

文章编号: 2095-4980(2015)03-0431-05

目标尺寸自适应压缩感知跟踪算法

苗青¹, 谭伟²

(1.上海交通大学 电子与通信工程系, 上海 200240; 2.复旦大学 通信科学与工程系, 上海 200433)

摘要: 压缩感知近年来在目标跟踪领域得到了广泛应用, 它对海量特征压缩降维, 在贝叶斯分类器模型下能取得很好的分类效果, 处理速度快, 具有实时性。但尺寸固定不变的跟踪窗口不能有效跟踪存在明显尺度变化的目标。本文采用多尺度和级联分类器机制, 选取最佳尺度下的窗口作为最终目标。实验结果表明, 本算法不仅在目标形态变化、光线变化、多目标干扰、运动模糊等复杂场景下有较好跟踪效果, 在目标尺度变化时也有较强鲁棒性。

关键词: 压缩感知; 跟踪; 多尺度; 尺寸自适应

中图分类号: TN911.73; TP391.4

文献标识码: A

doi: 10.11805/TKYDA201503.0431

A target size adaptive compressive tracking

MIAO Qing¹, TAN Wei²

(1.Department of Electronics and Communication Engineering, Shanghai Jiaotong University, Shanghai 200240, China;
2.Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China)

Abstract: Compressive Sensing has been successfully applied to visual tracking for its high efficiency and robustness in dimension reduction. It can achieve amazing classification results by using naive Bayes classifier with high frame rate. Nevertheless, it performs poorly while tracking targets with obvious scale change. An improved algorithm based on compressive tracking is proposed, which adopts multi-scale and cascade classifier to determinate the best scale and position. Experiment results indicate this algorithm not only performs well in challenging scenes such as pose variation, abrupt illumination, multi-target interference, motion blur, but also prevails in scale variant sequences.

Key words: compressive sensing; tracking; multi-scale; size adaptive

目标跟踪一直是机器视觉研究的重要分支, 广泛用于军事制导、道路监控、工业自动化控制、客流统计等领域。尽管近年来跟踪算法^[1-3]层出不穷, 但由于目标形态变化、光照变化、遮挡、背景干扰等影响, 对目标持续稳定跟踪一直是难题。部分算法^[1,4]跟踪窗口由目标初始大小决定, 在整个过程中跟踪窗口保持不变。当运动目标尺寸变大或变小时, 如果跟踪窗口固定不变, 跟踪窗口内包含目标的同时会包含大量背景, 或者只聚焦于目标局部区域, 这都将导致跟踪漂移甚至失败。文献[5-6]中方法可以解决尺度变化问题, 具有较好的跟踪效果和较快的处理速度, 本文引入了该思路。

压缩跟踪(Compressive Tracking, CT)算法^[7]是近年提出的一种基于压缩感知的简单高效跟踪算法。该算法实现简单, 跟踪效果好, 且处理速度快, 能达到 40 frame/s。但该算法缺点是跟踪窗口尺寸固定不变。本文在 CT 基础上提出了基于压缩感知的目标尺寸自适应跟踪算法(Target Size Adaptive Compressive Tracking, TSACT)。采用多尺度机制, 在不同尺度下挑选出多个候选目标作为第 2 级分类器输入。第 2 级分类器采用外观模型, 用图像灰度信息作为特征, 用归一化颜色相关性作为匹配标准。实验结果表明, TSACT 不仅具有针对复杂环境的鲁棒性, 而且当目标尺度发生变化时, 跟踪窗口也能跟随目标尺寸变化自适应调整。

