

基于高阶鲁棒主成分分析的视频前景与背景分离

周至恺¹ 廖亮¹ 王月轮¹ 顾强¹ 魏平俊¹

ZHOU Zhikai LIAO Liang WANG Yuelun GU Qiang WEI Pingjun

摘要

运动目标检测的发展历史由来已久,出现了基于均值方法、中值方法、直方图方法和高斯混合模型等各种检测算法。然而,这些目标检测方法往往未考虑图像背景之间的时间相关性以及前景对象内的空间连续性,也未考虑视频中的其他结构信息。采用微观视角,将目标视频中每个帧像素的张量化进行了扩展,即将原始图像的像素邻域扩展为一定阶数的数组。因此,原始图像转化为一个形式上的高阶图像。所提出的模型旨在应用于诸如视频前景背景分离等任务,以增强图像算法的有效性。

关键词

运动目标检测;张量;邻域扩展;高阶图像;视频前景背景分离

doi: 10.3969/j.issn.1672-9528.2024.03.014

0 引言

视频前景与背景分离在计算机视觉中扮演着至关重要的角色,它是一项复杂的任务,涉及分析视频流中像素级的变化,以准确区分移动的前景对象与相对静止的背景环境。这项技术在许多领域都具有重要意义,如监控系统、自动驾驶汽车技术以及行为分析^[1]。

传统的方法,如高斯混合模型(GMM)^[2]、视差基于外观的背景建模(VIBE)^[3]等,旨在解决视频前景与背景分离问题。然而,这些方法仍面临挑战,例如在户外环境中对光照变化的敏感性以及对异常噪声的不稳定处理。近年来,鲁棒主成分分析(RPCA)作为一种方法被提出,能够分解视频数据矩阵为低秩背景和稀疏前景分量,但其对异常噪声过于敏感,限制了其实际应用。

针对这些挑战,本文提出了一种新的方法,即基于广义矩阵模型的高阶鲁棒主成分分析(HORPCA)。该方法利用了多维结构和邻域信息,通过引入循环卷积和广义标量,实现了更准确的视频前景与背景分离。HORPCA具有更强的可交换性和鲁棒性,在处理异常噪声和光照变化方面表现出更优越的性能。

本研究通过对比实验验证了HORPCA方法在视频前景与背景分离中的有效性。结果显示,相较于传统方法和RPCA^[4],HORPCA在多个数据集上表现出更高的精确度和稳健性。这一创新性方法为解决视频前景与背景分离问题提供了新的思路和可行性,对未来计算机视觉领域具有重要的

研究和应用价值。

1 T-矩阵

1.1 广义矩阵数学模型

一个 N 阶数组是由 $\mathbb{C}$ 上的复数构成,且该 N 阶数组为集合 $\mathbb{C}$ 上的元素,则有:

$$C \equiv \mathbb{C}^{I_1 \times \dots \times I_n}, I_n (1 \leq n \leq N) \in \mathbb{N}^+ \quad (1)$$

当然,若高阶数组的元素为实数,同样有:

$$C \equiv \mathbb{C}^{I_1 \times \dots \times I_n}, I_n (1 \leq n \leq N) \in \mathbb{N}^+ \quad (2)$$

定义均为交换环结构,用循环卷积代替乘积运算,集合 C 和集合 R 中的元素被定义广义标量(t -scalars)。

以广义标量 t -scalars为元素构成的二阶数组为广义矩阵 t -matrix。若用 X_{TM} 表示大小为 $D_1 \times D_2$ 广义矩阵,则有 $X_{TM} \in \mathbb{C}^{D_1 \times D_2} \equiv \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N \times D_1 \times D_2}$ 。

广义矩阵的形成如图1所示。

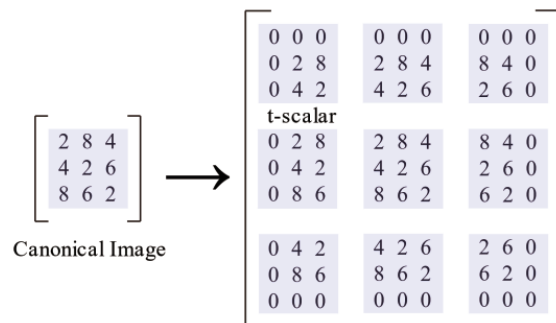


图1 使用中心邻域拓展策略对一个 3×3 的矩阵进行拓展

使用这种邻域拓展策略,可将一个大小为 3×3 的矩阵拓展为一个 $3 \times 3 \times 3 \times 3$ 的4阶数组,其中 t -matrix的大小为 3×3 。

