

文章编号: 1672-8785(2014)08-0032-06

## 基于多模型融合的红外序列跟踪算法

陈鹏飞<sup>1</sup> 史文中<sup>2</sup>

(1. 武汉大学遥感信息工程学院, 湖北 武汉 430079 ;

2. 武汉大学-香港理工大学空间信息联合实验室, 湖北 武汉 430079)

**摘 要:** 作为一种新型图像获取技术, 热红外技术能避免光照强度和背景颜色对目标提取的影响。当前利用热红外序列进行目标跟踪的算法大多结合可见光和热红外图像。针对这种局限性, 提出了一种纯粹基于热红外序列的跟踪算法。首先, 通过改进的码书算法和灰色预测模型, 对序列中的背景信息和运动目标的运动方向进行建模; 然后, 将建模结果结合到压缩跟踪算法中, 通过自适应的方法使背景信息、运动信息以及目标特征信息在跟踪过程中互相补充, 消除跟踪算法对于可见光图像的依赖性, 减少由于目标纹理特征不明显而对跟踪效果产生影响。实验证明, 经改进的跟踪算法处理速度快, 正确性得到了明显的提高, 能更好地适应热红外序列的特点。

**关键词:** 热红外; 压缩跟踪; 码书算法; 背景信息

**中图分类号:** TP391.4      **文献标志码:** A      **DOI:** 10.3969/j.issn.1672-8785.2014.08.007

## A Tracking Algorithm of Infrared Sequence Based on Multi-model Integration

CHEN Peng-fei<sup>1</sup>, SHI Wen-zhong<sup>2</sup>

(1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;

2. A Joint Research Laboratory on Spatial Information, The Hong Kong Polytechnic University and Wuhan University, Wuhan 430079, China)

**Abstract:** As a new image acquisition technology, the thermal infrared imaging can avoid the influence of illumination and background color on its image acquisition. Currently, the algorithms which use a thermal infrared sequence for target tracking mostly combine visible light images with thermal infrared images. In view of this limitation, a tracking algorithm based solely on the thermal infrared sequence is proposed. First, the background information and the moving direction of a target in the sequence are modeled by using a modified codebook algorithm. Then, the modeling result is incorporated into the compression tracking algorithm. The information on background, target motion and target characteristics is complemented one another in the tracking process. The experiment shows that the improved algorithm is fast in processing. Because its accuracy is improved greatly, it is adapted to the characteristics of thermal infrared sequences better.

**Key words:** thermal infrared; compressive tracking; codebook algorithm; background information

收稿日期: 2014-06-20

作者简介: 陈鹏飞(1989-), 男, 湖北武汉人, 硕士研究生, 主要从事 GIS 和热红外研究。E-mail: Pfcchen@whu.edu.cn

## 0 引言

目标跟踪一直是计算机视觉领域中的核心问题之一,它在智能交通、军事边防、机器人导航以及制导武器研究等问题上都有广泛的应用。目标跟踪算法通常基于可见光序列,其效果往往依赖于可见光序列的质量。而可见光序列的质量在很大程度上是依赖于光照强度、环境可见度以及背景的复杂程度的<sup>[1,2]</sup>。因此,这类跟踪算法的鲁棒性不好,在夜间几乎完全失效。

热红外序列是一种新型图像获取技术,它根据任何物体都会不断地散发热辐射这一原理,能快速地获取物体的热辐射强度,并将其作为视频跟踪的数据源。这不仅能大幅地消减光照强度的影响,而且对环境中的烟雾以及复杂背景都有很好的过滤作用。

目前,利用热红外序列进行目标跟踪的算法大多通过热红外图像与可见光图像融合来实现。这种融合方式既能保留可见光图像中的多数纹理特征,又可以结合热红外图像的热辐射强度信息,在实践中能很好地提高目标跟踪的效果<sup>[3-6]</sup>。可是这类算法在夜间同样会几乎完全失效。

