

文章编号:1007-2780(2019)03-0291-11

自适应多尺度颜色特征目标跟踪算法研究

李晓云¹, 何秋生^{1*}, 张卫峰¹, 梁慧慧¹, 陈伟²

(1. 太原科技大学 电子信息工程学院, 山西 太原 030024;

2. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

摘要:为了解决尺度变化对目标跟踪的影响,本文在颜色特征跟踪算法的基础下提出了一种多尺度目标跟踪算法。该算法通过训练位置和尺度两个相关滤波器以实现自适应尺度跟踪。首先通过最小二乘分类器学习获得位置相关滤波器,采用主成分分析法对颜色特征进行降维,计算响应的最大值作为下一帧目标中心位置;接着根据设定的尺度因子在中心位置周围形成多个大小不一的矩形区域,并计算每个区域的颜色特征;学习每个区域的颜色特征,获得尺度相关滤波器,并采用正交三角分解对尺度相关滤波器进行降维;然后根据响应的最大值确定跟踪目标的尺寸;最后对目标的位置和尺寸进行更新。通过对 13 组挑战性的视频序列进行测试,结果表明,本算法不仅对目标尺度变化具有一定的适应性,而且对光照变化、快速运动、运动模糊等复杂情况下,均具有鲁棒性,多项性能指标均优于目前跟踪性能先进的算法。

关键词:目标跟踪;颜色属性;主成分分析;尺度自适应

中图分类号:TP311 **文献标识码:**A **doi:**10.3788/YJYXS20193403.0291

Adaptive multi-scale color feature target tracking algorithm

LI Xiao-yun¹, HE Qiu-sheng^{1*}, ZHANG Wei-feng¹, LIANG Hui-hui¹, CHEN Wei²

(1. School of Electronic Information Engineering,

Taiyuan University of Science and Technology, Taiyuan 030024, China;

2. School of Computer Science and Technology, China University of Mining,
Xuzhou 221116, China)

Abstract: In order to solve the influence of scale change on target tracking, a multi-scale target tracking algorithm was proposed based on the color feature tracking algorithm. The algorithm realized adaptive scale tracking by training position and scale correlation filters. Firstly, the target center position of next frame was obtained by computing the maximum of the response, where the position correlation filter was learned by the least squares classifier and the dimensionality reduction for color features was analyzed by principal component analysis. The scale correlation filter was obtained by color

收稿日期:2018-09-20; **修订日期:**2018-12-24.

基金项目:山西省重点研发计划重点项目(No.201703D111027);国家自然科学基金委员会 G 山西省人民政府煤基低碳联合基金项目(No.U1510115);先进控制与智能信息系统山西省重点实验室(No.201805D111001);山西省研究生联合培养基地人才培养项目(No.2017JD32)

Supported by Key Research and Development Projects of Shanxi Province(No.201703D111027); Natural Science Foundation of China Coal Based Low Carbon Joint Fund Project of the People's Government of Shanxi Province(No.U1510115);Shanxi Key Laboratory of Advanced Control and Intelligent Information System(No.201805D111001);Shanxi Province Joint Student Training Base Talent Training Project(No.2017JD32)

* 通信联系人, E-mail: heqs2008@126.com

characteristics at 33 rectangular areas which was set by the scale factor around the central location and reduced dimensions by orthogonal triangle decomposition. Finally, the location and size of the target were updated by the maximum of the response. By testing 13 challenging video sequences, the results show that the algorithm has adaptability to the changes in the target scale and its robustness along with many other performance indicators are both better than the most state-of-the-art methods in illumination variation, fast motion, motion blur and other complex situations.

Key words: target tracking; color feature; principal component analysis; scale adaptation

1 引言

视觉目标跟踪是一项具有挑战性的任务,在光照和尺度变化、遮挡、背景杂乱以及快速运动、运动模糊的影响下,跟踪会变得非常困难。

跟踪算法由早期的均值偏移算法 (Mean Shift)^[1-3]、粒子滤波算法^[4-6]、支持向量机 (SVM)^[7-9]到多示例学习^[10-12],之后相关滤波器算法在跟踪过程中的快速性被大家关注。Bolme 等^[13]是最早提出将相关滤波用于目标跟踪,将设计的最小平方误差和输出 (Minimum output sum of squared error, MOSSE) 采用简单的灰度特征训练滤波器,然后根据响应的最大值确定跟踪目标位置,虽然跟踪速度很快,但是跟踪精度略有欠缺。之后大家开始致力于提高跟踪算法的精度,并取得了不错的效果。Henriques 等^[14]提出 CSK 算法,在相关滤波的基础上引入循环矩阵,解决脊回归问题,提高了算法的精确度,但是由于仍然采用灰度特征,对目标变形、尺度变化、背景杂乱以及快速运动等问题的鲁棒性较弱。文献^[15]提出的 KCF 算法,将 HOG 特征应用到跟踪算法中,使得跟踪精度提高,但是当目标快速运动、尺度发生变化及运动模糊时,容易导致目标跟踪失败。Bertinetto 等^[16]提出 Staple 算法,将颜色直方图特征和 HOG 特征以一定的融合因子进行融合,然后利用融合后的特征训练滤波器,跟踪精度提高,但是跟踪速度大幅度降低。Galoogahi 等^[17]提出 BACF 算法,在跟踪过程中考虑背景信息,解决了目标遮挡、快速运动等问题,且提高了跟踪精度,但对目标尺度变化的适应性不强,跟踪速度也有所下降。文献^[18]和文献^[19]将深度特征应用到跟踪算法中,虽然深度特征对跟踪精度有一定的提高,但是跟踪速度大幅降低。

