

文章编号:1007-2780(2018)01-0092-07

基于 Vibe 和时空上下文的运动手势跟踪算法

王 民,石新源*,王稚慧,李泽洋

(西安建筑科技大学 信息与控制工程学院, 陕西 西安 710055)

摘要:针对时空上下文算法易发生漂移以及在目标跟踪丢失后不能重新找回目标的问题,提出了一种融合 Vibe 前景检测和时空上下文的运动手势跟踪算法。首先使用时空上下文算法对手势预估计并进行干扰检测,当检测到干扰发生时,使用 Vibe 算法对时空上下文算法的预估计结果进行校准,并更新目标模型。该方法的优势在于,采用无参数模型的 Vibe 算法校准手势跟踪全过程。实验采用重叠度成功率和中心偏差作为评价体系,实验结果表明,改进算法比原算法跟踪成功率提高 60%。该方法增强了运动手势跟踪效果,提高了时空上下文算法的鲁棒性。

关 键 词:时空上下文;Vibe;手势跟踪;干扰检测;鲁棒性

中图分类号:TP391.41 **文献标识码:**A **doi:**10.3788/YJYXS20183301.0092

Moving gesture tracking algorithm based on vibe and space-time context

WANG Min, SHI Xin-yuan*, WANG Zhi-hui, LI Ze-yang

(School of Information and Control Engineering,
Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Aiming at the problem that the space-time context algorithm is prone to drift and can not retrieve the target after target tracking is lost, a moving gesture tracking algorithm combining Vibe foreground detection and space-time context is proposed. Firstly, the space-time context algorithm is used to estimate the gesture and detect the interference. When the interference is detected, the Vibe algorithm is used to calibrate the estimated results of the space-time context algorithm and update the target model. The advantage of this method is that the Vibe algorithm with no parameter model is used to calibrate the whole process of gesture tracking. The experiment adopts overlap success rate and center location errors as the evaluation system. The experimental results show that the improved algorithm is 60% higher than the original algorithm in tracking success rate. This method enhances

收稿日期:2017-06-16; **修订日期:**2017-08-28.

基金项目:住房和城乡建设部科学技术项目计划(No.2016-R2-045);陕西省教育厅专项基金(No.2013JK1081);陕西省科学技术研究发展计划项目(No.CXY1122(2));陕西省自然科学基金青年基金(No.2013JQ8003)

Supported by Ministry of Housing and Urban-Rural Development Science and Technology Project plan (No.2016-R2-045); Shaanxi Provincial Department of Education special fund(No.2013JK1081); Shaanxi Province Science and Technology Research and Development Project(No.CXY1122(2)); Shaanxi Provincial Natural Science Foundation Youth Fund(No.2013JQ8003)

* 通信联系人, E-mail: wangmin1329@163.com

the motion gesture tracking effect and improves the robustness of the space-time context algorithm.

Key words: space-time context; Vibe; gesture tracking; interference detection; robustness

1 引言

手语是听障者通过手势模拟不同形象或者音节来进行交流的语言^[1]。对于听障者来说,手语是主要的交际工具。运动手势识别则是通过检测跟踪视频序列中手势的运动变化来识别手语内容,并以此实现人机交互的一项技术。目前运动手势识别技术在远程控制、手语教学、互动游戏等方面有着较好的应用前景^[2]。而手势跟踪的高鲁棒性和准确性直接影响着运动手势识别的精准度,因此可通过提高手势跟踪效果来实现运动手势的快速、准确识别^[3]。

隐马尔可夫模型^[4]和动态时间规整^[5]是目前运动手势跟踪算法中比较常见的方法,其实质在于对时间轴的控制和把握^[6]。上述方法虽然抗干扰能力强,鲁棒性好,但运算量大,实时性不高,随着跟踪误差的不断引入或者目标形态的持续变化,单纯基于跟踪和检测来实现长时跟踪都不现实,最终会导致跟踪失败^[7]。时空上下文^[8](STC)算法完整采纳了目标周围背景空间上下文信息,并将其有效利用,同时配合FFT进行加速,具有较好的实时性和准确性。因此,可以采用STC算法来跟踪运动手势,但STC算法在目标模型更新时不验证更新样本的有效性,从而会导致跟踪失败。对于STC算法存在的问题,本文拟引入Vibe算法作为STC算法的后备校正算法,即当视频手势快速运动或出现运动模糊的情况下,Vibe算法能及时校正运动手势并进行模型更新,从而达到准确跟踪目标手势的效果。