1 基本原理

1.1 压缩感知和随机投影理论

由稀疏感知理论可知, 用满足有限等距性质(Restricted Isometry Property, RIP)^[8]条件的稀疏测量矩阵对原

图像特征空间投影,可得到低维压缩子空间,该低维压缩子空间可以很好地保留高维图像特征空间的信息。

一个维度为 $n \times m$ 的随机矩阵 $\mathbf{R}$,可以将一个高维图像空间 $\mathbf{x}(m \text{ 维})$ 变换到一个低维空间 $\mathbf{v}(n \text{ 维})$, $\mathbf{v} = \mathbf{R}\mathbf{x}(n < m)$ 。在理想情况下,低维 $\mathbf{v}$ 可以完全保留高维信息,或保持原始空间各样本 $\mathbf{x}$ 的距离关系。Johnson-Lindenstrauss 推论^[9]表明:可以随机选择一个适当高维子空间,原始空间 2 点的距离投影到这个子空间,能高概率保留这种距离关系。Baraniuk^[10]证明了满足 Johnson-Lindenstrauss 推论的随机矩阵同样满足压缩感知理论中的 RIP 条件。因此,如果随机矩阵 $\mathbf{R}$ 满足 Johnson-Lindenstrauss 推论, $\mathbf{x}$ 是可压缩的,则可以通过最小化误差来从 $\mathbf{v}$ 中高概率恢复 $\mathbf{x}$ 。

1.2 随机测量矩阵

一个比较典型的满足 RIP 条件的测量矩阵是随机高斯矩阵,矩阵元素满足 $N(0,1)$ 分布,如式(1)所示:

$$\gamma_{ij} = \sqrt{s} \times \begin{cases} 1 & \text{prob} = 1/2s \\ 0 & \text{prob} = 1 - (1/s) \\ -1 & \text{prob} = 1/2s \end{cases} \quad (1)$$

Achlioptas^[9]证明上式中 s 取 2 或者 3 时,矩阵满足 Johnson-Lindenstrauss 推论。

1.3 压缩算法

从图像角度看,将 $\mathbf{x}$ 看成图像块, $\mathbf{v} = \mathbf{R}\mathbf{x}$, 则 $\mathbf{v}$ 的每个元素 v_i 表示为 $v_i = \sum_{j=1}^m r_{ij}x_j$, 即用测量矩阵 $\mathbf{R}$ 每行非零元素去感知 $\mathbf{x}$ 中的元素,等价于一个方形窗口滤波器和输入图像某一固定位置的灰度卷积。由于测量矩阵 $\mathbf{R}$ 的系数可正可负,所以压缩特征可以像广义 Haar-like^[4]特征一样计算灰度差。稀疏感知理论保证:通过稀疏测量矩阵对高维 Haar-like 特征进行压缩后的特征能保留原有图像大部分信息。因此,可以直接对压缩空间的投影特征进行分类,避免维数灾难。

1.4 贝叶斯分类器

对每个样本 $\mathbf{x}$,低维表示为 $\mathbf{v}$,假定 $\mathbf{v}$ 中各元素独立,可用朴素贝叶斯分类器^[11]来建模,具体如式(2)所示。其中, $y = \{0,1\}$ 代表样本标签, $y=0$ 表示负样本, $y=1$ 表示正样本。假设 2 个类先验概率相等: $p(y=1) = p(y=0) = 0.5$ 。Diaconis 和 Freedman^[12]证明了高维随机向量的随机投影几乎都是高斯分布的。因此,假定分类器 $H(\mathbf{v})$ 中的条件概率 $p(v_i | y=1)$ 和 $p(v_i | y=0)$ 属于高斯分布,并且用 4 个参数来描述: $\mu_i^1, \sigma_i^1, \mu_i^0, \sigma_i^0$ 。

$$H(\mathbf{v}) = \log \left(\frac{\prod_{i=1}^n p(v_i | y=1)p(y=1)}{\prod_{i=1}^n p(v_i | y=0)p(y=0)} \right) = \sum_{i=1}^n \log \left(\frac{p(v_i|y=1)}{p(v_i|y=0)} \right) \quad (2)$$

4 个参数按式(3)和式(4)进行遗忘更新:

$$\mu_i^1 = \lambda \mu_i^1 + (1-\lambda) \mu^1 \quad (3)$$

$$\sigma_i^1 = \sqrt{\lambda(\sigma_i^1)^2 + (1-\lambda)(\sigma^1)^2 + \lambda(1-\lambda)(\mu_i^1 - \mu^1)^2} \quad (4)$$

式中: $\mu^1 = \frac{1}{z} \sum_{k=0|y=1}^{z-1} v_i(k)$; $\sigma^1 = \sqrt{\frac{1}{z} \sum_{k=0|y=1}^{z-1} (v_i(k) - \mu^1)^2}$; z 为样本个数。上式由最大似然估计^[13]得到。

