

doi:10.11835/j.issn.1000-582X.2014.04.013

DSPCA 在自适应视频跟踪算法中的应用

相入喜^{1,2}, 李见为¹, 冯海亮¹

(1. 重庆大学 光电技术及系统教育部重点实验室, 重庆 400044; 2. 常州工学院 光电学院, 常州 213002)

摘要:为了解决单一固定目标模型在复杂的场景中易产生跟踪漂移问题, 提出一种基于 DSPCA 的自适应粒子滤波跟踪方法, 通过稀疏主成分分解 (DSPCA) 在线获取互补图像集, 同时将其按照新的相似度 BRS 进行自适应融合作为新目标模型。与经典的粒子滤波跟踪算法、视觉分解跟踪算法和多特征自适应融合跟踪算法, 与有挑战性较高的场景视频相比, 提出的算法在形态、运动快速及严重遮挡的运动场景中, 都能鲁棒地跟踪到目标。

关键词: 视频跟踪; 粒子滤波; 稀疏主成分分解; BRS

中图分类号: TP391

文献标志码: A

文章编号: 1000-582X(2014)04-082-06

Adaptive visual tracking based on DSPCA

XIANG Ruxi^{1,2}, LI Jianwei¹, FENG Hailiang¹

(1. Key Laboratory of Optoelectronic Technology & Systems, Ministry of Education, Chongqing University, Chongqing 400044, China; 2. College of Optoelectronic Engineering, Changzhou Institute of Technology, Changzhou 213002, China)

Abstract: In order to solve the problem resulted from tracking drift of single reference color histograms, a sort of particle filtering algorithm based on DSPCA (decomposition-sparse-principal component-analysis) is presented. The image set with complementary is accessed by DSPCA, which is fused by means of the new similarity as the new object model. Compared with the basic particle filtering, visual tracking decomposition and adaptive multiple cues fusion tracking algorithm in challenging scenarios, the presented method can robustly track the object in fast motion and the object which is mostly blocked.

Key words: visual tracking; particle filter; DSPCA; BRS

视频跟踪是机器视觉研究热点之一, 近几十年来引起了众多研究者的关注, 提出了大量的跟踪算法, 并且已成功应用于视频监控、运动分析和人机交互等领域^[1-5]。现有的跟踪算法粗略可以分为 2 类: 确定跟踪算法和随机跟踪算法^[6]。确定跟踪算法是以核跟踪^[7]为代表, 核跟踪最先由 Comanicu 把均值移动成功应用到视频跟踪中, 具有简单、速度快的特点, 但易收敛到局部极小值, 且难以跟踪快速运动、姿态快速变化及严重遮挡的目标, 而随机跟踪算法则能有效处理该类问题, 粒子滤波^[8]是该类方

法中应用最广泛的方法之一, 它能有效地处理非线性非高斯分布的状态预测问题, 同时还是一种比较鲁棒的跟踪方法。众多跟踪算法采用固定不变的目标模板进行跟踪, 对于姿态快速变化和严重遮挡的目标极易产生漂移。为了避免漂移, Ross 等^[9]利用子空间方法获取持续变化的目标集代替固定不变的目标模型, 但该方法需要提前训练且计算量较大; Leichte Ido 等^[10]和 Bajramovic F 等^[11]都是采用先验多目标模型代替固定目标模型; Wang 等^[12]利用混合高斯分布在线构建目标模型, 但在实际场景中,

收稿日期: 2013-10-12

基金项目: 重庆市科委科技计划攻关重点项目 (CSTC 2009AB2231); 常州工学院科研基金资助项目 (YN1208)

作者简介: 相入喜 (1976-), 男, 博士, 主要从事目标跟踪、模式识别研究, (E-mail) xjmtinacq@gmail.com。

很多目标模型分布不适合高斯分布;Kwon J 等^[13]利用了 DSPCA^[14]在线构建多参照直方图,但其时间复杂度非常高,不适合实时跟踪。

为了解决上述问题,结合改进后 DSPCA 在线获取目标图像变化集,将图像集中的图像按照改进后的 BRD^[15]准则线性融合为单参照颜色直方图。在粒子滤波框架下,使用随机漂移模型作为系统状态转换函数,将相似度函数 BRS (bin-ratio-similarity) 作为各粒子的权值。实验证明,提出的方法鲁棒地跟踪到快速形态变化及严重遮挡的目标。