上文列举的跟踪算法及现存的大多数跟踪算法都没有解决目标尺度变化的问题。如果目标缩小,滤波器就会学习到大量的背景信息,如果目标扩大,滤波器就会受到目标局部纹理的影响,这两种情况都有可能出现非预期的结果,导致跟踪漂移和失败^[20]。Danelljan 等^[21]首次将 CN (Color name) 应用到目标跟踪中,通过将 11 维特征用 PCA 进行降维抽取,再采用 CSK 跟踪算法实现跟踪,跟踪精度相对较高,但它在目标尺度变化、部分遮挡以及发生形变时,效果不太理想^[22]。本文提出基于颜色属性的多尺度目标跟踪算法,解决尺度变化问题,实现自适应颜色特征的多尺度目标跟踪。

2 颜色视觉跟踪

颜色特征拥有丰富的表现力和较高的辨识度,CN 算法将原始的 RGB 颜色空间映射到 11 维颜色属性空间^[23],对目标在跟踪过程中易受环境变化的问题具有一定鲁棒性。算法融合了灰度特征和颜色特征,对每一个特征通道,都经过 Hann 窗做预处理,最终得到灰度和颜色的综合特征表示^[24]。

为了缩短计算时间,算法采用主成分分析 (PCA) 降维技术将 11 维的颜色空间降到 2 维。令 D_1 、 D_2 分别代表 11 维和 2 维颜色空间, x 为 D_1 维的颜色特征,通过寻找一个 $D_1 \rightarrow D_2$ 列向量正交的映射矩阵 B_P 实现降维。通过线性映射 $\tilde{x} = B_P^T x$, 得到 $\tilde{x}$, 它就是 D_2 维的颜色特征。

颜色特征提取完成后,然后对特征进行处理,实现跟踪^[10]。CN 算法是判别式跟踪方法,根据分类器响应最大值确定新一帧目标的位置。所以跟踪问题最主要的就是分类器设计^[25]。

颜色特征分类器采用 RLS 分类器,该分类器是一种递归最小二乘算法,CN 算法直接在由核

定义的再生核希伯特空间上最小化一个线性正则化函数。通过训练目标周围大小为 $M \times N$ 的图像块获得分类器,训练样本 $\tilde{x}_{m,n}$ 采用循环移位

来获得,其中, $m \in \{0, \dots, M-1\}, n \in \{0, \dots, N-1\}$ 。该分类器通过最小正则化风险泛函进行训练,风险泛函如式(1)所示。

$$\varepsilon = \sum_{j=1}^p \beta_j \left(\sum_{m,n} |\langle \varphi(\tilde{x}_{m,n}^j), \omega^j \rangle - h^j(m,n)|^2 + \lambda \langle \omega^j, \omega^j \rangle \right), \quad (1)$$

其中, $\varphi(\cdot)$ 表示图像块通过核函数 k 映射到希伯特空间的函数,这里核函数采用高斯核。 $\tilde{x}_{m,n}^j$ 为第 j 帧的目标样本, h^j 为 $\tilde{x}_{m,n}^j$ 的期望高斯函数输出, $\lambda \geq 0$ 为正则化参数。此风险泛函考虑所有帧误差, β 为每一帧误差的加权系数。

式(1)的解可表示为输入的线性组合,即

$$\omega^j = \sum_{m,n} a(m,n) \varphi(\tilde{x}_{m,n}^j), \quad (2)$$

损失函数最小时应满足:

$$A^p = \frac{A_N^p}{A_D^p} = F\{a\} = \frac{Y^p}{U_x^p + \lambda} = \frac{\sum_{j=1}^p \beta_j H^j U_x^j}{\sum_{j=1}^p \beta_j U_x^j (U_x^j + \lambda)}, \quad (3)$$

其中,大写字母表示相应的傅里叶变换, $U_x^j = F\{u_x^j\}$, $F\{\cdot\}$ 为傅里叶变换, $u_x^j(m,n) = k(x_{m,n}^j, x^j)$, 则目标模型用式(4)来更新, $\hat{x}^p$ 为更新后估计的第 p 帧目标特征。

$$\begin{cases} A_N^p = (1-\beta)A_N^{p-1} + \beta Y^p U_x^p \\ A_D^p = (1-\beta)A_D^{p-1} + \beta U_x^p (U_x^p + \lambda) \\ \hat{x}^p = (1-\beta)\hat{x}^{p-1} + \beta \tilde{x}^p \end{cases} \quad (4)$$

这种机制使得模型更新不需要存储冗余信息,只需要存储上一帧的信息 $\{A_N^p, A_D^p, \hat{x}^p\}$, 确保跟踪速度。

分类器的输出响应为 $\hat{y}^p = F^{-1}(A^p U_z^p)$, 其中 $U_z^p = F\{u_z^p\}$, $u_z^p(m,n) = k(z_{m,n}^p, \hat{x}^p)$, $z_{m,n}^p$ 是第 p 帧提取的目标特征, $\hat{x}^p$ 是通过分类器学习之后估计的第 p 帧目标特征。计算输出响应 $\hat{y}^p$ 并通过其最大值来确定目标在下一帧中的位置。

3 改进的 CN 跟踪算法

原始的 CN 算法是在 CSK 算法上进行的改进,但是没有解决尺度变化的问题。文献[26]提出一种基于 MOSSE 滤波器的尺度估计方法,本

文受其启发在 CN 跟踪算法的框架下,增加了尺度估计策略,以实现自适应尺度目标跟踪。

3.1 尺度估计策略

在跟踪过程中,目标因运动其尺度经常发生变化,如果算法不能适应目标的尺度变化,那么分类器的输出就会受到一定的影响,从而使得跟踪效果不佳。本文所采取的方法是,首先对目标位置进行估计,其次根据估计位置区域的信息进行尺度估计,最后根据尺度估计结果更新目标的跟踪结果。而尺度估计是通过相关滤波器去检测目标尺度的变化,从而能够依据目标尺度变化合理地限定搜索区域。尺度估计的特征提取过程如图1所示。