2 时空上下文算法

STC算法建立在贝叶斯框架下,用式(1)定义了一个估计目标位置 x 似然的置信图:

$$c(x) = P(x | o) = \sum_{c(z) \in X^c} P(x, c(z) | o) = \sum_{c(z) \in X^c} P(x | c(z), o) P(c(z) | o), \quad (1)$$

式中: $c(\cdot)$ 是目标置信图, X^c 是时空上下文特征集合, o 表示跟踪目标出现在当前帧中, $c(x)$ 最大的那个位置 x^* 就是目标的位置。从公式(1)可以看到,似然函数可以分解为两个概率部分。一个是建立局部上下文各个点 x 的上下文先验概率 $P(c(x) | o)$ 。另一个是建立周围上下文信息的空间关系的条件概率 $P(x | c(z), o)$,而条件概率 $P(x | c(z), o)$,也是空间上下文和目标位置的对应关系。

STC算法步骤为:

在 t 帧时刻,根据该帧图像和得到的目标位置 x^* 。顺序进行以下计算:

学习空间上下文模型:

$$h^{sc}(X) = F^{-1} \left(\frac{F(b e^{-|\frac{x-x^*}{\sigma}|^\beta})}{F(I(X)\omega_\sigma(X-X^*))} \right). \quad (2)$$

视频帧数更新后,跟踪下一帧待跟踪目标需要的时空上下文模型:

$$H_{t+1}^{stc} = (1-p)H_t^{stc} + p h_t^{sc}. \quad (3)$$

更新尺度等参数:

$$\begin{cases} S'_t = \sqrt{\frac{c_t(X'_t)}{C_{t-1}(X_{t-1}^*)}} \\ \bar{S} = \frac{1}{n} \sum_{i=1}^n S'_{t-i} \\ S_{t+1} = (1-\lambda)_{st} + \lambda \bar{S}_t \\ \sigma_{t+1} = S_t \sigma_t \end{cases}, \quad (4)$$

在 $t+1$ 帧,计算置信图:

$$C_{t+1}(X) = F^{-1}(F(H_{t+1}^{stc}(X)) \otimes F(I_{t+1}(X)\omega_{\sigma_t}(X-X_t^*))), \quad (5)$$

找到最大值,其位置也就是最终目标确定位置:

$$X_{t+1}^* = \arg \max C_{t+1}(X) \quad x \in \Omega_c(X_t^*). \quad (6)$$

式(3)中 p 为学习因子。学习因子用来减小短时干扰对跟踪目标模型的影响、减少图像噪声。可是在持续干扰中,该方法也存在一定问题,会把偏差的目标信息更新到模型中,从而发生漂移情况。其根本原因是STC算法对每一帧的输出结果都会进行目标模型更新。STC算法在Praise测试序列的部分运行结果如图1所示。

Praise测试序列的跟踪目标为运动的拳头部

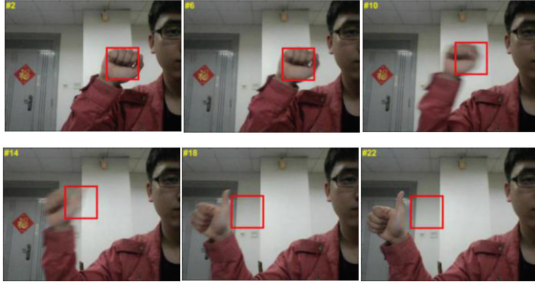


图 1 时空上下文跟踪算法的漂移问题

Fig.1 Drifting problem of STC algorithm

分。目标在运动过程中以正常的手速表达赞扬的手势含义。图 1 演示了赞扬手势出现漂移的全过程。在目标快速移动(第 10 帧)和目标运动初始(第 2 帧)之间算法都可以有效跟踪目标,但(第 10 帧)之后由于手势运动后期速度加快,跟踪出现了漂移现象。STC 算法可以通过目标邻域背景的情况最终明确目标位置,但 STC 算法是通过搜寻极值的形式没有对目标输出结果进行有效性的检验从而会导致跟踪失败。

根据以上分析,STC 算法的相关性目标模型会出现错误更新,也就是无法在干扰(快速运动、运动模糊)结束后重新确定跟踪目标位置。这个问题会严重影响算法的可靠性。因此本文引入 Vibe 算法干预以上问题。Vibe 算法是一种内存占用较少,计算复杂度较低的一种像素级背景建模算法。最重要的是其基本符合运动场景的情况,采用无参数模型,能在持续更新中有效地识别目标,从而弥补原算法相关性目标模型缺陷的问题。同时 Vibe 算法也能用来评估 STC 算法输出目标跟踪结果的有效性,以及为 STC 算法在跟踪失败后重新锁定目标位置。