1 粒子滤波器简介

视频目标跟踪可视作为带有隐式状态的马尔科夫状态估计问题,为了有效解决非线性、非高斯状态估计问题,Isard^[6]引进了粒子滤波。假设目标状态向量 $X_t = [x, y]$ 是目标的中心坐标,且状态方程满足一阶马尔科夫 $p(X_t | X_{1:t-1}) = p(X_t | X_{t-1})$, 量测方程可用 $p(Y_t | X_t)$ 表示,其中 Y_t 是量测值, $1:t-1$ 表示直到 $t-1$ 时刻对应值的集合。根据贝叶斯理论,状态的估计值可以通过式(1)迭代求解

$$p(X_t | Y_{1:t}) \approx p(Y_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | Y_{1:t-1}) dX_{t-1}. \quad (1)$$

后验分布 $p(X_t | Y_{1:t})$ 可以用蒙特卡罗方法近似表示 $p(X_t | Y_{1:t}) \approx \{x_t^i, w_t^i\}_{i=1}^N = 1$, 其中, X_t^i 表示 t 时刻对应的第 i 个粒子的状态, w_t^i 表示第 i 个粒子状态所对应的权重,可以通过式(2)迭代获得。

$$w_t^i \approx_{i=1}^{i-1} p(Y_t | X_t^i) \quad \text{S. T.} \quad \sum_{i=1}^N w_t^i = 1. \quad (2)$$

具体的采样、量测似然及重采样的理论可以参考文献[16]。

2 多参照颜色直方图

固定参照颜色直方图易产生跟踪漂移,为了改善这一问题,文中仍采用单参照直方图,但它是从不同目标形态的颜色直方图中计算。不同目标形态是利用 DSPCA^[14] 构建,具体将在 2.1 中讨论。2.2 介绍多参照颜色直方图。

2.1 稀疏主成分分解

假设 A_t 是 Gramian 矩阵, e 是稀疏主成分且满足(5)式。

$$\text{Maximize } e^T A_t e - \lambda \|e\|_0 \quad \text{s. t.} \quad \|e\|^2 = 1, \quad (5)$$

其中, $\|e\|_0$ 是 0 阶范数,也就是列向量 e 中非零的个数; λ 是控制参数,用来控制 e 的稀疏程度, λ 值越

大, e 越稀疏。在文中, t 时刻的 Gramian 阵 A_t 可以通过式(6)计算。

$$A_t = s^T S, \quad s = [f_t^1 \cdots f_t^i \cdots f_t^{N_s}], \quad (6)$$

其中: f_t^i 是 t 时刻的灰度图; N_s 是灰度图像像素数; 为了求解稀疏主成分 e , 利用文献[4]的优化工具, 可以获得式(6)近似解。在所有成分中, 按照式(7)选出 r 个成分向量 e_r 。

$$I_t = \{i_a^y | i_a^y = s(y), e_r(y) = 0\}. \quad (7)$$

如果 $e_r(y)$ 的第 y 个元素是一个非零值, I_t 就包含式(6)中的矩阵 s 所对应列的第 y 个灰度图所对应的 HSV 图。具体的分解过程如图 1 所示。

