

doi: 10.3788/gzxb20154407.0710001

# 空间相位相关的快速视觉目标跟踪方法

查宇飞, 张立朝, 覃兵, 黄宏图

(空军工程大学 航空航天工程学院, 西安 710038)

**摘 要:**在相位相关目标跟踪中,由于响应函数为冲击函数,不能适应目标的变化,且易受噪音和杂波的影响,导致跟踪失败.而均值合成滤波器利用符合期望的相关输出分布,可获得较好的滤波效果.本文考虑跟踪过程中目标的上下文信息,设计与目标空间位置相关的响应函数取代冲击函数,从而获得更为鲁棒的滤波器;利用峰值旁瓣比评估跟踪结果并自适应更新滤波器,适应目标的变化;用快速傅里叶变换进行加速算法,对目标实现快速鲁棒跟踪.实验结果表明,本文方法平均每秒处理 50 帧图像,在准确度和精确性方面优于当前算法.

**关键词:**空间相位相关;快速傅里叶变换;峰值旁瓣比;自适应更新

中图分类号: TP391

文献标识码: A

文章编号: 1004-4213(2015)07-0710001-10

## Fast Visual Tracking with Spatial Phase Correlation

ZHA Yu-fei, ZHANG Li-chao, QIN Bing, HUANG Hong-tu

(School of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China)

**Abstract:** The visual tracking based on phase correlation is always failure, because the response is dirac delta function disturbed by the noise and cluster. The desired correlated output distribution is adopted in average of synthetic exact filters, which obtains the excellent filtered result. This paper designs the response related with the object's spatial position to replace the dirac delta function, which can achieve a robust filter. In order to adapt the changed object, peak-to-sidelobe ration was used to evaluate the results to update the filter. The fast Fourier transform was used to accelerate the algorithm to track the object fast and robustly. The experiments show that the proposed method can process 50 frames per second, and excels the state-of-the-art methods in both accuracy and precision.

**Key words:** Spatial phase correlation; Fast fourier transform; Peak-to-sidelobe ration; Adaptively updating

**OCIS Codes:** 100.0100; 100.4999; 100.5010; 330.5000

## 0 引言

视觉目标跟踪是计算机视觉的一个重要问题,在许多领域发挥着重要作用,尤其是在人机交互、侦察监视、机器人等方面.其难点在于如何适应由光照变化、形变、遮挡、相机移动等因素所引起的目标外观变化.

为获得目标的空间位移矢量,光流法<sup>[1]</sup>假定目标在相邻两帧之间亮度不变,然而这一假设具有非常大的局限性,导致在实际场景中,微小的变化将使算法失效.密度光流法<sup>[2]</sup>采用前后向轨迹计算光流预测的误

差,误差小的点作为匹配点计算仿射变换参量.为解决目标丢失的问题,利用随机森林检测目标,与该方法相结合<sup>[3]</sup>,取得了较好的跟踪效果.而基于分类的方法将目标跟踪看作分类问题来处理.文献[4]利用线性判别分析选取区分性最强的特征在背景中检测目标.支持向量机(Support Vector Machine, SVM)跟踪算法<sup>[5]</sup>和集成跟踪算法<sup>[6]</sup>寻求最优分类器判别目标和背景.这些方法都采用灰度值、梯度等原始特征,导致其鲁棒性较差.在人脸检测中获得成功的 Haar 特征,也被用于跟踪<sup>[6]</sup>.

**基金项目:**国家自然科学基金(Nos. 61472442, 61203628, 61202339),博士后基金(Nos. 2012T50879, 2012M512144)资助

**第一作者:**查宇飞(1980—),男,副教授,硕导,主要研究方向为视频目标跟踪. Email: zhayufei@126.com

**收稿日期:**2014-10-22; **录用日期:**2015-05-08

<http://www.photon.ac.cn>

另一类方法是在频域进行处理. 相位相关(Phase Correlation)<sup>[7]</sup>将信号变换到频域,通过相关运算获得目标在空域的平移量. 但是传统相位相关算法的响应函数为狄克来  $\delta$  函数,容易受噪音和杂波影响,且不适应目标的变化,鲁棒性较差. 均值合成滤波器(Average of Synthetic Exact Filters, ASEF)<sup>[8]</sup>采用类高斯函数形状(Gaussian-function-like shape),更符合期望的相关输出分布,在人眼定位上取得了成功. 最小输出平方误差和(MOSSE)滤波器<sup>[9]</sup>寻求与目标模板平方误差最小的候选样本作为当前目标估计,利用快速傅里叶变换(FFT)实时跟踪目标. 为了解决训练样本的问题,循环结构核跟踪算法(CSK)<sup>[10]</sup>利用循环卷积计算核函数,实现对候选样本的快速计算. 文献[11]将其拓展到高维颜色空间,通过 PCA 降维提取颜色主空间,获得更好的跟踪效果. 在 2014 年的视频跟踪竞赛(VOT2014)中,基于相关滤波器的跟踪算法获得巨大成功<sup>[12]</sup>. 这类方法的核心是模板匹配,快速傅里叶变换只是用来加速计算目标空域位移偏量. 事实上,目标频谱由幅度谱和相位谱两部分组成,而目标位移偏量仅仅与相位谱相关,与幅度谱无关. 因此,对目标模板和候选目标进行相位相关运算,即可获得目标的位移偏量. 本文改进传统相位相关算法,利用目标上下文信息,依据其空间分布,设计考虑噪音模型的响应函数,提出一种基于空间相位相关的视觉目标跟踪算法.

## 1 空间相位相关

时域信号  $x$  与滤波器  $h$  的卷积为

$$x * h = g \tag{1}$$

式中,  $*$  为卷积运算,  $g$  为输出响应. 根据卷积定理,空域的相关滤波可以通过频域进行计算,式(1)可以转

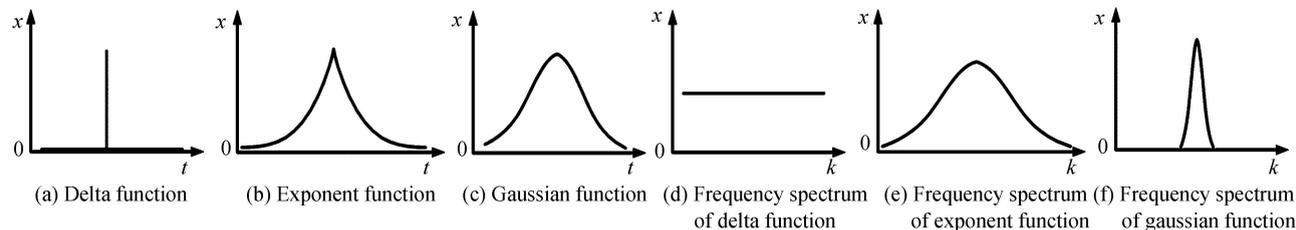


图 1 表示不同空间分布和频谱

Fig. 1 Represent different spatial distribution and spectrumm

(b)、(c)所示,图 1(e)、(f)为其对应的频谱. 分布函数可以写为

$$g \propto \exp(-\alpha \|p - pc\|^\beta), \beta = \{1, 2\} \tag{8}$$

式中,  $\alpha$  为分布参量,控制曲线陡峭程度. 当  $\beta$  等于 2 时,为高斯分布,通常也称为贝尔曲线,概率响应与距离的平方成比例. 当  $\beta$  等于 1 时,偶双边指数分布与距离成比例.

