

文章编号 1004-924X(2017)02-0529-08

张量低秩表示和时空稀疏分解的视频前景检测

隋中山^{1*}, 李俊山¹, 张 姣¹, 樊少云², 孙胜永³

(1. 火箭军工程大学 信息工程系, 陕西 西安 710025;

2. 96518 部队, 湖南 怀化 418000;

3. 南京炮兵学院 战役战术教研室, 江苏 南京 211132)

摘要:针对视频中前景检测的问题,提出了一种基于张量低秩表示(Tensor Low-Rank Representation, TLRR)和时空稀疏分解的检测方法。由于视频序列中的前景除具有稀疏性外,本身还具有空间上的连续性以及时间上的持续性,本文提出采用时空稀疏范数对前景特性进行深入发掘。利用张量低秩表示方法将原始视频用张量形式进行分解,充分利用了原始数据的行信息和列信息,且将原始的背景、前景二分解泛化为背景、前景和噪声的三分解,使用非精确增广拉格朗日乘子(Inexact Augmented Lagrange Multiplier, IALM)方法进行最优化求解,并对算法进行了分析。设计实验对本文新方法的有效性进行了验证和比较,并对影响算法效果的重要参数 ρ 进行了进一步研究实验。实验结果表明:该方法能够有效检测出视频中的运动前景,其准确性相对已有方法有一定提高。

关键词:视频;前景检测;低秩;时空稀疏分解;张量低秩表示;非精确增广拉格朗日乘子

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.3788/OPE.20172402.0529

Video foreground detection of tensor low-rank representation and spatial-temporal sparsity decomposition

SUI Zhong-shan^{1*}, LI Jun-shan¹, ZHANG Jiao¹, FAN Shao-yun², SUN Sheng-yong³

(1. Department of Information Engineering, Rocket Force University of Engineering, Xi'an 710025, China;

2. 96518 Army Force; Huaihua 418000, China;

3. Operational Tactics Teaching and Research Section, Nanjing Artillery Academy, Nanjing 211132, China)

* Corresponding author, E-mail: zclszs@163.com

Abstract: A detection method based on Tensor Low-Rank Representation (TLRR) and spatial-temporal sparsity decomposition was proposed to detect foreground targets in video sequences. Since foreground in video sequence has sparsity inherently besides spatially continuous and temporally continuous, this paper put forward spatial-temporal sparsity-inducing norm to perform deep research on property of foreground. Original video was decomposed in tensor representation formed by tensor low-rank representation method, line information and column information of original data were fully used, and

收稿日期:2016-07-21;修订日期:2016-10-12.

基金项目:国家自然科学基金资助项目(No. 61175120)

two-stage decomposition of original background and foreground was generalized to three-stage decomposition of background, foreground and noises. Optimization solution was performed with Inexact Augmented Lagrange Multiplier (IALM) method. Verification and comparison experiment was established, and further research experiment was performed to research how ρ affecting performance of algorithm. Experimental results show that the method can detect moving foreground in video effectively and improve accuracy when compared with existing methods.

Key words: video; foreground detection; low rank; spatial-temporal sparsity decomposition; tensor low-rank representation; inexact augmented Lagrange multiplier

1 引 言

从视频流中分割出前景对象是高级计算机视觉任务中最基本的一步,其在交通控制、视频编码、人机交互、监控安防等领域有着广泛的应用。前景分割精度是其中的重要指标。背景减除法是一类较为常见的前景检测方法。该方法首先建立背景模型,然后通过目标图像中减除背景模型获得前景,然而在复杂环境中建立背景模型本身就是一项艰巨的任务。STAUFFER 等利用高斯混合模型法针对每个点进行建模,结果证明其可以有效捕捉背景变化^[1]。然而高斯混合模型法存在参数难以调节的问题,针对这一问题,Elgammal 提出了非参数的 KDE(Kernel Density Estimate)方法^[2]。之后,Yaser 等人^[3]对 Elgammal 算法进行扩展,增加了对前景目标的建模能力,同时利用马尔可夫随机场模型对同一视频帧像素之间的邻域关系建模,再通过求解马尔可夫随机场模型实现前景分割。Heikkila 提出了基于纹理的动态背景建模方法^[4],他是用 LBP(Local Binary Patterns)算子提取纹理特征。Kim 提出了^[5]将图像中每个点都用码本表示的码本模型。

近期,低秩稀疏分解方法在前景检测领域越来越受到重视,著名的基于主成分追踪的鲁棒主成分分析^[6](Robust Principal Component Analysis,RPCA)方法就是其中的一种。但是目前大多数方法都是对原始视频的背景部分研究较为深入,对前景部分研究不足,且大多为二分解,只考虑背景和前景,并不单独考虑噪声。因此本文提出一种同时深入发掘背景特性和前景特性并独立考虑噪声的三分解方法来解决视频中的前景检测问题。

