

基于时空上下文模型的 RGB-D 序列目标跟踪方法

孔德慧^{1,2}, 荣子豪^{1,2}, 贾思宇^{1,2}, 王少帆^{1,2}, 尹宝才^{1,2}

(1. 北京工业大学信息学部, 北京 100124; 2. 北京工业大学多媒体与智能软件技术北京市重点实验室, 北京 100124)

摘要: 为了实现更为精确的视频目标跟踪, 提出一种以时空上下文模型为基础的 RGB-D 序列目标跟踪算法. 通过引入更新模板的深度信息, 该模型精准地区分了输入序列的目标区域与背景区域, 实现了深度权值和颜色权值的融合; 基于目标序列的深度及目标动量计算, 该模型有效地实现了尺度更新与遮挡处理. 通过在 RGB-D 图像序列数据集上的详细实验评估, 该时空上下文模型相对于其他先进的同类方法表现出更好的性能. 因此, 该方法实现了更为精确可靠的视频目标跟踪.

关键词: 计算机视觉; 目标跟踪; 机器学习; RGB-D; 时空上下文; 目标动量

中图分类号: U 461; TP 308

文献标志码: A

文章编号: 0254-0037(2021)03-0224-07

doi: 10.11936/bjutxb2020100005

Object Tracking in RGB-D Sequences Using a Spatio-Temporal Context Model

KONG Dehui^{1,2}, RONG Zihao^{1,2}, JIA Siyu^{1,2}, WANG Shaofan^{1,2}, YIN Baocai^{1,2}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: To improve the precision of object tracking in videos, this work presented an RGB-D tracking method using a spatial-temporal context (STC) model. By introducing depth data, STC can clearly distinguish target from background in the context, and perform effective fusion of the depth weights and color weights. At the same time, based on the depth information and the target momentum, the proposed method is capable of adjusting scale and handling occlusions. As a result, the proposed tracker is able to produce precise prediction of target locations even when the target object is under severe occlusion. Comprehensive evaluations on challenging datasets demonstrate that the proposed tracker gives favorable performance over several state-of-the-art counterparts. Consequently, the proposed method in this work is capable of achieving more precise and reliable object tracking in videos.

Key words: computer vision; target tracking; machine learning; RGB-D; spatial-temporal context (STC); target momentum

面向 RGB 图像序列的目标跟踪方法在计算机视觉领域的诸多问题上应用广泛. 然而, 考虑到现

实应用环境中的很多影响因素, 例如目标对象被遮挡、外观形变、光照条件多变等, 面向 RGB 图像

收稿日期: 2020-10-15

基金项目: 国家自然科学基金面上资助项目(61772049); 北京市自然科学基金资助项目(4202003)

作者简介: 孔德慧(1968—), 女, 教授, 主要从事计算机图形学、模式识别、人工智能方面的研究, E-mail: kdh@bjut.edu.cn

序列的目标跟踪仍然存在精度低、鲁棒性差的问题。近年来,伴随深度传感器技术的进步,许多高效捕获场景深度信息的设备(如微软 Kinect 相机)可使人们用较低的成本高效、准确地获得场景的深度信息。相应地,面向 RGB-D 图像序列的目标跟踪技术也随之发展。

目标跟踪技术大体可分为生成式目标跟踪和判别式目标跟踪。生成式目标跟踪算法通过学习一种表观模型来表达跟踪目标,并在图像范围内搜索最好的匹配结果。此类的经典算法包括基于均值漂移(mean-shift)的跟踪算法^[1]和分块跟踪算法^[2]等^[3-6]。而判别式方法将目标跟踪问题归为二分类问题,通过局部搜索,并判断图像块是目标还是背景,实现目标定位。这其中囊括了基于 online-boosting 的跟踪算法^[7]和基于在线多实例的跟踪算法等^[7-11]。目前,生成式和判别式算法逐渐混合,自适应学习模板匹配算法也逐渐兴起。