1. 中原工学院电子信息学院 河南郑州 450007

[基金项目] 中原工学院科研团队发展项目(K2022TD001), 河南省杰出外籍科学家工作室(GZS2022012)

广义标量的实质是以高阶数组为数据结构的广义复数组。利用抽象代数中的模理论,可以对经典矩阵的相关概念进行推广,从而引入兼容对应传统概念的广义范数(*t-matrix norm*)、广义秩(*t-matrix rank*)、广义正交矩阵(*orthogonal t-matrix*)等概念。广义矩阵向下兼容经典矩阵,同样具有秩(*rank*)、范数(*norm*)、迹(*trace*)等基本概念以及广义加法、广义乘法、广义傅里叶变换、广义共轭转置等基本运算。

1.2 广义矩阵数学模型的部分定义与定理

定义1 广义矩阵加法: 定两个任意广义矩阵 $A_{TM} \in C^{D_1 \times D_2}$ 和 $B_{TM} \in C^{D_1 \times D_2}$, 则广义矩阵加法记为:

$$C_{TM} \doteq A_{TM} + B_{TM} \in C^{D_1 \times D_2} \quad (3)$$

定义2 广义矩阵乘法: 给定两个任意广义矩阵 $A_{TM} \in C^{D_1 \times D_2}$ 和 $B_{TM} \in C^{D_1 \times D_2}$, 则广义矩阵乘法记为:

$$C_{TM} \doteq A_{TM} \circ B_{TM} \in C^{D_1 \times D_2} \quad (4)$$

定义3 广义矩阵的共轭转置: 给定一个任意的广义矩阵 $X_{TM} \in C^{D_1 \times D_2}$, 它的共轭转置记为 $X_{TM}^H \in C^{D_1 \times D_2}$, 具体定义为:

$$X_{TM, \beta, \alpha} = \text{conj}(X_{TM, \beta, \alpha}) \in C, 1 \leq \alpha \leq D_1, 1 \leq \beta \leq D_2 \quad (5)$$

2 鲁棒主成分分析

RPCA 方法是一种用于分析来自复杂非线性系统数据的技术。其主要目标是在数据中识别稀疏结构,同时将其余条目拟合到低秩基础上。如果给定的数据确实存在于低维子空间并包含稀疏成分,由 Candes 等人证明的 RPCA 算法可以准确地将数据分离为两个部分: 一个低秩矩阵(表示为 L)和一个稀疏矩阵(表示为 S), 其中 $X=L+S$ 。

RPCA 算法将问题表述为一个可解的、非光滑凸优化问题,称为主成分追踪(*principal component pursuit, PCP*)。其目标是极小化核范数 $\|L\|_*$ (表示低秩结构)和 $L1$ 范数 $\|S\|_1$ (表示稀疏结构)的组合,其中通过正则化参数 λ 控制权衡。当 λ 接近 0 时, L 包含所有数据,使得 S 具有稀疏元素。相反,当 λ 增加时, S 捕捉更多原始数据,同时 L 逐渐接近零矩阵其中, X 是一个 $n \times m$ 的矩阵。

λ 的选择至关重要, Candes 等人建议了一个具体的数值:

$$\lambda = \frac{1}{\sqrt{\max(n, m)}} \quad (6)$$

这个 λ 值的选择在正确分离 L 和 S 方面具有很高的成功概率。微调 λ 可能会略微改善分离结果。

有各种方法可以解决凸形式的 PCP 问题,其中增广拉格朗日乘子(ALM)方法以其简单性、稳定性和高效性而脱颖而出,与其他方法相比, ALM 在较少迭代次数内实现了高

精度。为了将 RPCA 应用于视频,本质上是二维的帧被重塑为一维的列向量,并合并成一个单一的数据矩阵 X 。然后, RPCA 将 X 分解为 L 和 S , 其中 L 代表背景视频,而 S 代表移动的前景物体。

3 HO-RPCA 方法

3.1 数据预处理和张量表示

一个视频由 n 帧图像组成,每个灰度图像用矩阵 I_i 表示,其中 i 表示第 i 帧。每个图像矩阵的尺寸为 $M \times N$, 其中 M 表示图像的高度, N 表示图像的宽度。将这 n 帧图像转换成一个矩阵 X , 其中每一行包含一幅图像的所有像素值。矩阵 X 的大小为 $n \times (M \times N)$ 。利用张量代数算法, 本文将 X 升级为 X_{extended} , 其维度为 $3 \times 3 \times n \times (M \times N)$ 。

