

# 运动目标检测和跟踪算法综述

王慧

(安徽三联学院, 安徽 合肥 230601)

摘要: 运动目标检测和跟踪在多媒体图像、视频监控等方面有普遍应用。多年来, 人们在该领域进行了大量的、深入的研究, 发表了大量显著性的成果。但目标被遮挡、尺度变化以及光照变化仍然对跟踪结果存在较大影响。为解决这些问题, 研究人员仍然在研究如何能构造一个鲁棒性好的跟踪算法。该文主要对近年来常见的运动目标检测和跟踪算法的发展做了简单叙述。简单分析了运动目标检测的帧间差分法、背景减法、光流法的原理; 描述了运动目标跟踪所用的 Meanshift 算法、Kalman 滤波器、粒子滤波方法; 最后对压缩感知理论进行简介, 并对基于该理论的两种算法: 稀疏表示的目标跟踪算法和实时压缩感知目标跟踪算法做出系统的描述。

关键词: 运动目标检测; 运动目标跟踪; 压缩感知; 稀疏表示; 特征提取

中图分类号: TP391 文献标识码: A 文章编号: 1009-3044(2018)24-0194-04

DOI: 10.14004/j.cnki.ckt.2018.2862

A summary of Moving Target Detection and Tracking Algorithm

WANG Hui

(Anhui Sanlian University, Hefei 230601, China)

Abstract: Moving target detection and tracking are widely used in multimedia images, video surveillance, etc. Over the years, people have done a lot of in-depth research in this field, and published a lot of significant results. However, target occlusion, scale change and illumination change still have a great impact on tracking results. To solve these problems, researchers are still studying how to construct a robust tracking algorithm. In this paper, we briefly describe the development of common algorithms for moving object detection and tracking in recent years. The principle of frame difference method, background subtraction and optical flow method for moving object detection is simply analyzed. The Meanshift algorithm, Kalman filter and particle filter method for moving target tracking are described. In the end, a brief introduction to the theory of compressed sensing is made. And two algorithms based on this theory are described: sparse representation tracking and compressive tracking.

Key words: moving target detection; moving target tracking; compressed sensing; sparse representation; feature extraction

## 1 引言

运动目标检测和跟踪是计算机视觉学科主要的研究方向之一, 涵盖了自动控制、图像处理、智能识别等学科, 目前在多个领域比如视频监控、人工智能、军事等方面有非常广泛的应用。

运动目标检测与跟踪是分目标检测和跟踪两方面。目标检测指的是在监视的场景中对运动的目标做实时检测, 并提取该目标。目标跟踪是在将目标检测出来后, 根据该目标的特征, 利用合适的算法, 在一系列图像中寻找目标的位置。运动目标跟踪可以提供目标的运动轨迹, 可以对目标准确定位, 为下一步对目标行为分析与理解提供可靠的数据来源, 同时也可以为运动目标检测提供帮助, 形成一个良性的循环。

## 2 运动目标检测算法

运动目标检测是指在视频图像序列中检测出变化区域并将运动目标从背景图像中提取出来。由于目标分类、跟踪和分析等后续处理过程只是对运动的目标像素区域做分析, 因

此如何能正确的将运动目标检测与分割出来非常重要。

通常检测运动物体的有三种方法, 第一种是用帧间差分法; 第二种是背景减法; 第三种是光流法。

### 2.1 帧差分法

帧间差分法简单来说就是将前后两帧图像对应像素值相减。如果差值很小, 认为此处静止, 如果差值很大, 则认为是由物体运动引起的。把所有差值很大的像素标记出来, 就是运动目标在图像中的位置, 当然前提条件是前后两帧图像的环境亮度基本不变。如果背景亮度的差别很大, 以及目标在光照下的阴影处理, 帧间差分法还需要考虑更多的内容。

设差分图像为  $Y_k(i, j)$ , 第  $k+1$  帧与第  $k$  帧图像在  $(i, j)$  点的像素分别为  $T_{k+1}(i, j)$  与  $T_k(i, j)$ , 对其二值化后产生的结果为  $I_k(i, j)$ , 帧间差分法运算公式为:

$$Y_k(i, j) = |T_{k+1}(i, j) - T_k(i, j)| \quad (1)$$

$$I_k(i, j) = \begin{cases} 1 & Y_k(i, j) \geq I \\ 0 & Y_k(i, j) < I \end{cases} \quad (2)$$

收稿日期: 2018-07-08

基金项目: 安徽三联学院校级科研基金自然科学基金重点项目: 中医经络按摩机器人视觉系统研究(KJZD2017009)