为了解决上述问题,本文提出一种结合跟踪算法、背景提取与轨迹预测的多模型跟踪算法。该算法通过改进后的码书(简称 Codebook)模型与灰色预测模型,分别对样本的背景信息和运动信息进行建模,并将建模结果融入到压缩跟踪(Compression Tracking, CT)算法中,通过自适应方法使背景信息、运动信息以及目标特征信息在跟踪过程中互相补充,以提高跟踪精度。实验结果表明,该算法准确度高,能很好地弥补传统 CT 算法在跟踪红外目标方面的不足。

## 1 Codebook 模型

Codebook 算法<sup>[7]</sup>是一种像素级背景分离算法。其背景模型由一系列码本组成,每个码本由若干个码元组成,每个码元又由一个包含像素 R、G、B 通道信息的三维矢量  $V_i = (R_i, G_i, B_i)$  和一个包含亮度最大最小值、出现频率和出现

时间等成员的六维矢量  $aux_i = (\check{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i)$  组成。Codebook 算法会对每一帧视频图像的每个像素点与对应位置的码元进行匹配。若像素值满足一定条件,则更新对应码元的矢量信息;若不匹配,则建立新的码元信息。

原始的 Codebook 算法适用于三通道的可见光序列。当视频序列为热红外序列时,由于通道数的减少,码元的匹配和更新方法不再适用。本文提出的改进型 Codebook 算法剔除了原始算法中的矢量  $V_i$ ,同时,将六维矢量  $aux_i$  改为  $aux_k = (I_k, f_k, \lambda_k, p_k, q_k)$ ,其中  $f_k, \lambda_k, p_k, q_k$  四个参数的含义与原始算法保持一致,分别代表该码元出现的频率、最大重复出现的时间间隔、首次出现时间和最后出现时间。参数  $I$  代表该码元的亮度值,新进像素的亮度值若与  $I$  的差值在匹配阈值  $\omega$  内,则更新  $I$ 。更新方法与原始的 Codebook 算法一致。这里修改原始 Codebook 的  $I_{low}$  与  $I_{high}$  为固定的匹配阈值  $\omega$ ,是因为对热红外序列而言,运动目标的亮度值大小比较稳定,不会出现可见光序列中由于光照或者颜色明暗程度造成亮度剧烈变化的现象。

## 2 灰色预测模型

灰色预测模型是灰色系统理论的一个部分,它通过对历史数据进行挖掘,掌握系统的发展规律,对系统的未来状态进行预测<sup>[8]</sup>。GM(1,1)模型是灰色模型中最常用的一种,通过累加原始数据生成的新数据模型,能够在一定程度上弱化原始数据的随机性。

灰色预测模型的原理是用序列  $x^{(0)}(k)$ ,  $k = 0, 1, 2, \dots, N$  代表原始数据,经过累计后的序列为

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, N \quad (1)$$

灰色微分方程  $x^{(0)}(k) + az^{(1)}(k) = u$  即为 GM(1,1)模型,其中

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad (2)$$

通过多个原始数据组成方程组, 方程组参数的最小二乘估计为

$$U = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T y \quad (3)$$

式中

$$y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(N) \end{bmatrix}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(N) & 1 \end{bmatrix} \quad (4)$$

求得估计参数后, 求解灰色微分方程, 即可得到预测模型

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(1) - \frac{\hat{a}}{\hat{a}}]e^{-\hat{a}k} + \frac{\hat{a}}{\hat{a}} \quad (5)$$

当  $k \geq N$  时, 即可求得预测值。

相对于常用的运动预测方法如 Kalman 滤波等, 灰色预测模型需要的历史数据少, 不需要进行初值选取和参数设定, 计算量小, 预测速度快, 并且能够对行人运动等这类随机性较大的运动模式进行较准确的预测。

### 3 压缩跟踪算法

CT 算法是一种简单高效的新型跟踪算法, 在 2012 年由 Zhang 等人提出<sup>[9]</sup>。该算法利用稀疏矩阵对跟踪目标周围的多尺度图像样本进行降维并提取其特征, 利用朴素贝叶斯分类器对特征进行分类, 同时通过正负样本对该分类器进行更新, 最终得到被跟踪目标在下一帧的位置。

