

doi: 10.3788/gzxb20134201.0098

采用栅格特征及可视化技术感知跟踪中目标的遮挡状态

孟勃¹, 韩广良², 毛雯蓓³

(1 东北电力大学 信息工程学院, 吉林 吉林 132012)

(2 中国科学院长春光学精密机械与物理研究所, 长春 130033)

(3 北京航空航天大学 计算机学院, 北京 100191)

摘要:针对监控视频中运动目标跟踪问题,提出了一种栅格划分的梯度朝向直方图特征对目标进行描述,并通过可视化技术来实时观测目标局部区域特征量的变化,通过变化来感知目标在运动中可能发生的遮挡等异常情况.利用感知到的状态自适应调整参加融合计算的子模板的个数,指导跟踪过程,以防止异常情况对跟踪结果的影响,实现了对视频目标准确、稳健的跟踪.实验结果表明,提出的目标局部特征能够较准确地感知目标及周围环境发生的变化,对于提高跟踪算法的鲁棒性和准确性具有积极的作用.

关键词:视觉感知;栅格特征;可视化;子模板匹配;遮挡判断;目标跟踪

中图分类号:TP391

文献标识码:A

文章编号:1004-4213(2013)01-0098-6

Grid Feature and Visualization for Occlusion Perception in Object Tracking

MENG Bo¹, HAN Guang-liang², MAO Wen-bei³

(1 Northeast Dianli University, Academy of Information Engineer, Jilin, Jilin 132012, China)

(2 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China)

(3 Beihang University, Academy of Computer Science, Beijing 100191, China)

Abstract: A grid-based histogram of the orientation gradient feature and visualization technique were used in object tracking algorithm to solve the problem of tracking objects moving in surveillance video. The histogram of orientation gradient based on grid was designed as the local feature to describe the object. And it can apperceive the abnormalities of objects such as occlusion, disturbance and deformation etc. The visualization was used to real-time observe the changes of the characteristics through the changes of the local feature. The weights of the sub-model was adjusted adaptively to keep the stability of the tracking algorithm. The experimental results show that the proposed local feature can perceive the abnormity correctly, and can improve the robustness and stability of the tracking algorithm. The self-perception matching algorithm based on this local feature can tracking object accurately and stably in surveillance video.

Key words: Vision perception; Grid feature; Visualization; Sub-model matching; Occlusion judgment; Object tracking

0 引言

视频目标跟踪作为模式识别、人工智能、计算机视觉等的重要研究课题,被广泛应用于智能监控、机

器人视觉、军事导航制导、医疗诊断等领域.视觉目标跟踪的目的就是模拟人类视觉感知的功能,赋予机器感知和分辨视频中运动目标的能力,为后续视频内容的分析和理解提供重要依据.

基金项目:国家自然科学基金(No. 61172111)和东北电力大学博士科研启动基金(No. BSJXM-201111)资助

第一作者:孟勃(1980—),女,副教授,博士,主要研究方向为数字图像处理、视频目标特征提取和识别跟踪等. Email: mengbo_nannan@163.com

收稿日期:2012-04-24; **录用日期:**2012-07-15

近几十年来,关于视频目标跟踪已经开展了大量的研究,并提出了一些有效的方法.但是由于背景的复杂性以及目标的多变性,对运动目标进行稳健的跟踪还十分困难.现有的跟踪方法缺乏与人类视觉感知的结合,而人类视觉系统能较好地适应各种非刚体(比如行人)的运动^[1-4],但计算机对物体的各种形变及遮挡等都比较敏感,物体的某些无意识变化都会导致跟踪的失败,因此,观察和模拟人类视觉的跟踪模式和智能方法,通过感知目标在运动中发生的各种变化,对于设计更准确、稳定的跟踪算法具有十分重要的意义^[5-8].

当目标发生类似于遮挡、干扰、自身形变等复杂情况时,从人类视觉的角度来观察,通常是目标的某些部分或者某些区域发生了变化,也就是目标的局部区域的特征量发生了变化,但是,从机器计算的角度如何能够及时、准确地“感受”到这个渐变过程呢?

