文章编号:1005-0523(2012)01-0043-05

基于压缩感知和字典学习的背景差分法

郭厚焜,吴 峰,黄 萍

(华东交通大学信息工程学院,江西 南昌,330013)

摘要:针对当使用背景差分法时,背景存在突变和渐变、图像数据的冗余和伪前景对目标检测的干扰等问题,提出一种基于稀疏表示和字典学习的背景差分法。该方法首先训练视频流得到其数据字典,并根据数据字典学习与稀疏表示理论建立背景模型,可以有效减少数据的冗余。然后根据目标及其邻域的密集度进行目标分割,以排除前景的干扰。最后再根据数据字典的更新算法,有效解决了背景的突变和渐变问题。实验结果表明,该方法具有可行性。

关键词:稀疏表示;字典学习;背景差分;前景分割

中图分类号:TP399 **文献标志码**:A

智能视频监控技术已被广泛应用到国民经济的各领域,在军事、工业、智能人机交互、智能交通等方面具有重要的意义。运动目标的检测是智能视频监控系统中一个重要组成部分,如何实现对运动目标的检测,是一个需要关注和研究的重要问题。

目前运动目标检测通常采用的方法有光流场法[1-2]、帧间差分法[3]和背景差分法[4]。光流场法运算复杂,不便实时实现。帧间差分法数据计算量大,并且帧与帧之间运动目标存在相对运动,会造成被检测处的运动目标位置与大小不准确。背景差分法运算简单,容易实现,但是容易受树叶的摆动、水面的波动和外界光线的突变等自然环境的影响,这就要求不断自适应更新背景模型。为此,各国学者已经提出了各种各样的新算法来力图解决现有算法出现的问题。如文献[5-6]提出的背景建模算法,以每一帧图像为一整体,发掘变化区域的内部结构特征,缺点在于不能较好地排除伪前景的干扰。文献[7-8]提出的数据字典学习算法,当输入的背景图像中还有运动目标时,不能很好地建立背景模型。

针对这一问题,在背景差分法的基础上,使用压缩感知原理,用一组数据字典线性稀疏表示背景区域[9-10],建立背景模型。根据分割算法,排除伪前景的干扰,得到正确的运动目标区域。最后,采用数据字典更新算法,建立实时更新的背景模型。

1 传统背景差分法

背景差分法由于其算法实现简单被广泛应用,一般适合在静态场景中(摄像头固定)。背景差分法的基本算法是利用当前帧灰度值与背景帧灰度值的差分值的差异来实现运动目标检测。具体步骤如下:

1)读取一段视频序列,根据统计平均法获得背景模型 $B(x,y)$。

2)利用当前帧 $I(x,y)$ 与背景模型 $B(x,y)$ 做差分运算得到 F,然后利用Ostu局部二值化算法对得到的绝对差分图像进行二值化,公式如下

$$F=\left|I(x,y)-B(x,y)\right|>T \tag{1}$$

式中:F 为前景图像的表现形式;T 为阀值。

3)采用数学形态学开、闭运算,排除背景区域抖动与噪声的影响,平滑运动目标区域的轮廓,去掉图像

收稿日期:2011-11-01

基金项目:江西省研究生创新专项基金项目(YC2011-X013)

作者简介:郭厚焜(1965—),男,教授,研究方向为图像处理与模式识别和三维图像检测。

内部小孔、微小噪声及小面积非运动目标部分。

4）背景更新，返回第2步。

2 基于压缩感知的背景差分法

该文提出了一种改进的背景差分法，用一组通过学习得到的基向量对背景模型进行特征表示，利用前景图像、背景图像的稀疏性和数据字典的学习，可以准确检测出运动目标，并且自动排除前景的干扰。该方法主要内容有背景模型的建立、前景检测、运动目标的确定和背景模型的更新。算法过程如下：首先输入一段时间的视频流作为训练样本，通过数据字典学习得到背景的特征表示矩阵 $\boldsymbol{D}$，然后根据前景与背景的稀疏性得到背景模型与前景区域，最后采用分割算法准确检测出前景中的运动目标。算法流程图见图1。