1.5 归一化颜色相关

归一化颜色相关(Normalized Color Correlation, NCC)^[14]经常被用来做模板匹配或特征匹配。其定义如式(5):

$$r(u, v) = \frac{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}] [t(x-u, y-v) - \bar{t}]}{\sqrt{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}]^2 \sum_{x,y} [t(x-u, y-v) - \bar{t}]^2}} \quad (5)$$

式中: f 表示图像; t 为模板; x, y 为图像坐标; u, v 为窗口求和中心; 窗口尺寸为 $w \times h$ 。

2 改进算法

2.1 基本思想

CT 算法尺度不变性是通过 1.3 节计算特征时采用多尺度矩形滤波器实现的。虽然 Haar-like 特征具有尺度不变性,但由于跟踪目标尺寸固定不变,没有真正意义上实现多尺度特性,不能有效跟踪尺度变化目标。本文思想之一是对该算法尺度不变性加以利用。以 1.05 为尺度变化步长,在 9 个不同尺度进行 CT 跟踪。多尺度的实现与高斯金字塔类似^[15],但不图像缩放,而是以固定因子对初始目标尺寸缩放,并提取矩形框内特征。采用 1.05 为步长保证候选目标尺寸变化较小,提高 CT 分类器准确率。采用 9 个尺度能覆盖目标尺寸 1.55 倍变化,对于视频连续 2 帧来说,可以处理剧烈尺度变化。尺度越多,计算量越大,因此取 9 个尺度来寻求性能和处理速度的折衷。通过仿真发现,在每个尺度下,CT 分类器都能找到最佳位置,如图 1 所示。将不同尺度下的多个候选目标通过二级分类器再次分类,以获得最终输出。二级分类器采用外观模型,用图像灰度作为特征,用归一化颜色相关系数进行精确匹配定位。这样,既能保持原算法的高效准确性,又能弥补它只能处理固定尺度的不足。

与 TLD(Tracking Learning Detection)算法^[5-6]类似,采用级联分类器机制。第 1 级为 CT 分类器。训练时,在目标附近采样足够的正负样本并训练朴素贝叶斯分类器。检测时,在上一帧检测位置附近,在多个尺度上获取图像块并用贝叶斯分类器分类,分数最高的多个候选位置进入下一级分类器。第 2 级为 NN(Nearest Neighbor)分类器。该分类器维持一个正样本库,用来存储目标多个姿态。将第 1 级分类器输出与二级样本库所有样本进行 NCC 匹配,匹配最好的结果作为最终输出。只有当候选目标与样本库中样本相似度较低时,才学习该样本,表明当前目标形态发生了变化。为了防止学到错误的样本,相似度太低的样本不予学习。为避免样本库无限扩大,同时考虑计算复杂度,采用遗忘机制,删除老样本。

2.2 具体算法

输入:第 t 帧图像,NN 分类器正样本库 Φ 。

1) 在 $t-1$ 帧目标附近采样图像块 p 。采样时采用 r 个尺度,相邻 2 个尺度间隔为 s 。得到不同尺度的图像块集合 Ω 。对 Ω 中的所有元素进行压缩降维,得到每个图像块特征向量 $\mathbf{v}_p$ 。

2) 用 CT 分类器分类,选取分数最大的 N 个图像块作为当前帧候选目标 $C_1, C_2, \dots, C_N$ 。

3) 对于 $C_1, C_2, \dots, C_N$,与 Φ 中所有正样本进行 NCC 匹配。匹配采用归一化颜色相关系数,具体见式(5)。选取分数最大的候选样本 C_k 作为当前帧最终目标框输出 B_t ,对应分数为 $Score_k$,并找出 Φ 中对应正样本 Q_j ,以及当前尺度 s_k 并更新 NN 分类器样本库。

4) 在尺度 s_k 下目标框 B_t 附近采样正负 2 个样本集,用于 CT 分类器训练。2 个样本集分别为 $D_\alpha = \{p \mid \|L(p) - B_t\| < \alpha\}$ 和 $D_{\zeta, \beta} = \{p \mid \|L(p) - B_t\| < \beta\}$ 。其中, $\alpha < \zeta < \beta$, $L(p)$ 为图像块的位置。

