



OPEN

Moving objects detection based on tensor ring low rank decomposition

Ruixuan Chen^{2,7}, Xusheng Li^{3,7}, Chenda Chen⁴, Shiyu Zhu⁵, Qipeng Chen^{1✉} & Jianting Cao^{2,6✉}

The advancement of high-quality camera technology has increased the demand for efficient video analysis methods. Current methods mostly rely on matrix-based approaches, which break data structures and lose some spatial information. This paper proposes a novel approach (TRLRTTV) that combines Low Rank Tensor Ring decomposition and Tensor Total Variation regularization for moving objects detection (MOD). For static background detection, the tensor ring (TR) decomposition is utilized to extract low rank information, and low rank assumption is placed on tensor factors instead of the original data. For moving objects, a tensor total variation model with $l_{1/2}$ regularization is employed to ensure the representation of foreground information along the spatio-temporal direction. The results demonstrate that, compared to existing methods, the proposed algorithm achieves a 3%-8% performance improvement in background separation and the comprehensive performance metric f . Furthermore, the proposed method ensures robustness against noise interference, such as Gaussian and salt-and-pepper noise and more suitable for higher-dimensional video processing.

Keywords Machine learning, Tensor ring decomposition, Moving objects detection, Low rank decomposition

Video surveillance, autonomous driving, and face recognition are widely applied in the contemporary society, and the demand for video analysis technology is also increasing¹. Video analysis tasks can be broadly classified into basic tasks and advanced tasks. Object detection, video de-noising, video compression, etc., can be regarded as basic tasks, and Moving Objects Detection (MOD) is one of the most vital research directions in basic tasks. Some advanced tasks, such as people re-identification and video semantic understanding, are based on MOD. The accuracy of the MOD algorithm significantly determines the performance of these advanced algorithms.

With constant advances in sensor technology, high-quality cameras capture more detailed information from the scene, which increases the difficulty of MOD research. Furthermore, challenges such as moving backgrounds, rainy weather², camouflage, tiny object detection, and varying lighting conditions further complicate MOD research³. MOD approaches generally fall into two categories: pixel-based approaches⁴ and frame-based approaches⁵. Pixel-based approaches include classic approaches such as Mixture of Gaussian (MOG)⁶, Support Vector Machine (SVM)⁷, K-means clustering, Fuzzy C-means clustering (FCM)⁸, and others⁹. However, pixel-based approaches often exhibit inconsistent performance and tend to misclassify foreground objects under noisy conditions, making frame-based methods more favorable than pixel-based methods. Candès et al. proposed Robust Principal Component Analysis (RPCA)¹⁰, which is an influential contribution among different MOD methods. This method decomposes video into two parts: a low rank background component and a sparse foreground component. However, this method fails to detect dynamic background effectively. The GRASTA method¹¹, based on a subspace model, was proposed to overcome this issue by using an l_1 -norm model to detect moving objects. The robust subspace tracking method has been proven successful in dynamic object detection, but there are still some problems that need to be addressed¹².

Another MOD method, which combines Total Variation(TV) and RPCA, was proposed¹³. It utilizes the TV model to capture spatial and temporal relationships. However, it fails to detect tiny moving objects and struggles to capture spatial information effectively, as it tends to lose some information during the conversion from video frames into low-dimensional matrices. Subsequently, LR- l_1 TV method¹⁴, which is based on the total variation model, was proposed. This method can effectively detect static backgrounds and point objects but is less adept at handling noisy videos. A noise-robust model that combines tensor low rank approximation and tensor total

¹China Jiliang University, College of Information Engineering, Hangzhou 310018, China. ²Graduate School of Engineering, Saitama Institute of Technology, Fukaya 369-0293, Japan. ³Zhejiang XCC Group CO., Ltd., Xinchang 312500, China. ⁴Wen Zhou Special Equipment Inspection & Science Research Institute, Wen Zhou 325038, China. ⁵Department of Computing Science, Curtin University, Perth 6102, Australia. ⁶RIKEN Center for Advanced Intelligence Project (AIP), Tokyo 103-0027, Japan. ⁷Ruixuan Chen and Xusheng Li contributed equally to this work. ✉email: 22a0303184@cjlu.edu.cn; cao@sit.ac.jp

variation (TTV) regularization was proposed. It employs $l_{1/2}$ regularization and TTV regularization to suppress dynamic backgrounds and extract foreground information. However, this method exhibits limited precision in background separation¹⁵.

A tensor-based model has been proven to be effective in extracting low rank information from high-order data¹⁶. In the same year, a Tucker-based model named RTCUR was proposed. To solve this large-scale nonconvex problem, CUR decomposition was employed to reduce the computational complexity¹⁷. However, its number of parameters scales exponentially with the tensor order. A multi-mode outlier-robust tensor ring decomposition(ORTRD) method was proposed. It demonstrates a strong low rank representation capability based on high-order data¹⁸. Tensor nuclear norm(TNN) is a conventional method to solve the low rank model, the FC-TNN method provides new inspiration to solve the tensor-based RCPA model. It incorporates Chebyshev polynomial approximation(CPA) method into the alternating direction method of multipliers (ADMM) algorithm, and results show the efficiency¹⁹.

