基于稀疏表示和特征选择的 LK 目标跟踪*

潘　晴, 曾仲杰
(广东工业大学 信息工程学院, 广州 510006)

摘　要: 为了实现复杂场景中的视觉跟踪,提出了一种以 LK(Lucas-Kanade)图像配准算法为框架,基于稀疏表示的在线特征选择机制。在视频序列的每一帧,筛选出一些能够很好区分目标及其相邻背景的特征,从而降低干扰对跟踪的影响。该算法分别构造前景字典和背景字典,前景字典来自于第一帧的手动标定,并随着跟踪结果不断更新,而背景字典则在每一帧重新构造。同时,一种新的字典更新策略不仅能有效应对目标的外观变化,而且通过特征选择机制,能避免在更新过程中引入干扰,从而克服了漂移现象。大量的实验结果表明,该算法能有效应对视角变化、光照变化以及大面积的局部遮挡等挑战。

关键词: 视觉跟踪; 稀疏表示; LK 图像配准算法; 特征选择
中图分类号: TP301.6;TP391　　**文献标志码**: A　　**文章编号**: 1001-3695(2014)02-0625-04
doi:10.3969/j.issn.1001-3695.2014.02.075

LK tracking based on sparse representation and features selection

PAN Qing, ZENG Zhong-jie
(*School of Information Engineering*, *Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: In order to tracking object in complex scenes, the paper proposed an online features selection mechanism based on sparse representation in the Lucas-Kanade image registration framework. To reduse the impact of interference on the tracking, it selected the features that owned best discrimination betweem object and adjacent background in each frame of the video sequence. The algorithm was composed of forward dictionaries and background dictionaries, the former which would be created manually from the first frame and updated with the tracking results, the background dictionary would be reconstruct by evrey frame. Meanwhile, a new dictionary updating strategy not only can effectively cope with the appearance changes, but also handle drift. Experiment shows that the proposed algorithm can effectively deal with pose change, illumination change and large partial occlusion.

Key words: visual tracking; sparse representation; Lucas-Kanade image registration algorithm; feature selection

近三十年来,视觉跟踪一直是计算机视觉领域的一个重要课题,随着众多学者不断研究和探索,视觉跟踪已经在视频监控、智能交通、人机交互等领域得到广泛的应用。目前大量的跟踪算法被提出以解决跟踪过程中遇到的各种难题,包括目标的快速运动、视角变化、遮挡、复杂背景干扰以及光照变化等[1]。

视觉跟踪算法主要分为判别法和统计生成法两类。判别法是将跟踪视做一个分类问题,通过分类器实现对目标的定位。统计生成法则用一个或多个模型来表示目标,然后在一定范围内寻找最佳的匹配结果。Grabner 等人[2]提出一种基于在线 Adaboost 的特征选择算法,其思想是用每一帧的跟踪结果去更新分类器,同时进行特征选择,但是这种方法也会把干扰学习进去,随着误差的累加很可能会导致漂移现象。为了应对漂移问题,Grabner 等人[3,4]随后引入了半监督思想,用无标签数据和有标签数据来更新分类器。Ross 等人[5]利用主成分量分析技术(PCA)对多模板提取二阶不相关特征,目标则由这些不相关特征的线性组合来描述。通过不断对外观进行学习更新,Ross 提出的方法能够很好地应对视角变化。但是在遇到长时间大面积遮挡的时候,这种全局更新机制容易导致漂移现象。Mei 等人[6,7]在粒子滤波的框架下,利用稀疏表示技术,使跟踪目标能用尽可能少的模板线性表示出来。这种方法不仅能有效应对光照变化,而且在发生大面积遮挡时依然保持很好的鲁棒性。但是该方法也是用跟踪结果来对模板进行全局更新,在求解稀疏表示的时候,模板中一些属于背景的特征会产生很大的重构误差,这些误差可能会影响求解的精度,因此 Mei 的方法并不能应对极端的视角变化。