本文提出了一种栅格划分的梯度朝向直方图特征对目标进行描述,并通过可视化技术来实时观测目标局部区域特征量的变化,以此来表现目标特征对于环境变化的感知能力.算法通过分析目标局部特征发生的变化以及与环境或其他对象之间的关系,判断当前目标出现的异常情况,利用感知到的状态来指导跟踪过程,自适应调整参加融合计算的子模板的个数,实现了对视频目标准确、稳健的跟踪.

1 目标特征的提取及描述

1.1 基于栅格划分的梯度朝向直方图

基于栅格划分的梯度朝向直方图特征是一种基于梯度的直方图统计特征^[9],它的核心思想是:物体的边缘或梯度的方向密度分布可以很好地描述物体的形状和纹理.本文采用 HSI 空间的 H 分量和 I 分量计算基于栅格划分的颜色梯度朝向直方图. H 为目标本身的颜色,不会发生较大改变,色度分量是相对稳定的, I 分量是亮度分布,即灰度分布, S 分量是饱和度分量,会随外界环境的变化而发生较大的改变,因此,不予采用.

LT:	LT	MT	RT
LC:	LC	MC	RC
LD:	LD	MD	RD

LT: 左上, MT: 中上, RT: 右上
LC: 左中, MC: 中间, RC: 右中
LD: 左下, MD: 中下, RD: 右下

图 1 区域编码

Fig. 1 Region coding

特征提取的方法是:在目标模板区域内沿水平和垂直方向划分栅格区域,对每个区域进行编码,并在每个子区域中分别计算 H 和 I 颜色分量的梯度

朝向直方图.其中 MC 为相对稳定区域.

对每个子区域中的像素分别计算该像素点的梯度和幅值,其中水平和垂直方向的梯度分别为 I_x 和 I_y ,则幅值 M 和梯度朝向角度 θ 分别为

$$M = \sqrt{I_x^2 + I_y^2} \quad (1)$$

$$\theta = \arctan(I_y / I_x) \quad (2)$$

梯度幅值 M 反映了该像素点周围灰度变化的大小,而梯度朝向 θ 则反映了该点周围灰度变化的方向.

将幅值 M 加权量化至由梯度朝向角度 θ 定义的具有 O 个 bin ($O=4$) 值的直方图中.

$$O_{x,y} = [O \times \arctan(I_y / I_x) / 2\pi + 0.5] \quad (3)$$

根据梯度朝向角,将区域内所有点分为 4 类: $(0, \pi/2)$ 、 $(\pi/2, \pi)$ 、 $(\pi, 3\pi/2)$ 、 $(3\pi/2, 2\pi)$. 每一点量化为: $M \cdot e^{-\frac{(\theta-u)^2}{2}}$, 其中 u 分别为 $\pi/4$ 、 $3\pi/4$ 、 $5\pi/4$ 、 $7\pi/4$, 将每个像素的方向累加到对应的 bin 中,从而得到每个子区域的梯度朝向直方图.