作者简介: 王慧(1980—), 女, 安徽合肥人, 讲师, 学士, 主要研究方向为图像处理, 计算机网络。

上式中“0”代表第k+1帧与第k帧图像间没有发生运动变化的地方,“1”代表图像间发生运动变化的地方。

帧间差分法的主要优点是:算法简单,程序设计复杂度低,实时性强。但在实际应用中,这种最基本的帧间差分法精度不高,要求背景绝对静止或很少有变化,对目标很难精确标记出来。

如果背景是动态的,有很多改进方法,常见的有双差分法,区域差分法等。双差分方法是在第一次差分图像的基础上再将像素相乘,由于噪声在时间域上很难重复,相乘后就可以除去孤立的噪点。区域差分方法是在某一模板上做差分,利用空间信息将噪声消除。

## 2.2 背景差法

背景差法<sup>[1]</sup>是用图像序列中的视频帧与背景图像相减而检测出目标的方法,之后选择一个合理阈值对差值图像进行二值化,从而获取运动目标。

简单常用的方式为:直接把视频序列中某一幅图像提取出来,或以多幅图像的平均值作为背景。最简单的背景模型是时间平均图像,提供的目标特征数据也很完全。但这种背景对场景变化,比如光照等干扰敏感度较强。

背景差法中将当前帧与背景帧的差值图像用于检测运动目标。设视频序列像素(i,j)第K帧处的灰度值为 $M_k(i,j)$ ,此处的背景灰度值为 $B_k(i,j)$ ,则背景差图像 $Y_k(i,j)$ 与二值化结果 $I_k(i,j)$ 分别为:

$$Y_k(i,j) = |M_{k+1}(i,j) - B_k(i,j)| \quad (3)$$

$$I_k(i,j) = \begin{cases} 1 & Y_k(i,j) \geq T \\ 0 & Y_k(i,j) < T \end{cases} \quad (4)$$

背景差分法计算也很简单,得到的运动目标也比较完整,并且对视频中停止运动的物体也可以检测出来。缺点是如果出现了较大的光照和场景变化,可能会出现较多的噪声点,这就需要及时更新背景,因此算法的复杂性增加,实时性变差;而且阈值的选取对检测效果影响较大,一个好的阈值可以大大提高检测的精度。

## 2.3 光流法

在空间中,运动可以用运动场描述。在一个图像平面上,物体的运动往往通过图像序列中不同图像灰度分布的变化体现。空间中的运动场转移到图像上就表示为光流场,光流场反映了图像上每一点灰度在空间位置的变化趋势。光流是研究图像灰度在时间上的变化与景象中物体结果及其运动的关系。

光流法的优点是不需要预先知道场景的任何信息,就能够检测出独立运动的对象,可用于静态、动态背景。缺点是如果运动目标的亮度变化较大,会影响结果准确性;难以检测暂时不动的目标;计算复杂,抗噪性能差,实时性差。

## 3 运动跟踪

目标跟踪技术的基本思想是在序列视频图像中,已知需要跟踪的目标信息,通过提取出目标的各种特征来建模得到目标模板,再用候选模板与目标模板进行相似匹配,从而得到目标的新位置。由于目标可能因光照、遮挡、变形等各种影响导致目标特征发生变化,所以跟踪技术很重要的一点是要实时更新目标模板,从而达到精准跟踪定位的目的。

运动目标跟踪算法中常用的数学工具有卡尔曼滤波(Kal-

man Filtering)及其改进算法、动态贝叶斯网络(Dynamic Bayesian Network)算法以及粒子滤波(Particle Filtering)算法。

### 3.1 基于 Mean Shift 的跟踪算法

Meanshift 是 1975 年提出的,是先计算出当前点的偏移均值,再移动该点到其偏移均值,然后以此为一个新起点,持续移动迭代,直到最后达到一定的条件要求结束。

Meanshift 算法是一种非参估计方法,它根据运动目标的颜色概率分布直方图,在每一帧中,对目标邻近区域中做最大化寻找,完成目标的跟踪。MeanShift 算法对跟踪目标初始化时,需要人工选定目标区域,首先在第一帧目标区域,按像素点计算特征空间的每个特征值的概率,这样将候选目标模型描述出来;然后在后续每帧图像描述出候选模型,即在可能存在目标的候选区域中,对特征空间的每个特征值的概率进行计算,使用相似性函数来判断目标模型与当前候选模型的相似程度,相似性函数一般用 Bhattacharyya 系数,通过反复的迭代,最终在当前候选帧中得到目标最优位置。该算法的优点是对目标边缘遮挡、旋转、变形不敏感,计算量不大。