2014年,文献[12]在时空上下文(spatio-temporal context, STC)学习算法基础上提出了加权的时空上下文(weighted spatio-temporal context, WSTC)学习算法。该算法借助前后帧的光流估计,通过在时空上下文间进行搜索实现上下文信息的分级,达到了上下文信息中其他物体和目标的分离,提升了算法辨别目标和背景的能力。

最近几年,深度信息在计算机视觉领域的物体检测、物体分割等方面广泛应用,但是在基于深度信息的视频跟踪领域中,目前成熟的算法还很少^[13],且基于深度图的跟踪算法^[14-15]刚开始兴起。如何更好地利用深度信息这一新的数据,并和 RGB 信息融合,成为摆在人们面前的问题。

2015年 Chen 等^[16]提出了在 RGB 图中加入深度信息,形成三维空间上下文信息,从而实现在三维空间的快速上下文学习跟踪算法。本文也采取基于上下文信息搜索的做法,在原有基于 RGB 信息的跟踪算法的基础上,通过获取深度图上下文和目标的信息并加入深度权重,实现了目标和背景的分离,并因此在背景变换、杂波发生、物体交叉、连续遮挡的情况下,仍可保证目标的准确跟踪。

本文的主要贡献如下:

- 1) 实现了一个基于自适应深度信息模板的上下文先验模型(context prior model)的跟踪算法。
- 2) 实现了一个基于深度时域变化估算的遮挡检测及处理算法。

- 3) 实现了一个基于深度信息的尺度更新算法。

1 STC 模型

在 STC 算法中,通过计算彩色图的时空上下文,采用自适应学习的方法将上下文信息转换为概率置信图,再搜索概率置信图得到目标位置,并进行尺度更新。

一般来说,跟踪方法需要计算出一个用以估计目标物体可能位置的置信图

$$m(x) = P(x|o) \quad (1)$$

式中: $m(\cdot)$ 为目标位置置信图,可表达为条件概率分布 $P(x|o)$; $x \in \mathbb{R}^2$ 代表目标方位; o 代表目标在场景中是否存在。在最近的已处理帧中,目标的位置 x^* 已定。据此,局部上下文特征 X^c 定义为

$$X^c = \{c(z) = (I(z), z) | z \in \Omega_c(x^*)\}$$

式中: $I(z)$ 为位置 z 处的图像灰度; $\Omega_c(x^*)$ 是位置 x^* 的近邻域,为目标物体大小的 2 倍。通过边缘化联合概率 $P(x, c(z)|o)$, 物体位置的可能性函数(1)可转化为

$$m(x) = P(x|o) = \sum_{c(z) \in X^c} P(x, c(z)|o) = \sum_{c(z) \in X^c} P(x|c(z), o)P(c(z)|o) \quad (2)$$

式中: $P(x|c(z), o)$ 描述目标物体位置和它上下文的空关系; $P(c(z)|o)$ 为上下文的先验概率。按此计算求得概率置信图,进而得到目标物体的位置。然而,仅仅依靠低层次的彩色上下文特征,不足以将目标和背景完全区分,因而在处理复杂的遮挡和背景杂乱等情况时效果并不理想。

本文通过在上下文模型构建中引入深度信息,实现了基于自适应深度信息模板的上下文先验模型;在原有 STC 跟踪算法流程的基础上,添加了基于深度信息的尺度更新算法及遮挡检测、处理机制。通过将目标跟踪状态分为未被严重遮挡和被严重遮挡 2 种情况,采用不同的方法分别处理,以实现更好的适应能力。本目标跟踪方法的总体流程如图 1 所示。

1.1 基于自适应深度信息模板的上下文先验模型

本跟踪算法的核心是自适应的深度信息模板和上下文模型,其中深度信息模板又是计算上下文模型的关键。对于深度信息模板,依照第 1 帧给定的目标检测框的方位和尺寸,求出目标检测框范围内的深度区间 $[d_1, d_2]$,从而推算出跟踪目标所处深度范围。在 $D_1 = [d_1, d_2]$ 范围内,建立深度信息直方图,直方数为 n 。分别计算每个直方元组中元素个