为了应对目标外观的变化,上述方法都提出了各自更新策略。但是这些方法仅仅把目标视做单一整体进行更新,更新的同时不可避免地会引入干扰。在复杂场景中,对跟踪起积极作用的往往是那些最能够区分目标与其相邻背景的特征。鉴于以上分析,本文提出一种新的跟踪算法。相对于现有的算法,本文提出的算法主要有以下几个优点:a)同时构造前景字典和背景字典,通过稀疏表示,筛选出最能够区分目标和背景的特征;b)一种新的字典更新策略,既能应对外观的变化,同时避免更新造成的漂移;c)将稀疏表示应用于 LK 图像配准,减少干扰对配准的影响,提高算法的鲁棒性。

1　算法原理

该算法主要由三个部分构成:a)输入第 t 时刻的跟踪结

收稿日期: 2013-03-19; **修回日期**: 2013-05-09　　**基金项目**: 广东省自然科学基金资助项目(9451009001002667)

作者简介:潘晴(1975-),男,湖北武汉人,副教授,主要研究方向为图像处理、模式识别;曾仲杰(1986-),男,广东湛江人,硕士,主要研究方向为模式识别、视觉跟踪(zhzhji440@163.com).

果,用稀疏表示来进行特征选择;b)根据特征选择的结果,对前景字典和模板进行更新;c)用 Lucas-Kanade 图像配准算法去预测目标在 $t+1$ 时刻的状态。

算法 1 本文提出的算法流程

初始化:手动标定目标的初始状态,然后在目标区域均匀地采样一系列子块,用于构造前景字典 $\boldsymbol{\Phi}_{\text{target}}$。

a)输入 t 时刻的目标状态 $\boldsymbol{\tau}_t=[x_t,y_t,\theta_t,S_t]^{\mathrm{T}}$;

b)根据目标状态 $\boldsymbol{\tau}_t$,在第 t 帧目标的周围随机采样一系列局部子块,用这些子块构建背景字典 $\boldsymbol{\Phi}_{\text{bac}}$;

c)利用稀疏表示,计算出模板的每一维特征的概率权重;

d)根据概率权重对特征进行排序,选择权重最大的前 k 个特征以及相应的子块,用于更新前景字典 $\boldsymbol{\Phi}_{\text{target}}$ 和模板 $\boldsymbol{T}$;

e)以 t 时刻的跟踪结果 $\boldsymbol{\tau}_t$ 为起始状态,用 LK 图像配准算法迭代地求解出目标在 $t+1$ 时刻的状态 $\boldsymbol{\tau}_{t+1}$;

f)输出跟踪结果 $\boldsymbol{\Phi}_{t+1}=[x_{t+1},y_{t+1},\theta_{t+1},S_{t+1}]^{\mathrm{T}}$。

1.1 特征选择

稀疏表示在计算机视觉领域已经取得显著成绩[8~13]。稀疏表示基本思想是从超完备字典 $\boldsymbol{\Phi}$ 中,选择具有最佳线性组合的若干原子来表示观察信号 $\boldsymbol{y}$,实际上是一种逼近过程,形式描述如下:

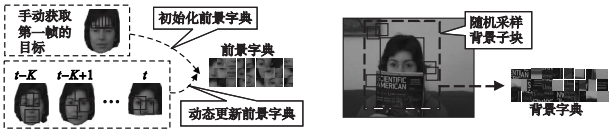
$$\hat{\boldsymbol{a}}=\underset{\boldsymbol{a}}{\operatorname{argmin}}\|\boldsymbol{\Phi}\boldsymbol{a}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{a}\|_1 \tag{1}$$