3.2 低秩矩阵和稀疏矩阵分解方法

3.2.1 问题表述

RPCA 的核心问题是极小化以下函数:

$$\min_{L, S} \|L\|_* + \lambda \|S\|_1 \quad (7)$$

式中: $\|L\|_*$ 表示矩阵 L 的核范数, $\|S\|_1$ 代表矩阵 S 的 $L1$ 范数。参数 λ 控制稀疏性。

3.2.2 ADMM (交替方向乘子法) 迭代步骤

(1) 更新 L

$$L^{k+1} = \arg \min_L \|L\|_* + \frac{\mu}{2} \|L - (X_{\text{extended}} - S^k + \frac{Y^k}{\mu})\|_F^2 \quad (8)$$

这是一个核范数阈值问题。通过奇异值分解(SVD),核范数阈值问题可以转化为保留前 r 个最大奇异值并将其余奇异值设为零的问题,其中 r 是通过硬阈值或软阈值等方法确定的阈值。

(2) 新 S

使用软阈值法解决的 $L1$ 范数问题:

$$S^{k+1} = \arg \min_S \lambda \|S\|_1 + \frac{\mu}{2} \|S - (X_{\text{extended}} - L^{k+1} + \frac{Y^k}{\mu})\|_F^2 \quad (9)$$

(3) 更新 Y

$$Y^{k+1} = Y^k + \mu(L^{k+1} + S^{k+1} - X_{\text{extended}}) \quad (10)$$

这个更新步骤是 ADMM (交替方向乘子法) 算法的关键部分,用于保持矩阵 L 和 S 之间的一致性。它不断调整矩阵 Y , 使得 $L^{k+1} + S^{k+1}$ 近似于原始的扩展数据矩阵 X_{extended} 。

在每次迭代中,基于当前的 $L^{k+1} + S^{k+1}$ 和 X_{extended} 计算出一个新的 Y^{k+1} 。随着迭代的进行,根据 $L^{k+1} + S^{k+1}$ 和 X_{extended} 之间的差异,持续调整 Y , 旨在实现更好的分解结果。

(4) 更新 μ

$$\mu^{k+1} = \min(\rho \mu^k, \max \mu) \quad (11)$$

在每个迭代步骤中, $L^{k+1} + S^{k+1}$ 分别是更新的低秩和稀疏矩阵。 Y^{k+1} 是一个辅助变量, μ^{k+1} 是逐渐增加的参数。迭代停

止的条件是收敛，即变量的变化变得小于阈值 ε 。

算法 1 对方程 (7) ~ (11) 的优化算法

输入: $X_{extended}, \lambda, \varepsilon = 1e-8, N = 1000, \rho = 1e-3,$

$\mu = 1e-3, \mu_{max} = 1e10, \lambda$

输出: L, S

1. For $k=1$ to K do

2. $L^k = L, S^k = S$

3. 迭代更新

(1) 更新 L :

$$L^{k+1} = \arg \min_L \|L\|_* + \frac{\mu}{2} \|L - (X_{extended} - S^k + \frac{Y^k}{\mu})\|_F^2$$

(2) 更新 S :

$$S^{k+1} = \arg \min_S \lambda \|S\|_1 + \frac{\mu}{2} \|S - (X_{extended} - L^{k+1} + \frac{Y^k}{\mu})\|_F^2$$

(3) 更新 Z : $Z \leftarrow L + S - X_{extended}$

(4) 更新 Y : $Y \leftarrow Y + \mu \times Z$

(5) 更新 μ : $\mu \leftarrow \min(\rho \times \mu, \mu_{max})$

4. L, S, Z 变化的计算

(1) $L: chgL \leftarrow \max(|L^k - L|)$

(2) $S: chgS \leftarrow \max(|S^k - S|)$

(3) $Z: chgZ \leftarrow \max(|Z|)$

(4) 总体变化: $chgL \leftarrow \max(chgL, chgS, chgZ)$

5. If $change < \varepsilon$ then

break

6. End If

7. End For

4 实验

本文使用 DAVIS (密集标注视频分割) 和 CDnet14 (2014 年变化检测数据集) 来评估本文的方法, 与 DMD^[5]、RPCA^[6] 和来自 LRS Library^[7] 的 R2PCP^[8]、ManhNMF^[9]、IALM^[10] 在同一组数据集上获得的实验结果进行比较。为了保持公平性, 本文选择对应方法作者发布的默认参数, 并将所有帧的分辨率调整为 320×240 。