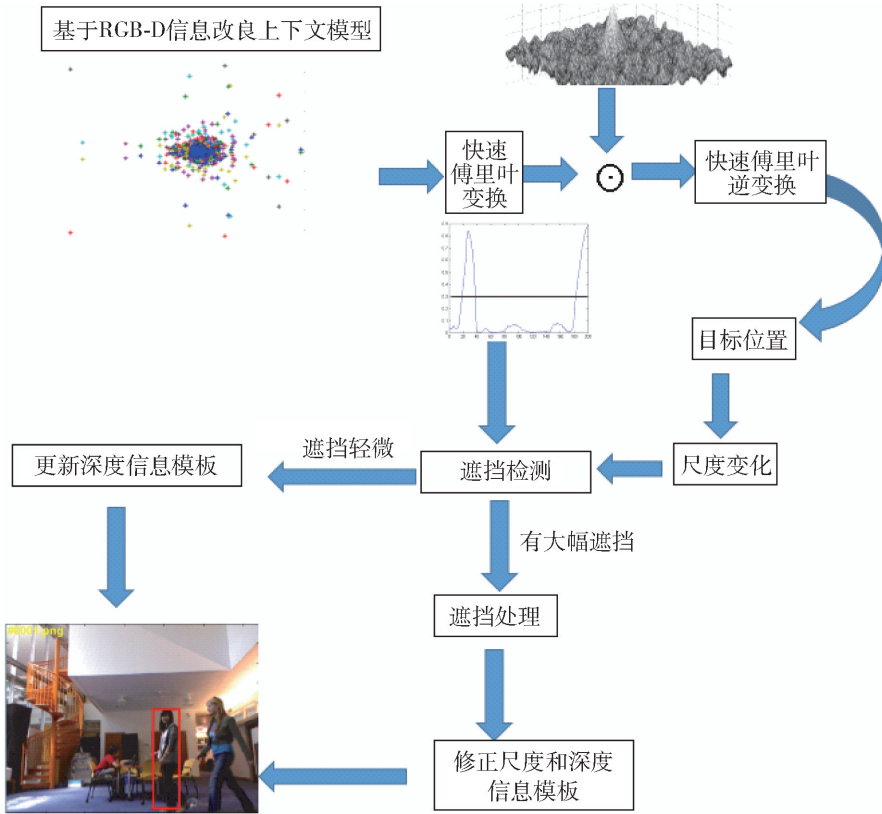


图1 时空上下文模型的算法流程

Fig. 1 Flowchart of spatio-temporal context model

数,并寻找出元素最多者(其元素个数计为 t_k),并求出其相对元素总数 t 的占比,记为

$$\text{rate} = \frac{t_k}{t} \quad (3)$$

如果 $\text{rate} \geq c_3$ (c_3 是超参数,实验中取0.5),则保存范围为 $[d'_1, d'_2] = [d_1, d_2]$,作为目标的初始深度模板. 如果 $\text{rate} < c_3$,则需要对该范围进行一定程度的放缩,即

$$D_1 = [d''_1, d''_2] = [d'_1 - ck', d'_2 - ck'] \quad (4)$$

式中 c 是常数, k' 从1开始渐增并重复以上直方图统计,直到 $\text{rate} \geq c_3$,算法停止.

而在之后的跟踪过程中,本方法在每帧跟踪计算结束后,根据当前跟踪目标检测框内的深度情况,对深度信息模板进行更新. 在获取当前帧后,计算目标检测框内深度均值 avg_t ,再计算前后2帧均值深度差

$$\text{dvg}_{t-1} = \text{avg}_t - \text{avg}_{t-1} \quad (5)$$

则更新后的深度信息模板有

$$D_t = D_t + \frac{1}{n} \sum_{k=t-1}^{t-n} \text{dvg}_k \quad (6)$$

通过式(6)可以得到每帧更新后的深度信息模板;而在跟踪计算中,由于并不清楚当前帧目标实际

情况,因而采用上一帧 $D_{t-1} = [d_1^{t-1}, d_2^{t-1}]$ 的先验深度模板信息.