CT 算法中最大的特点就是其对图像进行特征提取时运用到了压缩感知理论。根据压缩感知理论, 用满足 RIP 条件的非常稀疏的矩阵对图像进行降维压缩后, 可以很好地保留原本高维图像中的信息<sup>[10]</sup>。另外, 利用朴素贝叶斯分类器对跟踪目标和背景进行分离也是 CT 算法中的

特点之一。朴素贝叶斯分类器算法简单, 能够快速选取待分类项最大概率从属的类别, 满足实时目标跟踪的需求。其在算法中具体体现为

$$H(\nu) = \log\left(\frac{\prod_{i=1}^n p(\nu_i|y=1)p(y=1)}{\prod_{i=1}^n p(\nu_i|y=0)p(y=0)}\right)$$

$$= \sum_{i=1}^n \log\left(\frac{p(\nu_i|y=1)}{p(\nu_i|y=0)}\right) \quad (6)$$

从式 (6) 可以看到, CT 算法中的分类过程实际上是计算待分类项的每个特征值属于正样本和负样本的概率, 最终选取所有样本中贝叶斯值最大的那一项作为跟踪结果。很明显, 这种概率计算方式在很大程度上依赖于样本特征值的大小。但是对于红外序列而言, 由于每一帧图像都是单通道图像, 其样本特征值的差异性会大幅小于可见光序列, 若在红外序列的跟踪中仅仅考虑特征值的差异性, 跟踪结果的可靠性会大幅降低。

### 4 结合多模型的改进压缩跟踪算法

为了解决上述问题, 本文对 CT 算法中的分类器进行了改进, 在其中加入了背景信息与被跟踪目标的历史运动信息。通过考虑样本与前景信息和被跟踪目标运动信息的关系, 计算样本在这两种条件下的概率, 最后结合 CT 算法中的特征值, 并运用自适应的方法赋予三种信息一定的权重值, 选出最优的样本作为最终结果。

假设样本集合  $S = \{s_1, s_2, \dots, s_N\}$ , 待检样本为  $s_k$ , 前帧利用 Codebook 算法进行前景分离的结果为  $X_t$ 。对跟踪窗口而言, 最优的跟踪结果应当尽可能地处于窗口的中间位置。若  $z_i$  在  $X_t$  中对应的区域为  $N$ , 则样本  $s_k$  在背景信息中的概率定义为

$$P(X_t|s_k) = \exp\left[\sum_{i \in N} \delta(x_t^i - f)\omega(z_i - z^*)\right] \quad (7)$$

式中,  $\hat{x}_t^N = \{x_t^i|i \in N\}$  为  $X_t$  中对于区域  $N$  的像素集合,  $f$  为前景标签,  $\delta$  为克罗内克函数,  $z^*$  为样本  $s_k$  的中心位置,  $z_i$  为  $x_t^i$  对应的位置,  $\omega$  为空间权重函数, 其定义为

$$\omega(z) = e^{-\frac{|z|^2}{\sigma^2}} \quad (8)$$

式中,  $\sigma$  为尺度参数, 在本文中取窗口  $N$  的长与宽的和的一半。

除背景信息外, 本文利用灰色预测模型对被跟踪目标的运动位置进行预测。为了保证跟踪的实效性, 本文通过前三帧检测到的跟踪结果建立 GM(1,1) 模型, 然后根据预测的目标位置与上一帧的跟踪位置计算出运动矢量信息,  $\hat{D} = (\hat{x}, \hat{y})$ , 其中  $\hat{x}, \hat{y}$  分别代表目标预测的水平方向和垂直方向的位移。据此, 若样本  $Z$  的运动方向矢量为  $D = (x, y)$ , 历史运动信息为  $\hat{Z}_t$ , 则运动信息的相似度可以定义为

$$d_{\hat{Z}_{t-1}, Z} = 1 / \sqrt{(x^2 + y^2) - \frac{\hat{x}_{t-1}x + \hat{y}_{t-1}y}{\hat{x}_{t-1}^2 + \hat{y}_{t-1}^2}} \quad (9)$$