2 低秩稀疏分解

近年来,低秩稀疏分解方法在前景检测领域已经取得了可喜的成果。其对背景的唯一假设是:背景的任何变化均可以通过低秩矩阵捕获。其原理为:将观察到的视频序列组成的矩阵分解成一个低秩矩阵代表的背景和一个稀疏矩阵代表的前景。它可以用式(1)表示:

$$\mathbf{D} = \mathbf{L} + \mathbf{S}, \quad (1)$$

其中 $\mathbf{D} \in R^{m \times n}$ 是所观察到的视频 (n 帧), \mathbf{L} 和 \mathbf{S} 分别表示背景和前景。RPCA 提出用 l_1 -范数来约束前景矩阵 \mathbf{S} , 因为前景矩阵是包含一小部分非零元素的稀疏矩阵,并假设背景图像是线性相关的,从而可用低秩矩阵 \mathbf{L} 表示。该矩阵分解可以通过以下凸优化解决:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s. t.} \quad \mathbf{D} = \mathbf{L} + \mathbf{S}, \quad (2)$$

其中 $\|\mathbf{L}\|_*$ 是矩阵 \mathbf{L} 的核范数,是其奇异值的总和, $\|\mathbf{S}\|_1$ 表示 \mathbf{S} 的 l_1 -范数。 λ 为一个正的系数。RPCA 方法前提假设比较简单,且对只有局部区域受到严重腐蚀的图像效果较好,因此受到广泛关注。但是 RPCA 是一种直推式方法,无法直接处理新数据,而且 RPCA 默认所有背景数据基本处于一个子空间,无法处理来自多个子空间的数据。

为了解决上述问题,学者们对 RPCA 进行了有效的改进和泛化,提出了 IRPCA^[7](Inductive Robust Principal Component Analysis)方法:

$$\min_{\mathbf{U}, \mathbf{S}} \|\mathbf{U}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s. t.} \quad \mathbf{D} = \mathbf{X} + \mathbf{S}, \mathbf{X} = \mathbf{U}\mathbf{D}, \quad (3)$$

其中 \mathbf{U} 为一个低秩投影矩阵 $\mathbf{U} = [u_1, u_2, \dots, u_m] \in R^{m \times m}$ 。如果知道 \mathbf{U}^* , 就可通过计算 $\mathbf{U}^* \mathbf{D}$ 得到背景数据。该方法可以处理来自多个子空间的背

景数据,而且对于一帧新数据 y , $U^* y$ 代表背景, $y - U^* y$ 代表前景,但由于 IRPCA 方法是沿着矩阵的列方向进行误差修正的,所以原始数据矩阵行方向包含的信息就得不到体现。

LRR^[8] (Low-Rank Representation) 方法是与 IRPCA 类似的一种方法。LRR 方法的核心思想是找到一个低秩表达矩阵 $V = [v_1, v_2, \dots, v_n] \in R^{n \times \pi}$, 将数据向量用数据字典 A 的基的线性组合进行表达, 如果将原始数据 D 直接作为数据字典(即 $A = D$), LRR 方法则可用如下最优化问题表示:

$$\min_{V, S} \|V\|_* + \lambda \|S\|_{2,1} \quad \text{s. t.} \quad D = AV + S, A = D, \quad (4)$$

当计算出最优解 (V^*, S^*) 后, 背景即为 $D - S^*$ 。由于 LRR 方法沿着行方向恢复或分割数据, 所以原始数据矩阵列方向上包含的信息将丢失。另外 LRR 方法也是一种直推式方法, 不能高效处理新数据。由于该方法直接将原始数据 D 作为数据字典, 所以要求原始数据 D 必须包含足够多的纯净信息, 但这在现实场景中较难满足。

为了能同时使用原始数据 D 的行信息和列信息, 在 LRR 方法的基础上, 提出了 LatLRR^[9] (Latent Low-Rank Representation) 方法。LatLRR 方法构造了一个与 LRR 方法不一样的数据字典 $A = [D_o, D_H]$, 式中 D_o 为观察到的数据(即原始数据 D), D_H 为隐藏数据, A 为 D_o 、 D_H 的矩阵连接。LatLRR 方法可以用如下优化问题表示:

$$\min_{U, V, S} \|U\|_* + \|V\|_* + \lambda \|S\|_{2,1}, \quad (5)$$

$$\text{s. t.} \quad D = UD + DV + S$$

由式(5)可以看出, LatLRR 方法是 IRPCA 和 LRR 方法的融合。它同时使用了原始数据的行信息和列信息。但由于 D_H 是从相同的子空间采样得到的, 因此 LatLRR 方法也要求原始数据中必须包含足够多的纯净信息, 而且引入所谓的隐藏数据 D_H 对整个算法的具体影响还不明了。