Supervised learning is also one of the solutions for MOD research. Convolutional Neural Networks (CNNs), You Only Look Once (YOLO), and Long Short-Term Memory (LSTM)²⁰⁻²³ are well-known deep learning models and semi-supervised learning has demonstrated strong potential in MOD research²⁴, these algorithms necessitate the establishment of large-scale databases and their annotation. Furthermore, the training process demands considerable time, computational and electrical resources.

Drawing inspiration from the aforementioned works, we develop a new MOD model, with the specific contributions as follows:

- This paper proposes a novel tensor ring low rank decomposition-based method for moving objects detection, which enhances the ability to estimate the low rank information (static backgrounds).
- The proposed method employs $l_{1/2}$ regularization and tensor total variation model to capture dynamic foreground information. The TTV model is used to ensure the smoothness of moving objects, and the $l_{1/2}$ regularization model is utilized to separate moving objects and dynamic background. To solve the proposed minimization model, we employ the augmented Lagrange multiplier (ALM) method in conjunction with the alternating direction method of multipliers (ADMM). The model is decomposed into several subproblems, each of which can be efficiently solved to iteratively approach the optimal solution.
- We conducted a series of experiments under various real-world scenarios to evaluate the performance of the proposed method. The results verify its superiority over existing state-of-the-art (SOTA) approaches. Furthermore, the proposed method demonstrates better suitability for high-dimensional video data and exhibits increased robustness against different types of noise.

Tensor ring

Notation description

In this paper, handwritten letters are used to represent an N -order tensor, such as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, where I_n denotes the dimension of the tensor along mode n , for $n = 1, 2, \dots, N$. The nuclear norm $\|X\|_*$ represents the sum of the singular values of the matrix X . Uppercase letters (e.g., A, B) are used to indicate matrices, while lowercase letters (e.g., x, y) are used to represent vectors.

Tensor ring low rank decomposition

Tensor ring decomposition decomposes a high-dimensional N -order tensor into a series of low-dimensional tensors. Figure 1 illustrates that an N -order tensor can be represented by a sequence of third-order tensors, where the edges representing the tensor dimensions. Each mode is indicated by the edge numbers, which determine the multilinear product between two tensors, also known as tensor contraction.

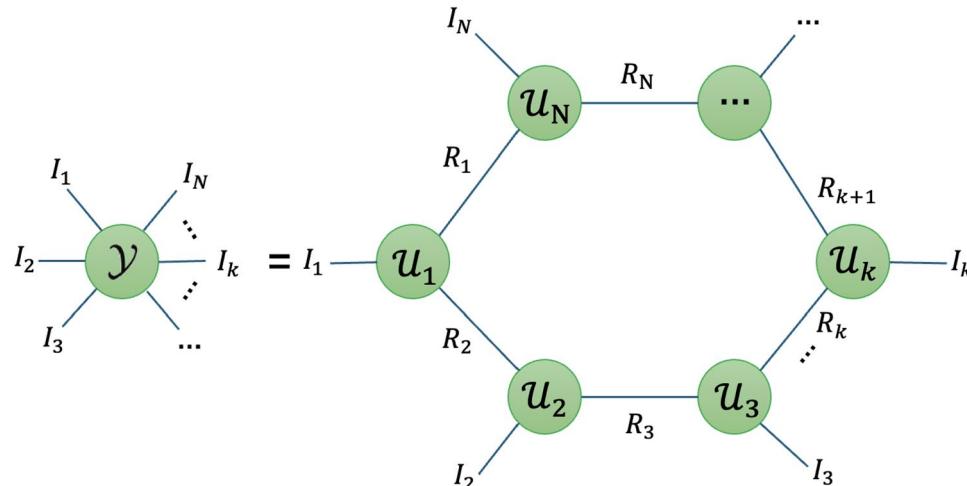


Fig. 1. Tensor ring decomposition diagram.

With (I_1, I_2, \dots, I_N) representing the ranks of the tensor ring, the decomposition of the tensor ring results in a series of third-order tensors, denoted as $\mathcal{U}_{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_{n+1}}$, where $R_1 = R_{N+1}$. The elements of tensor \mathcal{Y} can be expressed using the following equation: $\mathcal{Y}(i_1, i_2, \dots, i_N) = \text{Trace}((U_{i_1}^{(1)} U_{i_2}^{(2)} \cdots U_{i_N}^{(N)})^T)$, $U_{(i_n)}^{(n)} \in \mathbb{R}^{R_n \times R_{n+1}}$ represents the i_n th slice of $\mathcal{U}^{(n)}$. Trace (\mathbf{Y}) represents the matrix trace operation and $Y_{(n)}$ represents the mode- n unfolding of the tensor \mathcal{Y} . For the n -th tensor core $\mathcal{U}^{(n)}$, Yuan et al.²⁵ defined another standard matrix unfolding form as $Y_{(n)} \in \mathbb{R}^{I_n \times I_{n+1} \cdots I_N I_1 I_2 \cdots I_{n-1}}$, where $Y_{(n)} = U_{(2)}^{(n)} (U_{(2)}^{(n)})^T$. In this context, $U_{(i)}^{(n)} \in \mathbb{R}^{I_n \times R_{n-1} R_n}$ represents the mode- i unfolding of the n -th core tensor. $(U_{(2)}^{(n)})^T$ represents the concatenation of all core tensors along mode-2 except the n -th one. For all $n = 1, 2, \dots, N$, the rank relationship between tensor ring and corresponding core is as follows:

$$\text{Rank}(Y_{(n)}) \leq \text{Rank}(U_{(2)}^{(n)}) \quad (1)$$

From the equation, it follows that the rank of the mode- n unfolding of the tensor \mathcal{Y} is bounded by the corresponding core tensor's dimensional rank unfolding. In this paper, we explore the low rank structure of tensors by imposing a low rank constraint on the tensor cores U .