### 3.2 Kalman 滤波器

Rudolph E. Kalman 在 1960 年发表了一种最优化的递归方法——卡尔曼滤波,可以用于解决离散数据的线性滤波问题,是一种递归的估计方法。利用该方法,有了系统前一时刻状态就能估计出系统下一时刻的状态,同时还可利用当前的观测值修正系统状态,估计的均方误差最小。因为卡尔曼滤波器计算时具有无偏、稳定和最优的特点,因此在自动控制系统中用的比较多。

### 3.3 粒子滤波

粒子滤波是一种通过非参数化的蒙特卡罗方法(Monte Carlo methods)来实现递推的贝叶斯滤波。粒子滤波的基本原理是通过先验概率和当前观测值估计后验概率。

## 4 基于压缩感知的目标跟踪方法

目标跟踪领域有很多算法鲁棒性都很好,例如文献[2][3]在处理目标跟踪时引入了 Super Pixel 概念,处理目标的遮挡、纹理变化方面效果很好;而文献[4]中对非刚性目标跟踪有很好的鲁棒性;而文献[5]将 Distribution Fields (DFS) 引入到目标跟踪中,针对的是目标遮挡和大幅度运动的情况;等等。但是这些鲁棒性很好的算法的计算量都很大,因此实时性不高。

在 2009 年, Xue Mei 等人提出 (Robust Visual Tracking using L1 Minimization) 跟踪算法<sup>[6]</sup>,首次将压缩感知理论引入到目标跟踪领域中。

### 4.1 压缩感知理论简介

传统的采样定理(香农定理)指出:采样频率不能低于模拟信号频谱中最高频率的 2 倍。以香农定理方式采集的数据能够无失真的恢复原始信号,但是采样频率很高,导致采样后的信号仍然存在着大量的冗余,这样造成传输和存储的成本的增加,同时也降低了采样效率。

压缩感知 (Compressive Sensing, CS) 是由 D. Donoho、E. Candes 及华裔科学家 T. Tao (菲尔兹奖获得者) 等人提出<sup>[7-9]</sup>,该理论可以描述为:可压缩的或者在某个变换域是稀疏的信号,可以通过少量的采样,把高维信号通过一个与变换域的基不相关的测量矩阵投影到低维的空间中,然后利用一种恢复算法把



低维的信号还原或重构出原始的高维信号,如果重构出的信号包含了原始高维信号足够的信息,则证明了投影的低维信号或者说较少的采样就能够高精度表示原模拟信号。

压缩感知理论打破了经典的Nyquist采样定理,依据这个理论,信号的采样取决于信号本身所具有的稀疏性或可压性,极大降低了信号的采样频率,减少了数据的传输和存储代价,同时也提高了信号处理的运算速度,为信号处理领域带来革命性的影响,被评为美国2007年十大科技进步奖之一。

由上述内容可知,通过压缩感知采样后的信息去除了原始信息的数量庞大的冗余信息,同时还包含了重建原始信号的所有信息,采样信息和原始信息具有一对一的特性。而在目标跟踪时需要处理的大量的图像信息正是一种典型的具有稀疏性的信息,因此,目标跟踪领域现在也已经使用压缩感知理论<sup>[10-11]</sup>。

2009年之后的前几年,学者们提出的基于压缩感知的算法在本质上都没有避开粒子滤波,都是对粒子进行稀疏表示的目标跟踪,信号重建部分计算量很大。在2012年12月ICCV会议上,Kaihua Zhang, Lei Zhang提出了实时性压缩感知目标跟踪算法,算法中直接将目标的样本特征采样,然后把将采样后的特征作为目标的特征。

#### 4.2 基于稀疏表示的目标跟踪算法

稀疏表示的目标跟踪算法,其核心思想是将目标跟踪问题转成解决目标的稀疏逼近问题。跟踪目标时,用目标模板和正负小模板<sup>[12]</sup>组成字典,首先对每一个候选目标(粒子滤波产生的粒子),用该字典进行稀疏求解,求解方法是利用L1最小化来完成;然后选择具有最大概率的候选目标(粒子)作为当前帧的跟踪结果。下一帧继续使用粒子滤波器产生候选目标(粒子),重复上面的过程直至完成。