在式(2)中,上下文先验概率与上下文的表观模型有关. 在原STC当中,上下文先验概率被简单表示为

$$P(c(z) | o) = I(z) w_\sigma(z - x^*) \quad (7)$$

式中: $I(\cdot)$ 是上下文的图像灰度值; $w_\sigma(\cdot)$ 是高斯权重函数. 而在本文当中,为了利用深度信息实现对目标和背景的区分,将式(7)改写为

$$P(c(z) | o) = c_1 I(z) w_\sigma(z - x^*) + c_2 \mathbf{W}_{\text{de}[1]z[2]} \quad (8)$$

式中深度权重矩阵 \mathbf{W}_d 各分量定义为

$$W_{dij} = \alpha \left(\frac{\lambda}{\min(\|A_{ij} - d_1^{t-1}\|, \|A_{ij} - d_2^{t-1}\|)} + \frac{1 - \lambda}{\|A_{ij} - \text{avg}_{t-1}\|} \right) \quad (9)$$

式中: A_{ij} 为当前上下文中位置 (i, j) 的深度值; λ 为系数; α 为归一化参数. 通过融合深度和彩色权重,本方法进一步完善了上下文表观模型,使得目标和背景的区分更加明显,进而在背景复杂、剧烈变化或者背景和目標物体出现相似纹理特征时依然可以实现区分.

在定义了上述置信图涉及的诸组成概率分布

后,之后跟踪流程同STC算法一样.

时空上下文 H_t^{stc} 的更新方式为

$$H_{t+1}^{\text{stc}} = (1 - \rho) H_t^{\text{stc}} + \rho h_t^{\text{sc}} \quad (10)$$

式中: ρ 为更新率; h_t^{sc} 为第 t 帧的空间上下文.通过采用快速傅里叶变换(fast Fourier transform, FFT),第 $t+1$ 帧的置信图 $m_{t+1}(x)$ 的计算式可表示为

$$m_{t+1}(x) = F^{-1}(F(H_{t+1}^{\text{stc}}) \odot F(P(c(z)|o))) \quad (11)$$

式中: F 为傅里叶变换; F^{-1} 为傅里叶逆变换; \odot 为Hadamard积.

1.2 基于深度信息的尺度更新

本节讨论并分析目标物体在当前帧位置确定以后目标检测框尺度的变化.这里假设目标物体未被严重遮挡;而当物体可能被严重遮挡导致计算不可信时,则采取1.3节中介绍的遮挡检测和处理的办法.

在处理第1帧时,可以得到目标物体实际的尺度 S_1 ,以及根据式(4)得到 $D'_1 = (h', w')$.定义比值变量

$$d_1 = \frac{S_1}{D'_1} \quad (12)$$

在之后的跟踪计算中,第 t 帧的计算有

$$\begin{cases} S_t = (1 - \lambda) S'_t + \lambda \bar{S}_{t-1}^n \\ S'_t = D'_t * d_t \\ d_t = (1 - \lambda) d'_t + \lambda \bar{d}_{t-1}^n \\ d'_t = \frac{S_{t-1}}{D'_t} \\ \bar{S}_{t-1}^n = \frac{1}{n} \sum_{k=1}^n S_{t-k} \\ \bar{d}_{t-1}^n = \frac{1}{n} \sum_{k=1}^n d_{t-k} \end{cases} \quad (13)$$

式中: S_t 为第 t 帧计算所得目标尺度; D'_t 为第 t 帧模糊尺度(模糊尺度是在上下文计算中,依据深度信息模板划分出的目标大概的尺度).

1.3 基于深度时域变化估算的遮挡检测及处理

为处理目标被遮挡的问题,本方法在STC算法中获取RGB上下文信息,使得整个算法对于局部遮挡有了一定的容错性.但是若出现明显的大幅遮挡、交叉遮挡甚至长时间连续遮挡,包括STC算法在内,绝大多数基于RGB信息的跟踪算法都会预测错误.