5) 提取上述 2 个样本集的特征,用式(3)~式(4)更新 CT 分类器参数。

输出:跟踪目标位置 B_t ,更新后的 CT 分类器参数,NN 分类器样本库。

其中,NN 分类器样本库更新机制如下:

- 1) 如果 $Score_k$ 低于阈值 T ,学习该样本,表明目标形态发生变化。
- 2) 如果 $Score_k$ 高于阈值 T ,不学习该样本,但将正样本 Q_j 移到样本库中最新样本位置。
- 3) 如果样本个数超过阈值,删除最老的样本,保留新样本。

3 实验设置

初始目标框由人工选定。产生正样本半径 $\alpha=4$,产生负样本的内半径和外半径分别为 8 和 30,搜索半径为 20。投影空间维数为 50,学习因子 $\lambda=0.85$ 。尺度 $r=9$,尺度间隔 $s=1.05$ 。CT 中选取的候选目标个数 $N=100$ 。归一化尺寸 $w=25, h=25$,阈值 $T=0.65$ 。 T 越大,越容易学习新样本,也容易学到错误样本,导致跟踪漂移; T 越小,

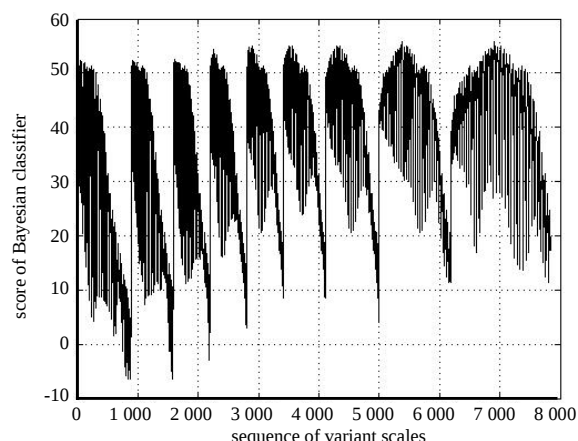


Fig.1 Multi-scale feature

图 1 多尺度特性

越难学习新样本，一旦目标形态快速发生变化，跟踪容易漂移。经过测试发现 $T=0.65$ 时跟踪效果较好。

4 实验结果

选取 3 段视频进行测试。其中 Twinnings 和 Cliff bar 目标尺度存在较大变化。由于随机性存在，每段测试视频运行 5 次。用 2 个指标来衡量跟踪性能：第 1 个指标为成功率(sr)。成功率定义为成功跟踪帧数 n_{succ} 与测试视频总帧数 n_{total} 比值，具体定义如式(6)。成功跟踪定义为跟踪框与真实目标框交集与并集比大于 0.5。

$$sr = \frac{n_{succ}}{n_{total}} \quad (6)$$

第 2 个指标为中心偏差(Center Location Error, CLE)，定义为跟踪框中心坐标与真实框中心坐标像素偏差，见式(7)。

$$CLE = \sqrt{(ROI_x - ROI_{gx})^2 + (ROI_y - ROI_{gy})^2} \quad (7)$$

式中： ROI_x 为跟踪框中心 x 坐标； ROI_y 为跟踪框中心 y 坐标。

从表 1 和表 2 可以看到，TSACT 比原算法在成功率上有提高，而中心偏差则降低。Twinnings 和 Cliff bar 跟踪目标由于存在明显尺度变化，在跟踪精确性上提升比较明显。Coupon book 在尺度上变化不大，两者精确度差不多。因此在尺度变化较大的场景，TSACT 算法更有优势。

表 1 成功率
Table1 Success rate

video sequence	success rate	
	CT	TSACT
Twinnings	0.550 2	0.670 7
Cliff bar	0.689 0	0.779 3
Coupon book	0.796 0	0.725 8
average	0.678 4	0.725 3

表 2 中心偏差
Table2 Center location error

video sequence	center location error	
	CT	TSACT
Twinnings	11.915 6	5.793 6
Cliff bar	7.912 0	6.651 7
Coupon book	16.8514	17.224 8
average	12.226 3	9.890 0