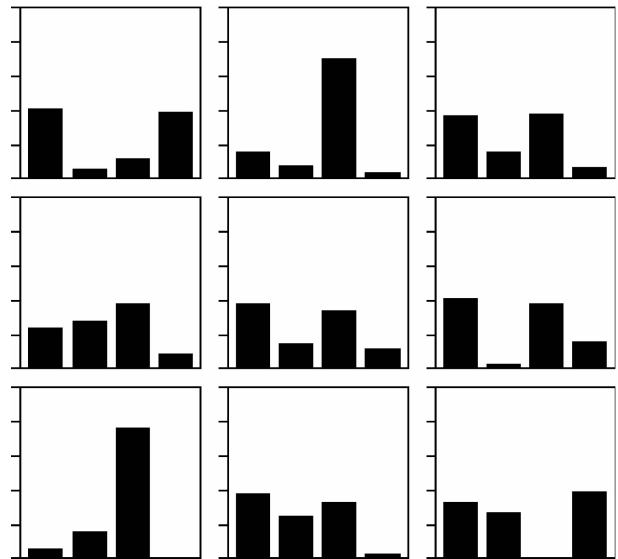


图 2 基于栅格划分的梯度朝向直方图

Fig. 2 Gradient-orientation histogram based on grid

这样分别计算出每个子区域的梯度朝向直方图之后,根据 (LT-LC-LD-MT-MC-MD-RT-RC-RD) 的次序构成局部纹理颜色特征的特征向量,并根据向量中每个参量的变化来判断目标所发生的变化.

1.2 特征感知能力的验证实验及结果分析

当目标发生类似于遮挡、干扰、自身形变等异常情况时,从人类视觉观察事物的角度来看,往往是目标的某些部分或者某些区域发生了变化,也就是目标局部区域的特征量发生了变化.但是,从机器计算的角度如何能够及时、准确地“感受”(状态感知)到这个渐变过程呢?为了实现这一目的,本文从数据变化的角度来验证所提出的局部区域特征对于这种变化的感知能力.

图3是从飞机过塔尖视频中选取出来的一组目标受遮挡的图像序列($160 \times 128 \times 8$ bit, 图3中(a_1)~(h_1)). 在每幅图像中提取目标 3×3 尺度的栅格梯度朝向直方图(图3中(a_2)~(h_2)), 并在图中对于每帧较前一帧变化较大(被遮挡)的区域的直方图

用圆圈做了标记. 从整个遮挡图像序列中可以观察到: 目标的被遮挡过程是从左侧区域——中间区域——右侧区域的变化过程, 反映在目标的子区域编码中是: RCRD-MCMD-MTMCMD-LCLD-LTLCLD的过程.