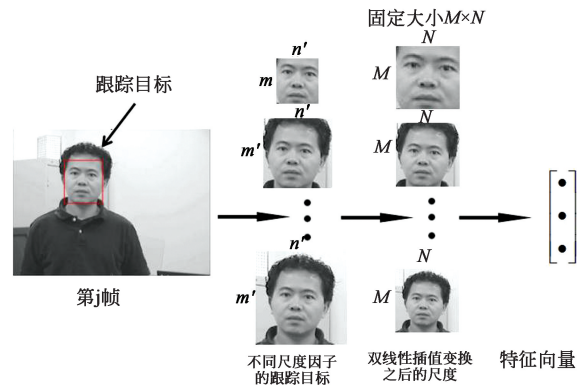


图1 尺度估计的特征提取过程

Fig.1 Feature extraction process of scale estimation

首先在目标周围确定一系列尺寸变化的矩形区域,并计算每个区域的颜色特征,再用提取的特征来训练 RLS 分类器,从而得到尺度相关滤波器。假设当前帧内目标大小为 $M \times N$, 尺度相关滤波器的大小为 $S \times 1$ 。依次提取目标周围大小为 $m' \times n'$ 的区域,获得的样本为 x_{scale}^i , $i=1, \dots, 33$, $m' = \theta^\tau M$, $n' = \theta^\tau N$, $\tau \in \left\{ -\frac{S-1}{2}, \dots, \frac{S-1}{2} \right\}$, θ 为尺度因子。

3.2 快速尺度跟踪

考虑到特征提取过程中计算量大的问题,本文采用双线性插值的方法将 3.1 节中得到的 33 个特征矩阵的维数统一为初始目标区域特征的维数,然后对双线性插值后的 33 个特征矩阵进行整合,得到目标的尺度特征矩阵 x_{scale} 。此外,采用正交三角分解对特征进行降维,减少计算量。

3.2.1 降维

在跟踪过程中,跟踪器的跟踪速度与特征维数成反比,考虑到计算的快速性,本文采用正交三角分解对特征进行降维处理。通过构造一个投影矩阵 $B_{p,scale}$ 使之与双线性插值变换之后的目标特征 x_{scale} 相乘,即可获得低维空间的目标特征。投影矩阵 $B_{p,scale}$ 的维数是 $\tilde{d} \times d$, 其中 d 为降维之前目标特征的维数, $\tilde{d}$ 是降维之后的维数, p 代表第 p 帧。尺度相关滤波器通过最小正则化风险泛函进行训练,风险泛函如式(5)所示。

$$\eta = \beta_p \sum_{m',n'} \| \hat{x}_{scale}^p(m',n') - B_{p,scale} B_{p,scale}^T \hat{x}_{scale}^p(m',n') \|^2, \quad (5)$$

要使风险泛函最小,则 $B_{p,scale} B_{p,scale}^T = I$ 。通过自相关矩阵的特征值分解求解投影矩阵。自相关矩阵为 $Q_p = \sum_{m',n'} \hat{x}_{scale}^p(m',n') (\hat{x}_{scale}^p(m',n'))^T$, 特征分解公式为 $Q_p = B_{p,scale} \Lambda_p B_{p,scale}^T$ 。投影矩阵 $B_{p,scale}$ 的每一行代表 Λ_p 中特征值所对应的特征向量。

3.2.2 降维尺度跟踪

通过降维处理,在不影响跟踪精度的情况下,跟踪速度得到很大提升。用求得的投影矩阵对特征进行降维时,分别对提取的目标特征 x_{scale}^p 和通过分类器学习估计的目标特征 $u_{z,scale}^p$ 计算两个投影矩阵 $B_{p,scale}^x$ 和 $B_{p,scale}^u$, 则降维之后提取的目标特征为 $\tilde{x}_{scale}^p = B_{p,scale}^x x_{scale}^p$, 降维之后通过分类器学习估计的目标特征为 $\tilde{u}_{z,scale}^p = B_{p,scale}^u u_{z,scale}^p$ 。在跟踪过程中,为了提高计算效率,并没有显式地构造自相关矩阵,而是分别对 x_{scale}^p 和 $u_{z,scale}^p$ 进行 QR 分解得到投影矩阵。

检测尺度滤波器的响应 $\hat{y}_{scale}^p$ 可由式(6)获得,找到使 $\hat{y}_{scale}^p$ 最大的尺度作为新一帧中目标的尺度。

$$\hat{y}_{scale}^p = F^{-1}(A^p G_{z,scale}^p), \quad (6)$$

其中 $G_z^p = F\{g_z^p\}$, $g_z^p(m,n) = k(B_{p,scale}^x z_{m,n}^p, B_{p,scale}^u \hat{x}_{scale}^p)$, $z_{m,n}^p$ 是第 p 帧提取的目标特征, $\hat{x}_{scale}^p$ 是分类器通过学习更新后的第 p 帧目标特征。计算输出响应 $\hat{y}_{scale}^p$ 并通过其最大值来确定目标在下一帧中的位置,然后对目标模型进行更新,更新采用式(7)。

$$\begin{cases} \hat{A}_{scale}^p = (1-\beta) \hat{A}_{scale}^{p-1} + \beta \hat{A}_{scale}^p \\ \hat{x}_{scale}^p = (1-\beta) \hat{x}_{scale}^{p-1} + \beta B_{p,scale}^x x_{scale}^p \end{cases}, \quad (7)$$

其中: β 为尺度学习因子, $\hat{A}_{scale}^p, \hat{A}_{scale}^{p-1}$ 分别表示当前帧和上一帧更新之后的系数矩阵, $\hat{x}_{scale}^p, \hat{x}_{scale}^{p-1}$ 分别表示当前帧和上一帧更新之后的目标特征。