样本  $Z$  在运动信息中的概率  $P(\hat{Z}_t|Z)$  定义为

$$P(\hat{Z}_t|Z) = \exp(d_{\hat{Z}_t, Z}) \quad (10)$$

为了将式 (6)、(7)、(10) 中三种信息的概率更有效地结合起来, 需要分别进行归一化处理。同时, 根据样本集合  $S$  中所有样本计算的这三种信息的概率值的分布情况, 对差异较大、样本间区分度大的信息赋予较大的权重; 对差异较小, 样本间区分度较小的信息赋予较小的权重。这样, 在跟踪的过程中, 本文算法能自适应地根据样本在三种信息下的区分度, 选择最适合当前环境的信息源作为搜索跟踪目标的依据。本文中, 以每一种信息概率值的方差作为描述其差异大小的指标。

若进行归一化的函数为  $F(x)$ , 假设对三种概率值的集合进行归一化后计算出的方差分别为  $\sigma_1, \sigma_2, \sigma_3$ , 则权重计算公式为

$$\omega_i = \frac{\sigma_i}{\sum_{i=1}^3 \sigma_i}, i = 1, 2, 3 \quad (11)$$

经归一化和赋予权重后, 对 (6)、(7)、(10) 分别进行修改, 样本  $s_k$  的特征信息下的概率为

$$P(s_k) = \exp(\omega_1 * F(H(\nu))) \quad (12)$$

背景信息下的概率为

$$P(X_t|s_k) = \exp[\omega_2 * F(\sum_{i \in N} \delta(X_t^i - f) \omega(z_i - z^*))] \quad (13)$$

运动信息下的概率为

$$P(\hat{Z}_t|s_k) = \exp(\omega_3 F(d_{\hat{Z}_t, Z})) \quad (14)$$

根据贝叶斯概率理论,  $s_k$  为跟踪结果的概率, 可以定义为

$$\begin{aligned} P(s_k|X_t, \hat{Z}_t) &\propto P(X_t, \hat{Z}_t|s_k)P(s_k) \\ &\propto P(X_t|s_k)P(\hat{Z}_t|s_k)P(s_k) \end{aligned} \quad (15)$$

最终, 计算出  $P(s_k|X_t, \hat{Z}_t)$  值最大的时所对应的样本  $s_k$ , 即为当前这一帧的跟踪结果。

## 5 实验分析

实验的运行环境为 Intel Core I5, 主频为 3.1GHz, 内存为 4GB。软件环境为 VS2012 和 Opencv2.4.6。第一组实验使用的视频图像下载于 OSU Color-Thermal Database<sup>[11]</sup>, 视频大小为  $320 \times 240$ , 帧频为 30 Hz, 一共有 538 帧。第二组实验使用由 FLIR-SC660 录制的视频图像, 视频大小为  $640 \times 480$ , 帧频为 30 Hz, 一共有 1153 帧。

实验中, 利用原始的 CT 算法以及本文的改进算法对视频中的红外目标同时进行跟踪。图 1 中的绿色框代表原始 CT 算法, 红色框代表改进后的算法。

首先利用视频的前 60 帧进行背景建模, 得到初始化的码书模型。第 68 帧时的原始视频、前景以及跟踪结果如图 1 所示。

可以看到, Codebook 算法能够很好地对红外视频进行背景分离, 提取出的前景目标稳定且无遗漏。刚开始对目标进行跟踪时, 原始的 CT 算法与本文算法在效果上差异不大, 均能很好地对前景进行跟踪。

图 2(a)、(b)、(c) 分别代表第 285 帧、第 333 帧和第 448 帧时的跟踪结果。可以看到, 从 285 帧开始, 由于目标 1 和目标 3 的距离太近, 特征过于相似, 原始的 CT 算法对目标 1 的跟踪