目前最新的背景求解方法是 TLRR^[10] (Tensor Low-Rank Representation) 方法。该方法通过双线性矩阵恢复在列方向和行方向上同时学习出 D 的张量表示形式 UDV 中低秩矩阵 U 、 V 。TLRR 方法可以用如下最优化问题表示:

$$\min_{U, V, S} \|U\|_* + \|V\|_* + \lambda \|S\|_{2,1} \quad (6)$$

$$\text{s. t.} \quad D = XV + S, X = UD.$$

TLRR 方法充分利用了原始数据的行信息和列信息。但是包括 TLRR 方法在内的上述方法都存在两个问题: 首先, 上述方法没有考虑前景中任何特定的结构或任何子集之间可能存在的关系, 而在现实场景中, 前景一般具有空间的连续性和局部性等特点。其次, 它们不考虑前景在视频前后帧中的顺序关系, 而前景事实上一般具有时间上的持续性, 即前景总是持续出现或存在的。

3 TLRR 方法的泛化与改进

通过分析低秩稀疏分解方法的发展脉络和存在问题, 可以得到如下启示: TLRR 方法针对低秩数据 L 做了深入发掘, 同时利用了原始数据的行信息和列信息, 本文可以用同样思路从行和列两个角度对前景数据 S 做深入研究; 同时针对其不足, 可将 TLRR 方法进行泛化, 即将二分解推广为三分解。

3.1 模型建立

在单独考虑背景、前景和噪声的情况下, 模型可用如下公式表示:

$$D = L + S + E, \quad (7)$$

其中 $D \in R^{m \times n}$ 为观察到的视频 (n 帧), L 和 S 分别表示背景和前景, E 为噪声。

TLRR 方法可以初步改写为:

$$\min_{U, V, S, E} \|U\|_* + \|V\|_* + \lambda \|S\|_{\epsilon} + \beta \|E\|_{\ell}, \quad (8)$$

$$\text{s. t.} \quad D = XV + S + E, X = UD$$

对于噪声 E , 如果是少量稀疏大幅值噪声, 则可利用最小化噪声数据的 $\|\cdot\|_0$ (即 $\ell = 0$) 来保证噪声的稀疏性; 如果是高斯噪声, 则使用 Frobenius 范数; 对于列奇异点噪声则使用 $l_{2,0}$ 范数 $\|\cdot\|_{2,0}$ 。

对于背景 L , 采用 TLRR 方法默认模型, 即 L 的秩较小, 且背景数据可以来自多个不同的低秩子空间。

对于前景 S , 除考虑其稀疏性以外, 本文拟发掘前景的空间连续特性和时间持续特性, 因为在现实视频场景中, 以上两个特性是普遍存在的。

3.2 时空稀疏范数

受结构稀疏范数^[11]启发, 为了利用前景的空间连续性和时间持续性的先验知识, 本文提出时空稀疏范数 Φ , 其定义如下:

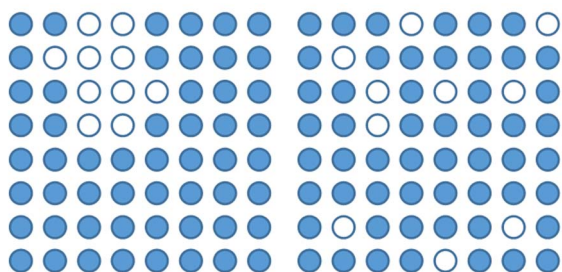
$$\Phi(\mathbf{S}) = (1 - \rho)\Psi(\mathbf{S}) + \rho\Omega(\mathbf{S}), \quad (9)$$

$$\Psi(\mathbf{S}) = \sum_j^n \sum_{g \in \mathcal{N}} \|s_g^j\|_\infty, \quad (10)$$

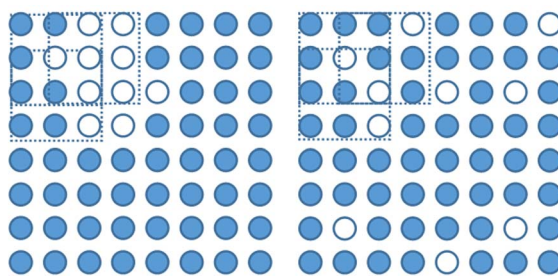
$$\Omega(\mathbf{S}) = \sum_i^m \sum_{h \in \mathcal{H}} \|\omega_h^i\|_\infty, \quad (11)$$