其中:λ 是正则化系数,它决定了表达式的稀疏性;$\hat{\boldsymbol{a}}$ 为所求的表示系数。

1.1.1 字典的构造

在第一帧手动截取目标并归一化到模板尺度,然后按 $m\times n$ 的大小等间隔地采样一列子块 $\boldsymbol{p}^i_{\text{target}}\in\mathbb{R}^{l\times 1}$,$i=1:M$,其中 $l=m\times n$ 为子块的维数。将这些子块拉成一维向量后就成为构造前景字典 $\boldsymbol{\Phi}_{\text{target}}\in\mathbb{R}^{l\times M}$ 的原子。由于目标的外观会不断变化,因此前景字典会随着跟踪结果不断动态更新,如图 1(a)所示。

类似地,在目标相邻的周围,随机采样一系列子块 $\boldsymbol{p}^i_{\text{bac}}\in\mathbb{R}^{l\times 1}$,$i=1:N$,背景字典 $\boldsymbol{\Phi}_{\text{bac}}\in\mathbb{R}^{l\times N}$ 由这些子块构成。因为背景的变化远比目标的外观变化迅速,因此在每一帧得到跟踪结果之后,都需要重新构造背景字典,如图 1(b)所示。



(a) 前景字典的初始化以及更新 (b) 构造背景字典

图1 字典构造

1.1.2 基于稀疏表示的特征选择

本文用一个归一化尺度的模板 $\boldsymbol{T}\in\mathbb{R}^{d\times 1}$ 来表示目标的外观,其中 d 为模板的特征维数。在得到跟踪结果后,首先将目标归一化到模板尺度后,然后以每个像素为中心采样子块 $\boldsymbol{p}^i\in\mathbb{R}^{l\times 1}$,$i=1:d$。按式(1)求出这些子块在字典中的最稀疏表示系数。图 2 展示了子块在两个字典所求出的稀疏表示系数。对比发现,与背景区分度高的子块,在前景字典中得到的稀疏解,其非零项只集中分布在少数原子上。同时,该子块在背景字典中的稀疏解,其非零项则广泛分布在多个原子上。而那些受到干扰的特征,其子块的稀疏解则具有相反的特性。本文按式(2)计算出第 i 维特征的概率权重为

$$\rho_i=\frac{1}{1+\mathrm{e}^{(\varepsilon^i_{\text{target}}-\varepsilon^i_{\text{bac}})}} \tag{2}$$

其中:$\varepsilon^i_{\text{target}}=\|\boldsymbol{p}^i-\hat{\boldsymbol{a}}^I_{\text{target}}\|_2$ 为子块在前景字典的重建误差,$I=\underset{j}{\operatorname{argmax}}(\hat{\boldsymbol{a}}^j_{\text{target}})$,$\hat{\boldsymbol{a}}_{\text{target}}=[\hat{a}^1_{\text{target}},\hat{a}^2_{\text{target}},\cdots,\hat{a}^M_{\text{target}}]^{\mathrm{T}}$ 为观察子块在前景字典的稀疏解;$\boldsymbol{\Phi}^I_{\text{target}}$ 为前景字典的第 I 个原子。同样地,对 $\varepsilon^i_{\text{bca}}$ 也有相似定义。如果 ρ_i 为 1,表示该特征区分目标和背景的能力最好,ρ_i 为 0 则表示很可能是一项干扰,该特征将会被屏蔽。ρ_i 同时表示子块 p^i 的权重。

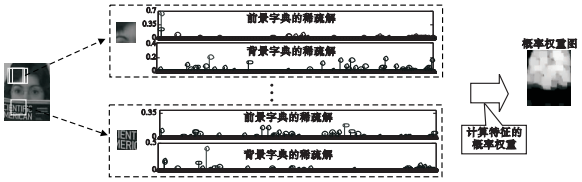


图2 特征选择

1.2 字典和模板的更新

为了应对目标外观的变化,在跟踪结束时需要对前景字典和模板进行实时更新。在跟踪过程中,纹理信息越丰富的特征,对结果影响就越大,因此前景字典应该尽可能保留纹理信息丰富的子块。本文用子块水平、垂直梯度的绝对值之和来表示该子块的纹理信息量。在特征选择阶段所提取的一系列子块 $\boldsymbol{p}^i(i=1:d)$,首先按概率权重对这些子块进行排序,并从中筛选出权重最高的前 k 个子块,然后用筛选出来的子块替换前景字典中纹理信息量最少的前 k 个子块。这样在每一帧更新后,字典的规模都不会变化。同时,将跟踪结果中概率权重最高的前 k 维特征相应地更新到模板中,从而实现模板的更新。这种更新机制不仅能较好地表示最近一段时间目标的外观,同时避免引入干扰而导致的漂移问题。