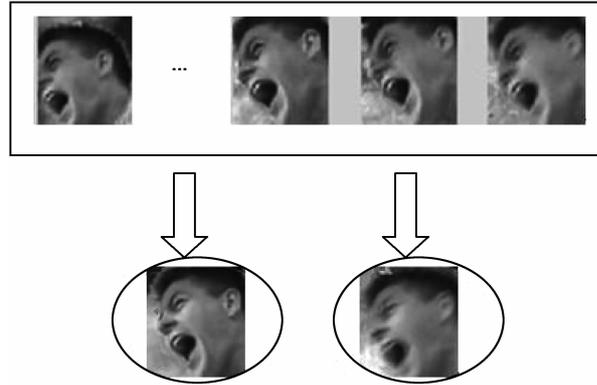


图 1 灰度图像的 DSPCA 分解过程

2.2 基于多参照直方图的单参照直方图

通过稀疏主成分分解获得目标形态集, 需要将形态集中的图像合成一个参照直方图。在合成时, 采用改进 BRD^[15] 准则(也就是 BRS)进行线性系数化。由于跟踪初始化时, 已确定参照直方图 q_u 是一个 U bin 的直方图。先将目标形态集中的图像按照文献[7]分别进行计算对应的颜色直方图 p_m^u ($m = 1, \dots, M$), 接着对所有的直方图归一化, 即满足。

$$\sum_{u=1}^U \|p_m^u\|^2 = 1. \quad (8)$$

对于目标直方图 q 和候选目标直方图 p , 他们的相似性可以通过式(9)计算。

$$d = \frac{\|p + q\|_2^2 \sum_{u=1}^N \frac{p_u q_u}{(p_u + q_u)^2}}{n_e}, \quad (9)$$

其中: N 是直方图量化的 bin 数; N_e 是有效 bin 数, 即也就是 2 个直方图量化空间和不为零的个数。

假设形态集中有 M 幅图像, 其对应的直方图 q_k , 对于 t 时刻的参照直方图可以利用下式进行计算。

$$q_i = \sum_{k=1}^M \alpha_k q_k, \quad (10)$$

其中: α_m 是稀疏图像集中的第 m 个图像直方图和 q 的 BRS 距离, 并且满足, $\sum_{m=1}^M \alpha_m = 1$ 。其中 M 是可变的, 文中最大值设为 5, 如果形态集中的图像多于 5, 就采用先进先出的原则替换形态集中的目标图像。

3 基于 DSPCA 的跟踪算法伪码

上面主要讨论了基于多参照图跟踪算法的粒子滤波框架和在线获取的目标参照直方图。这一节将给出算法的伪码和复杂性分析。

3.1 算法伪码

1) 初始化, 令 $t=0$, 在初始帧中, 手动选取参考视觉目标, 并将颜色空间变换成 HSV 空间, 计算参考目标的颜色分布 $\{q^u\}_{u=1 \dots U}$, 根据先验分布 $p(x_0)$, 建立初始状态样本集 $\{X_0^i, 1/N\}_{i=1 \dots N}$, $\text{num} = 1$;
For $t=1, 2, 3, \dots$

2) 粒子采样: 根据预测高斯分布采样

$$\tilde{X}_t^i \sim q(X_t^i | X_{t-1}^i, Y_t) = N(X_t^i, P_t^i). \quad (11)$$

3) 粒子的权值计算: 根据式(4)计算样本 \tilde{X}_t^i 的权值 w_t^i , 并归一化。

4) 视觉目标状态输出: 计算 t 时刻目标状态的 MMSE 估计 $\tilde{X} = \sum_{i=1}^N w_t^i \tilde{X}_t^i$, 获取其灰度图像, $\text{num} = \text{num} + 1$

5) 如果 num 等于 25, 按照式(7)进行 DSPCA 分解, 获得其稀疏主成份目标形态集, 通过式(10)获取新参照直方图;

6) $k=k+1$, 返回 2)。

3.2 算法复杂性

针对文中算法, 复杂性主要受到 DSPCA 目标更新和粒子滤波跟踪两方面制约, 因此可以从这 2 个方面进行分析: 1) 在 DSPCA 目标更新阶段, 所需复杂度为 $O(n \sqrt[4]{\log(n)}/e)^{[14]}$, 其中, n 是协方差矩阵 A_t 的大小, 是理想的精确值; 2) 在粒子滤波跟踪阶段, 粒子滤波的复杂度主要受粒子数 N_{pt} 和观测似然复杂度 L 的影响, 其复杂度为 $O(N_{\text{pt}}L)$, 影响 L 的是相似度的选择, 文中相似度采用的是鉴别力强的 BRS 相似度, 而经典粒子滤波采用的是普通相似度 Bhattacharyya。BRS 所需的颜色量化级数远远小于 Bhattacharyya 所需颜色量化级数, 因此总的复杂度是 $O(N_{\text{pt}}N\lambda n \sqrt[4]{\log(n)}/e + N_{\text{pt}}L)$, 其中 λ 是目标更新标志。