其中 $\mathbf{S} \in R^{m \times n}$, $\Psi(\mathbf{S})$ 为结构稀疏范数^[11] 如式(10)所示, \mathbf{S} 的第 j 列 $s^j \in R^m$ 包含 m 个元素, 下标为 $\{1, \dots, m\}$, 将第 j 列 s^j 重新排列成一幅完整图像, 其元素点可通过步长为 1 的滑动窗口划分成互相重叠的分组, 每个分组 $g \in \mathcal{N}$ 为元素集合的一个子集。本文提出时间持续范数 $\Omega(\mathbf{S})$ 如式(11)所示, \mathbf{S} 的第 i 行 $\omega^i \in R^n$ 包含 n 个元素, 下标为 $\{1, \dots, n\}$, 其元素点可通过步长为 1 的滑动窗口划分成互相重叠的分组, 并且每个分组 $h \in \mathcal{H}$ 为元素集合的一个子集。 ρ 为参数 ($0 \leq \rho \leq 1$), 可调节结构稀疏范数和空间连续范数所占的比例。 $\|\cdot\|_\infty$ 表示 l_∞ 范数, 是一组元素中的最大值。

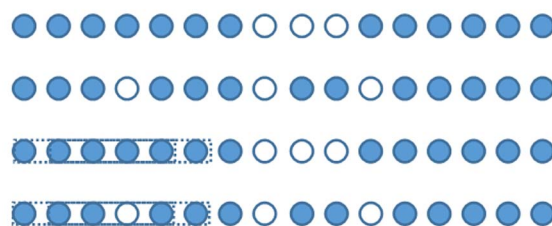
现实场景中, 前景目标在空间上通常是连续的(图 1(a))。假设稀疏前景目标为大小是 8×8 ($m=64$) 的两个不同分布(图 1(a), 1(b)), 其中白像素对应前景, 黑像素对应背景。由于 l_1 -范数是所有像素绝对值的和, 在图 1(a), 1(b) 两种不同分布下得到的类似 l_1 -范数, 从而无法有效区分两种分布。而结构稀疏范数只取各分组中最大值并求和(分组取 3×3 大小, 8×8 帧中有 36 个重叠分组), 因此可明显区分两种不同分布: 图 1(c) 的 $\Psi(\mathbf{S})$ 值比图 1(d) 的 $\Psi(\mathbf{S})$ 值小得多。以上可以看出, 由于 l_1 -范数对每个像素进行单独处理, 故不能很好地区分图 1(a), 1(b) 两种情况, 而结构稀疏范数则可以更好地反映出元素子集之间可能存在的空间连续关系。



(a) 空间连续分布 (b) 空间非连续分布
(a) Spatial-continuous distribution (b) Non-spatial-continuous distribution



(c) 3×3 重叠分组 I (d) 3×3 重叠分组 II
(c) 3×3 overlapping grouping I (d) 3×3 overlapping grouping II



(e) 时间持续与非持续分布及 1×5 重叠分组
(e) Temporal-continual and non-temporal-continual distribution and 1×5 overlapping

图 1 时空稀疏范数

Fig. 1 Spatial-temporal sparsity-inducing norms

前景矩阵 \mathbf{S} 第 i 行 ω^i 蕴含着某一固定位置在时间维度上的信息。在现实视频场景中, 前景目标通常持续出现在某一区域, 即目标的出现在时间上具有持续性, 因此当存在前景目标时, 前景矩阵 \mathbf{S} 某特定行 \mathbf{W}^x 中也应连续出现非零元素。但是这种信息在以往的处理方法中往往被忽略。为了利用这种时间持续信息, 本文提出时间持续范数 $\Omega(\mathbf{S})$ 。在实际运算中使用 1×5 的分组, 各分组与其邻接分组有 4 个元素重叠(如图 1(e)所示)。在图 1(e)中, 每一行为一组单独数据, 第一组表示某一位置持续被前景目标占据的场景, 第二组为该位置随机出现前景元素的情形, 显然第一组的元素分布比第二组更符合实际情况, 而此时采用时间持续范数 $\Omega(\mathbf{S})$ 能很好地区分这两种情况: 第一组的时间持续范数 $\Omega(\mathbf{S})$ 明显小于第二组。

将以上两种范数进行融合就可以得到时空稀疏范数, 并通过调节 ρ 的大小来改变两者所占权重。

3.3 模型求解

本文采用非精确增广拉格朗日乘子(Inexact

Augmented Lagrange Multiplier, IALM)方法来求解,首先引入两个辅助变量 \mathbf{J}, \mathbf{K} ,将式(8)转变为:

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{S}, \mathbf{E}, \mathbf{J}, \mathbf{K}} \|\mathbf{J}\|_* + \|\mathbf{K}\|_* + \lambda \|\mathbf{S}\|_1 + \beta \|\mathbf{E}\|_k, \quad (12)$$

$$\text{s. t. } \mathbf{D} = \mathbf{XV} + \mathbf{S} + \mathbf{E}, \mathbf{X} = \mathbf{UD}, \mathbf{U} = \mathbf{J}, \mathbf{V} = \mathbf{K}$$