1.3 跟踪

本节将详细介绍一种基于稀疏表示的 LK 图像配准算法。以前一帧的跟踪结果作为当前帧的初始点,用 LK 算法对目标进行跟踪。

1.3.1 LK 图像配准算法

LK 算法属于一种模板匹配方法[14],通过最小化式(3),迭代地求出目标状态的最优解。

$$\sum_i[Y(W(x_i;\tau))+\nabla Y\frac{\partial W}{\partial\tau}\Delta\tau-T(\boldsymbol{x}_i)]^2 \tag{3}$$

其中:Y 是当前帧的观测区域;$\boldsymbol{x}_i=(x_i,y_i)^{\mathrm{T}}$ 为模板的像素坐标;$\tau=[x,y,\theta,S]^{\mathrm{T}}$ 是一组描述跟踪目标状态的仿射参数;$W(\boldsymbol{x}_i;\tau)$ 为一仿射变换函数,该函数由参数 τ 的决定,实现对下标 x 的仿射变换。最小化式(3)实际上是一个最小二乘问题。通过计算式(3)关于 τ 的偏导数并让其等于 0,最终可求得增量 $\Delta\tau$ 为

$$\Delta\tau=H^{-1}\sum_i[\nabla Y\frac{\partial W}{\partial\tau}]^{\mathrm{T}}[T(\boldsymbol{x}_i)-Y(W(\boldsymbol{x}_i;\tau))] \tag{4}$$

其中,$H=\sum_i[\nabla Y\frac{\partial W}{\partial\tau}]^{\mathrm{T}}[\Delta Y\frac{\partial W}{\partial\tau}]$ 为 Hessian 矩阵。

1.3.2 鲁棒的 LK 图像配准算法

一旦目标发生局部遮挡或者存在相似背景的干扰,LK 算法就不能准确求解出目标状态。为了减少上述因素对 LK 配准结果影响,本文将式(3)改写成

$$\sum_i\rho_i[Y(W(\boldsymbol{x}_i;\tau))+\nabla Y\frac{\partial W}{\partial\tau}\Delta\tau-T(\boldsymbol{x}_i)]^2 \tag{5}$$

其中,概率权重 ρ_i 按式(2)计算得到。同样地,通过最小化式(5),可以迭代地求得增量 $\Delta\tau$:

$$\Delta\tau=H_\rho^{-1}\sum_i\rho_i[\nabla Y\frac{\partial W}{\partial\tau}]^{\mathrm{T}}[T(\boldsymbol{x}_i)-Y(W(\boldsymbol{x}_i;\tau))] \tag{6}$$

其中:$H_\rho = \sum_i \rho_i [\nabla Y \frac{\partial W}{\partial \tau}]^T [\Delta Y \frac{\partial W}{\partial \tau}]$。因为干扰的概率权重一般都很低,所以这种带权重的 LK 配准算法能极大降低干扰对配准结果的影响,从而保证了跟踪结果精度。

2　实验对比

为了说明本算法长时间跟踪的鲁棒性以及应对各种挑战的综合能力,本文将用几个富有挑战性的视频进行实验。表 1 归纳了这些视频的挑战,如视角变化、尺度变化已经遮挡等。本文将与文献[5]中提出的 IVT 算法和文献[6]中提出稀疏跟踪算法(简称 L1)进行对比。