从循环卷积的角度来说,狄克来  $\delta$  函数作为响应

变为

$$X \circ H = G \tag{2}$$

式中,  $X$ 、 $H$  和  $G$  分别为  $x$ 、 $h$  和  $g$  所对应的频域信号,  $\circ$  为矩阵点积(Hadamard)运算.

对时域信号  $y$  进行滤波,得

$$R = Y \circ H = Y \circ \frac{G}{X} = \frac{Y}{X} \circ G \tag{3}$$

式中  $R$  是频域信号  $Y$  的滤波结果. 如果  $y$  是  $x$  的循环移位信号,即

$$y_l = x_{(n-m) \bmod N} \tag{4}$$

则  $y$  的离散傅里叶变换在相位上可以相应地转化为

$$Y_k = X_k e^{imk} \tag{5}$$

则式(3)可写成幅度谱和相位谱的形式,即

$$R = \frac{\|Y\|}{\|X\|} \circ \|G\| e^{-i(\angle Y - \angle X + \angle G)} = \|G\| e^{-i(\angle G - 2\pi/Nmk)} \tag{6}$$

若采用狄克来  $\delta$  函数作为响应输出,如图 1(a)所示,即时域响应为

$$r(n) = \delta(n+m) \tag{7}$$

通过对频域输出  $R$  进行反傅里叶变换,计算最大响应即可获得目标在空域的位移  $m$ . 快速傅里叶变换使得运算复杂度由原来的  $O(N \cdot N)$  减少为  $O(N \log N)$ . 在跟踪过程中,目标外观会随着时间发生变化,而传统相位相关由于采用狄克来  $\delta$  函数,最大值判断极易出错,导致滤波效果很差. 因此,强约束的  $\delta$  分布不能适应跟踪过程中变化的目标,降低相位相关的性能. 为了增强相关算法的鲁棒性,受 ASEF 滤波器启发,本文考虑目标的空间分布特性,采用高斯分布或偶双边指数分布作为期望的响应输出,抵消目标变化对响应函数的影响. 偶双边指数分布和高斯分布如图 1

输出求解滤波器,认为只有中心信号是所需信号,而其它信号不是所需信号. 这种强约束使得当信号稍微发生变化,就会导致得到的滤波器输出变化很大. 而采用与空间位置相关的响应分布后,期望的响应输出离中心位置越近,则属于期望信号的概率越大,离中心位置越远,则属于期望信号的概率越小. 这样当信号发生变化时,对滤波器的影响有限,增强了滤波器的鲁棒性.

## 2 基于空间相位相关的视觉目标跟踪

从滤波角度看,目标跟踪首先利用目标和响应函数学习滤波器,然后用滤波器对下一帧图像滤波,取响应最大的点作为跟踪结果,最后利用该结果对滤波器进行更新,适应目标变化.本文重点关注目标在图像中的位置,将跟踪过程分为预处理、目标跟踪和滤波器更新三个过程.

### 2.1 预处理过程

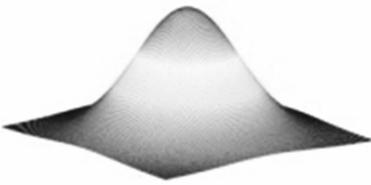
与图像配准不同,目标跟踪是在目标框内标定目标位置而非整个图像.图像边界会严重影响估计模板的性能,并且傅里叶变换的周期性要求不考虑图像边界,因此需要采用窗函数(比如汉明窗、高斯窗)减少图像边界效应.在此基础上,计算图像的二维傅里叶变换.根据目标空域一致性,目标搜索区域根据上一帧目标位置进行扩展,但不能超出图像边界.



(a) Image object and search window



(b) Raw



(c) Window



(d) Data



(e) Frequency

图2 预处理流程

Fig. 2 Pretreatment process

跟踪方法的初始化如图2.在第一帧中,采用虚线人工标定目标,如图2(a).目标状态表示为

$$t = [C_x, C_y, M, N] \quad (9)$$

式中,  $C_x$  和  $C_y$  为中心坐标,  $M$  和  $N$  分别为目标的宽度和高度.同时,采用搜索窗定位目标并检测目标位移.

搜索窗以目标中心坐标为基准进行扩展:  $s = [C_x, C_y, \eta M, \eta N]$ , 如图2(a),  $\eta$  值越大,搜索窗越大.为有效消除边界效应并获得周期图像,采用窗函数  $\kappa$  (如图2(c)所示)与原始图像相乘

$$x = I(s) \circ \kappa \quad (10)$$

式中,  $I$  为原始图像,  $x$  为预处理图像.预处理消除了边缘细节,得到的图像背景平滑,这样线性位移即等价于循环位移,如图2(d).将预处理图像  $x$  进行傅里叶变换如图2(e),则滤波器为

$$H = G/X \quad (11)$$

式中,  $G$  为空间分布的频谱.