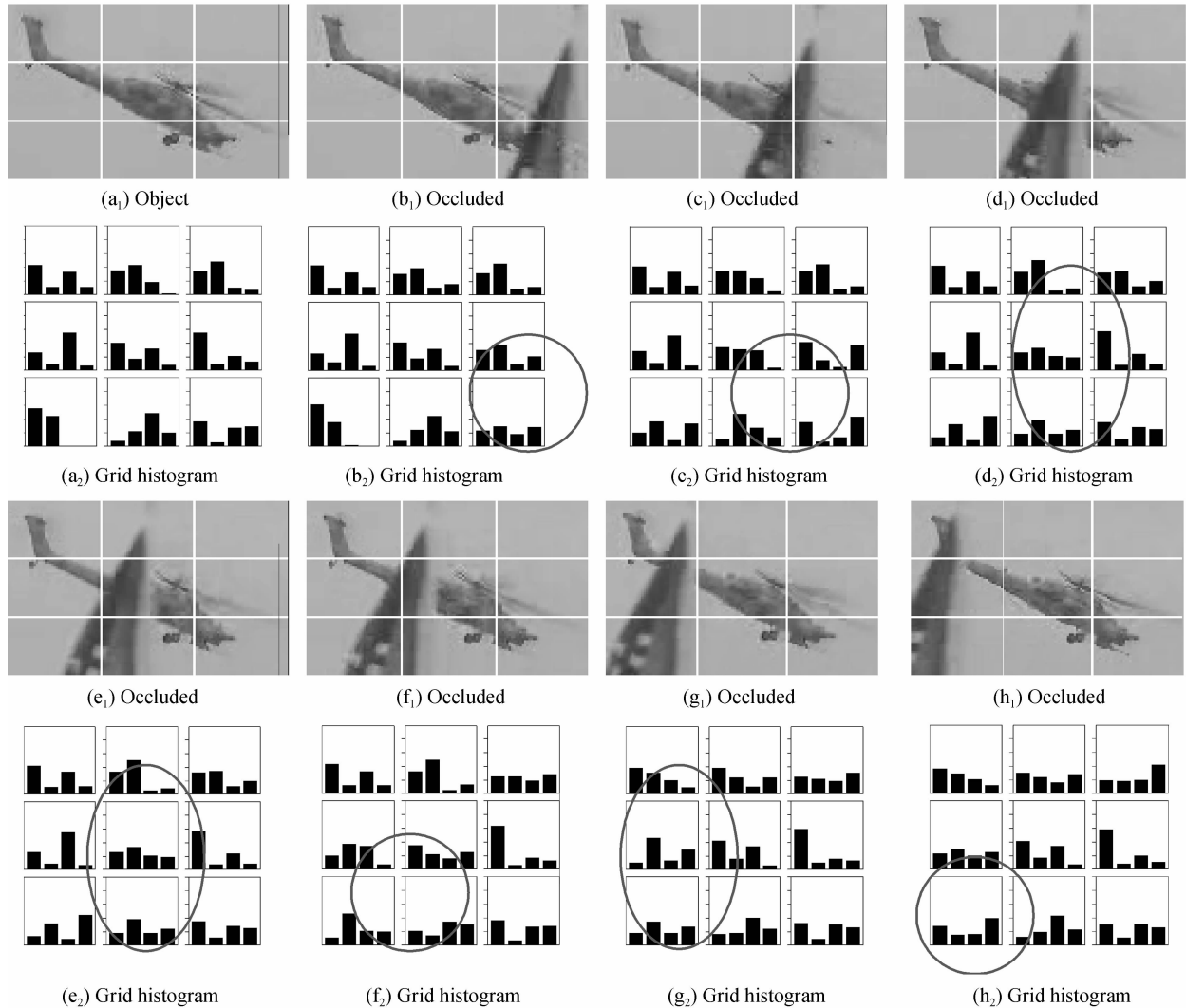


图3 目标的栅格梯度朝向直方图对遮挡情况的感知实验

Fig. 3 Perception experiment of Grid-gradient-orientation histogram towards occlusion

从图3可以分析出: 当目标发生遮挡时, 提出的目标局部特征——基于栅格的梯度朝向直方图能够比较准确地体现出目标在整个遮挡过程中的渐变过程, 并将这种变化进行了有效的量化, 使得我们从机器计算的角度可以自感知到这种渐变, 因此, 当遮挡发生时, 可以通过提出的局部特征对目标及周围环境的变化进行自感知. 需要说明的是, 本文提出的感知遮挡的方法, 仅适用于目标发生部分遮挡的情况, 当目标被全部遮挡时, 由于目标完全消失, 也就意味着所有的子区域的匹配值都在一个很小的范围之内, 所以, 很难得到正确的匹配结果. 如果想处理全部遮挡的情况, 还需要加入相应的运动预测方法.

2 跟踪算法的设计及可视化技术的应用

2.1 跟踪算法的设计

本文采用模板匹配的思想来设计跟踪算法. 采取子区域匹配的方法进行目标特征的描述, 并对子区域的权值进行加权融合, 融合后的结果作为最终的目标位置. 算法的优势是: 当目标的某个子区域受到遮挡或干扰时, 对应位置的子模板的匹配度虽然会降低, 但是那些未受干扰或遮挡的子区域仍然可以保持较高的匹配度, 经过对子模板的自适应融合可以得到比较好的跟踪效果, 避免遮挡给跟踪结果

带来的影响,从而达到感知目标当前状态并保持准确、稳定跟踪的目的。

跟踪过程概括为:

1)在初始帧中选取目标区域作为目标模板,并将目标模板划分为 3×3 的栅格区域,每个栅格代表一个目标的子区域,分别计算子区域的梯度朝向直方图作为目标子模板的特征量;

2)后续的跟踪中,对于每一帧图像的目标区域都采取同模板相同的处理方法,即分别计算每个子区域的梯度朝向直方图,得到每一帧的目标子区域的特征量;