在深度图信息引入之前,大多数的跟踪算法

并不会直接判断遮挡,而是采取自适应的方法来处理遮挡的情况.在引入深度信息之后,在深度图上对于遮挡的检测甚至处理成为了可能.快速上下文跟踪算法通过保存连续多帧深度信息并依据深度变化的连续性来断定遮挡是否发生^[16].若目标在某帧出现遮挡,则将最近的深度连续的位置用作新的目标中心.若出现长时间连续遮挡,且目标存在移动动量,则这种找寻方法有不稳定性.

考虑到STC原始算法对于遮挡有一定的容错性,所以在目标没有被大幅遮挡时,由算法自适应解决即可.本方法通过计算遮挡率(∂)来评估遮挡.遮挡率定义为

$$\partial = \frac{n'_t}{n_t} \quad (14)$$

式中: n'_t 是复合深度信息模板中目标检测框内像素的个数; n_t 是目标检测框像素总个数.当 ∂ 小于常数 σ 时,不使用遮挡处理算法,因为计算得到的当前目标物体的位置和尺度可信.而当 $\partial \geq \sigma$ 时,目标位置跟踪计算所得值的不可信度大增,需要进行遮挡处理.由于物体的移动轨迹和运动状态具有连续性,因而可以根据连续多帧间目标物体移动的动量来预测第 t 帧内被遮挡目标的位置.

定义第 t 帧的动量 Mt_t 为

$$Mt_t = x_{t+1}^* - x_t^* \quad (15)$$

则可大概推测当前第 t 帧被遮挡目标的位置为

$$\begin{aligned} x_t^* &= x_{t-1}^* + \frac{(Mt_{t-2} + Mt_{t-3})}{2} = \\ x_{t-1}^* &+ \frac{(x_{t-1}^* - x_{t-2}^* + x_{t-2}^* - x_{t-3}^*)}{2} = \\ &x_{t-1}^* + \frac{(x_{t-1}^* - x_{t-3}^*)}{2} \end{aligned} \quad (16)$$

同时由于被遮挡目标的尺寸变化无法准确预测,当前帧的目标尺寸继承上一帧的尺寸,且保持之前的尺度变化参数和深度模板更新变化量,即

$$\begin{cases} S_t = S_{t-1} \\ d_t = d_{t-1} \\ \text{dvg}_{t-1} = \text{dvg}_{t-2} \end{cases} \quad (17)$$

经过式(16)的处理后,跟踪目标的长时连续遮挡、短时交叉遮挡问题均可得到解决;而当 $\partial < \sigma$ 后,则可以继续执行跟踪算法流程.图2展示了本算法遮挡检测和处理的实验结果,图(a)反映了目标被遮挡的遮挡率在视频序列上的变化过程,相应主观跟踪效果如图(b)所示.

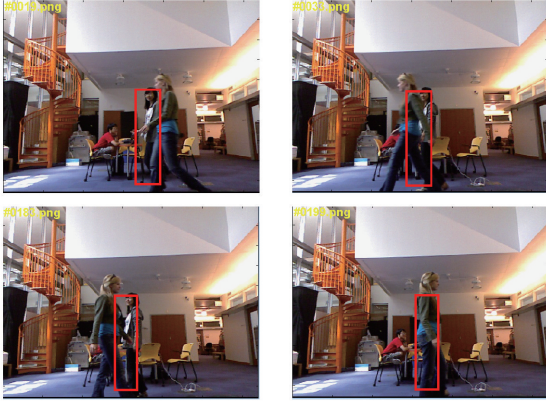
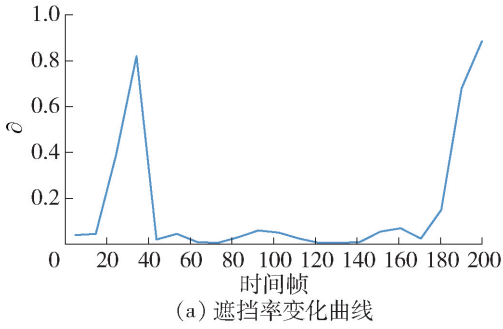