视频 Sylvester 共有 1 344 帧,目前大部分算法只能跟到 600 帧左右,而本文的算法则可以完整跟踪整个视频,结果如图 3 所示。即使在 922、1001、1282 帧这些极端的视角变化下,本文提出的算法依然能十分准确地跟踪到目标。从 690 ~ 711 帧,目标迅速地靠近光源,在视角和光照的双重挑战下,本文提出的算法仍然能够准确地定位目标。在 610 帧之前,IVT 都能很好地跟踪目标。第 610 帧,由于目标突然翻转,同时光照也发生变化,外观的剧烈变化导致 IVT 无法找到与模板匹配的结果。在 613 帧前,L1 都能跟踪到目标,但从 300 帧之后,L1 的跟踪效果开始变得不精准。

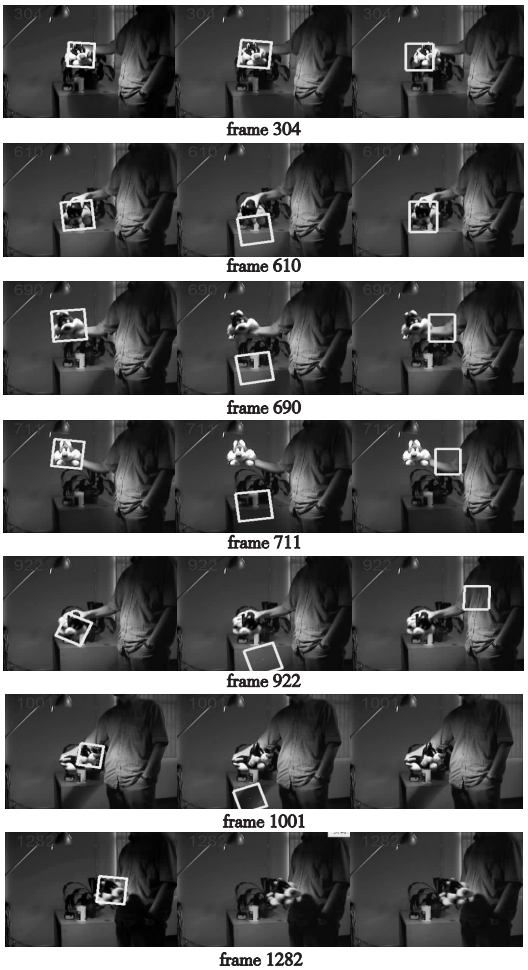


图3　本文算法(第一列)、IVT(第二列)和L1(第三列)对Sylvester视频的跟踪结果

对于视频 David,虽然三种算法都能跟踪完整个视频,但是在 152 ~ 181 帧之间,IVT 和 L1 的跟踪精度都明显出现下降。这是因为在这段时间里,目标外观同时发生了视角变化、尺度变化以及由于摄像机抖动引起的画面模糊。由于 IVT 和 L1 在整个过程都不加区分地进行全局更新,使得两种算法在 197 帧都出现了漂移。而本文提出的算法在整个跟踪过程始终表现得十分稳定,跟踪结果如图 4 所示。

图4　本文算法(第一列)、IVT(第二列)和L1(第三列)对David视频的跟踪结果

表 1　本文的实验视屏

视频	主要挑战
Sylvester	极端的视角变化,快速运动,光照变化
David	极端的尺度变化,极端的视角变化,光照变化,局部遮挡
faceocc	长时间的大面积局部遮挡
girl	大面积局部遮挡,同类目标干扰

视频 girl 和 faceocc 都会出现大面积的局部遮挡,其中 faceocc 的特点是遮挡持续的时间很长,而 girl 则是受到外观相似的同类目标遮挡。在 faceocc 视频中,随着遮挡面积的逐渐增大,IVT 变得相当不稳定,其跟踪精度会明显变差,而且渐渐出现了漂移现象。而对于 girl 视频,当目标受到同类的遮挡后,IVT 会把干扰误认为目标进行跟踪。这是因为同类目标的 PCA 特征存在极大的相似性,所以 IVT 会误把遮挡当成目标进行跟踪。而本文提出的算法和 L1 在应对局部遮挡时,都表现出很强的鲁棒性,跟踪结果如图 5(a)和(b)所示。值得注意的是,即使目标已经几乎被完全遮挡,L1 依然会对此跟踪结果进行学习,因此必然会加入了大量的干扰,这干扰很可能会导致漂移。