4 实验结果

为了验证上述跟踪算法的有效性和鲁棒性, 选用 4 段来自韩国首尔大学计算机视觉研究室^[13]的视频序列。4 个实验都是在带有优化的 Mex C++ 的 Matlab2009a, 主频为 AMD 2.9 GHz 四核 PC 机中执行。实验中, 先手动获取 25 帧的目标图像, 然后进行 DSPCA 分解, 其稀疏控制参数 λ 取 100; 每个颜色通道取 8 bin, 粒子采样状态个数为 50。经典粒子滤波每个颜色通道量化成 16 bin, 粒子采样数也为 50; 视觉分解跟踪算法采用作者的源码; 多特征自适应融合算法用 Matlab2009a 实现。

第 1 个实验是篮球赛视频序列, 被跟踪的目标是穿绿衣的运动员。在整个序列中, 被跟踪的运动员在运动过程中频繁出现遮挡、快速运动、相似目标干扰等变化。该序列长度为 686 帧, 每帧图像的大小为 640×352 , 运动员大小为 40×38 。对于该实验, 文中跟踪算法 (MPF) 处理速度平均为 10 fps (帧/秒), 基本粒子滤波跟踪算法 (BPF) 处理速度平均为 20 fps (帧/秒), 图 2 显示了该实验的跟踪结果, 其中 (a) 显示了 MPF 跟踪结果, (b) 是 BPF 跟踪结果。

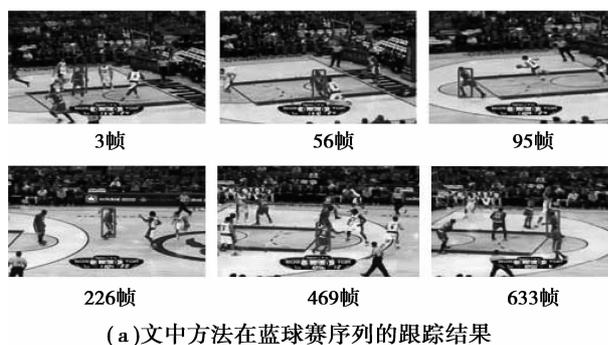


图 2 实验的跟踪结果

采用 BPF 进行跟踪, 在 423 帧时由于其附近队友经过时, 使得跟踪模板发生了严重漂移, 到 423 帧后, 跟踪模板几乎完全偏离目标, 这是因为目标在运

动过程中目标形态不断变化及相似目标的影响, 固定目标模型就不能适应这些变化, 使得基于固定目标模型的方法失效。相反基于多参照颜色直方图的跟踪性能优于 BPF, 特别是在目标快速运动造成的形态变化及相似目标的影响时, 如 226 帧和 653 帧, 就能一直稳健地跟踪到目标, 直至视频结束。主要原因在于: DSPCA 获取到互补的形态集, 使得能适应变化的目标。

第 2 个实验是双人滑冰比赛的视频序列, 被跟踪的女滑冰运动员经常被男滑冰运动员遮挡、其形态经常发生变化且运动比较快。该序列长度为 707 帧, 每帧图像的大小为 640×352 , 被跟踪的女滑冰员的大小为 152×52 , 对于该实验, MPF 处理速度平均为 10 fps (帧/秒), BPF 处理速度平均为 20 fps (帧/秒)。图 3 显示了该实验的跟踪结果, 其中(a)显示了 MPF 跟踪结果, (b)是 BPF 跟踪结果。BPF 只能跟踪到 290 帧, 由于女滑冰移动比较快, 跟踪器慢慢漂移, 到 298 帧时已偏移到男滑冰的身上, 而 MPF 就能鲁棒准确跟踪到 707 帧。