在目标跟踪中,由于相邻两帧中目标的尺度变化很小,为此先利用位置核相关滤波器检测出目标位置,然后在周围采集样本,利用尺度核相关滤波器检测出目标尺度,这样就完成了目标位置和尺度的检测。

3.3 本文算法流程

本文跟踪算法中在进行位置滤波和尺度滤波时,均采用高斯函数输出,目标特征选用多通道的颜色特征和灰度值。首先对特征通道采用 Hann 窗进行预处理,然后提取目标特征,进行降维处理,计算响应值,最后更新目标位置和尺度。具体的算法流程如下所示:

Input:

输入图像 patch I_t ;

上一帧的位置 P_{t-1} 和尺度 S_{t-1} ;

位置模型 A_{t-1}^p, x_{t-1}^p 和尺度模型 A_{t-1}^s, x_{t-1}^s 。

Output:

估计的目标位置 P_t 和尺度 S_t ;

更新位置 A_t^p, x_t^p 和尺度模型 A_t^s, x_t^s 。

其中:

位置评估:

(1) 根据视频前一帧的位置,在当前帧中按照前一帧目标尺度的 2 倍提取颜色特征 Z ;

(2) 利用 Z 和 A_{t-1}^p, x_{t-1}^p , 计算 y ;

(3) 计算 $\max(y)$, 得到目标准确的位置 P_t 。

尺度评估:

(4) 以目标当前位置为中心, 提取 33 中不同尺度的颜色特征 Z' ;

(5) 降维到 17 维, 再使用 Z' 和 A_{t-1}^S, x_{t-1}^S 计算出 $\hat{y}_S$;

(6) 计算 $\max(\hat{y}_S)$, 得到目标准确的尺度 S_t 。

模型更新:

(7) 提取样本 f_P 和 f_S ;

(8) 更新位置模型 A_t^P, x_t^P ;

(9) 更新尺度模型 A_t^S, x_t^S 。

4 实验评估及结论

为了验证本文算法的跟踪效果, 选取了 OTB2015 数据集中 13 个挑战性的视频进行测试, 此数据集提供精准的 Ground truth 和初始框的位置和大小, 以便计算跟踪精度。本文与近几年来跟踪性能优异的 10 个算法进行对比, 包括 CSK、KCF、SAMF、SRDCF、DCF_CA、MOSSE_CA、STAPLE_CA 和 SAMF_CA 以及 CN。

本文主要从跟踪的准确性和尺度变化上对算法的跟踪效果进行评估, 为了方便对比, 选取的对比算法均采用默认参数。本次实验机器配置是: 中央处理器为 Intel Pentium CPU G3250 @3.20 GHz, 运行内存 4 GB, 64 位操作系统。软件环境为: Win7 + Matlab2014。

4.1 快速尺度估计

在 3.2 节中, 对原始的尺度估计进行降维处理, 下面将具体分析尺度相关滤波器的维数对跟踪性能的影响。

如图 2 所示, 当维数从 33 维逐渐减小的过程中, 一段范围内跟踪性能基本保持一致。Y 轴的跟踪精度表示在跟踪过程中的成功率, 即算法计算所得的目标区域与真实目标区域的比值。结果表明, 当维数设置为 17 时, 不仅能够保证跟踪性能, 而且还可以显著降低特征的维数, 提高计算

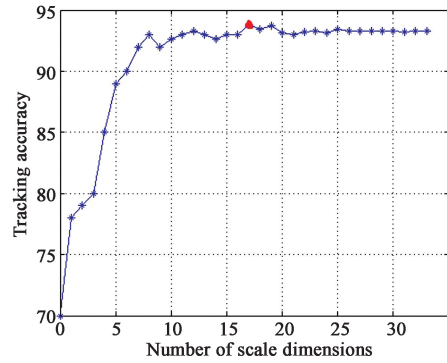


图2 尺度滤波器维数与跟踪精度的关系

Fig.2 Relationship between dimension of scale filter and tracking accuracy

速度。

4.2 性能分析

性能分析主要分为 3 个部分: 中心位置误差、跟踪成功率和距离精度。

4.2.1 中心位置误差

中心位置误差计算公式为: $K_{CLE} = \sqrt{\|O - O_t\|^2}$, 其中, O 和 O_t 分别目标的真实中心坐标和算法求得的中心坐标。中心位置误差表示算法求得的中心与真实目标中心之间的误差, 误差越小, 则跟踪精度越大。中心位置误差结果如表 1 所示。综合来看, 本文算法相较于对比算法在中心位置误差方面具有较好的性能。

4.2.2 距离精度

距离精度(DP)的公式为: $K_{DP} = \frac{m}{n}$, 其中, m 为中心位置误差小于某确定阈值的视频帧数, n 为测试数据集的总帧数。本文阈值取 20 个像素。从表 2 可看出, 与其他算法相比, 本文算法具有较好的跟踪精度。

4.2.3 跟踪成功率

跟踪的成功率定义为 $K_{OP} = \frac{\text{area}(R_T \cap R_G)}{\text{area}(R_T \cup R_G)}$,

其中: R_T 为算法求得的目标区域, R_G 为 ground-truth 标定的真实目标区域, $\text{area}(R_T \cap R_G)$ 为两个区域的重叠面积, $\text{area}(R_T \cup R_G)$ 为两个区域并集的面积。显然, K_{OP} 值越高, 表示算法求得的目标区域与真实目标区域越相近, 即目标被跟踪成功的正确率越高, 算法效果越好。跟踪成功率的结

果见表 3。显然,与其他算法相比,本文算法具有更高的跟踪成功率。

表 1 中心位置误差结果汇总表(最优结果用粗体显示)