3)采用 Bhattacharyya 方法计算当前帧子模板与目标模板之间的相似度,得到各子区域的权值^[10-11]

$$\rho_i = \sum_{n=1}^m \sqrt{\hat{p}_u \hat{q}_u} \quad (4)$$

式中 $\{\hat{q}_u\}_{u=1, \dots, m}$ 为目标子模板的直方图特征量, $\{\hat{p}_u\}_{u=1, \dots, m}$ 为当前帧目标子区域的直方图特征量。

对式(3)进行 Taylor 展开,得到 Bhattacharyya 系数的近似值,得到相似系数

$$\rho_i [p(y), q] \approx \frac{1}{2} \sum_{n=1}^m \sqrt{p_u q_u} + \frac{C}{2} \sum_{x_i \in \mathbf{R}} w_i K(\|d(x_i)\|^2) \quad (5)$$

式中

$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{p_u}} \delta(b(x_i) - u) \quad (6)$$

$$d_i = \sqrt{1 - \rho(q, p_i)} \quad (7)$$

d_i 为 Bhattacharyya 距离。

4)子区域的自适应融合计算:在得到所有子区域的权值之后,对所有权值大于 85% 的子区域进行加权平均的融合运算,最后得到一个加权结果,作为最终的匹配结果。

5)根据最终匹配的结果得出目标的跟踪位置。

2.2 可视化技术的应用

由于本文采取的子区域匹配的方法进行目标跟踪,所有每个子区域的匹配值的大小(即子区域的权值)对于匹配结果有着至关重要的影响,如果多个子区域的权值都很小时,必然带来跟踪失败的结果。因为,遮挡是个渐变的过程,如果能够在遮挡的开始阶段就能够“预知”或者“感知”到变化的话,可以采取有效的措施来预防跟踪失败的发生。因此,本文采用可视化技术来达到预知和感知的目的,期望通过对子区域权值变化的实时观测,来感知遮挡状态,并来有效避免遮挡对跟踪结果的影响。

图 4 是本文设计的感知匹配跟踪平台,平台上

面区域显示的是对目标实时跟踪的结果,中间的栅格框是各个子区域的匹配位置,栅格框外框(目标最外面的框)为目标的结果位置,是多个子区域(栅格框)融合运算得到的最终跟踪位置。平台下面区域显示的是跟踪的中间和结果数据:曲线显示的是各个子区域的匹配度实时曲线(曲线的个数等于子区域的个数),它代表对应着的每个子区域匹配在目标运动过程中的匹配结果,横坐标表示帧数,纵坐标表示当前帧各个子模块的匹配度大小。当目标在运动过程中发生异常时,受到干扰的子区域的匹配值就会相应的降低,而未受到干扰的子区域的匹配值则会保持较高的数值。由于多个子区域同时进行匹配,那么这些较高匹配度的子区域用于保持正确的运动位置,而那些较低匹配度的子区域就反应了目标所在的位置发生的诸如遮挡、干扰等变化,在进行状态感知匹配跟踪实验时,实际上总有一些曲线是保持较高匹配度的,而另一些曲线会随着实际的遮挡和变形而下降,检测到这种趋势的变化,可以得到相应的感应结果,进而采取相应的处理措施,也反映出算法对异常情况的感知能力。



图 4 感知匹配跟踪平台

Fig. 4 Tracking platform of perception matching

2.3 算法时间复杂性分析

本文提出的算法是对传统相关匹配算法的改进,并借鉴了局部特征的思想,利用 VC 平台的优势使算法的结果可以实时并观测。整个算法中没有过于复杂的运算,只是对于子区域的结果进行了 Bhattacharyya 计算,得到相似度,因此,算法完全可以满足实时要求。

3 实验结果及分析

首先采用 Pets 标准视频来验证所提出的栅格特征及可视化技术对遮挡状态的感知能力,并通过与粒子滤波算法的比较来验证算法的跟踪性能。

实验中选取一段车辆运动的视频场景进行试验 (Pets 标准视频: $384 \times 288 \times 32$ bit, 2 219~2 519 帧), 图 5 是采用本文提出的栅格特征及可视化技术对目标遮挡状态的感知结果 (Pets 标准视频: $384 \times 288 \times 32$ bit, 2 219~2 519 帧), 目标的模板大小为: 27×35 . 如图 5 所示, 目标 (白色面包车) 在静止并未有任何的干扰和遮挡时, 实时曲线 (框格外框) 比较稳定, 只有少数曲线发生了跳变, 各个子区域的匹配