设已知  $B = [b_1, \dots, b_n] \in R^{m \times n} (m \gg n)$ ,  $B$  为目标模板的集合,每个模板为  $b_i$ ,利用线性约束逼近表示跟踪结果  $y \in R^m$  为:

$$y \approx BS = s_1 t_1 + s_2 t_2 + \dots + s_n t_n \quad (5)$$

$S$  为目标稀疏表示向量系数,  $S = (s_1, \dots, s_n)^T \in R^n$ 。

因为光照,噪声,遮挡的存在或者目标自身形变等因素的影响,会导致目标跟踪的鲁棒性不强,产生不可预期的误差。为了消除这些误差,稀疏表示的目标跟踪算法使用了向量误差  $d$ ,因此公式(5)转成如下形式:

$$y \approx BS + d \quad (6)$$

误差向量  $d$  包含的一部分非零元素,是用于表示跟踪结果  $y$  中遮挡或其他干扰情况。为对抗遮挡,使用对角矩阵  $\Phi = (\phi_1, \dots, \phi_n)^T \in R^{m \times m}$ ,  $\Phi$  也称为正琐碎模板,其元素  $\phi_i \in R^m$  只有一个非零向量,公式如下:

$$Y = [B, \Phi] \begin{bmatrix} S \\ d \end{bmatrix} \quad (7)$$

$d = (d_1, \dots, d_m)^T \in R^m$  称为琐碎系数向量。为方便计算,再定义一条对角线元素均为-1的单位矩阵作为负琐碎模板,是与灰度图具有相同大小的只有一个非零元素的矩阵。公式(7)可以写成如下形式:

$$Y = [B, \Phi, -\Phi] \begin{bmatrix} S \\ d^+ \\ d^- \end{bmatrix} = QW, \text{ s.t. } W \geq 0 \quad (8)$$

$d^+, d^- \in R^m$  分别被称为正琐碎系数向量和负琐碎系数向量。 $Q = [B, \Phi, -\Phi] \in R^{m \times (n+2n)}$ ,  $W = [S, d^+, d^-] \in R^{n+2n}$ , 是一个非负系数向量。在跟踪过程目标变化不大时,  $d^+$  和  $d^-$  的值也比较小;当在跟踪过程中目标出现遮挡、形变的影响时,  $d^+$  和  $d^-$  的值将会变大,这样通过  $d^+$  和  $d^-$  的值的变化的变化,就能感知目标出现的各种变化。

根据已知的目标模板  $B$  和琐碎模板  $\Phi$ ,只需将非负系数向量  $W$  解出就得到跟踪目标  $y$ ,但实际上,  $W$  没有确定解。我们知道,在一个好的候选目标和  $d$  为有限的非零系数条件下:线性表示的系数满足稀疏性约束,这时可以通过L1范数最小化求解稀疏系数,公式如下:

$$\min \|QW - y\|_2^2 + \lambda \|W\|_1 \quad (9)$$

$\|\cdot\|_1$  和  $\|\cdot\|_2$  分别代表  $L_1$  和  $L_2$  最小范数,求解  $L_1$  范数最小化问题后,  $\|y - BS\|_2$  的最小结果就是目标跟踪的结果。

稀疏表示的目标跟踪算法对于遮挡问题成功解决,但是粒子滤波产生的候选目标通常有很多个,需要对每一个候选目标求解一次  $L_1$  范数最小化,这样会大大增加算法的时间复杂度。

#### 4.3 实时压缩感知目标跟踪算法 (Compressive Tracking)

2012年 Kaihua Zhang 第一次提出一种提取特征进行跟踪的算法——实时压缩感知目标跟踪算法 (Compressive Tracking)。算法首先使用一个非常稀疏的测量矩阵,该矩阵满足约束等距性 restricted isometry property (RIP),将高维的图像特征空间降维到一个低维的压缩子空间,然后分别将目标和背景的特征提取出,作为正样本和负样本,这两个样本用于在线更新分类器,分类器使用朴素贝叶斯分类器,最后利用分类器分类下一帧图像的目标。

实时压缩感知目标跟踪算法提取样本特征公式如下:

$$B = \Phi S \quad (10)$$

$S \in R^{n \times 1}$  为原始图像信号,在算法中  $S$  被转化为一维信号,  $\Phi \in R^{m \times n} (m \ll n)$  是测量矩阵,  $B$  是压缩采样后的数据。由压缩感知理论可知,  $B$  可以看成是跟踪目标的特征。

在第  $k$  帧时,在第  $k-1$  帧中目标上采样,得到若干正样本,再在目标周围采样,得到若干负样本,然后提取特征后,再通过对样本特征用测量矩阵采样,此时被采样后的特征,其维数要比原来的维数低很多,之后用被采样后的特征去训练分类器。