### 2.2 跟踪过程

跟踪过程的任务是根据上一帧目标状态估计当前帧目标状态.根据2.2节,采用空间相位相关在当前帧图像中标记目标.如果目标空域位移为  $(\Delta m, \Delta n)$ ,则当前帧图像  $y$  可以写为

$$y(m', n') = x(m + \Delta m, n + \Delta n) \quad (12)$$

式中,  $x$  为目标模板.空域位移  $(\Delta m, \Delta n)$  将改变当前帧图像  $y$  的频域相位谱,即

$$Y = X \circ e^{-i(2\pi/M\Delta m + 2\pi/N\Delta n)} \quad (13)$$

式(13)可以写成幅度谱和相位谱的结合,即

$$R = \frac{\|Y\|}{\|X\|} \circ \|G\| e^{-i(2\pi/M\Delta m + 2\pi/N\Delta n + \angle G)} = \|G\| e^{-i(2\pi/M\Delta m + 2\pi/N\Delta n + \angle G)} \quad (14)$$

式中,  $\angle$  为相位,  $G$  为空间响应分布  $g$  的频谱.因此,目标的空域位移只与相位差异有关而与幅度无关.实际上,目标外观模型不可能始终保持不变,这些变化对相位影响不大,但对幅度会有影响导致  $\|Y\| \neq \|X\|$ , 若响应函数为  $\delta$  函数,则  $\|Y\| / \|X\|$  会降低响应的性能.而空间响应分布频谱  $G$  采用本文提出的空间分布,能够消除图像之间目标外观模型差异所带来的影响.式(14)反映了响应可以分解为空间响应分布频谱  $G$  和两帧图像相位谱差异.

### 2.3 更新过程

在本文提出的算法中,模板更新对于提升跟踪性能起着十分关键的作用,正确的模板更新有利于后续帧的跟踪,而错误的模板更新则会导致跟踪失败.通常,对跟踪结果进行评估能够有效避免错误更新.

空间响应分布表明在目标中心位置处产生峰值,根据峰值尖锐程度,即峰值与其周围之间的相对高度,度量相似度.最普遍的峰值尖锐度度量为 PSR (peak-to-sidelobe ration)<sup>[8]</sup>, 定义为

$$PSR = \frac{r_{\max} - \mu}{\sigma} \quad (15)$$

式中,  $\mu$  和  $\sigma$  分别为空域响应分布的均值和标准差.该度量表示归一化的目标相似度, PSR 值越大,目标的置信度越高,非常适用于模板更新.

## 2.4 与相关研究的关系

视觉目标跟踪领域提出的相关滤波方法也是将目标跟踪转化到频域进行计算。

最小输出平方误差和 (MOSSE) 跟踪算法<sup>[9]</sup> 最小化期望相关响应与真实相关响应之间的均方误差 (MSE), 其输出响应写为

$$r = F^{-1} \left( \frac{\mathbf{X} \circ \mathbf{G}^*}{\mathbf{X} \circ \mathbf{X}^*} \circ \mathbf{Y} \right) \quad (16)$$

式中,  $\mathbf{X}$  和  $\mathbf{G}$  分别为模板和响应对应的频谱,  $\mathbf{G}^*$  为复数  $\mathbf{G}$  的共轭复数,  $\mathbf{Y}$  为候选目标的频谱. 通过对标记目标随机变换获得训练样本, 并且搜索窗与目标窗大小相等.

文献[10]提出基于循环数据的核回归循环结构跟踪算法, 其搜索窗大于目标窗并采用核卷积模板. 响应的输出写为

$$r = F^{-1} \left( \frac{F(k(x, y))}{F(k(x, x))} \circ \mathbf{G} \right) \quad (17)$$

式中,  $k(\cdot)$  为核函数. 核函数将线性模型转化为非线性模型. 基于该跟踪框架, 文献[11]将原始图像空间扩展到多维颜色空间, 以便更好地描述目标并获得优异的性能.

这些方法的本质都是空域模板匹配, FFT 只是用来提升算法效率. 本文提出的 SPC 跟踪算法只采用相位谱获得目标的空域位移, 设计目标的空域分布响应函数消除噪音和杂波的影响, 即

$$r = F^{-1} (e^{-i(\angle Y - \angle X)} \circ \mathbf{G}) \quad (18)$$

与空域分析方法相比, 本文方法根据目标的空域分布设计响应函数改进传统相位相关方法, 获得鲁棒的滤波器, 利用峰值旁瓣比自适应更新滤波器, 适应目标的变化, 将频域的相位谱差异映射到目标的空域位移, 实现目标定位, 从而能够实时鲁棒地跟踪目标.

## 3 实验和分析

### 3.1 算法实现

实验中, 目标初始值通过人工标定并只考虑目标发生相似变换. 运行平台为 Intel CoreDuo 3.6 GHz laptop with 6 G memory; 运行环境为 MATLAB2010a; 参量设置  $y=2, \alpha=1.5$ . 算法的伪代码为

算法 1: 空间相位相关视觉目标跟踪算法 (SPC)

输入:  $\{I_j\}_{j=1}^L, t = [C_x, C_y, M, N]$

预处理过程:

$$s = [C_x, C_y, \eta M, \eta N]$$

$$x = I_1(s) \circ \kappa$$

$$g \propto \exp(-\alpha \|p - p_c\|^\beta), \beta = \{1, 2\}$$

跟踪过程:

for  $j=2:L$  do

$$y = I_j(s) \circ \kappa$$

$$\mathbf{R} = \|\mathbf{G}\| e^{-i(\angle Y - \angle X + \angle G)}$$

$$[C'_x, C'_y] = \arg\{\max[F^{-1}(\mathbf{R})]\}$$

更新过程:

$$\text{PSR}_j = \frac{F^{-1}(\mathbf{R})_{\max} - \mu}{\sigma}$$

$$\mathbf{X} = \left(1 - \frac{\text{PSR}_j}{\text{PSR}_1}\right) \mathbf{X} + \frac{\text{PSR}_j}{\text{PSR}_1} \mathbf{X}'$$

End for

Return  $[C'_x, C'_y]$

### 3.2 数据设置和评估方法

挑选 20 个具有代表性的视频序列进行测试, 并与当前主流算法 (state-of-the-art) 进行比较. 实验中序列的难度因子包括: 光照变化, 快速移动, 图像模糊, 尺度变化, 严重遮挡, 三维运动, 姿态变化, 背景复杂, 相机移动, 共面旋转及异面旋转. 测试序列的难度因子统计如表 1 所示.

采用距离准确度 (Distance Precision, DP) 和成功率 (Success Rate, SR) 评估本文 SPC 算法和其他比较算法.

采用距离准确度时, 曲线斜率越陡表示跟踪算法准确度越高, 而如果跟踪算法跟丢目标, 则准确度曲线不会收敛于 1. 本文选取 20 像素作为准确度阈值, 在 20 个视频序列上比较 SPC 和 state-of-the-art 算法的优劣. 另外, 采用 FPS (frames per second) 比较跟踪算法的速度.

### 3.3 实验 1. 空间响应分布

通过相位谱对目标模板与搜索窗进行相关计算, 得到的响应分布表面包含一个尖锐的峰值, 即搜索窗中目标的位置. 然而, 在实际场景中响应分布表面会有一些错误的峰值, 有时候甚至高于真实峰值.