度大多保持较高的水平; 当黑色轿车逐渐向目标移动并对目标进行干扰时, 目标的某些子区域 (框格矩形框) 偏离了目标区域, 相应的匹配度实时曲线 (栅格外框) 也发生了较大的跳变; 当黑色轿车继续向目标接近并对目标进行部分遮挡时, 某些栅格外框所示实时曲线也随之发生了较剧烈的变化, 表明目标的特征发生了较大的变化, 对照视频图可以看到, 此时目标被黑色轿车部分遮挡了.



(a) Frame2219,matching:100 (b) Frame2268,matching:97.560 (c) Frame 2288,matching:97.557(d) Frame2343,matching:96.510 (e) Frame2384,matching:96.917



(f) Frame 2467,matching:95.633 (g) Frame2477,matching:96.707 (h) Frame2495,matching:96.202 (i) Frame 2519,matching:95.578

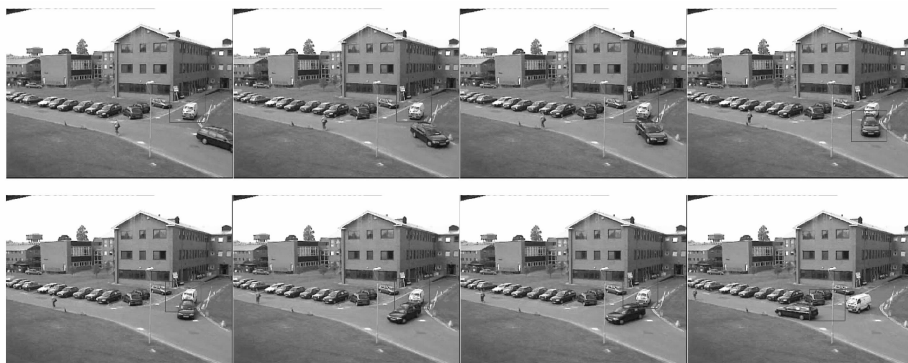
图 5 目标受遮挡及干扰的状态感知及匹配跟踪结果

Fig. 5 State perception and matching results of occlusion and disturbing

由这种曲线的渐变并实时跟踪效果可以看到: 本文提出的栅格特征及可视化技术对于目标的遮挡、干扰等状态具有较好的感知能力. 并且, 由于那些未受干扰和遮挡的区域仍然保持较高的匹配度, 经过对子区域的自适应融合, 仍然能够得到目标较准确的跟踪位置.

仍然采用 PETS 标准视频作为测试序列. 采用粒子滤波 (Particle Filter, PF) 算法进行跟踪实验, 实验参量为: 粒子数为 50, 目标模板大小 60×52 , 采用的目标特征量为灰度加权直方图特征, 比较结果如图 6.

从比较结果可以看出, 随着目标被遮挡和干扰的程度不断加剧, PF 跟踪算法在黑色轿车将目标的一半都遮住的时候丢失了目标, 跟踪框偏离到了干扰物上, 当黑色轿车开始运动时, 跟踪框随着干扰物脱离了跟踪目标. 这是因为, 随着遮挡和干扰的不断加剧, 使得跟踪算法的匹配度不断下降, 最后得到了错误的匹配结果. 本文算法在目标受到部分遮挡和干扰时, 在初始阶段, 算法具有较高的匹配度, 当目标不断受到遮挡和干扰时, 算法一直采用那些未受干扰和遮挡的区域, 进行位置计算, 由于这些子区域



(a) Tracking result of PF



(b) Tracking result of proposed method

图6 本文算法与粒子滤波(PF)算法跟踪结果比较

Fig. 6 Results of the comparison between proposed method and PF

仍然保持较高的匹配度,经过对子区域的自适应融合,仍然能够得到目标较准确的跟踪位置.因此,提出的算法能够保持跟踪的稳定性和准确性.