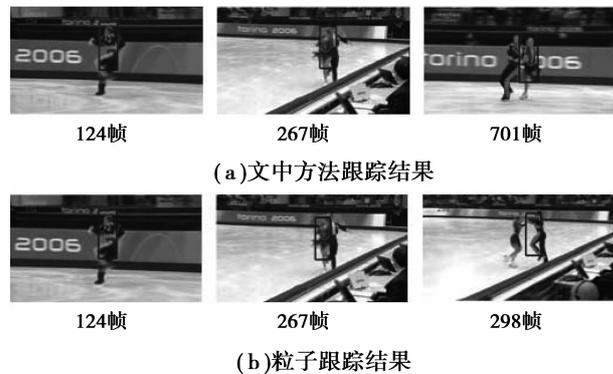


图 3 快速移动, 遮挡和形态变化
双人滑冰视频的跟踪结果

每个实验(a)中的参照直方图一直是在线更新, 因而所提出的算法都能稳健准确跟踪到目标, 特别是当目标被遮挡, 快速运动或者相似目标相似的干扰, 都能稳健跟踪。相反每个实验(b)所用的参照直方图一直保持不变, 因而(b)在目标遮挡、快速运动及相似目标的干扰, 就不能稳健跟踪到目标, 要么误跟踪, 要么跟丢目标。

在序列 3 中(见图 4), 被跟踪的运动员头部由于快速运动, 存在着遮挡, 及相似运动员的干扰。文中方法和视觉分解跟踪方法都能鲁棒地跟踪到目标, 但平均速度明显优于视觉分解跟踪速度, 为 20 帧/秒, 而视觉分解跟踪的平均速度仅为 1.5 帧/秒, 完全不能实时。

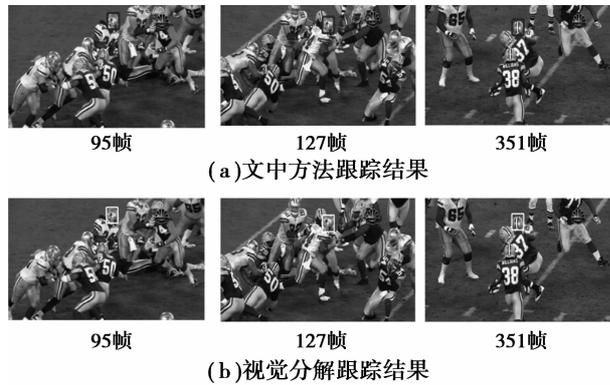


图 4 文中方法和视觉分解跟踪方法^[13]
在快速移动, 遮挡等足球序列中的跟踪结果

序列 4 是跟踪人脸结果(见图 5)。歌星在唱歌时头部不断发生变化, 同时还受到灯光的不断干扰, 在整个 330 帧序列中, 文中方法可以鲁棒地跟踪到目标, 而多特征自适应融合跟踪方法仅可跟踪到 80 帧, 然后就慢慢偏离目标。该方法在该序列的平均速度为 20 帧/秒, 而多特征融合跟踪是 15 帧/秒。