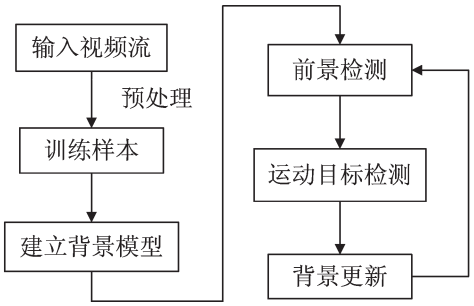

图1 本文算法流程图
Fig.1 Algorithm flow chart

2.1 基本原理

对于一幅图像 $\boldsymbol{X}$，可以看成由背景区域 $\boldsymbol{X}_{\mathrm{b}}$ 和前景区域 $\boldsymbol{X}_{\mathrm{f}}$ 组成

$$\boldsymbol{X}=\boldsymbol{X}_{\mathrm{b}}+\boldsymbol{X}_{\mathrm{f}} \tag{2}$$

式中：$\boldsymbol{X}_{\mathrm{b}}$，$\boldsymbol{X}_{\mathrm{f}}$ 和 $\boldsymbol{X}$ 都是列向量。

读入一个有 C 张图像的视频序列，假设每一张图像的背景区域 $\boldsymbol{X}_{\mathrm{b}}$ 可由一组基向量 $\boldsymbol{D}_i$ 特征化表示，就说背景区域 $\boldsymbol{X}_{\mathrm{b}}$ 中第 i 张图像的线性表达式为 $\boldsymbol{X}_{\mathrm{b}}=\boldsymbol{D}_i\boldsymbol{\alpha}_i$，其中 $\boldsymbol{\alpha}_i$ 是系数矩阵。定义一个新矩阵 $\boldsymbol{D}$ 为图像序列的基向量集合 $\boldsymbol{D}=(\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_c)$。因此，可以把 $\boldsymbol{X}_{\mathrm{b}}$ 写成

$$\boldsymbol{X}_{\mathrm{b}}=\boldsymbol{D}\boldsymbol{\alpha} \tag{3}$$

这里 $\boldsymbol{\alpha}=(0,0,\cdots,a_i^{\mathrm{T}},0,\cdots,0)^{\mathrm{T}}$ 是一个稀疏系数矩阵，除了在 $\boldsymbol{D}_i$ 相关处的值非零外，其它处均为零，$\boldsymbol{X}_{\mathrm{b}}$ 是背景区域矩阵。式(3)表明指定帧图像的背景区域可以通过一个数据字典 $\boldsymbol{D}$ 进行稀疏表示。由实际观察可知，在每幅图像中，运动目标只占少部分像素点，所以经过背景差分法后的候选前景区域 $\boldsymbol{X}_{\mathrm{f}}$ 也是稀疏的。

由上可知，$\boldsymbol{X}_{\mathrm{b}}$ 和 $\boldsymbol{X}_{\mathrm{f}}$ 都是稀疏的，因此，可以把背景差分问题分解成一个数学问题：已知某一帧图像 $\boldsymbol{X}$，求其背景稀疏编码 $\boldsymbol{X}_{\mathrm{b}}=\boldsymbol{D}\boldsymbol{\alpha}$ 和前景稀疏矩阵 $\boldsymbol{X}_{\mathrm{f}}=\boldsymbol{X}-\boldsymbol{D}\boldsymbol{\alpha}$：

$$\hat{\boldsymbol{\alpha}}=\arg\min_{\boldsymbol{\alpha}}\left\|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{\alpha}\right\|_0+\lambda\left\|\boldsymbol{\alpha}\right\|_0 \tag{4}$$

式中：$\left\|\boldsymbol{\alpha}\right\|_0$ 是 l_0 范数，用来统计非零元素个数；λ 是平衡参数；$\hat{\boldsymbol{\alpha}}$ 为 $\boldsymbol{\alpha}$ 的最优解。求解式(4)的最优解和传统的压缩传感问题有所不同，因为它包括两个稀疏矩阵：背景稀疏系数矩阵 $\boldsymbol{\alpha}$ 和前景稀疏矩阵 $\boldsymbol{X}-\boldsymbol{D}\boldsymbol{\alpha}$。把式(4)写成另一种形式，用 l_1 范数代替 l_0 范数，这样就转化为 l_1 正则化的凸优解问题

$$\hat{\boldsymbol{\alpha}}=\arg\min_{\boldsymbol{\alpha}}\left\|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{\alpha}\right\|_1+\lambda\left\|\boldsymbol{\alpha}\right\|_1 \tag{5}$$

这里 $\left\|\boldsymbol{\alpha}\right\|_1=\sum_i\left|\boldsymbol{\alpha}(i)\right|$。把式(5)写成等价于 l_1 逼近问题：

$$\hat{\boldsymbol{\alpha}}=\arg\min_{\boldsymbol{\alpha}}\left\|\begin{pmatrix}\boldsymbol{X}\\0\end{pmatrix}-\begin{pmatrix}\boldsymbol{D}\\\lambda\boldsymbol{E}\end{pmatrix}\boldsymbol{\alpha}\right\|_1 \tag{6}$$