为了验证算法的实用性,将本文提出的算法与PF算法进行了时间复杂性的分析和比较,如表1,采用两种算法分别对不同的视频进行了跟踪实验,从实验结果可以看出,本文提出的算法在时间复杂性上要优于PF算法.

表1 时间复杂性比较

Table 1 Comparison of time complexity

	Template size	PF		Proposed method	
		Matching	Time	Matching	Time
Experiment1	27×35	93	87 ms	96	23 ms
Experiment2	60×52	92	115 ms	97	26 ms
Experiment3	36×48	92	97 ms	96	24 ms

4 结论

作为神经心理学的概念,感知被认为是人类通过视听觉获取信息后的一个心理活动.一般来说,人类感知机制还没有被完全理解,对于机器来说要自动重复人类的感知过程是很困难的.因此,本文模拟了在目标跟踪时基于栅格划分的梯度朝向直方图特征的状态感知过程,通过可视化技术来实时观测目标局部区域特征量的变化,以此来表现目标特征对于环境变化的感知能力.通过目标局部特征的变化对目标可能发生的遮挡、形变等异常情况进行感知,以防止异常情况对跟踪结果的影响.并自适应调制参加融合计算的子模板的个数来得到目标较准确的跟踪位置,实现了对视频目标准确、稳健的跟踪.实验结果表明,该算法能够很好地克服目标的遮挡异常,实现较准确的目标跟踪.

参考文献

- [1] OLIVA A. Gist of the scene[J]. In: Itti, L. Rees. G. and Tsotsos, J. K. (Eds), *Neurobiology of Attention*, Elsevier, 2005, 251-256.
- [2] 罗四维. 视觉感知系统信息处理理论[M]. 北京: 电子工业出版社, 2006.
- [3] QIAN Lele. Researches on image understanding based on hierarchical visual perception mechanisms [D]. Hefei University of Technology, 2009.
钱乐乐. 基于视觉层次感知机制的图像理解方法研究[D]. 合肥工业大学, 2009.
- [4] DONG Li-geng, TAO Lin-mi, XU Guang-you. Head pose estimation based on a second order histogram of the orientation gradient[J]. *Journal of Tsinghua University (Science and Technology)*, 2011, 51(1): 73-79.
董力庚, 陶霖密, 徐光佑. 基于二阶梯度朝向直方图特征的头部姿态估计[J]. 清华大学学报(自然科学版), 2011, 51(1): 73-79.
- [5] DECO G & SCHURMANN B. A hierarchical neural system with attentional top-down enhancement of the spatial resolution for object recognition[J]. *Vision Research*, 2000, 40: 2845-2859.
- [6] AMIT Y, MASCARO M. An integrated network for invariant visual detection and recognition[J]. *Vision Research*, 2003, 43: 2073-2088.
- [7] DALAI N, TRIGGS B. Histograms of oriented gradients for human detection[C]. Proc IEEE Conf Computer Vision and Pattern Recognition. Washington DC, USA; IEEE Computer Society, 2005: 886-893.
- [8] WANG Hong-liang, ZHANG Mei-xian, DIAN Hai-fei. Application of the theory of D-S evidence in target identification[J]. *Automation & Instrumentation*, 2011(7): 14-17.
王红亮, 张美仙, 丁海飞. D-S证据理论在目标识别中的应用[J]. 自动化与仪表, 2011(7): 14-17.
- [9] GROSSBERG S, SWAMINATHAN G. A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention, and bistability[J]. *Vision Research*, 2004, 44(11): 1147-1187.
- [10] MENG Bo. Research and application of the local optimal particle filter target tracking algorithm [D]. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2008.
孟勃. 局部最优粒子滤波目标跟踪算法的研究和应用[D]. 中国科学院长春光学精密机械与物理研究所, 2008.