Data tensorization processing

Tensorization is an important pre-processing step that leverages local structures and low rank features of the data. High-order tensors provide more significant image structures through tensor ring decomposition. In this paper, the non-local coupled tensorization (NCT) method²⁶ is employed to transform 3-order video into a high-order tensor, allowing for better utilization of low rank representations while exploring non-local self-similarity and spatial correlations in local regions.

A third-order video data $\mathcal{T} \in \mathbb{R}^{M \times N \times B}$, M and N represent the spatial width and height, and B denotes the frame sequence. To represent the redundancy of the video data, it is divided into several small cubic patches C_i . The size of each is $s \times s \times B$. For each cubic patch, within the local temporal domain, the Euclidean distance equation is used to search for the $k^2 - 1$ nearest neighboring patches in the local window, where each neighboring patch is of size $s \times s \times p$. The $k^2 - 1$ neighboring patches are combined into a new cubic tensor of size $sk \times sk \times p$. In the same spatial location, there are $(2b/p - 1)$ cubes containing temporal information. As shown in Fig. 2, these cubes are combined into T fourth-order tensors of size $sk \times sk \times p \times h$, where $h = \frac{2b}{p} - 1$, and b is parameter to control the distance of searching area for local spatial information.

Overview of MOD algorithms

Tensor total variation model

MOD algorithms are used to detect objects located at salient positions of each frame, and these objects exhibit continuity in the temporal dimension. The focus of this study is on salient target detection, where the continuity of the consecutive frames of the moving objects is preserved in the temporal direction. In this paper, temporal-spatial continuity is utilized to regularize the detection of moving objects. We construct the model using the tensor total variation (TTV) framework. Specifically, we adopt anisotropic tensor total variation (TTV-A) as regularization model:

$$\|\mathcal{T}\|_{TTV-A} = \|\nabla_h \mathcal{T}\|_1 + \|\nabla_v \mathcal{T}\|_1 + \|\nabla_f \mathcal{T}\|_1 \quad (2)$$

where ∇_h , ∇_v and ∇_f represent the horizontal, vertical, and temporal difference operators, respectively, which can be expressed as:

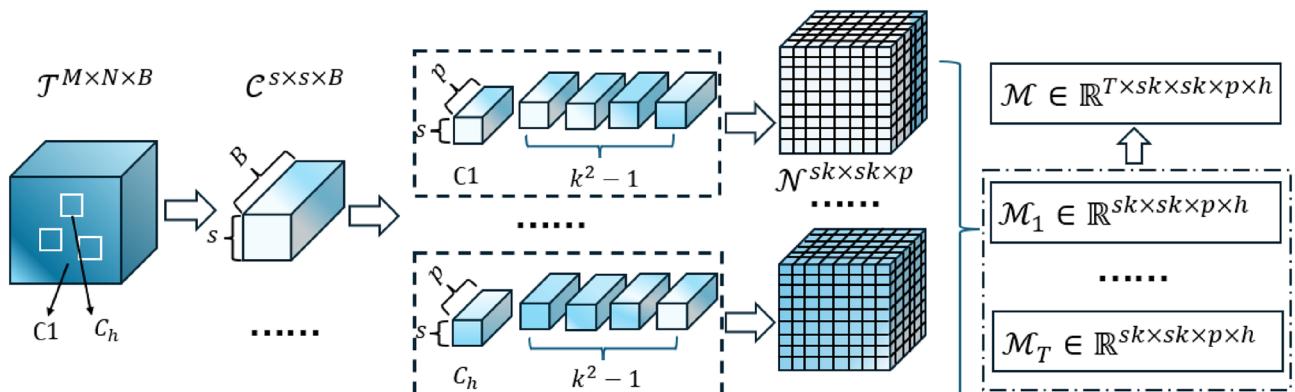


Fig. 2. High-dimensional tensorization method of tensor \mathcal{T} .

$$\begin{aligned}
\nabla_h \mathcal{T} &= \text{vec}(\mathcal{T}_h), \quad \nabla_v \mathcal{T} = \text{vec}(\mathcal{T}_v), \quad \nabla_f \mathcal{T} = \text{vec}(\mathcal{T}_f); \\
\text{where, } \quad \nabla_h \mathcal{T} &= \mathcal{T}(x, y, z) - \mathcal{T}(x+1, y, z), \\
\nabla_v \mathcal{T} &= \mathcal{T}(x, y, z) - \mathcal{T}(x, y+1, z), \\
\nabla_f \mathcal{T} &= \mathcal{T}(x, y, z) - \mathcal{T}(x, y, z+1).
\end{aligned} \tag{3}$$

$$\|\mathcal{T}\|_{TTV-A} = \sum_{x=1}^{p-1} (\mathcal{T}(x, y, z) - \mathcal{T}(x+1, y, z))^2 + \sum_{y=1}^{q-1} (\mathcal{T}(x, y, z) - \mathcal{T}(x, y+1, z))^2 + \sum_{z=1}^{v-1} (\mathcal{T}(x, y, z) - \mathcal{T}(x, y, z+1))^2$$