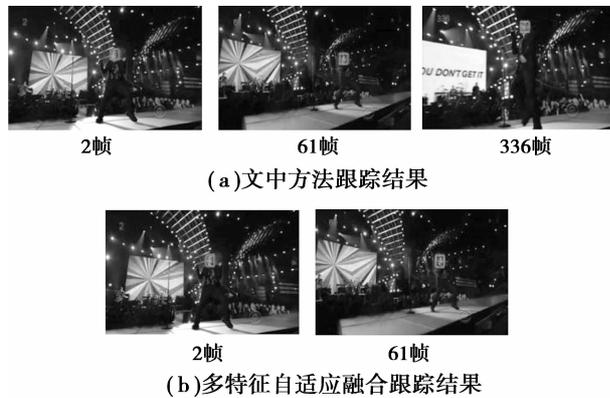


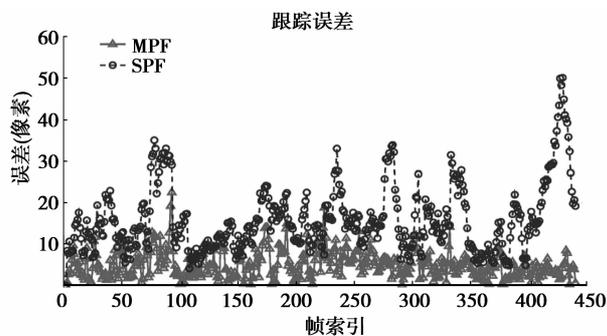
图 5 文中方法和多特征自适应融合
跟踪方法^[17]在演唱会序列的跟踪结果

为了进一步量化 2 种跟踪算法的性能, 利用目标跟踪误差进行量测。所谓跟踪误差就是指在二维图像上真实目标的中心位置坐标 p 和跟踪预测的中心位置坐标 p' 的距离^[16], 即

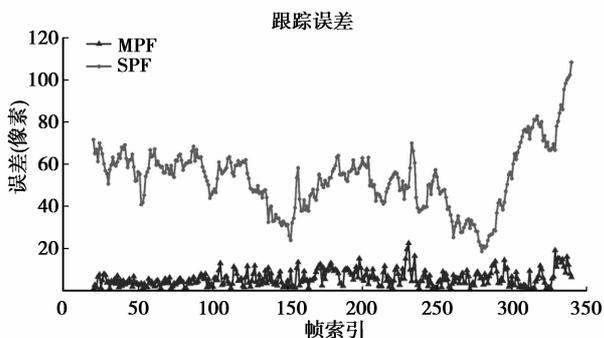
$$err = \sqrt{(p_x - p'_x)^2 + (p_y - p'_y)^2}$$

图 6 是这 2 种跟踪算法在这 2 个视屏序列中的跟踪误差。图 6(a) 是第 1 个实验的前 400 帧跟踪误差, 图 6(b) 是第 2 个实验的前 340 帧跟踪误差。从图 6(a) 中可以看出, 由于场景比较简单, 因而所提出跟踪算法与所经典粒子滤波跟踪算法都能跟踪到目标, 但 MPF 更精确, 误差更小。图 6(b) 是在复杂的场景中, 提出的跟踪算法与经典粒子跟踪算法

相比在稳健性方面有明显的优势,在 298 帧以后,MPF 就跟错目标,跟丢目标。



(a)第1个实验的前400帧跟踪误差



(b)第2个实验的前340帧跟踪误差

图 6 MPF 和 BPF 目标跟踪误差

5 结 论

提出了一种在线更新目标模型的跟踪算法。在跟踪中,借助稀疏主成分分解获取互补的图像集,同时将其按照 BRS 准则线性融合作为新的目标模型。通过 2 段运动视频序列的目标跟踪可以发现该方法的稳健性,尤其在目标严重遮挡、姿态不断变化和快速移动时,更能体现;同时,对比实验也表明该方法在复杂场景中优于经典粒子滤波跟踪算法和多特征自适应融合算法,和视觉分解跟踪方法相当,但平均速度明显优于视觉分解跟踪速度,经过优化以后,可以达到实时,而视觉分解跟踪完全达不到实时,平均每帧需 1.5~3 s 处理。此外,仍有待进一步提高的是多目标跟踪方法研究,这将是后续的研究课题。

参考文献:

[1] Nguyen Q D, Milgram M. Online subspace learning using a misalignment detection for tracking a talking face in video sequence [J]. *Neurocomputing*, 2010, 73(10/12):1624-1630.

[2] Candamo J, Shreve M, Goldgof D B, et al. Understanding transit scenes: a survey on Human

behavior-recognition algorithms [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2010, 11(1):206-224.