为了评估响应分布的性能, 本文选用 3 个都只含一个峰值的凸分布作为空间响应分布:  $\delta$  分布、偶双边指数分布和高斯分布 (见图 3). 如图 3 所示, 在 4 个视频上测试距离准确度, 偶双边指数分布和高斯分布的性能都好于  $\delta$  分布. 与高斯分布相比, 偶双边指数分布为更加平滑的卷积响应分布, 而获得更好的性能.

### 3.4 实验 2. 自适应更新

在相关研究中, 模板更新系数  $\rho$  是固定值. 本文中  $\rho$  通过 PSR 进行更新, 其中  $\rho_i = \text{PSR}_i / \text{PSR}_0$ . 目标与模板的相似度根据 PSR, 即峰值与其周围值的相对高度进行度量. 图 4 显示了固定系数  $\rho$  与自适应更新系数  $\rho$  的距离准确度比较结果. 结果表明自适应更新的跟踪优于固定系数的跟踪.

表 1 测试数据

Table 1 Test data

Sequence Name	Main difficulty factor	Resolution ratio	Frame number
Basketball	Rapid motion, Serious occlusion	$576 \times 432$	726
Car4	Scale change, Illumination change	$360 \times 240$	659
CarDark	Complicated background, Illumination change	$320 \times 240$	393
CarScale	Scale change, Rapid motion	$640 \times 272$	254
Coke	Serious occlusion, Illumination change	$640 \times 280$	291
Deer	Rapid motion, Image blurring	$704 \times 400$	71
Dog1	Scale change, Coplanar rotation	$320 \times 240$	1 351
Dudek	Serious occlusion, Scale change	$720 \times 480$	1 146
FaceOcc2	Serious occlusion, Coplanar rotation	$320 \times 240$	812
Fish	illumination change, Camera motion	$320 \times 240$	476
Freeman1	Complicated background, Three-dimensional motion	$360 \times 240$	326
Girl	Three-dimensional motion, Serious occlusion	$128 \times 96$	500
Ironman	Complicated background, Rapid motion	$720 \times 304$	116
Mhyang	Illumination change, Scale change	$320 \times 240$	1 490
Mountainbike	Rapid motion, Pose change	$640 \times 360$	228
Skating1	Rapid motion, Illumination change	$640 \times 360$	400
Suv	Serious occlusion Complicated background	$320 \times 240$	945
Sylvester	Pose change, Rapid motion	$320 \times 240$	1 346
Walking2	Serious occlusion, Scale change	$384 \times 288$	501
Woman	Serious occlusion, Pose change	$352 \times 288$	598

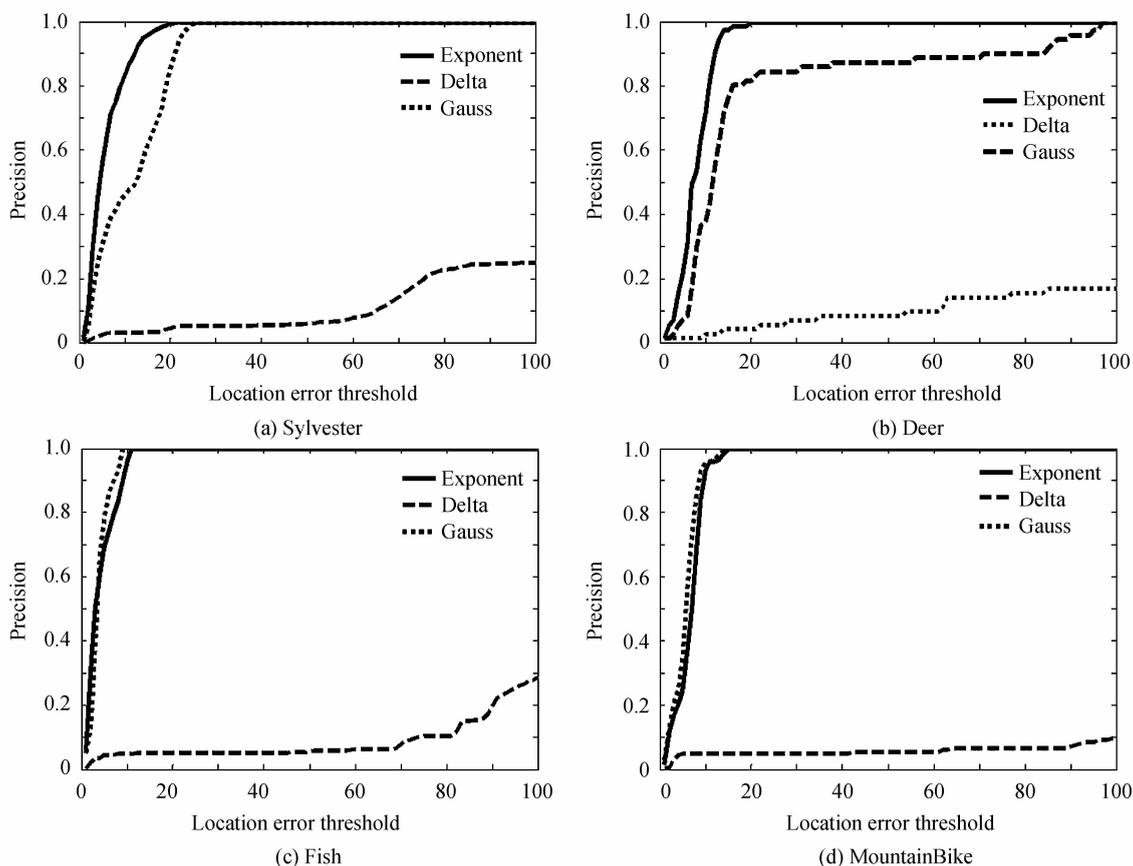


图 3 不同分布的距离准确度,4 个视频在 3 个不同的空间分布上进行测试,点线为  $\delta$  分布,实线和虚线分别为偶双边指数分布和高斯分布

Fig. 3 Distance precision of different distribution, four videos test in three different spatial distribution, the dotted line is  $\delta$  distribution, the full line and the imaginary line are accidentally bilateral exponential distribution and gaussian distribution respectively