Tab.1 Summary of central position error results (the optimal results are shown in bold)

视频帧 特点分类		CSK	KCF	SAMF	SRDCF	DCF_CA	MOSSE_CA	STAPLE_CA	SAMF_CA	CN	OUR
尺度变化	Dog1	3.81	4.15	3.92	3.95	4.25	3.62	5.35	4.06	3.73	3.43
	Doll	44.7	7.58	2.96	2.62	5.44	29.5	3.86	3.46	7.25	2.58
	FleetFace	36.5	30.4	22.7	31.1	25.7	28.1	28.8	20	57.2	16.8
	Mhyang	3.61	3.64	2.42	2.09	3.61	3.58	3.02	3.02	3.57	2.06
快速运动 与 运动模糊	Jumping	86	29	28.7	4.47	34	3.72	6.14	38	49.3	4.01
	Deer	4.79	21.3	11.6	3.97	15.9	4.65	3.94	4.7	5.04	4.68
	Boy	20.1	2.62	4.05	1.65	2.4	2.42	2.56	2.42	2.61	2.16
	Fish	41.2	4.37	4.89	3.42	4.24	8.49	3.79	4.01	43.1	5.64
光照变化 与 背景杂乱	CarDark	3.23	5.26	2.5	1.57	4.82	2.64	1.18	1.76	3.59	1.86
	Dudek	9.87	9.38	8.73	12.6	11.3	15.6	13.6	9.78	15.7	8.73
	Shaking	17.2	109	62.2	85.4	7.63	72.4	148	7.6	27	10.5
	Skating1	7.78	7.71	6.14	20.3	7.38	79.9	6.28	102	7.64	6.96
	Trellis	18.8	7.73	2.72	2.56	7.48	68	3.05	2.56	21.2	3.03

表 2 跟踪精度结果汇总表(最优结果用粗体显示)

Tab. 2 Summary of tracking accuracy results (optimal results shown in bold)

视频帧 特点分类		CSK	KCF	SAMF	SRDCF	DCF_CA	MOSSE_CA	STAPLE_CA	SAMF_CA	CN	OUR
尺度变化	Dog1	100%	100%	99.90%	100%	100%	100%	99.90%	100%	100%	100%
	Doll	57.90%	97%	99.30%	99.30%	98.10%	76.40%	99.10%	99.20%	98.70%	99.30%
	FleetFace	56.30%	60.50%	62.90%	59.70%	43.70%	38.20%	63.40%	67.90%	38.50%	62.20%
	Mhyang	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
快速运动 与 运动模糊	Jumping	5.11%	30.70%	27.20%	100%	69%	100%	97.10%	49.80%	18.50%	100%
	Deer	100%	81.70%	88.70%	100%	85.90%	100%	100%	100%	100%	100%
	Boy	84.40%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Fish	4.20%	100%	100%	100%	100%	100%	100%	100%	4.20%	100%
光照变化 与 背景杂乱	CarDark	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Dudek	84.30%	89.60%	91.90%	83.30%	87.10%	78.50%	81.80%	89.90%	75.10%	90.30%
	Shaking	56.40%	1.92%	2.74%	1.37%	94.50%	1.10%	3.29%	96.40%	20.80%	96.60%
	Skating1	98.80%	100%	100%	89.90%	100%	70%	100%	19.50%	100%	100%
	Trellis	81%	100%	100%	100%	100%	17.20%	100%	100%	65%	100%

表 3 跟踪成功率结果汇总表(最优结果用粗体显示)
Tab.3 Summary of track success rate results (best results shown in bold)

视频帧 特点分类		CSK	KCF	SAMF	SRDCF	DCF_CA	MOSSE_CA	STAPLE_CA	SAMF_CA	CN	OUR
尺度变化	Dog1	65.30%	65.20%	82.60%	100%	64.90%	65.30%	100%	81.30%	65.30%	100%
	Doll	21.80%	64.40%	86.40%	99.70%	72.50%	57.10%	99.70%	81%	71.40%	99.70%
	FleetFace	65.40%	63.50%	74.50%	66.30%	72.70%	61.40%	68.50%	79.30%	55.40%	100%
	Mhyang	100%	100%	100%	99.70%	100%	100%	99.50%	99.90%	100%	100%
快速运动 与 运动模糊	Jumping	4.79%	25.60%	24.30%	95.80%	68.10%	99.70%	70.60%	49.20%	6.39%	99.80%
	Deer	100%	81.70%	88.70%	100%	85.90%	100%	100%	100%	99%	100%
	Boy	84.20%	99.20%	99.30%	100%	99.30%	99.20%	100%	100%	99.20%	100%
	Fish	4.20%	100%	100%	100%	100%	100%	100%	100%	4.20%	100%
光照变化 与 背景杂乱	CarDark	99.20%	83.50%	100%	100%	92.90%	99.50%	100%	94.90%	94.70%	100%
	Dudek	87.30%	92.10%	100%	99.20%	97.50%	92.70%	79.10%	97.10%	91.50%	100%
	Shaking	58.10%	1.37%	1.37%	1.10%	92.10%	1.10%	3.01%	95.30%	21.10%	96.40%
	Skating1	36.80%	35.30%	80.30%	53.80%	35.30%	37.30%	62%	18.30%	35.30%	89.50%
	Trellis	59.10%	84%	99.30%	96.50%	84.20%	7.56%	99.30%	100%	56.40%	96.10%