图2 遮挡检测及处理效果

Fig.2 Illustration of occlusion detection and handling

2 实验结果与分析

本文基于 Princeton RGB-D 数据集的 22 组 RGB-D 数据进行跟踪算法的性能评估,这些数据涵盖遮挡 (heavy occlusion)、交叉遮挡 (cross shade)、背景杂波 (background clutter)、尺度变化 (scale variation)、剧烈形变 (drastic deformation) 等情况,总

计 4 335 帧. 此外,对该数据集中不含标注的数据进行了人工标注. 对于目标被遮挡的情况,本文以目标实际所在位置和尺度为准. 本文以 STC 算法为基准进行了对比,且与最近发表的基于 RGB 信息的跟踪算法^[9,7]进行了对比. 实验所用硬件配置为 Intel core i7-4790 CPU 及 8G RAM.

实验结果的量化评价采用当前跟踪算法测试常用的 2 个标准:中心位置误差 (center location error, CLE) 和跟踪成功率 (success rate, SR);评价结果依赖于每帧标注信息. CLE 度量跟踪结果的目标中心点与实际目标中心点距离. SR 首先定义每帧的得分 score,且

$$\text{score} = \begin{cases} \frac{\text{area}(\text{ROI}_{T_i} \cap \text{ROI}_{G_i})}{\text{area}(\text{ROI}_{T_i} \cup \text{ROI}_{G_i})}, & \text{其他情况} \\ 1, & \text{若 ROI}_{T_i} \text{ 与 ROI}_{G_i} \text{ 完全重叠} \\ 0, & \text{若 ROI}_{T_i} \text{ 与 ROI}_{G_i} \text{ 不重叠} \end{cases} \quad (18)$$

式中:ROI_{T_i}是第 *i* 帧目标跟踪框;ROI_{G_i}是第 *i* 帧实际目标框. 当 ROI_{T_i}与 ROI_{G_i}完全重合时,score 为 1;完全不重合时,score 为 0;部分重合时,score 为 0~1. 在所有实验中,重叠比例 (overlap ratio) 阈值为 $R_l = 0.5$. 若 $\text{score} > R_l$,则该帧跟踪成功,否则跟踪失败. SR 最终定义为整个视频序列上跟踪的成功率.

实验的量化结果见表 1. 本方法在绝大多数的情况下效果比其他方法好,特别在长时遮挡的数据上有显著提升. 在 22 组数据的 SR 对比中,本方法

表 1 不同序列目标跟踪量化评估结果

Table 1 Quantitative results of target tracking for different sequences

视频序列	不利因素	SR				CLE			
		STC	CT	CSK	本方法	STC	CT	CSK	本方法
bear_change	occlusion, cross shade, scale	0.081	0.255	0.257	0.375	127.550	69.759	114.483	35.099
bear_front	occlusion, cross shade, scale	0.121	0.135	0.146	0.470	164.986	116.268	126.006	23.548
cafe_occ1	occlusion, cross shade, background clutter	0.036	0.036	0.071	0.389	304.206	340.815	274.330	144.049
cc_occ1	occlusion, cross shade, scale	0.075	0.081	0.075	0.625	235.884	180.785	168.368	28.372
cf_difficult	occlusion, shade, scale, background clutter, deformation	0.140	0.198	0.182	0.405	145.611	107.347	144.951	19.448
cf_no_occ	background clutter	0.606	0.788	0.561	0.906	9.949	18.799	75.830	12.860
cf_occ2	occlusion, cross shade, background clutter, scale, deformation	0.156	0.222	0.200	0.244	194.230	232.306	173.732	210.015
cf_occ3	occlusion, cross shade, background clutter	0.348	0.696	0.348	0.706	148.736	19.237	162.308	15.675