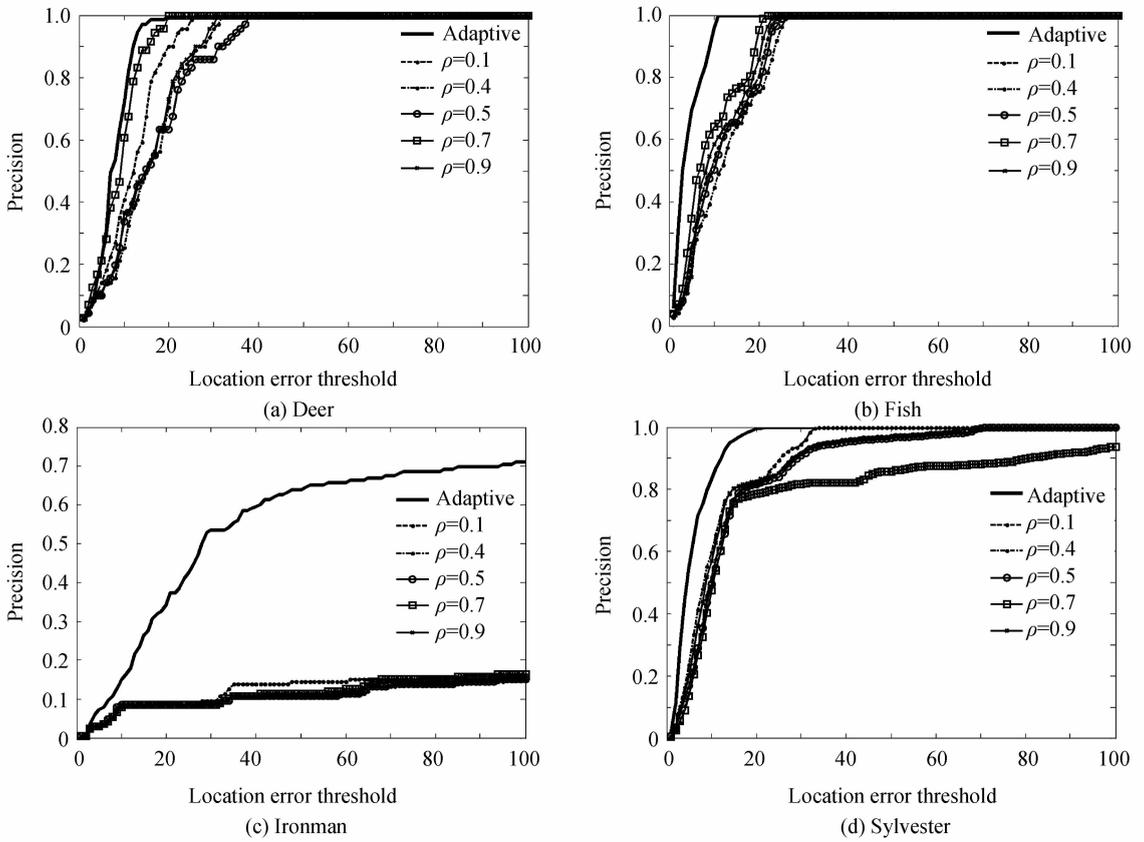


图 4 采用不同的参量进行更新,选取 4 个视频对固定阈值与自适应更新进行比较

Fig. 4 With different parameters are updated, choose four video for fixed threshold comparing with the adaptive update

### 3.5 结果讨论

将所有视频跟踪结果与 LOT<sup>[13]</sup>、CT<sup>[14]</sup>、TLD<sup>[3]</sup>、DFT<sup>[15]</sup>、VTD<sup>[16]</sup>、VTS<sup>[17]</sup>、LSK<sup>[18]</sup>、Struck<sup>[19]</sup>、CXT<sup>[20]</sup>、ACA<sup>[11]</sup>、STC<sup>[22]</sup>、KCF<sup>[10]</sup> 和 CSK<sup>[21]</sup> 算法进行比较,本文提出的算法表示为 SPC 跟踪算法.如算法 1 所示,本

文算法的实现过程非常简洁,而其余跟踪算法包含大量的复杂运算,例如,Struck 在很多不同种类的特征和不断增多的支持向量上运行,TLD 利用一组含有许多参量结构规则的级联分类器进行再检测.

表 2 和表 3 分别显示了阈值为 20 像素的距离准

表 2 距离准确度(DP) (%)

Table 2 Distance precision(DP) (%)

Video	LOT	DFT	TLD	VTD	VTS	LSK	Struck	CXT	ACA	STC	KCF	CSK	CT	SPC
CarDark	61	54	63	<b>73*</b>	<b>100</b>	<b>100</b>	<b>100</b>	71	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
Deer	18	31	<b>100</b>	6	4	32	<b>99*</b>	94	<b>100</b>	4	89	<b>100</b>	4	<b>100</b>
FaceOcc2	45	81	61	87	85	56	<b>94*</b>	85	61	86	84	<b>97</b>	51	65
Fish	11	86	86	66	93	33	<b>100</b>	<b>98*</b>	40	<b>100</b>	<b>100</b>	4	33	<b>100</b>
Girl	64	30	91	<b>94*</b>	85	47	<b>100</b>	82	86	92	86	58	68	<b>100</b>
Mountainbike	69	35	48	96	<b>100</b>	85	<b>100</b>	29	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>97*</b>	<b>100</b>
Suv	80	5	<b>92*</b>	47	52	80	57	91	53	57	<b>98</b>	57	63	<b>98</b>
Sylvester	75	42	93	81	82	27	<b>99*</b>	66	94	90	84	92	60	<b>100</b>
Walking2	43	<b>75*</b>	58	49	50	45	<b>100</b>	41	50	<b>100</b>	<b>100</b>	<b>100</b>	42	<b>100</b>
Women	14	<b>95*</b>	80	20	20	20	<b>96</b>	35	25	94	94	25	20	94
Basketball	96	89	4	<b>100</b>	<b>99*</b>	43	12	3	0	59	9	<b>100</b>	36	63
Car4	5	23	<b>71</b>	<b>37*</b>	36	0	32	25	21	24	27	21	9	24
CarScale	45	59	67	<b>72</b>	<b>68*</b>	65	53	66	56	58	59	58	44	58
Coke	10	9	69	15	23	24	<b>95</b>	53	62	80	13	<b>87*</b>	11	82
Dog1	<b>96*</b>	56	75	69	68	85	71	<b>99</b>	73	71	71	73	83	72
Dudek	54	61	55	<b>88</b>	<b>86*</b>	73	71	78	61	50	0	64	39	47
Freeman1	51	<b>93</b>	53	<b>72*</b>	<b>72*</b>	36	57	55	39	38	39	39	52	58
Ironman	11	10	<b>100</b>	25	23	14	7	3	17	14	16	16	11	<b>34*</b>
Mhyang	26	76	<b>100</b>	91	95	99	<b>100</b>	<b>100</b>	92	89	99	99	0	96
Skating1	23	19	36	55	<b>59*</b>	<b>68</b>	38	8	54	54	44	53	16	46
mean	45	51	67	70	62	65	52	<b>74*</b>	59	59	68	66	42	<b>77</b>