4.3 实验结果

从表 1~3 可以看出,在跟踪性能方面,本文提出的算法优于其他算法,在随后的图 3~5 中将给出本算法在跟踪过程中的部分视频帧截图,能更直观地体现跟踪效果。

4.3.1 尺度变化

在跟踪过程中,目标会发生尺度变化,影响跟踪精度。在图 3 所示的视频帧中,Dog1 视频在跟踪过程中尺度变化明显,从图 1 的第一行可以看到,运行到第 590 帧,所有算法基本跟踪正确,但是到了第 908 帧,目标尺度变大,可以看到,只有本文的算法以及算法 STAPLE_CA、SRDCF 能够自适应尺度的变化,尤其在第 1 039 帧表现最为明显。在后续帧中,目标尺度变小,本文算法也能自适应其尺度的变化,成功跟踪目标。在第 2 行视频帧 Doll 中,第 886 帧目标变大,已有大部分算法不能自适应其尺度的变化。在第 1 637 帧及之后,算法 CSK 和 MOSSE_CA 就已经开始偏离目标,跟踪失败。在第 3 行 Fleetface 视频帧中,第 614 帧,目标不仅尺度发生变化,而且目标具有一定程度的旋转,本文算法表现出良好的跟踪效果。在第 663 帧和第 693 帧中,算法 CN 已

跟丢目标。第四行视频帧 Dudek 中,大部分算法虽然成功跟踪,但未能很好的自适应目标尺度的变化,从而影响其跟踪成功率,而本文算法在该视频中仍能达到 100% 的跟踪成功率。

4.3.2 快速运动与运动模糊

在跟踪过程中,目标运动速度快,会使目标产生模糊的画面,对于跟踪是一个巨大的挑战。在图 4 所示的视频中,第 1 行是 Jumping 视频的跟踪结果截图,因目标范围小,运动速度快,运动过程中画面模糊,所以很多算法在此视频帧中跟踪失败。从图中可以看到,在第 25 帧、36 帧、48 帧、86 帧以及 186 帧中,除了 SRDCF 和本文的算法之外,其他算法均不能准确跟踪目标。SRDCF 在算法中对尺度做了相应的处理,并在算法中加入惩罚项,能够准确跟踪目标。第 2 行的 Deer 视频中,从第 25 帧可以看到,DCF_CA 和 SAMG 已经偏离目标,在第 28 帧时,KCF 算法也已跟踪失败,虽然之后又成功跟踪,但是跟踪成功率下降。第 3 行 Boy1 视频中,目标不仅具有快速运动的特点,而且在尺度方面也有一定程度的变化,第 508 帧中 CSK 开始偏离目标,第 513 帧已完全偏离。在第 4 行 Fish 视频中,大部分算法跟踪准

确,但是 CSK 和 CN 算法偏离目标。本文算法在这组视频中,跟踪效果均优于其他算法,在 Jumping 视频中表现最为明显,取得了 99.8% 的跟踪成功率。

4.3.3 光照变化与背景杂乱

视频中经常会有光照变化明显以及杂乱的背景信息,在跟踪过程中处理这些干扰是一个很大

的难题。图 5 所示的第一行是 Shaking 视频,该视频光照变化大,而且目标与背景颜色相似,背景也非常杂乱,仅在第 5 帧中,MOSSE_CA 已经偏离目标,第 22 帧中,只有 CSK、SAMF_CA 和本文算法跟踪成功。第 69 帧、162 帧和 239 帧中,DCF_CA 也表现出良好的跟踪效果,除此之外,其他算法均跟踪失败。第 2 行的 Skating1 视频