图 1 第 68 帧时的跟踪效果

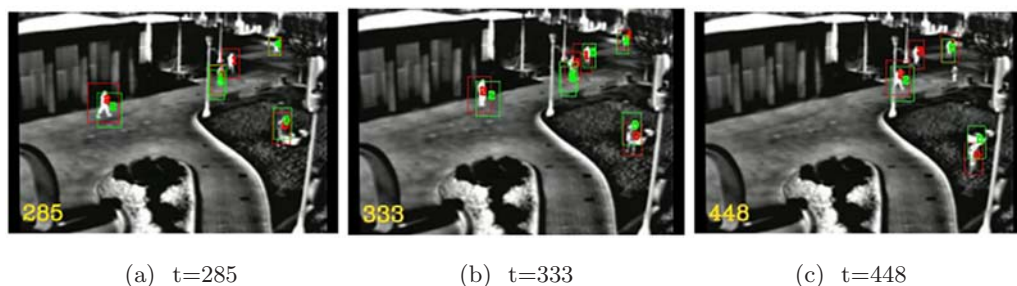
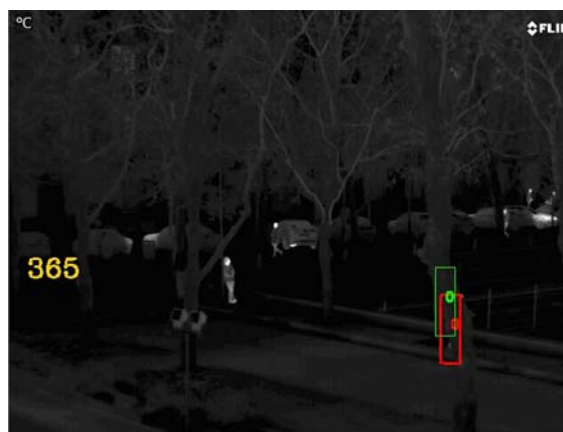


图 2 第一组实验中不同帧时的跟踪结果

出现了很明显的错误, 将目标 3 错判为目标 1; 同样, 在第 333 帧中, 对于目标 1 和目标 3 的跟踪同样出现了明显的偏移; 在第 501 帧中, 原始的 CT 算法将一个与目标 0 的温度相近的物体错判为目标 0, 导致跟踪结果出现很大的偏移。整个序列跟踪总耗时 21.39 s, 平均 25.15 帧/s, 基本满足视频监控 15 Hz 到 30 Hz 的水平。

第二组实验中, 目标跟踪从第 100 帧开始, 图 3(a) 显示其在第 171 帧时的跟踪结果。可以看到, 跟踪窗口差异并不大。但当目标在第 365 帧中被树木完全遮挡时, 原始 CT 算法中的跟踪窗口发生了明显的偏移, 如图 3(b) 所示。而本文算

(b)  $t=365$ (a)  $t=171$ (c)  $t=620$

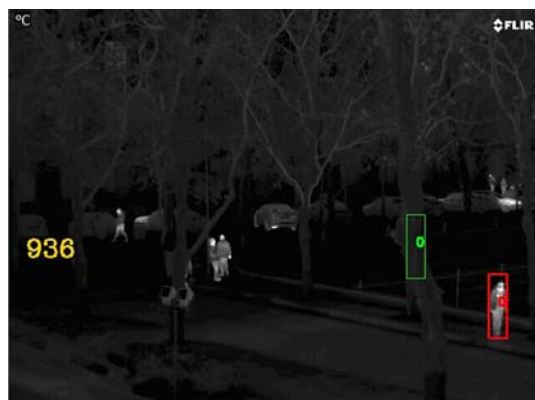
(d)  $t=936$ 

图 3 第二组实验中不同帧时的跟踪结果

法的跟踪窗口几乎未发生偏移。从图 3(c) 和 3(d) 可以很明显地看到, 由于之前树木的遮挡, 原始 CT 算法跟踪完全失效, 而本文算法的跟踪结果十分理想。