注:粗体代表最好的结果,\*粗体代表次好结果

表 3 成功率(SR)(%)  
Table 3 Success rate (SR)(%)

Video	LOT	DFT	VTD	VTS	LSK	ACA	STC	KCF	CSK	CT	SPC
CarDark	56	34	68	<b>100</b>	<b>100</b>	<b>99*</b>	<b>100</b>	72	<b>99*</b>	9	<b>99*</b>
Deer	6	31	4	4	34	<b>100</b>	4	<b>89*</b>	<b>100</b>	4	<b>100</b>
FaceOcc2	34	<b>100</b>	<b>99*</b>	97	72	62	90	<b>100</b>	<b>100</b>	83	86
Fish	11	86	64	<b>98*</b>	33	40	<b>100</b>	<b>100</b>	4	33	<b>100</b>
Girl	55	25	65	53	34	46	<b>73*</b>	67	40	15	<b>96</b>
Mountainbike	69	35	<b>100</b>	<b>100</b>	90	<b>100</b>	96	<b>98*</b>	<b>100</b>	<b>98*</b>	96
Suv	78	5	55	58	82	53	58	<b>98*</b>	57	31	<b>99</b>
Sylvester	68	41	80	81	26	71	70	<b>82*</b>	72	51	<b>99</b>
Walking2	39	38	<b>40*</b>	<b>40*</b>	<b>49</b>	38	<b>40*</b>	39	39	38	39
Woman	8	<b>93*</b>	18	17	19	24	91	91	24	14	<b>94</b>
Basketball	89	72	<b>92*</b>	<b>93</b>	34	0	55	9	87	28	41
Car4	5	26	35	35	6	28	<b>37*</b>	<b>38</b>	28	25	36
CarScale	46	45	48	<b>49*</b>	<b>63</b>	45	45	44	45	44	45
Coke	9	9	14	14	16	48	43	10	<b>74</b>	10	<b>51*</b>
Dog1	<b>99</b>	52	70	69	<b>91*</b>	65	65	65	65	58	65
Dudek	62	80	<b>100</b>	<b>100</b>	93	<b>96*</b>	80	0	95	85	86
Freeman1	19	18	<b>21*</b>	19	<b>33</b>	15	13	16	14	10	19
Ironman	8	4	<b>16*</b>	<b>20</b>	11	13	13	14	13	10	<b>20</b>
Mhyang	26	78	95	97	<b>100</b>	92	<b>98*</b>	<b>100</b>	<b>100</b>	0	<b>100</b>
Skating1	28	16	<b>57*</b>	53	<b>66</b>	37	36	36	37	12	35
mean	41	44	57	<b>60*</b>	53	54	<b>60*</b>	58	<b>60*</b>	33	<b>70</b>

注:粗体代表最好的结果,\*粗体代表次好结果

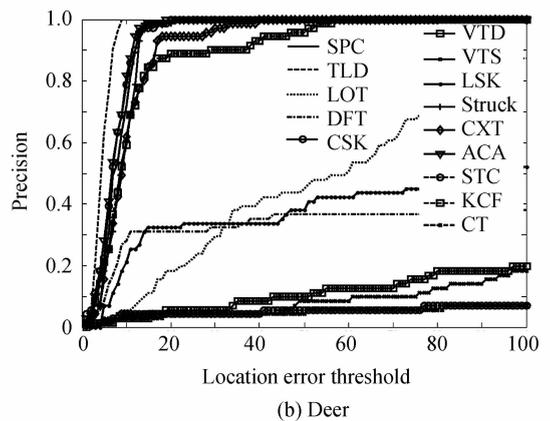
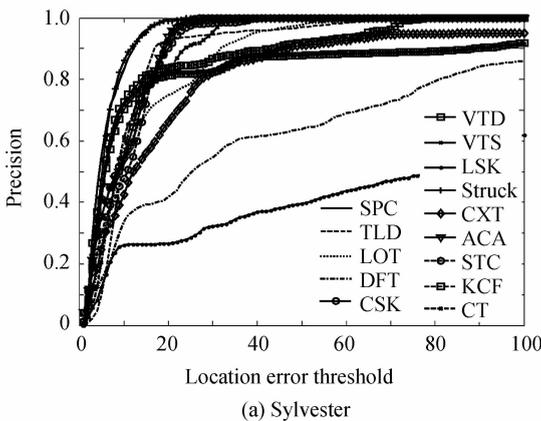
精度和成功率.与当前最先进跟踪算法相比,SPC跟踪算法在所有的视频中分别将以上三个指标从74%和60%提升到了77%和70%.Struck在最近评估中获得最好的成绩,同样在本文的评估中优于其他算法.但是SPC跟踪算法在20像素阈值距离准确度上超出Struck跟踪算法4.05%,并且运行速度高出25倍帧率.

图5显示了在20个测试视频中,阈值为20像素的距离准确度.在距离准确度曲线中,最好的跟踪算法是SPC跟踪算法.表2显示了在20个测试视频中,SPC算法与当前最先进算法在距离准确度上的比较.最好的两个结果分别表示为粗体和\*粗体.SPC跟踪算法在7个视频中打败了其他算法,并且在7个视频

中获得次好的成绩.由于视频序列长度不一致以及准确度计算公式差异,因此在一些视频中平均准确度小于阈值20像素的准确度.

表3和图6是跟踪成功率.为了评估的公平性,只与不估计尺度的跟踪算法进行比较.结果显示,本文提出的算法在所有视频中获得最好的成绩.

KCF、STC、SPC跟踪算法都采用FFT进行快速计算.但KCF跟踪算法采用非常耗时的核运算,STC跟踪算法在权重图像上计算相关滤波,也比较耗时.而SPC跟踪算法只计算一次空间相位相关,因此可以获得更快的速度.FPS如表4.