4.1 评估设置

4.1.1 数据集

DAVIS^[11] 数据集是一个专为视频对象分割 (VOS) 任务设计的计算机视觉数据集。它包含一系列高分辨率视频序列, 每个序列都包含一个或多个前景对象。

CDnet14^[12] 数据集是一个专为变化

检测研究设计的计算机视觉数据集。该数据集旨在为评估算法在图像序列中检测对象和变化的性能提供标准基准。

4.1.2 评价指标

精确度 (precision)、召回率 (recall) 和 F -度量 (F -measure) 是广泛使用的指标, 它们的定义为:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (12)$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (13)$$

$$F_measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

精确度 (precision) 衡量了模型正确预测为正例的样本数量与模型总共预测为正例的样本数量的比例。召回率 (recall) 则计算了模型正确预测为正例的样本数量与数据中总共的实际正例样本数量的比例。而 F 度量 (F -measure) 是精确度和召回率的调和平均数, 平衡了这两个指标, 是一种全面的评估方法。

4.1.3 参数设置

本文算法中有五个关键参数, 分别是 ε 、 ρ 、 N 、 μ 和 μ_{max} 。其中, ε 代表收敛的阈值。当算法状态的变化低于此阈值时, 认为算法已收敛至解。 N 定义了算法可以执行的最大迭代次数, 即使在多次迭代后未满足指定的收敛条件, 算法也会停止迭代。 ρ 是调整算法中步长的参数, 较小的 ρ 值可能导致非常小的步长, 潜在地减慢收敛速度, 较大的 ρ 值会产生较大的步长, 影响算法的稳定性。 μ 表示算法使用的初始步长, 步长影响每次迭代中更新的幅度, 较小的步长通常有助于稳定的收敛, 但可能会增加所需的迭代次数。 μ_{max} 限制最大步长, 防止快速增加, 确保算法的稳定性。本文根据经验设置为:

$$\{\varepsilon, \rho, N, \mu, \mu_{max}\} = \{1e-8, 1.1, 1000, 1e-3, 1e10\} \quad (15)$$

4.2 结果

图 2 展示了分别来自 DAVIS 和 CDnet14 视频剪辑的某

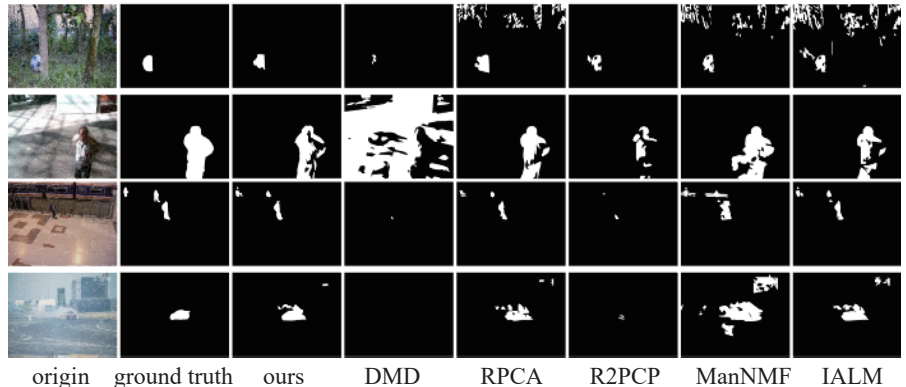


图 2 在 7 组数据集上不同算法的表现

一帧的检测结果。

从图2中可以看出,本文的方法可以产生更精确的边界信息,并更好地抑制了噪音的影响。DMD在人物或物体移动过快或过慢时难以捕捉前景。ManNMF和IALM未能有效抑制噪音,导致不必要的背景出现。

图3显示了在公共DAVIS数据集和CDnet14上的精度(precision)、召回率(recall)和F-measure。可以清楚地看到,本文的方法在某些数据集上在精度、召回率和F-measure方面明显优于现有技术水平。