在跟踪阶段,当目标发生极端视角变化(如视频 David、Sylvester)或受到遮挡干扰(如视频 girl 和 faceocc)时,目标的全局外观会发生明显变化,这时候仅有部分外观与模板保持一致。由于 IVT 和 L1 都是寻找全局最匹配的结果,因此发生上述情况时,这两种算法的精度都会下降。在本算法中,通过特征选择使干扰的权重很小,因此干扰对跟踪精度的影响被大大降低,而区分度高的特征都会得到较高的概率权重,这些可靠

的特征保证了跟踪的精度,所以本算法有效克服干扰造成的影响。在更新阶段,小权重的特征和子块都不会被加入到模板和前景字典中,因此相对 IVT 和 L1 这两种全局更新算法,本文算法能更好地避免漂移。

为了说明本文算法的跟踪精度,对上述视频进行了人工标定。通过计算每种算法跟踪结果和手动标定结果之间的误差来比较三种算法的精度。结果如图 6 所示(见电子版)。可以看出,在应对视角变化方面,IVT 的表现比 L1 要好,而 L1 应对遮挡的能力则十分出色。

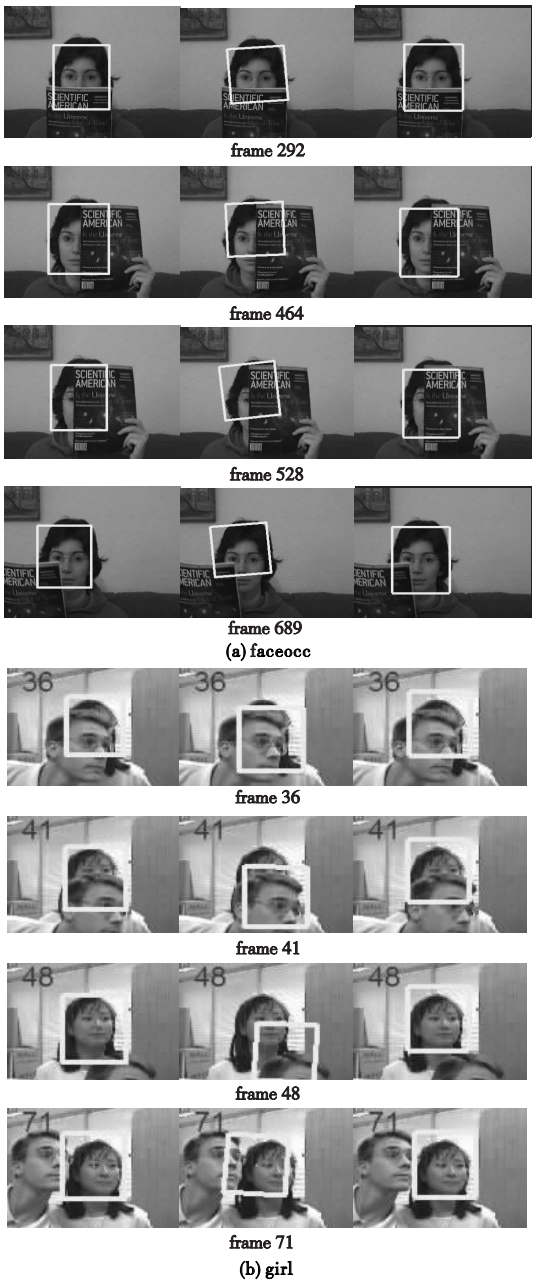


图5 本文算法(第一列)、IVT(第二列)和L1(第三列)对视频faceocc和girl的跟踪结果

3 结束语

大量实验表明,本文提出的算法能很好地应对视觉跟踪中的各种挑战环境。本算法的跟踪精度上都明显优于粒子滤波框架下的稀疏跟踪器。算法中引入的在线特征选择机制发挥了很强的作用,既能处理目标外观的变化,同时能避免漂移问题。可是目标一旦发生全局遮挡或者从视频中消失后再出现,本文算法将无法应对。在今后的工作中,希望能够加入在线检测机制,从而提高跟踪算法的性能。

