如表 1 所示. 由表 1 可知,本文算法在不同干扰源下的跟踪成功率均明显高于经典算法,且经典算法的跟踪成功率随着干扰源面积的增大而显著下降,而本文算法则相对稳定.

表 1 多场景人脸跟踪实验结果

视频 序号	干扰源	跟踪成功率/%	
		经典算法	本文算法
1	肢体	87.3	96.7
2	图片	62.5	92.1
3	家具	38.7	86.3

4 结论

针对经典 Camshift 算法在人脸跟踪中易受类肤色干扰而导致跟踪不准确的问题,本文在人脸跟踪匹配过程中利用分块加权来提高了面部器官的辨识度,并结合人脸形态约束来控制干扰所带来的过度形变,同时利用 Kalman 滤波器对视频图像中的人脸位置进行预测估计. 实验结果表明,相对经典 Camshift 算法,改进后的本文算法在人脸跟踪的准确性和类肤色干扰源的抵御能力上都有了较大的改善.

参考文献:

[1] Azarbayejani A, Horowitz B, Pentland A. Recursive estimation of structure and motion using relative orientation constraints[J]. Proc IEEE CVPR, 1993:294-299.

[2] Eleftheriadis A, Jacquin A. Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates[J]. Signal Processing Image Communication, 1995,7(3):231-248.

[3] Fukunaga K, Hostetler L D. The estimation of the gradient of a density function, with applications in pattern recognition[J]. IEEE Trans Information Theory, 1975,21(1):32-40.

[4] 梁路宏,艾海舟. 基于人脸检测的人脸跟踪算法[J]. 计算机工程与应用,2001(17):42-45.

[5] Bradski G R. Computer vision face tracking as a component of a perceptual user interface[C]//Proceedings Fourth IEEE Workshop on Applications of Computer Vision. Princeton, New Jersey, USA, 1998:214-219.

[6] 贾亮亮. 基于视频序列的运动目标检测与跟踪算法研究[D]. 重庆:重庆大学,2011.

[7] Kailath T. The divergence and Bhattacharyya distance measures in signal selection[J]. IEEE Transactions on Communication Technology, 1967,15(1):52-60.

[8] 张丽媛,梁凤梅. 改进的 CamShift 人脸跟踪算法[J]. 科学技术和工程,2014,14(13):231-235.

[9] 黄亚勤,董秀成,李郝,等. 改进的 CamShift 人脸跟踪算法[J]. 计算机工程,2011,37(2):180-182.

[10] 林建华,刘党辉,邵显奎. 多特征融合的 Camshift 算法及其进一步改进[J]. 计算机应用,2012,32(10):2814-2816.

[11] Baumberg A M, Hogg D C. Learning spatiotemporal models from training examples[C]//Proc British Machine Vision Conference, Birmingham, 1995:63-75.

[12] 胡钢,杨静宇. 基于分块颜色直方图的 MeanShift 跟踪算法[J]. 系统仿真学报,2009,21(10):2936-2939.

[13] 王丽,郝晓丽. 基于 Kalman 滤波器和改进 Camshift 算法的双眼跟踪[J]. 微电子学与计算机,2016,33(6):109-112.