Tensor robust principal component analysis (TRPCA)

RPCA is an algorithm that decomposes a matrix into two distinct components: a low rank part and a sparse part¹⁰. Typically, an original black-and-white video is represented as a 3-D tensor. In matrix-based RPCA, when applying the RPCA algorithm to process video data, the video is vectorized and then transformed into a matrix, resulting in the loss of some spatial and temporal information from the original data. To address this limitation, TRPCA²⁷ is proposed to decompose the video along the temporal direction into low rank and sparse tensors. For MOD algorithms, the background information is in the low rank tensor, while the moving foreground information is in the sparse tensor. This decomposition is mathematically expressed as follows:

$$\min_{\mathcal{L}} \text{rank}(\mathcal{L}) + \lambda \|\mathcal{T}\|_0 \quad \text{s.t.} \quad \mathcal{L} = \mathcal{L} + \mathcal{T} \tag{4}$$

where $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ represents the input video data, and $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ is the static background part, which exhibits low rank characteristics because the background has strong temporal correlations and resides in a low-dimensional subspace that slowly changes along time direction. λ is a parameter to control the weight of low rank and sparse terms.

$l_{1/2}$ regularization model

In addition to the moving foreground, real-world video scenarios may also include additional salient moving objects that are not the primary targets of MOD detection, such as falling raindrops. Another type of moving object exhibits irregular and discontinuous motion, such as water ripples or swaying trees. These small objects also belong to the sparse tensor and can be effectively analyzed using the $l_{1/2}$ regularization model.

The $l_{1/2}$ regularization is essentially a non-convex regularization model. Zhang et al.²⁸ proposed an iterative half-thresholding algorithm based on a matrix framework to solve the $l_{1/2}$ regularization problem. Inspired by this work, a tensor framework-based algorithm was proposed²⁹. The formula is shown as follows:

$$\min_{\mathcal{A}} \|\mathcal{A} - \mathcal{L}\|_F^2 + \lambda \|\mathcal{A}\|_{l_{1/2}}^{1/2} \quad \text{s.t.} \quad \mathcal{L} = \mathcal{A} + \varepsilon_o \tag{5}$$

where $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ represents the input data, $\mathcal{A} \in \mathbb{R}^{x \times y \times z}$ represents the sparse component of the input data, and ε_o represents the noise.

MOD algorithm based on tensor ring low rank decomposition

This part focuses on describing the development process of the proposed model. The original video data contains low rank tensor(background) and sparse tensor (foreground). Consider any black and white video data $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$, where x, y, z represent the width of the frame, height of the frame and the number of frames. The original video $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ can be divided into background part $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ and dynamic foreground part $\mathcal{T} \in \mathbb{R}^{x \times y \times z}$. Foreground part is usually comprised of dynamic background $\mathcal{S} \in \mathbb{R}^{x \times y \times z}$ and moving object $\mathcal{W} \in \mathbb{R}^{x \times y \times z}$, the real world video always contains dynamic background because of illumination changes, camera jitter, moving shadows, etc. By combining the low rank model and the $l_{1/2}$ norm, dynamic backgrounds can be effectively detected, and it requires adding a regularization term to analyze the dynamic nature of the background to prevent the misclassification of moving background components as target objects. This paper assumes that the dynamic background is temporally sparser than the moving objects in the temporal dimension. The formula can be expressed as follows:

$$\begin{aligned}
&\min_{\mathcal{L}, \mathcal{T}, \mathcal{S}, \mathcal{W}} \text{Rank}(\mathcal{L}) + \lambda_1 \|\mathcal{T}\|_1 + \lambda_2 \|\mathcal{S}\|_1 + \lambda_3 \|\rho(\mathcal{W})\| \\
&\text{s.t.} \quad \mathcal{L} = \mathcal{L} + \mathcal{T} \quad \mathcal{T} = \mathcal{S} + \mathcal{W}
\end{aligned} \tag{6}$$

where $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ represent the original video data, $\mathcal{L} \in \mathbb{R}^{x \times y \times z}$ and $\mathcal{T} \in \mathbb{R}^{x \times y \times z}$ represent the low rank tensor, and the sparse tensor of the input video. \mathcal{S} and \mathcal{W} represent the dynamic background tensor and the moving target foreground tensor. \mathcal{S} is sparser than \mathcal{W} , and exhibit different continuity in the temporal dimension. This paper employs the tensor ring low rank model to extract global low rank information, and the $l_{1/2}$ regularization is applied to enhance sparsity of dynamic background. For the moving foreground, this paper leverages temporal-spatial smoothness. To ensure the foreground data smoother, the $TTV - A$ regularization is combined with the proposed model:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{T}, \mathcal{S}, \mathcal{W}} & \sum_{n=1}^N \sum_{i=1}^3 \|\mathcal{U}_{(i)}^{(n)}\|_* + \lambda_1 \|\mathcal{T}\|_{l_{1/2}}^{1/2} + \lambda_2 \|\mathcal{S}\|_{l_{1/2}}^{1/2} + \lambda_3 \|\mathcal{W}\|_{TTV-A} \\ \text{s.t. } & \mathcal{Z} = \mathcal{L} + \mathcal{T} \quad \mathcal{T} = \mathcal{S} + \mathcal{W} \end{aligned} \quad (7)$$

where $\mathcal{U}_{(i)}^{(n)}$ represents the i -mode unfolded matrix of the n -th core tensor after tensorization of \mathcal{L} . λ_1 , λ_2 , and λ_3 are balancing weight parameters. The $TTV - A$ term denotes the TTV norm of the foreground \mathcal{W} .