- [3] Lucena M, Fuertes J M, Blanca N P, et al. Tracking people in video sequences using multiple models [J]. *Multimedia Tools And Applications*, 2010, 49(2):371-403.
- [4] 蒋蔚,伊国兴,曾庆双. 基于 SVM 数据融合的实时粒子滤波算法[J]. *系统工程与电子技术*, 2010, 32(6):1334-1338.
- JIANG Wei, YI Guoxing, ZENG Qingshuang. Real-time particle filter based on data fusion with support vector machines [J]. *Systems Engineering and Electronics*, 2010, 32(06):1191-1194.
- [5] 韩松,张晓林,陈雷,等. 基于改进高斯粒子滤波器的目标跟踪算法[J]. *系统工程与电子技术*, 2010, 32(6):1191-1194.
- HAN Song, ZHANG Xiaolin, CHEN Lei, et al. Object tracking method based on improved Gaussian particle filter [J]. *Systems Engineering and Electronics*, 2010, 32(6):1191-1194.
- [6] Yilmaz A, Javed O, Shah M. Object tracking: A survey[J]. *ACM Computing Surveys*, 2006, 38(4):352-355.
- [7] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(5):564-577.
- [8] Isard M, Blake A. Condensation- conditional density propagation for visual tracking [J]. *International Journal of Computer Vision*, 1998, 29(1):5-28.
- [9] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1/3):125-141.
- [10] Leichter I, Lindenbaum M, Rivlin E. Mean shift tracking with multiple reference color histograms [J]. *Computer Vision and Image Understanding*, 2010, 114(3):400-408.
- [11] Bajramovic F, Deutsch B, Grassl C, et al. Efficient adaptive combination of histograms for real-time tracking [J]. *Eurasip journal Journal on image Image and video Video processing Processing*, 2008(5):1-11.
- [12] Wang H Z, Suter D, Schindler K, et al. Adaptive object tracking based on an effective appearance filter [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(9):1661-1667.
- [13] Kwon J, Lee K M. Visual tracking decomposition[C]// *Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 13-18, 2010, San Francisco, CA. Piscataway: IEEE Press,

2010:1269-1276.

[14] Daspremont A, Ghaoui E L, Jordan M, et al. A direct formulation for sparse PCA using semidefinite programming [J]. *Siam Review*, 2007, 49(3): 434-448.

[15] Xie N H, Ling H B, Hu W M, et al. Use bin-ratio information for category and scene classification [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA. Piscataway: IEEE Press, 2010: 2313-2319.

[16] 程建, 周越, 蔡念, 等. 基于粒子滤波的红外目标跟踪[J]. *红外与毫米波学报*, 2006, 25(2): 113-117.

CHENG Jian, ZHOU Yue, CAI Nian, et al. Infrared object tracking based on particle filters [J]. *Journal of Infrared and millimeter Waves*, 2006, 25(2): 113-117.

[17] 尹宏鹏, 柴毅, 匡金骏, 等. 一种基于多特征自适应融合的运动目标跟踪算法[J]. *光电子·激光*, 2010, 21(6): 917-922.

YIN Hongpeng, CHAI Yi, KUAN Jinjun et al. A moving target tracking algorithm based on adaptive multiple cues fusion [J]. *Journal of Optoelectronics Laser*, 2010, 21(6): 917-922.

(编辑 陈移峰)

~~~~~  
(上接第 57 页)

[8] Lu J W, Tan Y P. Fusing shape and texture information for facial age estimation [C] // Proceedings of 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing, May 22-27, 2011, Prague. Piscataway: IEEE Press, 2011: 1477-1480.

[9] Dibeklioglu H, Salah A A, Gevers T. A statistical method for 2-D facial landmarking [J]. *IEEE Transactions on Image Processing*, 2012, 21(2): 844-858.

[10] Hardoon D R, Szedmak S, John S T. Canonical correlation analysis: An overview with application to learning methods [J]. *Neural Computation*, 2004, 16(12): 2639-2664.

[11] Daugman J. Face and gesture recognition; overview [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 19(7): 675-676.

[12] Déniz O, Bueno G, Salido J, et al. Face recognition

using histograms of oriented gradients [J]. *Pattern Recognition Letters*, 2011, 32(12): 1598-1603.

[13] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary pattern [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987.

[14] Lu J W, Tan Y P. Ordinary preserving manifold analysis for human age and head pose estimation [J]. *IEEE Transactions on Human-Machine Systems*, 2013, 43(2): 249-258.

[15] Chao W L, Liu J Z, Ding J J. Facial age estimation based on label-sensitive learning and age-oriented regression [J]. *Pattern Recognition*, 2013, 46(3): 628-644.

(编辑 侯 湘)