在  $k+1$  帧时,在  $k$  帧目标周围采样,得到若干个图像碎片,将其特征采样出后分类,分类分数最大框就认为是目标。

实时压缩感知目标跟踪算法思想采用了一种非常稀疏的随机测量矩阵  $\Phi \in R^{n \times m}$ ,其矩阵元素定义如下:

$$\phi_{ij} = \sqrt{z} \times \begin{cases} 1 & \text{概率为 } \frac{1}{2z} \\ 0 & \text{概率为 } 1 - \frac{1}{z} \\ -1 & \text{概率为 } \frac{1}{2z} \end{cases} \quad (11)$$

Achlioptas<sup>[13]</sup>证明:为满足约束等距性条件,公式(11)中的  $z$

等于2或者3,这时矩阵可以作为测量矩阵,即通过该矩阵采样的信号可以代表原始信号。

为产生该矩阵,需要一个均匀随机发生器。当 $z=3$ ,矩阵非常稀疏,计算量是原来的 $2/3$ ,并且有 $2/3$ 的概率为 $0^{[14]}$ 。并且 $z=O(m)(x \in R^m)$ 时,矩阵近似正态分布。在Compressive Tracking算法中,令 $z=m/3$ ,这样测量矩阵 $\Phi$ 的每一行只需要计算 $Y(Y \leq 4)$ 个非零稀疏元素的值,大大降低了计算量,并且只要存储 $Y$ 个非零元素,存储空间也很小。

实时压缩感知目标跟踪算法简单,将压缩感知和目标特征提取联系在一起,计算量小,具有很好的实时性,但是该算法采用的目标特征比较单一,如果在目标纹理或光照变化时,跟踪效果不稳定。

## 5 总结

运动目标检测与跟踪是计算机视觉技术领域中一个重要组成部分,目前应用在多个方面,与人们日常生活息息相关。在研究目标检测与跟踪算法时,目标被遮挡、形变等是需要解决的难题。经过多年的研究,以压缩感知为基础的算法慢慢发展起来。目前,基于稀疏表示的跟踪算法正在不断发展,实验效果良好,具有很好的发展前景。

## 参考文献:

- [1] University, Technical Report: CMU-RI-TR-00-12, 2000.
- [2] Wang S, Lu H, Yang F, et al. Superpixel tracking[C] 2011 IEEE International Conference on IEEE, Computer Vision (ICCV), 2011: 1323-1330.
- [3] Oron S, Bar-Hillel A, Levi D, et al. Locally orderless tracking [C] 2012 IEEE Conference on. IEEE, Computer Vision and Pattern Recognition (CVPR), 2012: 1940-1947.
- [4] Kwon J, Lee K M. Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive basin hopping monte carlo sampling[C] IEEE Conference on. IEEE, Computer Vision and Pattern Recognition, 2009. CVPR 2009. 2009: 1208-1215.
- [5] Sevilla-Lara L, Learned-Miller E. Distribution fields for tracking[C] 2012 IEEE Conference on. IEEE, Computer Vision and Pattern Recognition (CVPR), 2012: 1910-1917.
- [6] X. Mei and H. Ling, "Robust visual tracking using l1 minimization," in Proc. IEEE Int. Conf. Comp. Vis., Kyoto, Japan, 2009, 1436-1443.
- [7] Donoho D L, Elad M, Temlyakov V N. Stable recovery of sparse overcomplete representations in the presence of noise [J]. IEEE Transactions on, Information Theory, 2006, 52(1): 6-18.
- [8] Donoho D L, Tsaig Y, Drori I, et al. Sparse solution of under-determined systems of linear equations by stagewise orthogonal matching pursuit[J]. IEEE Transactions on, Information Theory, 2012, 58(2): 1094-1121.
- [9] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on, Signal Processing, 1993, 41(12): 3397-3415.
- [10] Li H, Shen C, Shi Q. Real-time visual tracking using compressive sensing[C] 2011 IEEE Conference on. IEEE, Computer Vision and Pattern Recognition (CVPR), 2011: 1305-1312.
- [11] Zhang K, Zhang L, Yang M H. Real-time compressive tracking[M] Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012: 864-877.
- [12] 张继等. 面向视频的快速稀疏编码算法. 计算机与科学. 2012, 6: 1673-9418.
- [13] Achlioptas D. Database-friendly random projections[C] Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. ACM, 2001: 274-281.
- [14] Zhang K, Zhang L, Yang M H. Real-time compressive tracking[M] Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012: 864-877.

【通联编辑:唐一东】