In this paper, the ALM method is used to solve the optimization problem in model (6). The augmented Lagrangian function is as follows:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{T}, \mathcal{S}, \mathcal{W}, \Lambda_1, \Lambda_2} & \sum_{n=1}^N \sum_{i=1}^3 \|\mathcal{U}_{(i)}^{(n)}\|_* + \lambda_1 \|\mathcal{T}\|_{l_{1/2}}^{1/2} + \lambda_2 \|\mathcal{S}\|_{l_{1/2}}^{1/2} + \lambda_3 \|\mathcal{W}\|_{TTV-A} \\ & + \frac{\beta}{2} \|\mathcal{Z} - \mathcal{L} - \mathcal{T}\|_F^2 + \langle \Lambda_1, \mathcal{Z} - \mathcal{L} - \mathcal{T} \rangle \\ & + \frac{\beta}{2} \|\mathcal{T} - \mathcal{S} - \mathcal{W}\|_F^2 + \langle \Lambda_2, \mathcal{T} - \mathcal{S} - \mathcal{W} \rangle \end{aligned} \quad (8)$$

where $\Lambda_1 \in \mathbb{R}^{x \times y \times z}$ and $\Lambda_2 \in \mathbb{R}^{x \times y \times z}$ are Lagrange multipliers, β is the penalty parameter, $\langle X, Y \rangle$ represents the inner product between matrices X and Y , and $\|\cdot\|$ denotes the Frobenius norm of the tensor. In this paper, we decompose Eq. (7) into subproblems for solution. Subproblem 1:

$$\begin{aligned} \min_{\mathcal{L}} & \sum_{n=1}^N \sum_{i=1}^3 \|\mathcal{U}_{(i)}^{(n)}\|_* + \frac{\beta}{2} \|\mathcal{Z} - \mathcal{L} - \mathcal{T}\|_F^2 + \langle \Lambda_1, \mathcal{Z} - \mathcal{L} - \mathcal{T} \rangle \\ & = \min_{\mathcal{L}} \sum_{n=1}^N \sum_{i=1}^3 \|\mathcal{U}_{(i)}^{(n)}\|_* + \frac{\beta}{2} \left\| \mathcal{L} - \left(\mathcal{Z} - \mathcal{T} + \frac{\Lambda_1}{\beta} \right) \right\|_F^2 \end{aligned} \quad (9)$$

Let $\mathcal{Z} - \mathcal{T} + \frac{\Lambda_1}{\beta} = \mathcal{M}$; then Equation (9) can be transformed into:

$$\min_{\mathcal{L}_{[n]}} \frac{1}{\beta} \sum_{n=1}^N \sum_{i=1}^3 \|\mathcal{U}_{(i)}^{(n)}\|_* + \frac{1}{2} \|\mathcal{L}_{[n]} - \mathcal{M}_{[n]}\|_F^2 \quad (10)$$

where $\mathcal{M}_{[n]}$ denotes the matrixization of tensor \mathcal{M} along the n mode, and $\mathcal{U}^{(n)}$ can be obtained by the following equation²¹:

$$\mathcal{U}^{(n)} = \text{fold} \left(\frac{\sum_{i=1}^3 \left(\beta U_{(2)}^{(n,i)} + \Lambda_1^{(n,i)} \right) + T_{(n)} U_{(2)}^{(\neq n)}}{U_{(2)}^{(\neq n)T} U_{(2)}^{(\neq n)} + 3E} \right) \quad (11)$$

where E represents the identity matrix. Equation (10) can be solved using the SiLRTC algorithm³⁰. Subproblem 2 can be solved using half-quadratic minimization combined with a shrinkage operator:

$$\begin{aligned} \min_{\mathcal{T}} & \lambda_1 \|\mathcal{T}\|_{l_{1/2}}^{1/2} + \frac{\beta}{2} \|\mathcal{Z} - \mathcal{L} - \mathcal{T}\|_F^2 + \langle \Lambda_1, \mathcal{Z} - \mathcal{L} - \mathcal{T} \rangle \\ & + \frac{\beta}{2} \|\mathcal{T} - \mathcal{S} - \mathcal{W}\|_F^2 + \langle \Lambda_2, \mathcal{T} - \mathcal{S} - \mathcal{W} \rangle \\ \mathcal{T} & = H_{\frac{\lambda_1}{\beta^k}} \left[\frac{\mathcal{Z} - \mathcal{L} + \mathcal{S} + \mathcal{W}}{2} + \frac{\Lambda_1 - \Lambda_2}{2\beta^k} \right] \end{aligned} \quad (12)$$

where $H[\cdot]$ represents the half-thresholding shrinkage operator³¹, and λ_1 denotes the regularization parameter. Equation (12) can be solved using the tensor-based half-quadratic alternate minimization algorithm. Similarly, subproblem 3 can also be solved using this method:

$$\begin{aligned} \min_{\mathcal{S}} & \lambda_2 \|\mathcal{S}\|_{l_{1/2}}^{1/2} + \frac{\beta}{2} \|\mathcal{T} - \mathcal{S} - \mathcal{W}\|_F^2 + \langle \Lambda_2, \mathcal{T} - \mathcal{S} - \mathcal{W} \rangle \\ \mathcal{S} & = H_{\frac{\lambda_2}{\beta^k}} \left[\mathcal{T} - \mathcal{W} + \frac{\Lambda_2}{\beta^k} \right] \end{aligned} \quad (13)$$