将 $\|\mathbf{E}\|_k$ 具体化为 $\|\mathbf{E}\|_1$, $\|\mathbf{S}\|_k$ 具体化为 3.2 节中提出的时空稀疏范数 $\Phi(\mathbf{S})$,则其增广拉格朗日方程为:

$$L(\mathbf{U}, \mathbf{V}, \mathbf{S}, \mathbf{E}, \mathbf{J}, \mathbf{K}, Y_1, Y_2, Y_3) = \|\mathbf{J}\|_* + \|\mathbf{K}\|_* + \lambda \Phi(\mathbf{S}) + \beta \|\mathbf{E}\|_1 + \langle Y_1, \mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E} \rangle + \langle Y_2, \mathbf{U} - \mathbf{J} \rangle + \langle Y_3, \mathbf{V} - \mathbf{K} \rangle + \frac{\mu}{2} (\|\mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E}\|_F^2 + \|\mathbf{U} - \mathbf{J}\|_F^2 + \|\mathbf{V} - \mathbf{K}\|_F^2), \quad (13)$$

其中, Y_1, Y_2, Y_3 是拉格朗日乘子, μ 为惩罚因子 ($\mu > 0$)。

其求解的具体步骤为:

(1)初始化:输入原始数据,对各变量赋初值,设定误差。

(2)固定其他变量更新 \mathbf{J} :

$$\mathbf{J} = \operatorname{argmin} \|\mathbf{J}\|_* + \langle Y_2, \mathbf{U} - \mathbf{J} \rangle + \frac{\mu}{2} \|\mathbf{U} - \mathbf{J}\|_F^2 = \operatorname{argmin} \frac{1}{\mu} \|\mathbf{J}\|_* + \frac{1}{2} \left\| \mathbf{J} - \left[\mathbf{U} + \frac{Y_2}{\mu} \right] \right\|_F^2, \quad (14)$$

式(14)的解可由奇异值收缩算子 \mathbf{SVT}_μ

$\left[\mathbf{U} + \frac{Y_2}{\mu} \right]$ 获得,奇异值收缩算子的具体公式为:

$$\mathbf{SVT}_\mu(\mathbf{Y}) = \mathbf{P} \operatorname{diag}[(\delta - \mu)_+] \mathbf{Q}, \quad (15)$$

$$(\delta - \mu)_+ = \begin{cases} \delta - \mu, & \delta > \mu \\ 0, & \delta \leq \mu \end{cases}, \quad (16)$$

其中 $\mathbf{P}, \mathbf{Q}, \delta$ 由 \mathbf{Y} 的奇异值分解获得,即 $\mathbf{Y} = \mathbf{P}\Sigma\mathbf{Q}$, $\Sigma = \operatorname{diag}(\delta)$ 。

(3)固定其他变量更新 \mathbf{K} :

$$\mathbf{K} = \operatorname{argmin} \|\mathbf{K}\|_* + \langle Y_3, \mathbf{V} - \mathbf{K} \rangle + \frac{\mu}{2} \|\mathbf{V} - \mathbf{K}\|_F^2 = \operatorname{argmin} \frac{1}{\mu} \|\mathbf{K}\|_* + \frac{1}{2} \left\| \mathbf{K} - \left[\mathbf{V} + \frac{Y_3}{\mu} \right] \right\|_F^2, \quad (17)$$

式(17)的求解方法与(14)相同。

(4)固定其他变量更新 \mathbf{E} :

$$\mathbf{E} = \operatorname{argmin} \beta \|\mathbf{E}\|_1 + \langle Y_1, \mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E} \rangle + \frac{\mu}{2} \|\mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E}\|_F^2 = \operatorname{argmin} \frac{\beta}{\mu} \|\mathbf{E}\|_1 + \frac{1}{2} \|\mathbf{E} - \mathbf{D} - \mathbf{UDV} - \mathbf{S} + \frac{Y_1}{\mu}\|_F^2, \quad (18)$$

式(18)的解可由收缩算子 $\delta_\mu^{\frac{\beta}{\mu}}(\mathbf{D} + \mathbf{UDV} + \mathbf{S} - \frac{Y_1}{\mu})$ 获得,收缩算子 $\delta_g(W)$ 的定义为:

$$\delta_g(W) = \begin{cases} \omega - \epsilon, & \omega > \epsilon \\ \omega - \epsilon, & \omega < -\epsilon \\ 0, & \text{otherwise} \end{cases}, \quad (19)$$