Methods		HORPCA	DMD	RPCA	R2PCP	ManhNMF	IALM
soccerball	P	0.7217	0.1537	0.0522	0.7162	0.1522	0.1114
	R	0.7359	0.4287	0.9224	0.7048	0.8103	0.8110
	F	0.7078	0.2198	0.0972	0.7033	0.2495	0.1935
peopleInShade	P	0.8050	0.0524	0.7926	0.9531	0.7210	0.8904
	R	0.8543	0.3549	0.8421	0.3611	0.8261	0.6427
	F	0.8246	0.0912	0.8105	0.5152	0.7674	0.7351
PETS2006	P	0.6476	0.5375	0.6519	0.7863	0.4361	0.7616
	R	0.8692	0.0953	0.8073	0.1177	0.8655	0.8213
	F	0.7401	0.1606	0.7178	0.2024	0.5795	0.7349
drift	P	0.3013	0.1271	0.2332	0.7424	0.1133	0.2573
	R	0.9556	0.0047	0.8428	0.1753	0.9342	0.9300
	F	0.4362	0.0030	0.3559	0.2714	0.1956	0.3928

图3 不同算法在4个数据集上的表现

5 结论

首先,本文提出了一种基于HORPCA的移动对象检测方法,通过使用t-matrix来提高视频数据的维度。然后,介绍了在t-matrix下的视频处理方法。最后,在不同数据集上比较了本文算法的性能与其他算法。未来将增强算法,使其对噪声更具鲁棒性,并能够在监控场景中实现实时应用。

参考文献:

[1] NEWSON A. On Several mathematical and data-driven models for image and video editing, synthesis and analysis [EB/OL]. [2023-10-29].<https://theses.hal.science/tel-04198797v1>.

[2] RAKESH S, HEGDE N P, GOPALACHARI M V, et al. Moving object detection using modified GMM based background subtraction[EB/OL].[2023-11-05].<https://www.sciencedirect.com/science/article/pii/S2665917423002349>.

[3] GUO D, ZHANG J, LOU W, et al. Moving vehicle detection in satellite videos by improved ViBe algorithm[J].Remote Sensing Letters, 2023,14: 844-853.

[4] CANDÉS E, LI X, MA Y, et al. Robust principal component analysis[EB/OL]. (2009-11-08)[2023-11-06]. <https://arxiv.org/abs/0912.3599>.

[5] HAQ I, FUJII K, KAWAHARA Y. Dynamic mode decomposition via dictionary learning for foreground modeling in videos[C]//International Joint Conference on Computer Vision and Computer Graphics Theory and Applications. New York: Curran Associates, 2021: 476-483.

[6] WANG Y, WEI H, DING X, et al. Video background/foreground separation model based on non-convex rank approximation RPCA and superpixel motion detection[J]. IEEE access, 2020(8): 157493-157503.

[7] BOUWMANS T, AYBAT N S, ZAHZAH E. LRSLibrary: low-rank and sparse tools for background modeling and subtraction in videos[M]//Handbook of Robust Low-Rank and Sparse Matrix Decomposition. New York: Chapman and Hall/CRC, 2016.

[8] CHEN R, TONG Y, YANG J, et al. Video foreground detection algorithm based on fast principal component pursuit and motion saliency[EB/OL]. (2019-03-03)[2023-11-20]. <https://www.hindawi.com/journals/cin/2019/4769185/>.

[9] DRAGANOV I, RUMEN M. Object motion detection in video by fusion of RPCA and nmf decompositions[J]. New approaches for multidimensional signal processing: proceedings of international workshop, 2020, 270: 35-47.

[10] FAN R, JING M, WEI Y, et al. Weighted Schatten p-norm and Laplacian scale mixture-based low-rank and sparse decomposition for foreground-background separation[J]. Journal of electronic imaging, 2023, 32: 21-23.

[11] CAELLES J, PONT-TUSET F, PERAZZ A, et al. The 2019 DAVIS challenge on VOS: unsupervised multi-object segmentation[EB/OL]. (2019-05-02)[2023-11-27]. <https://arxiv.org/abs/1905.00737>.

[12] GOYETTE N, JODOIN P-M, PORIKLI F, et al. Changedetection. Net: a new change detection benchmark dataset[EB/OL]. (2012-10-01)[2023-11-06]. <https://ieeexplore.ieee.org/document/6238919/?arnumber=6238919>.

【作者简介】

周至恺(1997—),男,江苏南通人,硕士研究生,研究方向:基于高阶RPCA的前景背景分离。

(收稿日期:2024-01-04)