续表 1

视频序列	不利因素	SR				CLE			
		STC	CT	CSK	本方法	STC	CT	CSK	本方法
face_occ2	occlusion, scale	0.196	0.111	0.791	0.883	232.479	111.269	18.786	14.351
face_occ5	occlusion	0.494	0.500	0.945	0.888	69.514	68.826	4.483	7.574
new_ex_occ1	occlusion, cross shade, background clutter, deformation	0.020	0.069	0.196	0.793	320.627	319.694	359.357	29.620
new_ex_occ3	occlusion, cross shade, background clutter	0.245	0.245	0.265	0.750	259.989	252.557	249.746	18.326
new_ex_occ4	occlusion, cross shade, background clutter, scale	0.490	0.529	0.510	0.627	94.990	98.683	85.338	16.565
new_ex_occ5_long	occlusion, cross shade, background clutter, scale, deformation	0.151	0.188	0.198	0.482	209.094	122.691	159.786	36.231
new_ex_occ6	occlusion, cross shade, background clutter	0.458	0.438	0.479	0.125	163.385	170.387	150.225	132.720
new_ex_occ7.1	occlusion, cross shade, background clutter, scale	0.121	0.431	0.448	0.655	201.023	172.448	151.611	20.409
studentcenter2.1	occlusion, cross shade, background clutter, scale	0.229	0.265	0.035	0.378	169.677	257.876	296.190	78.217
two_people_1.1	occlusion, cross shade, background clutter, deformation	0.027	0.207	0.230	0.619	313.145	186.969	160.501	20.406
two_people_1.2	occlusion	0.334	0.051	0.202	1.000	70.977	205.839	102.485	1.146
two_people_1.3	occlusion	0.693	0.059	0.114	0.209	26.246	178.224	242.460	19.422
walking_occ1	occlusion, cross shade, deformation	0.662	0.338	0.353	0.985	81.739	209.548	210.859	7.140
zcup_move_1	scale	0.830	0.827	0.824	0.851	7.662	13.207	13.662	31.916
	平均值	0.296	0.303	0.338	0.607	161.441	156.979	156.613	41.960
	帧率	28.169	19.481	226.614	13.184				

有 2 组比其他方法差, 并且在本方法实验结果好的组中, 有 13 组超过了 0.50, 有 6 组超过了 0.85, 有 1 组达到了 1.00; 平均 SR 比其他对比方法高出 1 倍. 而在 CLE 的对比中, 本方法有 19 组比其他方法好;

其中误差 20.0 以内的有 10 组, 10.0 以内的有 3 组, 且其中有 1 组误差接近 1.0. 平均 CLE 比其他对比方法低. 对于跟踪速度, 本算法比其他算法稍慢些, 但在可接受范围内. 部分主观的跟踪效果见图 3; 此