The equation for subproblem 4 is:

$$\begin{aligned} & \min_{\mathcal{W}} \lambda_3 \|\mathcal{W}\|_{TTV-A} + \frac{\beta}{2} \|\mathcal{T} - \mathcal{S} - \mathcal{W}\|_F^2 + \langle \Lambda_2, \mathcal{T} - \mathcal{S} - \mathcal{W} \rangle \\ &= \min_{\mathcal{W}} \lambda_3 \|\mathcal{W}\|_{TTV-A} + \frac{\beta}{2} \left\| \mathcal{W} - \left(\mathcal{T} - \mathcal{S} + \frac{\Lambda_2}{\beta} \right) \right\|_F^2 \end{aligned} \quad (14)$$

Equation (14) can be solved using the following equation:

$$\mathcal{W} = \text{sth} \left(\mathcal{T} - \mathcal{S} + \frac{\Lambda_2}{\beta}, \frac{1}{\beta} \right) \quad (15)$$

where $\text{sth}()$ represents the soft-thresholding operator, and its formula is as follows:

$$\text{sth}(x, t) = \text{sgn}(x) \max(|x| - t, 0) \quad (16)$$

The multipliers Λ_1 and Λ_2 can be updated using the following equations:

$$\begin{aligned} \Lambda_1 &= \Lambda_1 + \beta(Z - L - T) \\ \Lambda_2 &= \Lambda_2 + \beta(T - S - W) \end{aligned} \quad (17)$$

The proposed algorithm:

Input: Original video data $\mathbf{Z} \in \mathbb{R}^{M \times N \times B}$, $\lambda_1, \lambda_2, \lambda_3 > 0$, $\beta_{\max} > 0$

Initialization: $\mathbf{L}^0, \mathbf{T}^0, \mathbf{S}^0, \mathbf{W}^0, \Lambda_1^0, \Lambda_2^0 = 0$, $\beta^0 = 10^{-5}$

Output: Background \mathbf{L}^n and moving target \mathbf{W}^n

While the convergence condition is not met:

- Update \mathbf{L} via Equation (10)
- Update \mathbf{T} via Equation (12)
- Update \mathbf{S} via Equation (13)
- Update \mathbf{W} via Equation (15)
- Update multipliers Λ_1, Λ_2 via Equation (17)
- Check if $\|\mathbf{Z} - \mathbf{L} - \mathbf{T}\|^2 \leq \text{threshold}$

End

Experiment and result analysis

The MOD detection experiment is conducted using MATLAB 2024 software. The test equipment is a laptop equipped with an Intel Core i7 processor, a 2.2 GHz CPU, and 16 GB of 1600 MHz DDR3 memory. The parameter $\lambda = 1/\sqrt{\max(M, N) * B}$ is used to separate the low rank and sparse parts of the input tensor. In these experiment sequences, $\lambda_1 = 0.2/\sqrt{\max(M, N) * B}$, $\lambda_2 = 2/\sqrt{\max(M, N) * B}$, $\lambda_3 = 20/\sqrt{\max(M, N) * B}$ guarantee the desired detection result.

The background and foreground detection results are evaluated using the f_1 score and the f_2 score, and the formula is expressed as follows:

$$f_1 = \frac{2R_1P_1}{R_1 + P_1} \quad f_2 = \frac{2R_2P_2}{R_2 + P_2} \quad (18)$$

where $R_1 = TN/(TN + FP)$, $P_1 = TN/(TN + FN)$, $R_2 = TP/(TP + FN)$, and $P_2 = TP/(TP + FP)$. FP, FN, TP, and TN represent false positives, false negatives, true positives, and true negatives, respectively. Additionally, the parameter f is used to evaluate the algorithm³², and the formula for parameter f is expressed as follows:

$$f = \frac{2f_1f_2}{f_1 + f_2} \quad (19)$$

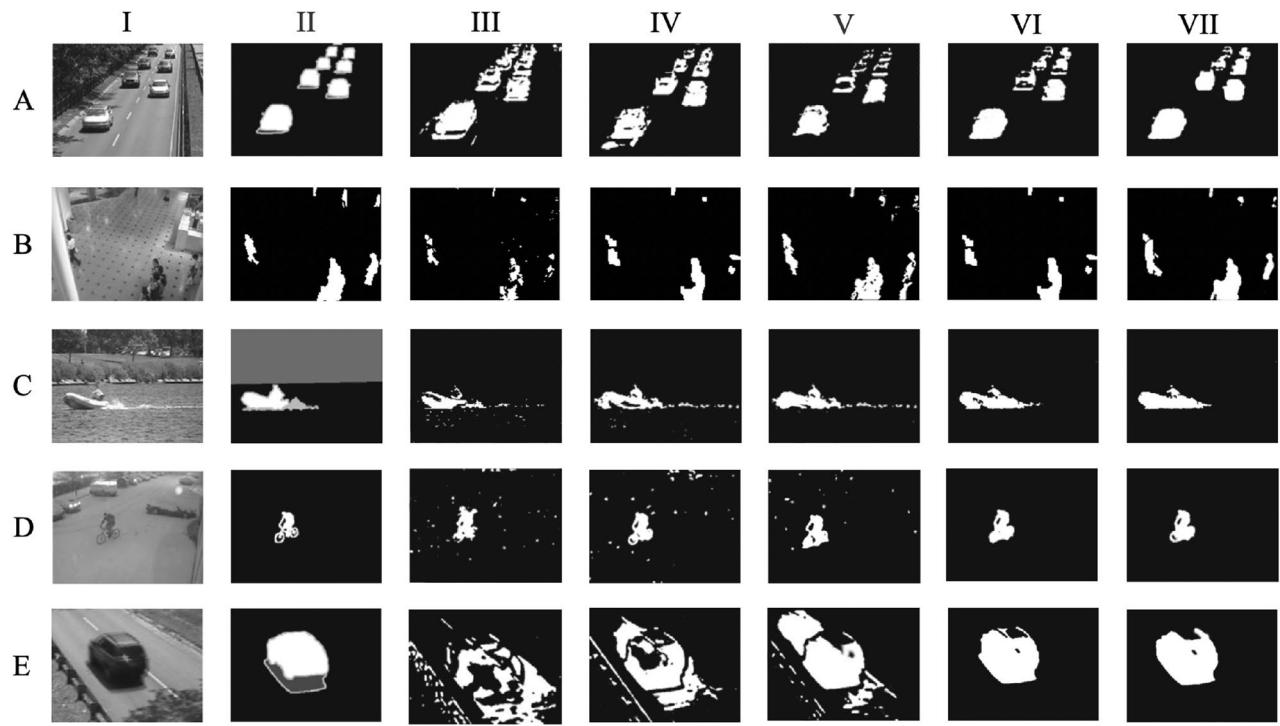


Fig. 3. The visual results of different algorithms: Columns I–VII represents the original data, ground-truth, TVPCRA, SCLR- $L_{1/2}$, DHLRTTV, TQR-SVD, and the proposed method. (A) Highway(static background), (B) Indoor corridor(static background), (C) Boat(dynamic background), (D) Snowfall(dynamic background), (E) Fast moving object(illusion).

| Algorithm | TVP-CRA | | | SCLR- $L_{1/2}$ | | | DNLRITV | | | DHLRTTV | | | TQR-SVD | | | TRLRTTV(<i>Ours</i>) | | |
|-----------|---------|-------|------|-----------------|-------|------|---------|-------|------|---------|-------|-------|---------|-------|------|------------------------|-------|--------------|
| | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f |
| Bird | 0.47 | 0.46 | 0.46 | 0.83 | 0.78 | 0.80 | 0.67 | 0.60 | 0.63 | 0.90 | 0.85 | 0.87 | 0.89 | 0.87 | 0.88 | 0.93 | 0.89 | 0.909 |
| Jumping | 0.68 | 0.42 | 0.52 | 0.74 | 0.67 | 0.59 | 0.65 | 0.55 | 0.52 | 0.96 | 0.85 | 0.90 | 0.95 | 0.83 | 0.89 | 0.97 | 0.90 | 0.933 |
| Ocean | 0.58 | 0.54 | 0.56 | 0.84 | 0.77 | 0.65 | 0.80 | 0.61 | 0.58 | 0.91 | 0.88 | 0.894 | 0.90 | 0.89 | 0.89 | 0.90 | 0.92 | 0.909 |
| Boat | 0.74 | 0.68 | 0.69 | 0.69 | 0.70 | 0.73 | 0.78 | 0.67 | 0.70 | 0.94 | 0.84 | 0.887 | 0.90 | 0.86 | 0.88 | 0.84 | 0.96 | 0.896 |
| Bridge | 0.57 | 0.43 | 0.48 | 0.43 | 0.33 | 0.37 | 0.66 | 0.62 | 0.63 | 0.85 | 0.77 | 0.83 | 0.87 | 0.82 | 0.84 | 0.89 | 0.84 | 0.864 |

Table 1. Comparison of f_1 , f_2 , f results for different algorithms.

By comparing TVPCRA¹³, SCLR- $L_{1/2}$ ¹⁴, DNLRITV³³, DHLRTTV³⁴, and TQR-SVD¹⁵, the results demonstrate the advantages of the proposed algorithm. To evaluate the proposed algorithm, two datasets, including the UCSD background subtraction dataset and Change Detection.net (CD)³⁵background subtraction dataset are tested in the experiment.

Fig. 3 illustrates the performance of different algorithms under multiple scenarios. Scenarios A and B involve static backgrounds, scenarios C and D contain dynamic background, scenario E features a fast-moving object with motion blur. From the results of scenarios A and B, we can conclude that only TQR-SVD and the proposed method can detect moving objects correctly, the TVPCRA method even misclassifies leaf shadows as moving objects. In scenarios C and D, the waves and the snow in the sky are recognized as dynamic background, and result C and D demonstrate the superiority of the proposed method in accurately detecting the exact profile of the moving object. Motion blur in images can be caused by camera jitter, slow shutter speed, or fast-moving objects. From the result of scenario E, the proposed method detects the moving object more accurately compared to other methods.