式中：$\boldsymbol{E}$ 为单位矩阵。选用 l_1 范数代替 l_0 范数形式的好处[14]：在背景差分法中的数据元素个数要远远小于背景图像的数据元素个数，因此，在实际程序运行过程中，计算量也会大大降低。

线性方程组

$$\begin{pmatrix} \boldsymbol{X} \\ 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{D} \\ \lambda \boldsymbol{E} \end{pmatrix} \boldsymbol{\alpha} \tag{7}$$

是由多个参数组成，如果直接求解 $\boldsymbol{\alpha}$，则无法解答此问题。但式(7)满足文献[11]提出的情况，所以有最优解。因此，能从某一帧图像 $\boldsymbol{X}$ 中分割出背景图像 $\boldsymbol{X}_{\mathrm{b}}$ 和前景图像 $\boldsymbol{X}_{\mathrm{f}}$。

2.2　图像分割

前景图像的像素值是当前帧与背景图像的差值，理想情况下前景区域的像素值非零，其余值均为零。但是在现实中，非零值的产生不止因为运动目标的出现，还有背景的变化、背景模型的误差和噪声等的影响。这时就要对分割出来的 $\boldsymbol{X}_{\mathrm{f}}$ 进行后处理，从候选目标中提取出运动目标。算法的关键在于，由于运动目标区域不仅稀疏而且还具有一定大小的连通区域，这就意味着，运动目标区域的像素是空间相关的。背景的变化和背景模型的误差相对而言是分散的和无结构联系的。受文献[7-8]的启发，该文提出了一种用于目标分割的算法，它判断前景区域 $\boldsymbol{X}_{\mathrm{f}}$ 属于运动目标的可能性，不但考虑了运动目标区域的密集度，还考虑了其邻域区域的密集度

$$S(i) = \boldsymbol{X}_{\mathrm{f}}^2(i) + \sum_{j \in N(i)} \boldsymbol{X}_{\mathrm{f}}^2(j) \tag{8}$$

式中：$S(i)$ 为运动目标区域的可能值；$N(i)$ 为某像素点的邻域。

2.3　通过数据字典学习建立背景模型

为了使数据字典学习算法既能分割运动目标，同时又能建立正确的背景模型，找到一个 $\boldsymbol{D}$ 的最优解并满足下列条件：它能以最小的误差代表所有训练的样本，同时产生一个最稀疏的特征矩阵。

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D},\{\boldsymbol{\alpha}_m\}} \sum_{m=1}^{M} \left\| \boldsymbol{X}_m - \boldsymbol{D}\boldsymbol{\alpha}_m \right\|_1 + \lambda \left\| \boldsymbol{\alpha}_m \right\|_1 \tag{9}$$

在实际应用中，被用来背景建模的图像可能存在以下情况：受到硬件的高斯噪声干扰，伪前景被误判为前景，自然环境中光强的变化和运动目标的渐变与突变。这些都直接或间接的影响背景模型的准确建立。通过式(9)，可以有效解决上述现象。

把式(9)写成矩阵的形式，得到等价问题：

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D},\boldsymbol{A}} \left\| \boldsymbol{x} - \boldsymbol{D}\boldsymbol{A} \right\|_1 + \lambda \left\| \boldsymbol{A} \right\|_1 \tag{10}$$

式中：$\boldsymbol{x}$ 是训练样本的列向量集合；$\boldsymbol{A}$ 是系数矩阵；$\left\| \boldsymbol{A} \right\|_1 = \sum_{i,j} \left| \boldsymbol{A}(i,j) \right|$ 是矩阵 $\boldsymbol{A}$ 的 l_1 范数，在这里求对其输入值的绝对值求和。由于式(10)含有两个未知数，不能直接求解，但依据文献[9-12]中提出的方法，反复交替优化 $\boldsymbol{D}$ 和 $\boldsymbol{A}$，可以求解 $\boldsymbol{D}$ 和 $\boldsymbol{A}$。

2.4　背景的更新

当建立数据字典后，$\boldsymbol{A}$ 可以可视为常量，忽略式(10)的第二项进行对 $\boldsymbol{D}$ 的自动更新

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D}} \left\| \boldsymbol{x} - \boldsymbol{D}\boldsymbol{A} \right\|_1 \tag{11}$$