图 2 和图 3 是测试视频部分检测结果。黑色框对应 TSACT，白色框对应 CT。TSACT 算法由于具有自适应尺度，比 CT 算法具有较好的跟踪准确性。2 个算法在跟踪目标形态变化、方向变化、背景干扰和相似物干扰的复杂场景下都能取得较好的跟踪性能。

最后对算法的实时性进行验证。实验在处理器为 1.83 GHz CPU 上进行，采用 Matlab 计算，平均每帧处理时间为 400 ms。如果转化为 C 代码，再加上 SIMD 流指令扩展(Streaming SIMD Extensions, SSE)优化、多线程优化，能达到实时性要求。



Fig.2 Test results for Twinning
图 2 Twinning 测试样本



Fig.3 Test results for Cliff bar
图 3 Cliff bar 测试样本

5 结论

本文提出了基于压缩感知的自适应尺寸跟踪算法。该算法基于压缩感知理论，用随机投影提取图像 Haar-like 特征，并用贝叶斯分类器进行分类。为了解决跟踪目标框尺寸不变的问题，运用多尺度机制及级联分类器思想，

在CT分类之后进行NN分类,挑选最佳尺度及位置作为最终输出。仿真结果表明,该算法既具有CT算法的高效及准确性,又有效地解决了跟踪目标尺寸固定不变的限制性。该算法在尺度变化较大的场景更能凸显其优势。

参考文献

- [1] 李才会,张群,陆起涌. 基于自适应背景信息的目标跟踪算法[J]. 太赫兹科学与电子信息学报, 2011,9(5):596-599.
(LI Caihui,ZHANG Qun,LU Qiyong. Object tracking algorithm based on adaptive background information[J]. Journal of Terahertz Science and Electronic Information Technology, 2011,9(5):596-599.)
- [2] Arulampalam S,Maskell S,Gordon N. A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking[J]. IEEE Transactions on Signal Processing, 2002,50(2):174-188.
- [3] Babenko B,YANG Minghsuan. Robust object tracking with online multiple instance learning[J]. IEEE Trans. on PAMI, 2011, 33(8):1619-1632.
- [4] MEI Xue,LING Haibin. Robust visual tracking and vehicle classification via sparse representation[J]. IEEE Trans. on PAMI, 2011,33(11):2259-2272.
- [5] Kalal Z,Mikolajczyk K,Matas J. Tracking-Learning-Detection[J]. Pattern Analysis and Machine Intelligence, 2012,34(7): 1409-1422.
- [6] Kalal Z,Matas J,Mikolajczyk K. P-N learning:bootstrapping binary classifiers by structural constraints[C]// Conference on Computer Vision and Pattern Recognition. San Francisco,CA:[s.n.], 2010:49-56.
- [7] ZHANG Kaihua,ZHANG Lei,YANG Minghsuan. Real time compressive tracking[C]// ECCV. Florence,Italy:[s.n.], 2012: 864-879.
- [8] Candes E,Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005,51(12):4203-4215.
- [9] Achlioptas D. Database-friendly random projections:Johnson-Lindenstrauss with binary coins[J]. Computer System Science, 2003,66(4):671-687.
- [10] Baraniuk R,Davenport M,Devore R,et al. A simple proof of the restricted isometry property for random matrices[J]. Constructive Approximation, 2008,28(3):253-263.
- [11] Ng A,Jordan M. On discriminative vs. generative classifier:a comparison of logistic regression and naive Bayes[J]. Neural Information Processing Systems(S2249-7110), 2002,52:841-848.
- [12] Diaconis P,Freedman D. Asymptotics of graphical projection pursuit[J]. Ann. Statist., 1984,12(3):228-235.
- [13] Myung I J. Tutorial on maximum likelihood estimation[J]. Journal of Mathematical Psychology, 2003,47(1):90-100.
- [14] Yong S H,Kyoung M L. Robust stereo matching using adaptive normalized Cross-Correlation[J]. IEEE Transactions on PAMI, 2011,33(4):807-822.
- [15] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

作者简介:



苗 青(1988-),男,浙江省衢州市人,工程硕士,主要研究方向为计算机视觉、图像处理.email:miaoqing883@163.com.

谭 伟(1986-),男,湖南省益阳市人,在读硕士研究生,主要研究方向计算机视觉、图像处理.