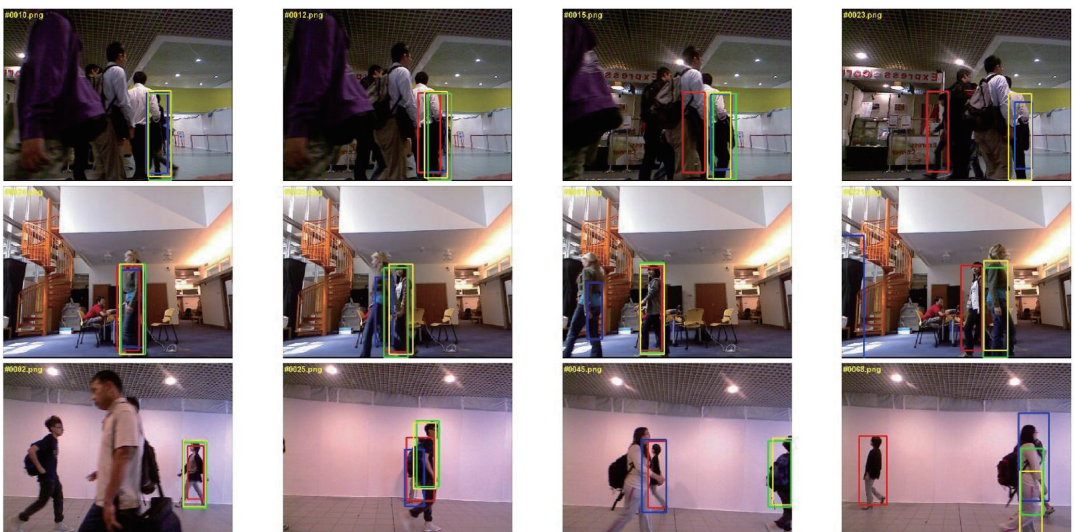


图 3 在 3 个 Princeton 序列上的主观跟踪效果

Fig. 3 Qualitative results of target tracking for three sequences in Princeton dataset

3组都是从 Princeton 数据集中选出;从上往下分别是“new_ex_occ3”“two_people_1.1”“walking_occ1”;其中红色目标检测框是本文方法所得,蓝色是 STC,绿色是 CT,黄色是 CSK.

3 结论

1) 通过对快速上下文 STC 算法进行改进,引入深度信息,利用实时更新的深度信息模板,从而更好地区分目标和背景.

2) 引入尺度更新算法和遮挡处理算法,使跟踪器在出现大幅遮挡、交叉遮挡、长时间遮挡的情况下表现出明显的性能提升.

但是,本文方法在保证较高的精确性和鲁棒性的情况下,处理速度略微降低. 未来工作中,可以在兼顾时间性能的情况下,利用上下文中 RGB 轮廓和深度轮廓融合进一步提升跟踪精确度和鲁棒性.

参考文献:

- [1] COLLINS R T. Mean-shift blob tracking through scale space[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2003: 234-240.
- [2] ADAM A, RIVLIN E, SHIMSHONI I, et al. Robust fragments-based tracking using the integral histogram[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2006: 798-805.
- [3] COLLINS R T, LIU Y, LEORDEANU M, et al. Online selection of discriminative tracking features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631-1643.
- [4] ROSS D A, LIM J, LIN R, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125-141.
- [5] KWON J, LEE K M. Visual tracking decomposition[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 1269-1276.
- [6] KWON J, LEE K M. Tracking by sampling trackers[C] // International Conference on Computer Vision. Piscataway: IEEE, 2011: 1195-1202.
- [7] ZHANG K, ZHANG L, YANG M, et al. Real-time

compressive tracking [C] // European Conference on Computer Vision. Zurich: ECVA, 2012: 864-877.

- [8] BABENKO B, YANG M, BELONGIE S, et al. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.
- [9] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C] // European Conference on Computer Vision. Zurich: ECVA, 2012: 702-715.
- [10] KALAL Z, MATAS J, MIKOLAJCZYK K, et al. P-N learning: bootstrapping binary classifiers by structural constraints[C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 49-56.
- [11] HARE S, SAFFARI A, TORR P H, et al. Struck: structured output tracking with kernels[C] // International Conference on Computer Vision. Piscataway: IEEE, 2011: 263-270.
- [12] XU J, LU Y, LIU J, et al. Robust tracking via weighted spatio-temporal context learning [C] // International Conference on Image Processing. Piscataway: IEEE, 2014: 413-416.
- [13] SONG S, XIAO J. Tracking revisited using RGBD camera: unified benchmark and baselines [C] // International Conference on Computer Vision. Piscataway: IEEE, 2013: 233-240.
- [14] ALAMSYAH D, FANANY M I. Particle filter for 3D fingertips tracking from color and depth images with occlusion handling [C] // International Conference on Advanced Computer Science and Information Systems. Piscataway: IEEE, 2013: 445-449.
- [15] ZHANG K, ZHANG L, LIU Q, et al. Fast visual tracking via dense spatio-temporal context learning[C] // European Conference on Computer Vision. Zurich: ECVA, 2014: 127-141.
- [16] CHEN Z, LUO L, WEN M, et al. Fast tracking via context depth model learning [C] // International Conference on Image Processing. Piscataway: IEEE, 2015: 4215-4218.

(责任编辑 吕小红)