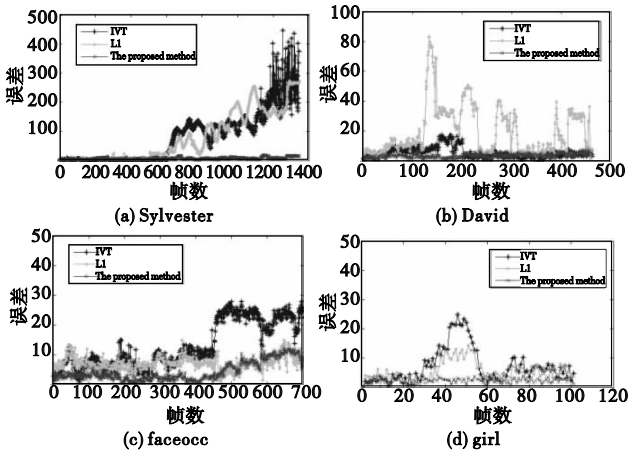


图6 三种算法对实验视频的跟踪误差

参考文献:

[1] YILMAZ A, JAVED O, SHAH M. Object tracking: a survey[J]. ACM Computing Surveys, 2006, 38(4): 13.

[2] GRABNER H, GRABNER M, BISCHOF H. Real-time tracking via on-line boosting[C]//Proc of British Machine Vision Conference. 2006: 47-56.

[3] GRABNER H, LEISTNER C, BISCHOF H. Semi-supervised on-line boosting for robust tracking[C]//Proc of European Conference on Computer Vision. 2008: 234-247.

[4] STALDER S, GRABNER H, Van GOOL L. Beyond semi-supervised tracking: tracking should be as simple as detection, but not simpler than recognition[C]//Proc of the 12th IEEE International Conference on Computer Vision. 2009: 1409-1416.

[5] ROSS D, LIM J, LIN R S, *et al*. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125-141.

[6] MEI Xue, LING Hai-bing. Robust visual tracking and vehicle classification via sparse representation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2259-2272.

[7] MEI Xue, LING Hai-bing. Robust visual tracking using L1 minimization[C]//Proc of the 12th IEEE International Conference on Computer Vision. 2009: 1436-1443.

[8] WRIGH J, YI M, MAIRAL J, *et al*. Sparse representation for computer vision and pattern recognition[J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044

[9] WANG Jin-jun, YANG Jian-chao, YU Kai, *et al*. Locality-constrained linear coding for image classification[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3360-3367.

[10] LING Hai-bin, BAI Li, BLASCH E, *et al*. Robust infrared vehicle tracking across target pose change using L1 regularization[C]//Proc of the 13th Conference on Information Fusion. 2010: 1-8

[11] YANG Meng, ZHANG Lei, YANG Jian, *et al*. Robust sparse coding for face recognition[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2011: 625-632.

[12] WAGNER A, WRIGHT J, GANESH A, *et al*. Toward a practical face recognition system: robust alignment and illumination by sparse representation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34(2): 372-386.

[13] DONG Wei-sheng, ZHANG Lei, SHI Guang-ming, *et al*. Nonlocally centralized sparse representation for image restoration[J]. IEEE Trans on Image Processing, 2013, 22(4): 1620-1630.

[14] BAKER S, MATTHEWS I. Lucas-kanade 20 years on: a unifying framework[J]. International Journal of Computer Vision, 2004, 56(3): 221-255.