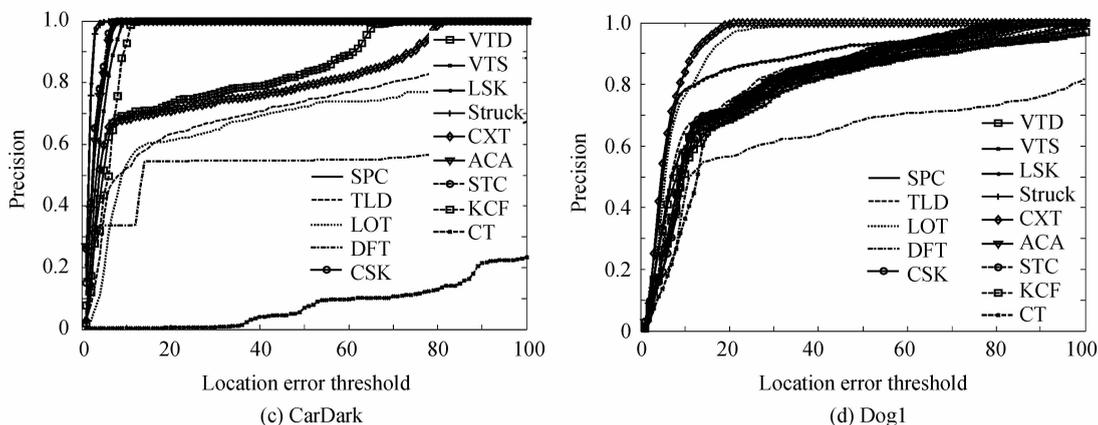


图 5 距离准确度. 横轴表示阈值, 纵轴表示距离准确度. (a)、(b)、(c) 分别为挑选的视频序列, (d) 为所有 20 个视频的距离准确度

Fig. 5 Distance precision. the horizontal axis shows the threshold, the vertical axis shows distance precision. (a), (b), (c) are video sequence selected respectively, (d) is the distance precision of the twenty videos

表 4 中值 FPS

Table 4 Median value FPS

Algorithm	LOT	DFT	TLD	VTD	VTS	LSK	Struck	CXT	ACA	STC	KCF	CSK	CT	SPC
FPS	0.51	10.31	22.7	3.0	3.0	3.8	10.4	10.3	41.97	67.77	29.52	80.24	76.93	49.14

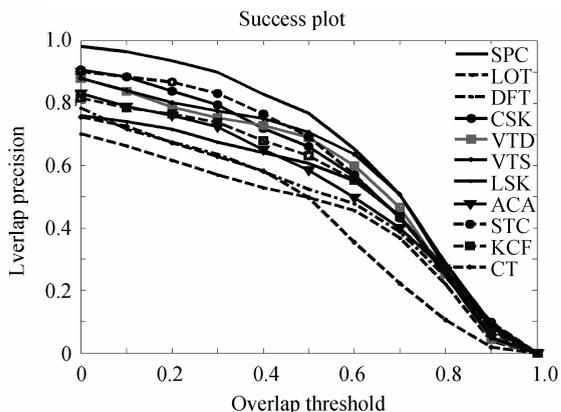
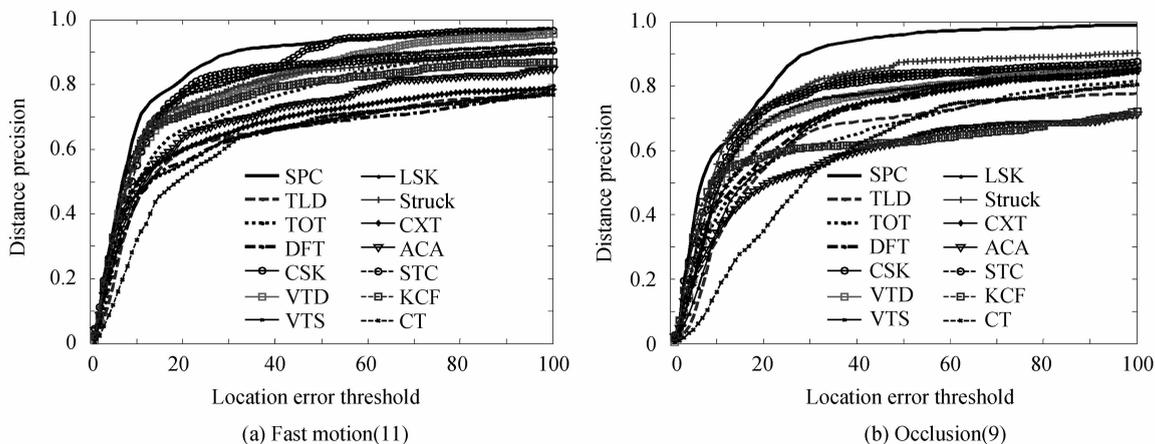


图 6 成功率. 横轴表示覆盖率阈值, 纵轴表示成功率

Fig. 6 Success rate, the horizontal axis shows the threshold of the coverage rate, the vertical axis shows success rate



序列属性实验: 鲁棒的跟踪算法能够适应目标的变化, 根据表 1 的难度因子, 将 20 个视频序列分为四个主要类型: 快速运动、遮挡、光照变化和背景复杂, 结果如图 7 所示.

快速移动: 当目标快速移动时, 基于空间卷积的 SPC 跟踪算法, 性能明显优于其他跟踪算法. 原因在于搜索区域比较大并且采用判别模型可以从复杂背景中区分出目标. 图 7(a) 显示了 11 个快速移动视频的距离准确度结果.

遮挡: 图 7(b) 显示了 9 个遮挡视频的距离准确度曲线. 在出现遮挡时, ACA、KCF 和 STC 的性能显著下滑, 主要是由于模板更新系数为固定值, 而 SPC 算法根据 PSR 自适应更新模板, 当目标被背景遮挡时, PSR 更新参量变小, 因此, 模板受非目标干扰较小.