(5)固定其他变量更新 \mathbf{S} :

$$\mathbf{S} = \operatorname{argmin} \lambda \Phi(\mathbf{S}) + \langle Y_1, \mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E} \rangle + \frac{\mu}{2} \|\mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E}\|_F^2 = \operatorname{argmin} \lambda \Phi(\mathbf{S}) + \frac{\mu}{2} \|\mathbf{E} - \mathbf{D} - \mathbf{UDV} + \frac{Y_1}{\mu} - \mathbf{S}\|_F^2, \quad (20)$$

式(20)的解可以通过二次最小代价流的方法获得。

(6)固定其他变量更新 \mathbf{U} :

令 $\partial L(\mathbf{U}, \mathbf{V}, \mathbf{S}, \mathbf{E}, \mathbf{J}, \mathbf{K}, Y_1, Y_2, Y_3) / \partial \mathbf{U} = 0$, 得:

$$\mathbf{U} = \left[\frac{1}{\mu} (Y_1 \mathbf{V}^T \mathbf{D}^T - Y_2) + (\mathbf{D} - \mathbf{S} - \mathbf{E}) \mathbf{V}^T \mathbf{D}^T + \mathbf{J} \right] \cdot (\mathbf{D} \mathbf{V} \mathbf{V}^T \mathbf{D}^T + \mathbf{D})^{-1}, \quad (21)$$

(7)固定其他变量更新 \mathbf{V} :

令 $\partial L(\mathbf{U}, \mathbf{V}, \mathbf{S}, \mathbf{E}, \mathbf{J}, \mathbf{K}, Y_1, Y_2, Y_3) / \partial \mathbf{V} = 0$, 得: $\mathbf{V} = (\mathbf{D}^T \mathbf{U}^T \mathbf{U} \mathbf{D} + \mathbf{I})^{-1} \left[\frac{1}{\mu} (\mathbf{D}^T \mathbf{U}^T Y_1 - Y_3) + \mathbf{D}^T \mathbf{U}^T (\mathbf{D} - \mathbf{S} - \mathbf{E}) + \mathbf{K} \right]$, (22)

(8)更新拉格朗日乘子 Y_1, Y_2, Y_3 :

$$Y_1 = Y_1 + \mu (\mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E}), \quad (23)$$

$$Y_2 = Y_2 + \mu (\mathbf{U} - \mathbf{J}), \quad (24)$$

$$Y_3 = Y_3 + \mu (\mathbf{V} - \mathbf{K}), \quad (25)$$

(9)更新参数 μ :

$$\mu = \min(\kappa \mu, \mu_{\max}). \quad (26)$$

其中 κ 为更新步长且 $\kappa > 1$, μ_{\max} 为指定的上限。

(10)判断是否达到收敛条件:

$$\|\mathbf{D} - \mathbf{UDV} - \mathbf{S} - \mathbf{E}\|_\infty < \epsilon, \quad \|\mathbf{U} - \mathbf{J}\|_\infty < \epsilon, \quad \|\mathbf{V} - \mathbf{K}\|_\infty < \epsilon, \quad (27)$$

其中 ϵ 为误差。

当达到收敛条件时算法结束,否则重复步骤(2)-(10)。

3.4 算法分析

3.2 节中提出的时空稀疏范数的前提条件较易满足,即所获得的数据矩阵 \mathbf{D} 必须是由一段或若干段视频序列组成。通过参数 ρ 调节空间连续性和时间持续性所占的权重,当前提条件无法保

证时间持续性时(数据矩阵 D 的相邻列不具有前后顺序关系),将 ρ 置为 0,时空稀疏范数将退化为结构稀疏范数,此时将只考虑前景的空间连续性。时空稀疏范数中 3×3 、 1×5 的分组大小可调节,当取较小值时体现的是较小尺度上前景的时间持续性和空间连续性,但抗干扰能力较差;当取较大值时体现的是较大尺度上前景的时间持续性和空间连续性,但是无法处理前景的内部空洞和突然缺失的情形。

3.3 最优化求解过程中矩阵变量个数为 6,且目标函数非平滑,因此从理论上证明其收敛性非常困难,然而在实际应用过程中,算法得到的局部最优解往往就是全局最优解,因此本文不做进一步证明。需要指出的是本文算法的复杂度为 $O(m^2 n + n^3)$ 。

4 实验和分析

为了验证本文算法的有效性,设计了两组实验。所有程序均在 Intel Core i5-3470 3.2GHz 处

理器上运行,8G 内存,采用 Matlab 2013b 编程实现。

实验 1:

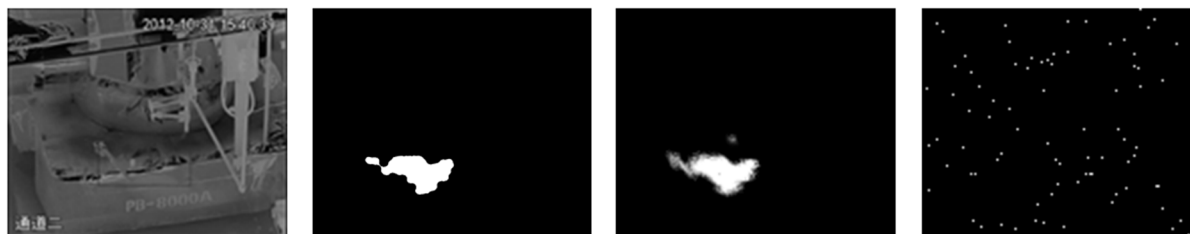
本文实验采用的是大小为 250×180 ,共 450 帧的视频序列。图 2(a)是原始视频,图 2(b)是加入了参数为 0.02 的椒盐噪声的视频。 λ 、 β 、 κ 、 ρ 、 μ 分别取为 0.08、0.03、1.2、0.3、 10^{-5} 进行实验。为了更清楚地将前景检测结果与真实前景做比较,做了二值化处理,结果如图 3 所示。



(a) 原始视频 (b) 噪声视频
(a) Original video (b) Video with noises

图 2 实验视频

Fig. 2 Experimental videos



(a) 背景 (b) 二值化真实前景 (c) 二值化前景 (d) 噪声
(a) Background (b) Binary real foreground (c) Binary foreground (d) Noises

图 3 实验结果

Fig. 3 Experiment results



(a) 二值化真实前景 (b) RPCA (c) TLRR (d) 本文算法
(a) Binary real foreground (b) RPCA (c) TLRR (d) Proposed algorithm

图 4 实验结果

Fig. 4 Experiment results

本文将经典的 RPCA^[7]方法、TLRR^[10]方法与本文算法做了比较,结果如图 4 所示。为了更客观地体现本文算法的优点,将检出的前景与真实前景做了量化比对,两者二值化之后的重叠区域至少要大于某一比例才认为是正确检出,否则即为误检。3 种方法的误检率(False Positive Rate, FPR)曲线如图 5 所示。图 5 中横轴为检出

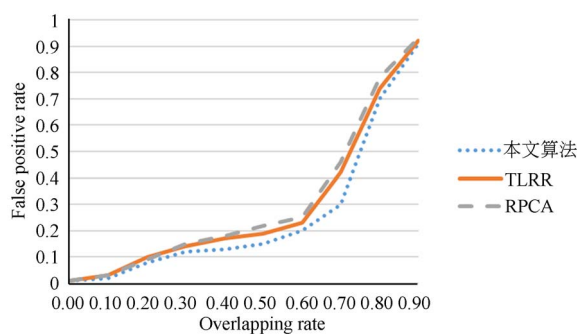


图 5 3 种方法的误检率曲线

Fig. 5 False positive rate curves for three methods

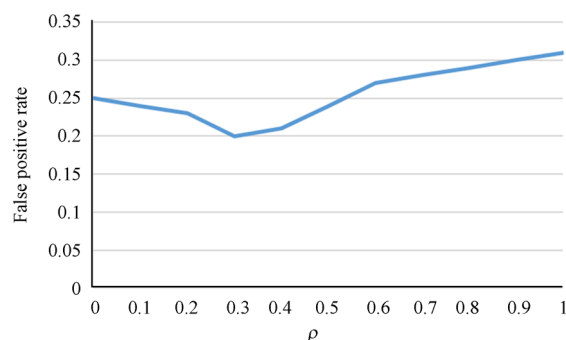


图 6 参数 ρ 对误检率的影响

Fig. 6 Influence of ρ on false positive rate

前景与真实前景重叠比例标准,重叠比例小于标准为误检。从图 5 中可以看出本文方法在 3 种方法中性能最优。

实验 2:

为了研究时空稀疏范数中参数 ρ 对实验结果的影响,固定其他参数,对 ρ 在 $(0, 1)$ 范围内取值,在检出前景与真实前景重叠 60% 作为误检率标准时,误检率曲线如图 6 所示。经过实验发现在本文所设定的条件下, $\rho=0.3$ 时效果最佳。需要指出的是 ρ 越大越着重考虑前景运动的持续性; ρ 越小越着重考虑前景的空间连续性。

5 结 论

针对视频中前景检测的问题,根据前景本身具有的空间连续特性和时间持续特性,提出了一种基于 TLRR 和时空稀疏分解的检测方法。首先,构造结构稀疏范数和时间持续范数,并将二者融合成时空稀疏范数;然后,将 TLRR 由 2 分解推广为 3 分解,并通过 IALM 方法进行最优化求解,直接分解出背景矩阵,前景矩阵和噪声矩阵;最后通过实验证明本文提出的方法具有较好的精度,并且研究了时空稀疏范数中参数 ρ 对算法性能的影响,证明 ρ 的取值将影响整个算法的精度,在本文条件下 ρ 取 0.3 效果较好, ρ 过大和过小都会降低检出的准确度。下一步的研究重点将考虑背景本身含有运动元素的情景以及对算法收敛性的证明。