式(11)中 $\boldsymbol{D}$ 中元素是线性无关的，因此可把 $\boldsymbol{D}$ 写成独立形式

$$\boldsymbol{d}_k = \arg\min_{\boldsymbol{d}_k} \left\| \boldsymbol{x} - \boldsymbol{d}_k \boldsymbol{\alpha}^k \right\|_1 \quad k = 1, 2, \cdots, K \tag{12}$$

$$\boldsymbol{d}_k^i = \arg\min_{\boldsymbol{d}_k^i} \left\| \boldsymbol{X}^i - \boldsymbol{d}_k^i \boldsymbol{\alpha}^k \right\|_1 \quad i = 1, 2, \cdots, N;\ k = 1, 2, \cdots, K \tag{13}$$

$$\boldsymbol{\alpha}_j^k = \arg\min_{\boldsymbol{\alpha}_j^k} \left\| \boldsymbol{X}_j - \boldsymbol{d}_k \boldsymbol{\alpha}_j^k \right\|_1 \quad j = 1, 2, \cdots, M;\ k = 1, 2, \cdots, K \tag{14}$$

式中：$\boldsymbol{d}_k$ 是 $\boldsymbol{D}$ 的第 k 个元素；$\boldsymbol{\alpha}^k$ 是 $\boldsymbol{A}$ 的第 k 行元素(数据矩阵和稀疏矩阵式对应的元素)。

在式(12)中，提取 $\boldsymbol{x}$ 的列对应的系数值 $\boldsymbol{\alpha}_k$ 在一个小门限值之上。这是因为 $\boldsymbol{\alpha}_k$ 的元素值只是接近零而不是真正等于零，门限操作只保留那些相关的元素而不必处理所有像素，这样大大加速了 d_k 的运算速度。

把式(12)进一步分解成一系列的 l_1 范式问题(13),可以通过迭代重加权最小二乘思想得到一个closed-form的解决方案[13]。通过实验发现,如果同时使用式(14),这样迭代的次数会更少,同时还能更新 α^k 。总之,反复使用式(13)与(14)直到 d_k 收敛。

3 实验结果

该文实验所采用的计算机硬件为CPU C2Q 2.5 GHz,内存8 G,在Matlab 2010b环境下,对算法进行了具体实现,验证了其可行性。

如图2(a)~(c)所示,读入一个视频序列前30帧图像作为训练样本。传统的背景差分法的背景是通过一段时间内的图像均值来获得的,由于图像中的前景像素值和背景像素值相差较大,得到的背景在这种情况下将有误差,会产生图2(d)所示的错误背景。图2(e)是采用该文算法得到的背景模型,排除了伪前景和运动目标对背景的干扰。如果用传统的背景差分法进行处理,由于该方法是基于颜色的亮度特性,即RGB 3个分量,它对亮度的变化很敏感。这里,运动目标图像与背景图像时间相差2小时,汽车表面金属反射光强变化明显,反映在图像上为汽车位置的像素灰度值发生突变,经过传统的背景差分法,静止的汽车会被检测出来,误认为是运动目标,如图2(f)所示。图2(g)所示本文提出的背景差分法,有效解决了亮度变化问题,并且排除了伪前景的干扰,提高了目标检测的准确性。