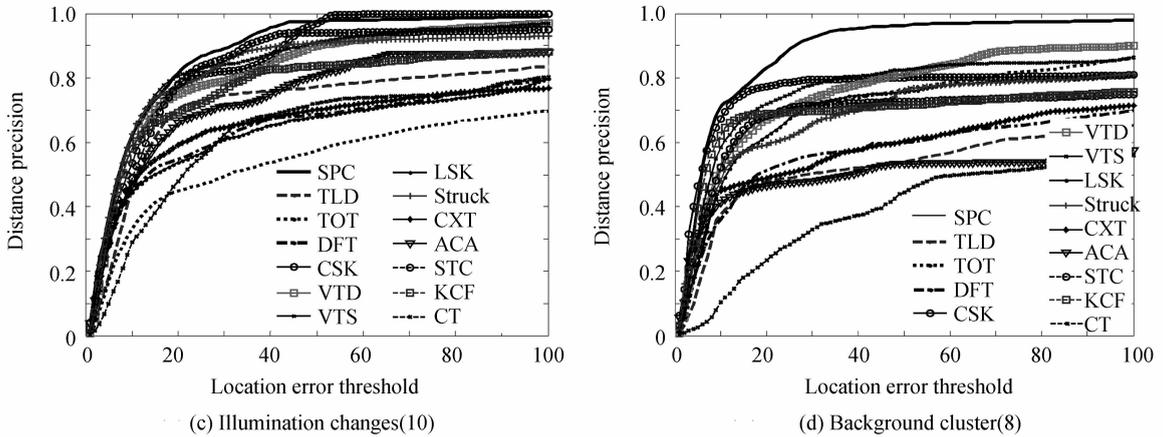


图7 不同难度因子的距离准确度曲线.将20个视频分成4个类型,横轴表示为帧数,纵轴表示为每一类难度因子的距离准确度

Fig.7 Distance precision curve in different difficulty factor. Twenty videos are divided in four types, the horizontal axis shows frame, the vertical axis shows distance precision of every degree of difficulty

光照变化:图7(c)显示了10个光照变化视频的距离准确度曲线.在评估视频中,有大量的光照变化.SPC算法在距离准确度和成功率上,均达到很好的效果.SPC算法采用相位相关,频域相位信息对光照不敏感,因此当光照和姿态变化使目标外观发生剧烈变化时,也能够准确定位目标.

背景复杂:图7(d)显示了8个背景复杂视频的距离准确度曲线.除了SPC跟踪算法,其他跟踪算法都受背景复杂影响,甚至是Struck.由于TLD和CT均采样很多的负样本,导致其跟踪性能下降.而SPC跟踪算法采用空间窗函数对目标周围背景进行平滑处理,并且在目标周围隐式包含负样本,因此能够获得优异的性能.

## 4 结论

本文考虑目标的空间上下文关系,设计相应的响应函数,改进传统的相位相关算法,消除噪音和杂波的影响,提出了空间相位相关,通过频域的相位差来计算目标在空域的位置平移.目标模板的自适应更新能够避免漂移问题.通过大量的实验,在最近公开的测试视频中,与当前最先进的算法进行比较,本文算法能够获得平均每秒50帧的跟踪速度,同时在准确度和精确度上都能够获得优异的性能.本文跟踪算法仅仅利用相位信息估计目标位移,然而频域幅度谱反映了目标的尺度和旋转,若并行考虑相位谱和幅度谱,可解决视觉目标跟踪中位置估计和尺度估计两方面问题,这将是下一步的工作.

### 参考文献

[1] BAKER S, MATTHEWS I. Lucas-kanade 20 years on: a unifying framework[J]. *International Journal of Computer Vision*, 2004, **56**(3): 221-255.  
 [2] KALAL Z, MIKOLAJCZYK K, MATAS J. Forward-backward error: automatic detection of tracking failures[C].

International Conference on Pattern Recognition, Istanbul, Turkey, IEEE, 2010: 23-26.  
 [3] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **6**(1): 1-14.  
 [4] COLLINS R T, LIU Y, LEORDEANU M. Online selection of discriminative tracking features[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1631-1643.  
 [5] AVIDAN S. Support vector tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 1064-1072.  
 [6] GRABNER H, GRABNER M, BISCHOF H. Real-time tracking via on-line boosting [C]. In Proceedings British Machine Vision Conference (BMVC), Edinburgh, UK, Springer, 2006, **1**(5): 47-56.  
 [7] LUO Hai-bo, SHI Ze-lin, CHEN Yong-hong. Phase congruency image and its application in target tracking[J]. *Acta Photonica Sinica*, 2010, **39**(3): 547-552.  
 罗海波, 史泽林, 陈永红. 相位一致性图像及其在目标跟踪中的应用[J]. *光子学报*, 2010, **39**(3): 547-552.  
 [8] BOLME D S, DRAPER B A, BEVERIDGE J R. Average of synthetic exact filters [C]. IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA, IEEE, 2009: 2015-2022.  
 [9] DAVID S, BOLME J, ROSS B B A, et al. Visual object tracking using adaptive correlation filters [C]. CVPR, San Francisco, USA, IEEE, 2010: 2544-2550.  
 [10] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(1): 996-1010.  
 [11] DANELLJAN M, SHAHBAZ KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, USA, IEEE, 2014.  
 [12] The official homepage of the VOT2014 benchmark, VOT2014 [EB/OL]. (2014-09-06) [2014-10-22]. <http://www.votchallenge.net/vot2014/>.  
 [13] ORON S, BAR-HILLEL A, LEVI D, et al. Locally orderless tracking [J]. *International Journal on Computer Vision*, 2014, **70**(6): 1573-1405.

- [14] ZHANG K, ZHANG L, YANG M H. Real-time compressive tracking [C]. Proceedings of 11th European Conference on Computer Vision, Firenze, Italy, Springer, 2012; 864-877.
- [15] SEVILLA-LARA L, LEARNED-MILLER E. Distribution fields for tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, RhodeIsland, USA, IEEE, 2012; 1910-1917.
- [16] KWON J, LEE K M. Visual tracking decomposition[C]. IEEE Conference on Computer Vision and Pattern Recognition, CVPR, San Francisco, USA, IEEE, 2010; 1269-1276.
- [17] KWON J, LEE K M. Tracking by sampling trackers[C]. International Conference on Computer Vision, Barcelona, Spain, IEEE, 2011; 1195-1202.
- [18] LIU Bai-yang, HUANG Jun-zhou, KULIKOWSK CASIMIR, *et al.* Robust visual tracking using local sparse appearance model and k-selection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(12): 2968-2981.
- [19] SAM H, AMIR S, PHILIP H S, Torr. struck; structured output tracking with kernels[C]. ICCV. Barcelona, spain, IEEE, 2011; 263-270.
- [20] DINH T B, VO N, MEDIONI G. Context tracker: Exploring supporters and distracters in unconstrained environments[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, IEEE, 2011; 1177-1184.
- [21] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[C]. ECCV, Firenze, Italy, Springer, 2012(4): 702-715.
- [22] ZHANG Kai-hua, ZHANG Lei, Ming-hsuan Yang, *et al.* Fast visual tracking via dense spatio-temporal context learning [C]. ECCV, Firenze, Italy; Springer, 2014(5): 127-141.