3 融合 Vibe 的时空上下文跟踪算法

针对 STC 算法在目标手势运动中容易被错误更新的问题,本文提出一种将 Vibe 与 STC 相互融合的算法。该算法在 STC 算法的基础上,结合了 Vibe 算法作为修正算法。通过研究发现:STC 模型对运动手势目标和背景的信息依赖程度较高,而对手势目标自身特征建模却存在不足。因此,当干扰发生时,跟踪结果就可能出现偏差,模型不能够有效的跟踪目标^[9]。故本文首先需要

对 STC 算法输出结果的有效性进行判断,一旦跟踪结果失效,Vibe 算法立即对目标进行检测,并重新更新 STC 模型。

改进算法 STC-Vibe 的形式化描述如图(2)所示。

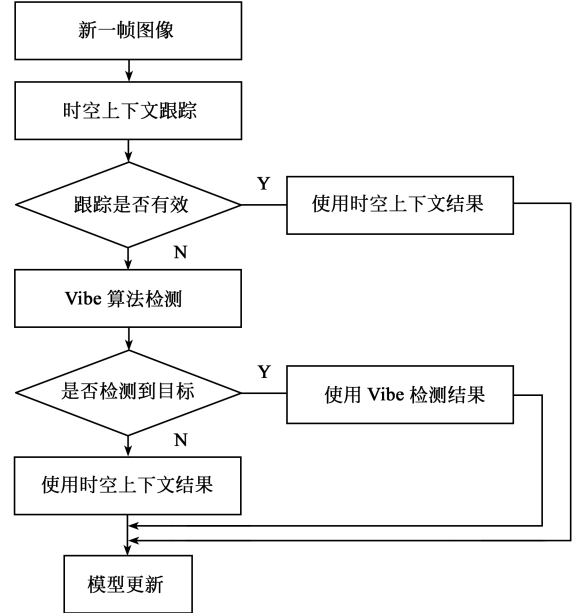


图 2 改进算法形式化框图

Fig.2 Formalized block diagram of proposed algorithm

由图 2 可知,STC-Vibe 算法进行目标跟踪时,首先对每一帧输入图像进行相关目标预估计,同时进行干扰检测评估。如果评估有效,就可以作为算法的跟踪输出结果;如果评估未达到标准,那么 STC 算法目标跟踪出现较大偏差,此时就要使用 Vibe 算法对目标进行搜索。当检索到目标时,通过检索结果重新初始化 STC 模型。如果未检索到目标,则仍然使用 STC 算法的跟踪结果作为输出。

3.1 强干扰检测

文献[10]KMCT 算法中判定目标是否被严重遮挡主要采用分类器判别度的变化趋势,对应公式为:

$$\frac{H^{k-1} - H^k}{H^{k-1}} > \xi, \quad (7)$$

其中: H^{k-1} 表示已经得知第 $k-1$ 帧的判别度,同理 H^k 表示预测的当前帧的判别度, ξ 是设定的阈值,为 30%。该公式表示如果待跟踪的第 k 帧的判别度相比于第 $k-1$ 帧的判别度下降幅度超过了 30%的阈值,即认为发生了严重遮挡。本文

基于公式(7)加以改进采用置信度作为检测基准,使其不仅可以对遮挡检测,而且可以作为干扰的通用检测。改进后公式为:

$$\frac{C^{k-1} - C^k}{C^{k-1}} > \xi, \quad (8)$$

其中: C^{k-1} 为 STC 对第 $k-1$ 帧估计得到的置信度的平均值, C^k 表示 STC 对第 k 帧进行估计得到的置信度, ξ 是设定阈值,为 20%。公式(8)表示如果待跟踪的第 k 帧的置信度相比于第 $k-1$ 帧的置信度下降幅度超过 20% 的阈值,即认为发生了强干扰。本文即使用公式(8)用于干扰情况的检测。

3.2 手势目标修正

在 STC 跟踪结果出现较大偏差时,使用 Vibe 算法进行检索目标,其检索结果用来重新更新目标偏差的跟踪结果。Vibe(视觉背景提取算法),用来实现运动目标检测。算法首次即对单帧图像初始化建模,随机化引入背景模型中,随机跟进背景模型。