从两组实验中分别随机选取 300 帧对两种算法的错误率进行统计, 其中把有目标但未进行跟踪的划定为遗漏, 把对目标进行了跟踪但跟踪位置有偏离的划定为偏移。统计结果见表 1。

表 1 两种算法错误率的统计

| 实验组      | 实验组一  |      | 实验组二  |      |
|----------|-------|------|-------|------|
| 算法       | 原始 CT | 本文算法 | 原始 CT | 本文算法 |
| 存在遗漏时的帧数 | 25    | 2    | 155   | 10   |
| 存在偏移时的帧数 | 134   | 13   | 10    | 0    |
| 错误率      | 53%   | 5%   | 55%   | 3%   |

可以看到, 通过增加目标运动的方向信息和背景信息之后, 改进的 CT 算法在跟踪结果的遗漏和偏移上有显著提高, 错误率由 53% 和 55% 分别降低到了 5% 和 3%。新增的运动背景信息和目标运动方向信息能很好地消减红外序列由于通道数少、特征不明显和目标特征相似等固有特性对跟踪结果的不良影响。

## 6 结论

提出了一种结合码书算法、灰色预测模型以及压缩跟踪的红外目标跟踪算法, 可以很好地处理静态红外序列中的运动目标跟踪。该算法的核心思想是将码书提取的背景信息以及目标的运动信息融入到原始 CT 算法中, 以弥补红

外序列由于通道数过少而导致 Haar 特征不显著的缺陷。实验结果证明, 这种思路是可行的, 对红外序列的目标跟踪效果能够起到很好的提升作用。今后的工作可以结合特征识别技术, 对红外序列中的运动目标进行提取, 并进一步结合运动轨迹进行时空分析, 以便更加有针对性地结合实际运用跟踪技术。

## 参考文献

- [1] Roos D, Lim J, Yang M H. Adaptive Probabilistic Visual Tracking with Incremental Subspace Update [C]. *Proceedings of the Eighth European Conference on Computer Vision*, 2004 ,
- [2] Wu Y, Huang T S. Robust Visual Tracking by Integrating Multiple Cues based on Co-Inference Learning[J]. *International Journal of Computer Vision*, 2004, **58**(1):55-71.
- [3] Goubet E, Katz J, Porikli F. Pedestrian Tracking Using Thermal Infrared Imaging[C]. *Infrared Technology and Applications XXXII*, 2006, Proceedings of SPIE 6206, 797-808
- [4] Leykin A, Hammoud R. Pedestrian Tracking by Fusion of Thermal-visible Surveillance Videos[J]. *Machine Vision and Applications*, 2010, **21**(4): 587-595.
- [5] Bunyak F, Palaniappan K, Nath S K, et al. Geodesic Active Contour Based Fusion of Visible and Infrared Video for Persistent Object Tracking[C]. *IEEE Workshop on Applications of Computer Vision*, 2007:35-35.
- [6] Leykin A, Ran Y, Hammoud R. Thermal-visible Video Fusion for Moving Target Tracking and Pedestrian Classification[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007 : 1-8.
- [7] Kim K, Chalidabhongse T H, Harwood D, et al. Real-time Foreground-background Segmentation Using Codebook Model[J]. *Real-time Imaging*, 2005, **11**(3): 172-185
- [8] 邓聚龙. 灰预测与灰决策[M]. 武汉: 华中科技大学出版社, 2002
- [9] Zhang K, Zhang L, Yang M H. Real-time Compressive Tracking[M]. Berlin Heidelberg: Springer, Computer Vision-ECCV 2012, 2012: 864-877.
- [10] Candès E J, Wakin M B. An Introduction to Compressive Sampling[J]. *Signal Processing Magazine, IEEE*, 2008, **25**(2): 21-30.
- [11] Davis J and Sharma V. Background-Subtraction Using Contour-based Fusion of Thermal and Visible Imagery[J]. *Computer Vision and Image Understanding*, 2007, **106**(2-3):162-182.