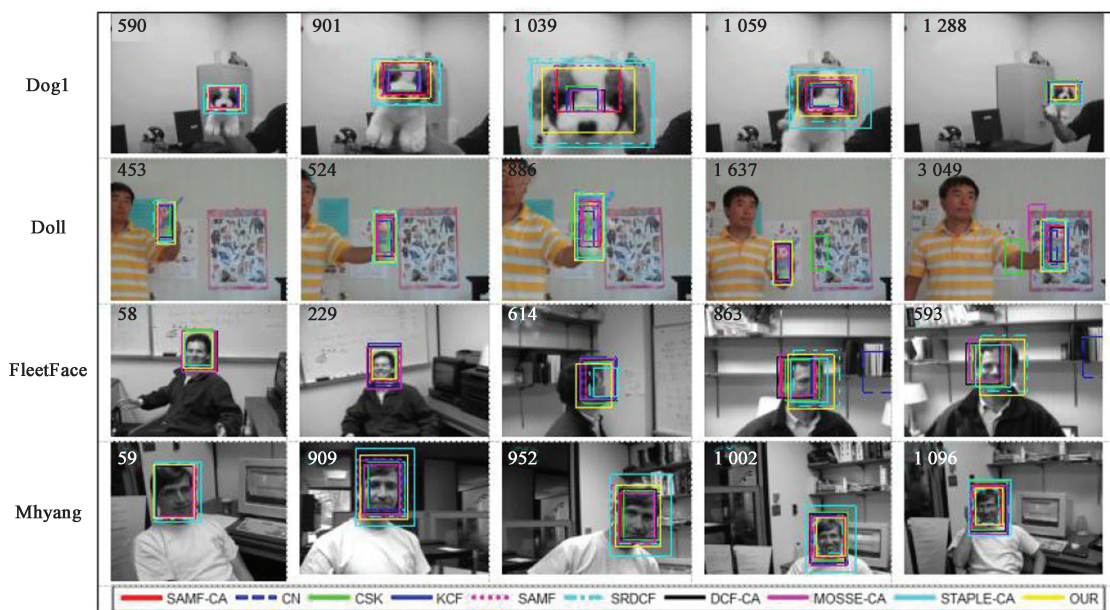


图 3 部分尺度变化的跟踪结果截图

Fig.3 Screenshots of partial scale changes tracking results

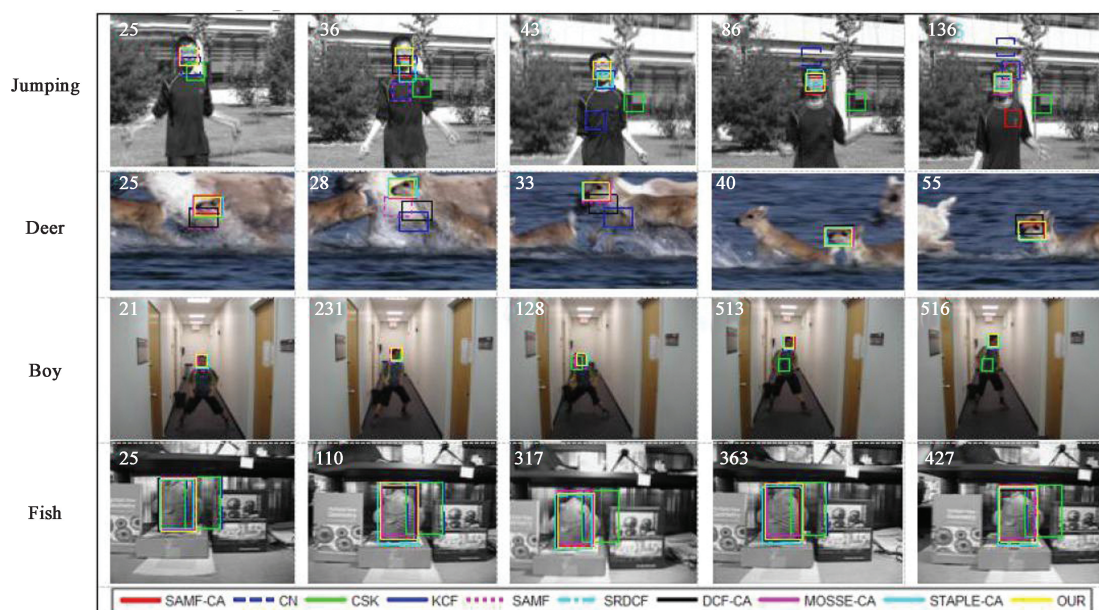


图 4 部分快速运动与运动模糊的跟踪结果截图

Fig.4 Screenshots of fast motion and motion blur tracking results

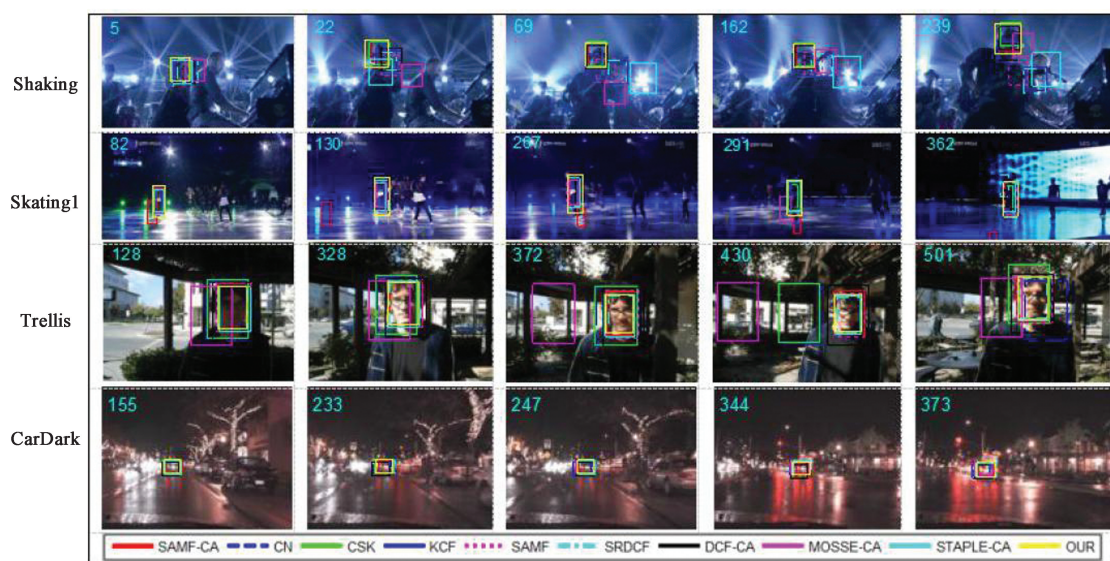


图5 部分光照变化与背景杂乱的跟踪结果截图

Fig.5 Screenshots of light variations and background clutter tracking results

中,不仅光照变化明显,背景复杂,而且目标还有一定程度的尺度变化,对于不能自适应尺度变化的算法来说,跟踪成功率不高。第82帧时,SAMF_CA 就已经开始偏离目标,第30帧已完全偏离目标。在第267帧、291帧和362帧中,本文算法能够自适应其尺度的变化,跟踪效果优于其他算法。在第3行的Trellis视频中,第128帧MOSSE_CA 开始偏离目标,第430帧MOSSE_CA、CSK 和CN 都跟踪失败。第四行的CarDark 视频中,大部分算法都成功跟踪。在这组视频中,本文算法均取得了很好的跟踪效果,尤其在Shaking 和Skating1 视频中表现最为明显,以96.4%和89.5%的跟踪成功率优于其他算法。

5 结 论

本文提出自适应颜色特征的多尺度目标跟踪

算法,将原始RGB 图像映射到11 维的颜色属性特征空间,从而实现了颜色特征的鲁棒表示。利用基于颜色特征的核跟踪方法对目标进行快速准确地跟踪。在尺度估计上,通过在目标周围矩形区域计算颜色特征,设计尺度相关滤波器,解决跟踪过程中目标尺度变化的问题。本文提出的自适应颜色特征多尺度跟踪算法拥有以下几个优点:

- (1)对于目标尺度变化具有较强的适应性(如Dog 和Doll 序列的结果);
- (2)对于快速运动与运动模糊的视频,本算法运算速度快,能够以高精度成功跟踪;
- (3)对于光照变化和背景杂乱具有较好的抵抗能力,这是因为将原始的RGB 颜色映射到CN 特征空间,提高了颜色表示的鲁棒性。

实验结果表明,本文算法既保持了颜色特征跟踪快速准确的优点,又很好地实现了尺度变化、快速运动以及背景干扰下的鲁棒跟踪。

参 考 文 献:

- [1] RAZAVI S F, SAJEDI H, SHIRI M E. Integration of colour and uniform interlaced derivative patterns for object tracking [J]. *IET Image Processing*, 2016, 10(5): 381-390.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [3] 胡良梅,段琳琳,张旭东,等.融合颜色信息与深度信息的运动目标检测方法[J].电子与信息学报,2014,36(9): 2047-2052.

- HU L M, DUAN L L, ZHANG X D, *et al.* Moving object detection based on the fusion of color and depth information [J]. *Journal of Electronics & Information Technology*, 2014, 36(9): 2047-2052. (in Chinese)
- [4] WANG S, CLARK R, WEN H K, *et al.* DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks [C]//2017 *IEEE International Conference on Robotics and Automation*. Singapore: IEEE, 2017: 2043-2050.
- [5] 刘林涛,董雪莹,刘俊,等.基于时空上下文和随机森林的人眼跟踪定位算法研究[J].液晶与显示,2018,33(5): 443-449.
- LIU L T, DONG X Y, LIU J, *et al.* Human eye locating and tracking using space-time context and random forest [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(5): 443-449. (in Chinese)
- [6] AMMIRATO P, POIRSON P, PARK E, *et al.* A dataset for developing and benchmarking active vision [J]. arXiv preprint arXiv:1702.08272, 2017.
- [7] 龙思源,张葆,宋策,等.基于改进的加速鲁棒特征的目标识别[J].中国光学,2017,10(6):719-725.
- LONG S Y, ZHANG B, SONG C, *et al.* Object detection based on improved speeded-up robust features [J]. *Chinese Optics*, 2017, 10(6): 719-725. (in Chinese)
- [8] KHAN F S, VAN DE WEIJER J, VANRELL M. Modulating shape features by color attention for object recognition [J]. *International Journal of Computer Vision*, 2012, 98(1): 49-64.
- [9] KHAN F S, ANWER R M, VAN DE WEIJER J, *et al.* Color attributes for object detection [C]//2012 *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, 2012: 3306-3313.
- [10] DANELLJAN M, KHAN F S, FELSBERG M, *et al.* Adaptive color attributes for real-time visual tracking [C]//2014 *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014: 1090-1097.
- [11] COSTANTE G, MANCINI M, VALIGI P, *et al.* Exploring representation learning with CNNs for frame-to-frame ego-motion estimation [J]. *IEEE Robotics and Automation Letters*, 2016, 1(1): 18-25.
- [12] NING J, ZHANG L, ZHANG D, *et al.* Scale and orientation adaptive mean shift tracking [J]. *IET Computer Vision*, 2012, 6(1): 52-61.
- [13] VOJIR T, NOSKOVA J, MATAS J. Robust scale-adaptive mean-shift for tracking [J]. *Pattern Recognition Letters*, 2014, 49: 250-258.
- [14] WU Y, LIM J, YANG M H. Online object tracking: A benchmark [C]//2013 *IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 2411-2418.
- [15] 王民,石新源,王稚慧.基于 Vibe 和时空上下文的运动手势跟踪算法[J].液晶与显示,2018,33(1):92-98.
- WANG M, SHI X Y, WANG Z H. Moving gesture tracking algorithm based on vibe and space-time context [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(1): 92-98. (in Chinese)
- [16] MOJSILOVI A. A computational model for color naming and describing color composition of images [J]. *IEEE Transactions on Image Processing*, 2005, 14(5): 690-699.
- [17] VAN DE WEIJER J, SCHMID C, VERBEEK J, *et al.* Learning color names for real-world applications [J]. *IEEE Transactions on Image Processing*, 2009, 18(7): 1512-1523.
- [18] CLARK R, WANG S, WEN H K, *et al.* VINet: Visual-inertial odometry as a sequence-to-sequence learning problem [C]//Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, California: AAAI, 2017.
- [19] COMANICIU D, MEER P, RAMESH V. Kernel-based object tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(5): 564-577.
- [20] CHATFIELD K, SIMONYAN K, VEDALDI A, *et al.* Return of the devil in the details: Delving deep into convolutional nets [EB/OL]. arXiv preprint arXiv:1405.3531, 2014.
- [21] ORON S, BAR-HILLEL A, LEVI D, *et al.* Locally orderless tracking [C]//2012 *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, 2012: 1940-1947.
- [22] VAN DE WEIJER J, SCHMID C, VERBEEK J. Learning color names from real-world images [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, Minnesota, USA: IEEE, 2007: 1-8.
- [23] CLARK R, WANG S, MARKHAM A, *et al.* VidLoc: 6-dof video-clip relocalization [EB/OL]. arXiv preprint

- arXiv:1702.06521, 2017.
- [24] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Hilton Head Island, SC, USA: IEEE, 2000: 142-149.
- [25] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al*. Spatial transformer networks [EB/OL]. arXiv:1506.02025, 2015.
- [26] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 770-778.

作者简介:



李晓云(1991—),女,山西太原人,硕士研究生,2015年于太原科技大学获得学士学位,主要研究方向:机器学习理论及应用。E-mail: 1910049018@qq.com;



何秋生(1974—),男,博士,副教授,2007年于中国矿业大学(北京)获得博士学位主要研究方向:控制理论及应用、机器视觉。E-mail: heqs2008@126.com