Vibe 对视频第一帧图像的每一个像素点建立一个包含 N 个样本的像素模型。随机选取每个像素的 8 个邻域中某个像素值来初始化背景模型中该像素的值。根据相邻位置的像素值具有相同空间分布特性。定义 $v(x)$ 为 RGB 彩色空间中 x 处的像素值, v_i 为选取的样本,则像素 $v(x)$ 的背景模型记为:

$$M(x) = \{v_1, v_2, \dots, v_N\}, \quad (9)$$

定义一个圆域 $S_R(v(x))$, 如图 3 所示,中心为像素 $v(x)$, 半径为 R , 统计落在圆域内的样本的个数,记为 $\#$, 给定一个阈值 $\# \min$, 若满足:

$$\# \{S_R(v(x)) \cap \{v_1, v_2, \dots, v_N\}\} \geq \# \min, \quad (10)$$

且在 RGB 通道上,样本像素的色彩差满足公式(3)时,则将 $v(x)$ 分类为背景。如果不是那就是前景。

$$\{|v^c(x) - v_i^c| \leq R, i = 1, \dots, N\}, \quad (11)$$

其中, c 代表 R, G, B 这 3 个通道。

3.3 更新策略

STC 模型、Vibe 模型进行在线更新,主要是为了保证目标的可靠性。STC 算法在手势目标受到干扰之后继续使用错误结果更新模型,导致发生漂移不能改正。本文在干扰发生之后通过 Vibe 模型重新检测目标并更新模型,从而解决

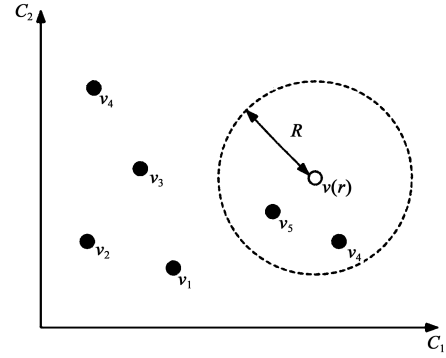


图 3 Vibe 背景模型

Fig.3 Background model of Vibe

STC 算法在干扰结束后不能够恢复跟踪的问题。Vibe 算法采取保守更新策略,在视频图像每一帧中,如果 STC 算法跟踪结果有效,目标处于未遮挡状态,可把结果用来更新 Vibe 模型。如果 STC 跟踪结果无效,那么表示目标处于干扰中,此时对目标进行搜索需要用 Vibe 算法。如果检索到目标,那么 STC 模型应使用检索目标进行更新。

3.4 算法流程

(1)初始化跟踪目标,手动选取目标。

(2)使用 STC 算法对待跟踪的第 t 帧图像的矩阵数据进行预估计。

(3)将预估计置信度 C 代入公式(8)中计算是否发生了漂移,若未发生漂移,则转入(4),否则转入(5)。

(4)STC 算法预跟踪结果可靠,采纳 STC 的预跟踪结果作为最终结果,并对 Vibe 模型进行更新。

(5)STC 算法预估计过程中检测到漂移现象,此时启用 Vibe 对结果进行校准,若 Vibe 检测到跟踪目标,则使用 Vibe 检测结果重新初始化 STC 模型如果没有检测到跟踪目标,那么应该用 STC 结果作为输出。

(6)跟踪结束。

4 实验结果与分析

实验所用电脑为 Intel Core i5-2450 @ 3.3 GHz CPU 与 4G 内存,操作系统为 Windows10,软件为 Matlab 2015a。测试使用数据来自自行采集 5 种手势样本(Praise、Sorry、Watch news、

Myself、Dizzy),除了 STC 算法以外,其他对照跟踪算法包括 CPF^[11], VTS^[12] 和 TLD^[13], OAB^[14]。在上文的 STC 模型中, ρ 是学习因子,目的是对模型更新进行控制,为了降低图像噪声对模型产生的影响, ρ 应该取较小的值。本文中与原 STC 算法中一致, ρ 取 0.075。为对算法进行比较说明,用重叠度成功率(Overlap Success Rate,OSR)和中心偏差(Center Location

Error,CLE)作为评价体系。前者主要在计算算法输出的目标区域与真值区域的重合部分使用较多,后者主要在真值区域中心点的像素距离计算与计算算法输出的目标区域中心点使用的较多。计算重叠度成功率时,重叠度大于 0.5 就说明该次跟踪结果有效,跟踪成功率如表 1 所示。表 2 为中心偏差结果。其中 Ours 为本文所用算法。