参考文献:

- [1] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking [C]. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins*; IEEE, 1999: 246-252.
- [2] ELGAMMAL A, HARWOOD D, DAVIS L. Non-parametric model for background subtraction [C]. *Proc. Eur. Conf. Comput. Vis. Springer*, 2000: 751-767.
- [3] YASER S, MUBARAK S. Bayesian modeling of dynamic scenes for object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(11): 1778-1792.
- [4] HEIKKILA M, PIETIKAINEN M. A texture-based method for modeling the background and detecting moving objects [J]. *IEEE Trans on Pattern Anal. and Machine Intelligence*, 2006, 28(4): 657-662.
- [5] KIM K, CHJIDABHONGSE T H, HARWOOD D, et al.. Real-time foreground-background segmentation using codebook model [J]. *Real-Time Imaging*, 2005, 11(3): 172-185.
- [6] BOUWMANS T, ZAHZAH E H. Robust PCA via principal component pursuit; A review for a comparative evaluation in video surveillance [J]. *Comput. Vis. Image Underst.*, 2014, 122: 22-34.
- [7] BAO B K, LIU G C, XU C S, et al.. Inductive robust principal component analysis [J]. *IEEE Trans on Image Processing*, 2012, 21(8): 3794-3800.

- [8] LIU G C, LIN Z C, YAN S C, *et al.*. Robust recovery of subspace structures by low-rank representation [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2013, 35(1): 171-184.
- [9] LIU G C, YAN S C. Latent low-rank representation for subspace segmentation and feature extraction [C]. *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011: 1615-1622.
- [10] ZHANG Z, YAN S C, ZHAO M B, *et al.*. Robust bilinear matrix recovery by tensor low-rank representation [C]. *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014: 2945-2951.
- [11] LIU X, ZHAO G Y. Background Subtraction Based on Low-rank and Structured Sparse Decomposition *IEEE Trans on Image Processing DRAFT*, 2015.
- [12] 何玉杰, 李敏, 张金利, 等. 基于低秩三分解的红外图像杂波抑制[J]. *光学 精密工程*, 2015, 23(7): 2070-2078.
HE Y J, LI M, ZHANG J L, *et al.*. Clutter suppression of infrared image based on three-component low-rank matrix decomposition [J]. *Opt. Precision Eng.*, 2015, 23(7): 2070-2078. (in Chinese)
- [13] YAO J W, LIU X, QI CH. Foreground detection using low rank and structured sparsity [C]. *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014.
- [14] 穆治亚, 魏仲慧, 何昕, 等. 采用稀疏表示的红外图像自适应杂波抑制[J]. *光学 精密工程*, 2013, 21(7): 1850-1857.
MU ZH Y, WEI ZH H, HE X, *et al.*. Adaptive clutter suppression of infrared images by using low-rank representation [J]. *Opt. Precision Eng.*, 2013, 21(7): 1850-1857. (in Chinese)
- [15] ZHANG Z, ZHAO M B, CHOW T W. Binary- and multi-class group sparse canonical correlation analysis for feature extraction and classification [J]. *IEEE Trans. Knowledge and Data Engineering*, 2013, 25(10): 2192-2205.
- [16] 赵九龙, 马瑜, 李爽, 等. 三维医学图像的混合噪声去除方法[J]. *液晶与显示*, 2015, 30(2): 340-346.
ZHAO J L, MA Y, LI SH, *et al.*. Mixed noise removing method for three-dimensional medical images [J]. *Chinese Journal of Liquid and Displays*, 2015, 30(2): 340-346. (in Chinese)
- [17] 杨亚威, 李俊山, 张士杰, 等. 基于生物视觉标准模型特征的无参考型图像质量评价方法[J]. *液晶与显示*, 2014, 29(6): 1016-1023.
YANG Y W, LI J SH, ZHANG SH J, *et al.*. Non-reference image quality assessment approach based on standard model features of biological vision [J]. *Chinese Journal of Liquid and Displays*, 2014, 29(6): 1016-1023. (in Chinese)

作者简介:



隋中山(1985—),男,山东平度人,博士研究生,2007年、2009年于第二炮兵工程学院分别获得学士、硕士学位,主要从事图像目标识别方面的研究。E-mail: zclszs@163.com



李俊山(1956—),男,陕西白水人,教授,博士生导师,1981于国防科技大学获得学士学位,1988年于第二炮兵工程学院获得硕士学位,2001年于西安微电子技术研究所获得博士学位,主要从事智能图像处理与目标感知识别,电子对抗模拟与仿真等。