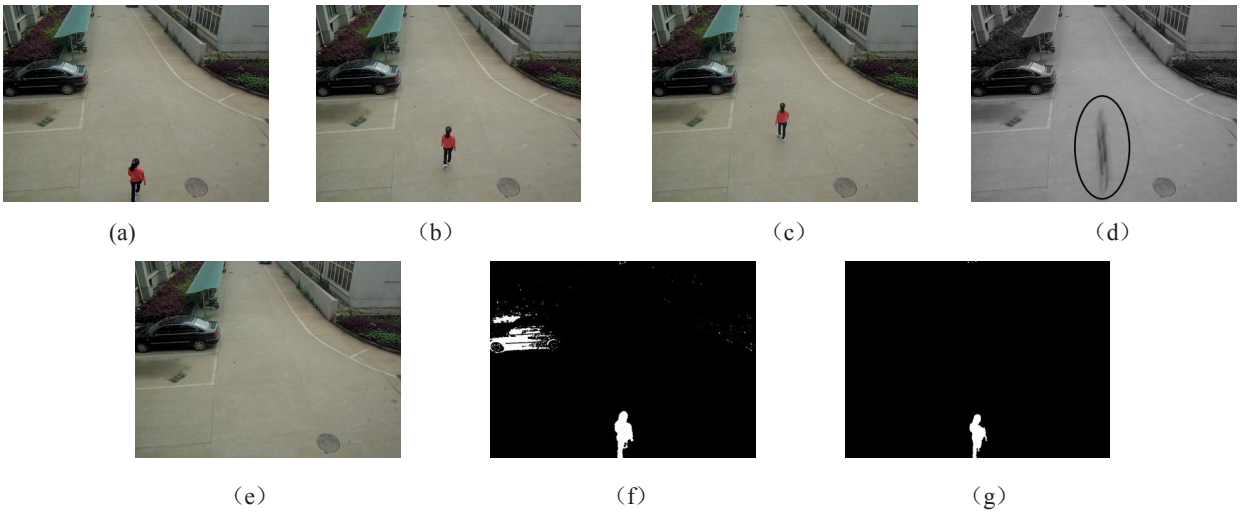

图2 数据字典的学习即更新
Fig.2 Update of data dictionary

4 结语

提出了一种基于稀疏表示和字典学习的背景差分法,该方法利用字典学习来建立背景模型,通过优化 l_1 正则化问题来分割前景区域。实验结果表明,与现有背景差分法相比,该算法能较好处理背景的突变和高频变化,有效消除了数据在学习阶段的离群值,减少噪声干扰,建立正确的背景模型。

参考文献:

[1] HORN B K P, SCHUNCK B G. Determining optical flow[J]. Artificial Intelligence, 1981, 17(1): 185-203.
[2] ALI S, SHAH M. Human action recognition in videos using kinematic features and multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(2): 288-303.
[3] HUI K C, SIU W C. Extended analysis of motion-compensated frame difference for block-based motion prediction error[J]. IEEE Transactions on Imaging Processing, 2007, 16(5): 1232-1245.

[4] TSAI D M，LAI S C. Independent component analysis-based background subtraction for indoor surveillance[J]. IEEE Transactions on Imaging Processing，2009，18(1)：158-167.
[5] OLIVER N M，ROSARIO B，PENTLAND A P. A bayesian computer vision system for modeling human interactions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2000，22(8)：831-843.
[6] MONNET A，MITTAL A，PARAGIOS N. Scene modeling and change detection in dynamic scenes：a subspace approach[J]. Computer Vision and Image Understanding，2009，113(1)：63-79.
[7] AHARON M，ELAD M，BRUCKSTEIN A. K-SVD：an algorithm for designing over complete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing，2006，54(11)：4311-4322.
[8] MAIRAL J，BACH F，PONCE J，et al. Online learning for matrix factorization and sparse coding[J]. Journal of Machine Learning Research，2010，11(3)：19-60.
[9] JI S H，YA X，CARIN L. Bayesian compressive sensing[J]. IEEE Transactions on Signal Processing，2008，56(6)：2346-2356.
[10] DO T T，GAN L，NGUYEN N，et al. Fast and efficient compressive sensing using structurally random matrices[J]. IEEE Transactions on Signal Processing，2012，60(1)：139-154.
[11] CANDES E J，TAO T. Decoding by linear programming[J]. IEEE Transactions on Information Theory，2005，51(12)：4203-4215.
[12] RUBINSTEIN R，BRUCKSTEIN A M，ELAD M. Dictionaries for sparse representation modeling[J]. Proceedings of the IEEE，2010，98(6)：1045-1057.
[13] CABRIEL K R，ZAMIR S. Lower rank approximation of matrices by least squares with any choices of weights[J]. Technometrics，1979，21(4)：489-498.
[14] WRIGHT J，YANG A Y，GANESH A，et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2009，31(2)：210-227.

Background Subtraction Based on Sparse Representation and Dictionary Learning

Guo Houkun，Wu Feng，Huang Ping

（School of Information Engineering，East China Jiaotong University，Nanchang 330013，China）

Abstract：In this paper，we propose a CS-based background subtraction approach based on the theory of sparse representation and dictionary learning，to handle sudden and gradual background changes and the redundancy of excessive image data and the interference of prospect. This method gets their data dictionary according to the video stream and establishes the background model based on the theory of dictionary learning and sparse representation to effectively reduce data redundancy. Then，the moving objects correctly depending on the intensity of the target and its neighbors are segmented so as to rule out interference of the foreground. Finally，the problem of sudden and gradual background changes is solved through the update algorithm of data dictionary. Experiments show that this method is feasible.

Key words：sparse representation；dictionary learning；background subtraction；foreground segmentation