表 1 跟踪成功率
Tab.1 Tracking success rate

测试序列	CPF	VTS	TLD	OAB	STC	Ours
Praise	0.51	0.69	0.65	0.77	0.49	0.93
Sorry	0.49	0.63	0.71	0.70	0.61	0.79
Watch news	0.53	0.74	0.69	0.73	0.56	0.83
Myself	0.44	0.68	0.61	0.69	0.48	0.78
Fever	0.38	0.56	0.59	0.66	0.41	0.77
平均	0.47	0.66	0.64	0.71	0.51	0.82

表 2 中心偏差
Tab.2 Center location errors

测试序列	CPF	VTS	TLD	OAB	STC	Ours
Praise	40	22	18	12	31	5
Sorry	51	28	32	18	38	8
Watch news	43	31	37	11	29	7
Myself	44	26	20	15	33	9
Fever	62	43	28	24	44	11
平均	48	30	27	16	35	8

由结果看出:在精度方面,改进算法和 STC 算法相差不多,和其它跟踪算法相比较而言,也是存在较大的优势。改进算法对数据集进行测试,得到的平均跟踪成功率为 82%,在和 STC 算法进行比较之后,提升近 60%。

图 4 展示了选用 VTS 算法和 OAB 算法、STC 算法作为对照算法的跟踪结果,以下是部分视频序列跟踪结果对比图。

图 4(a)是 Fever 序列上本文 STC-Vibe 算法与 STC 算法的跟踪情况。Fever 序列跟踪的目标是手语视频我正在发烧的手掌部分。由图可见,在手势运动初始阶段(第 10 帧),算法都较好地跟踪到

手势运动中手掌的位置。但是随着视频帧数的跟进,第 35 帧时,本文算法的跟踪效果较好,但是跟踪结果出现了偏差。手势运动后期(第 60 帧)STC 算法出现漂移现象且不能恢复,本文算法引入 Vibe 检测算法辅助更新目标模型能在跟踪出现较大偏差时重新检测目标并恢复更新,较好地跟踪了手势运动的轨迹。

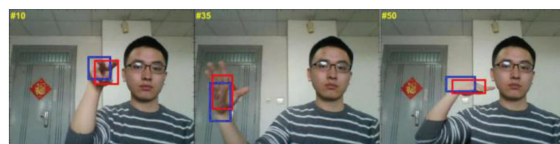
图 4(b)为 VTS 算法和本文算法在 Watch news 序列上的表现。Watch news 序列跟踪目标是手语视频看新闻的手掌部分。在视频序列中 STC-Vibe 算法和 VTS 都可以较好的对手势运动轨迹进行跟踪。从实验数据可以看出,VTS 跟踪成功

率 0.74 很接近本文算法 0.83,但中心偏差较大。

图 4(c)为 OAB 算法和本文算法在 Sorry 序列上的表现。Sorry 序列跟踪目标是手语视频对不起的手掌部分。视频序列中,在跟踪成功率方面 STC-Vibe 算法和 OAB 效率都比较高,有着较好的跟踪效果。



(a)发烧序列(红:改进算法;绿:时空上下文)
(a) Fever sequence (Red: Improved algorithm; Green: STC)



(b)看新闻序列(红:改进算法;蓝:VTS)
(b) Watch news sequence (Red: Improved algorithm; Blue: VTS)



(c)对不起序列(红:改进算法;黄:OAB)
(c) Sorry sequence (Red: Improved algorithm; Yellow: OAB)

图 4 部分算法跟踪结果

Fig.4 Some algorithms track results

如表 3 所示,算法平均运行速度方面 STC 算法主要使用图像序列测试,平均运行速度为 95.1 021 fps, TLD 算法为 17.462 4 fps, STC-Vibe 算法为 24.267 3 fps。

表 3 算法平均运行速度

Tab.3 Algorithm average running speed

算法名称	STC	TLD	STC-Vibe
平均运行速度	95.102 1 fps	17.462 4 fps	24.267 3 fps

本文改进算法在跟踪精度方面有着较大的进步,尤其是在手势运动过程中有着明显的优势,但是和传统 STC 算法速度相比有所降低。同时在精度以及速度上,改进算法和 TLD 算法效率都是明显较高的。

5 结 论

本文对运动手势进行跟踪识别,提出了一种把 Vibe 前景检测和时空上下文算法相融合的跟踪算法。本文实验表明:和传统时空上下文算法相比本文提出的方法在解决手势目标漂移问题上表现较好。在实验数据集上,改进算法跟踪成功率为 82%,比原算法提高 60%,能快速、准确地对动态手势进行跟踪识别,基本满足人机交互的需求。此外为了进一步提高跟踪准确率、实现更稳定的手势跟踪识别、更好地应用于人机交互中,在以后的实践研究中会采用多种摄像头相结合的方式。

参 考 文 献:

- [1] 伍玉霞.基于深度图像信息的手势识别技术研究与应用[D].沈阳:东北大学,2013.
WU Y X. Research and application of hand gesture recognition technology based on depth image information [D]. Shenyang: Northeastern University, 2013. (in Chinese)
- [2] 史东承,倪康.基于 YCbCr 颜色空间背景建模与手势阴影消除[J].中国光学,2015,8(4):589-595.
SHI D C, NI K. Background modeling based on YCbCr color space and gesture shadow elimination [J]. *Chinese Optics*, 2015, 8(4): 589-595. (in Chinese)
- [3] 王彩虹,沈燕飞,王毅,等.基于时空上下文的手势跟踪与识别[J].计算机工程与应用,2016,52(9):202-207.
WANG C H, SHEN Y F, WANG Y, et al. Gesture tracking and recognition based on spatio-temporal context[J]. *Computer Engineering and Applications*, 2016, 52(9): 202-207. (in Chinese)
- [4] STARNER T, PENTLAND A. *Real-time American Sign Language Recognition From Video Using Hidden Markov Models* [M]//SHAH M, JAIN R. *Motion-Based Recognition*. Dordrecht, Netherlands: Springer, 1997: 227-243.
- [5] CORRIGINI A. Dynamic time warping for off-line recognition of a small gesture vocabulary [C]//*Proceedings of*

- IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-time Systems*, Washington, DC, USA: IEEE, 2001: 82-89.
- [6] 高喆.多重空间特征融合的手势识别[J].小型微型计算机系统,2016,37(7):1577-1582.
GAO Z. Hand gesture recognition using multiple spatial features fusion [J]. *Journal of Chinese Computer Systems*, 2016, 37(7): 1577-1582. (in Chinese)
- [7] 孙保基,张葆,宋策,等.基于角点增强改进的 TLD 目标跟踪算法[J].液晶与显示,2016,31(9):921-928.
SUN B J, ZHANG B, SONG C, *et al.* Improved TLD object tracking algorithm based on corner reinforced [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(9): 921-928. (in Chinese)
- [8] ZHANG K H, ZHANG L, LIU Q S, *et al.* Fast Visual Tracking Via Dense Spatio-temporal Context Learning [M]//FLEET D, PAJDLA T, SCHIELE B, *et al.* *Computer Vision-ECCV 2014*. Cham: Springer, 2014: 127-141.
- [9] 张雷,于凤芹.基于置信图特性的改进时空上下文目标跟踪[J].计算机工程,2016,42(8):277-281,288.
ZHANG L, YU F Q. Improved object tracking via spatial-temporal context based on confidence map property [J]. *Computer Engineering*, 2016, 42(8): 277-281, 288. (in Chinese)
- [10] 赵龙,肖军波.一种改进的运动目标抗遮挡跟踪算法[J].北京航空航天大学学报,2013,39(4):517-520.
ZHAO L, XIAO J B. Improved algorithm of tracking moving objects under occlusions [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2013, 39(4): 517-520. (in Chinese)
- [11] PÉREZ P, HUE C, VERMAAK J, *et al.* Color-based probabilistic tracking [C]//*Computer Vision-ECCV 2002*. Antibes: Springer, 2002, 2350: 661-675.
- [12] KWONJ, LEE K M. Tracking by sampling trackers [C]//*Proceedings of 2011 IEEE International Conference on Computer Vision*. Barcelona: IEEE, 2011: 1195-1202.
- [13] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.
- [14] GRABNER H, GRABNER M, BISCHOF H. Real-time tracking via on-line boosting [C]//*Proceedings of the British Machine Conference*. Edinburgh: British Machine Vision Association, 2006: 6.

作者简介:王民(1959—),男,江苏常州人,教授,硕士生导师,主要从事智能信息处理方面的研究工作。E-mail: wang-min1329@163.com

石新源(1992—),男,陕西西安人,硕士研究生,主要研究智能信息处理相关技术。E-mail: 15529561263@163.com