A comparison of the proposed algorithm with TVPCRA and SCLR- $L_{1/2}$ is shown in Table 1, and the experiment results indicate that the proposed method achieves higher accuracy in background detection than other algorithms, particularly in detecting fast-moving objects in high-jump videos. The proposed algorithm demonstrates superior performance in terms of f_1 , f_2 , and f . The redundant information in the background occupies a large portion of the data in the ocean video. The proposed method achieves better results in background detection and overall performance compared to other algorithms. When compared to the DHLRTTV algorithm, the proposed algorithm achieves a MOD detection performance parameter f of 0.91, which is

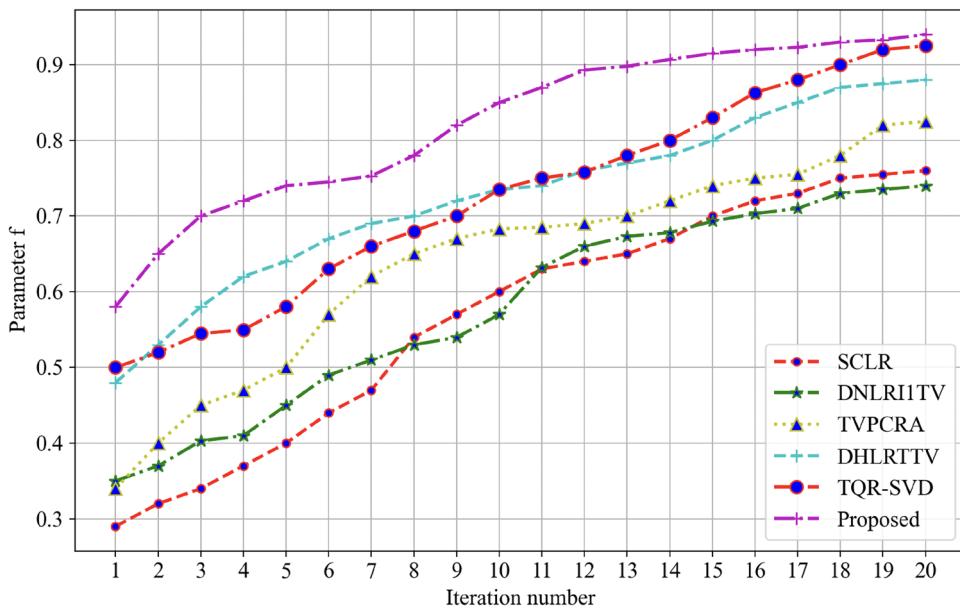


Figure 4. The relationship graph of the f parameter for different algorithms with different iteration counts.

| Algorithm | Gaussian Noise | | | | | | Salt-and-Pepper Noise | | | | | |
|-----------------------------|------------------|-------|-------------|-----------------|-------|-------------|-----------------------|-------|-------------|-----------------|-------|-------------|
| | $\sigma = 0.001$ | | | $\sigma = 0.01$ | | | $\sigma = 0.001$ | | | $\sigma = 0.01$ | | |
| | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f | f_1 | f_2 | f |
| <i>TVPCRA</i> | 0.56 | 0.49 | 0.42 | 0.51 | 0.27 | 0.35 | 0.66 | 0.30 | 0.53 | 0.42 | 0.31 | 0.35 |
| <i>SCLR-L_{1/2}</i> | 0.65 | 0.28 | 0.39 | 0.53 | 0.39 | 0.46 | 0.64 | 0.35 | 0.45 | 0.46 | 0.21 | 0.35 |
| <i>DNLRITV</i> | 0.58 | 0.43 | 0.49 | 0.55 | 0.35 | 0.40 | 0.68 | 0.48 | 0.45 | 0.46 | 0.36 | 0.40 |
| <i>DHLRTTV</i> | 0.78 | 0.72 | 0.74 | 0.76 | 0.75 | 0.75 | 0.88 | 0.80 | 0.83 | 0.81 | 0.76 | 0.79 |
| <i>TQR-SVD</i> | 0.80 | 0.72 | 0.76 | 0.78 | 0.79 | 0.78 | 0.88 | 0.84 | 0.86 | 0.77 | 0.80 | 0.78 |
| <i>TRLRTTV(Our)</i> | 0.88 | 0.78 | 0.82 | 0.85 | 0.79 | 0.82 | 0.91 | 0.83 | 0.87 | 0.87 | 0.80 | 0.83 |

Table 2. Comparison of f_1 , f_2 , f results for different algorithms under different noise levels.

comparable to DHLRTTV. Experimental results demonstrate that the proposed algorithm outperforms other algorithms in terms of background data detection. The tensor ring low rank constraint enhances the ability to capture global low rank background information in video data. By combining the tensor total variation model and the $l_{1/2}$ norm regularization model, the MOD detection performance is significantly improved in dynamic background video data.

Figure 4 illustrates the results based on highway video data. This experiment compares the comprehensive parameter f for moving object detection, calculated over the first 20 iterations of different algorithms. As shown, our algorithm achieves the highest accuracy during initial iterations and maintains a performance advantage across subsequent iterations.

To test the robustness of the proposed algorithm, Gaussian noise and salt-and-pepper noise with variances of 0.001 and 0.01 were introduced into the dataset. As shown in Table 2, the proposed algorithm can effectively detect moving objects from noisy data, demonstrating that the tensor low rank model has strong noise separation ability. The proposed algorithm also effectively detect the foreground in noisy data. The results confirm that the method is robust against different types of noise.

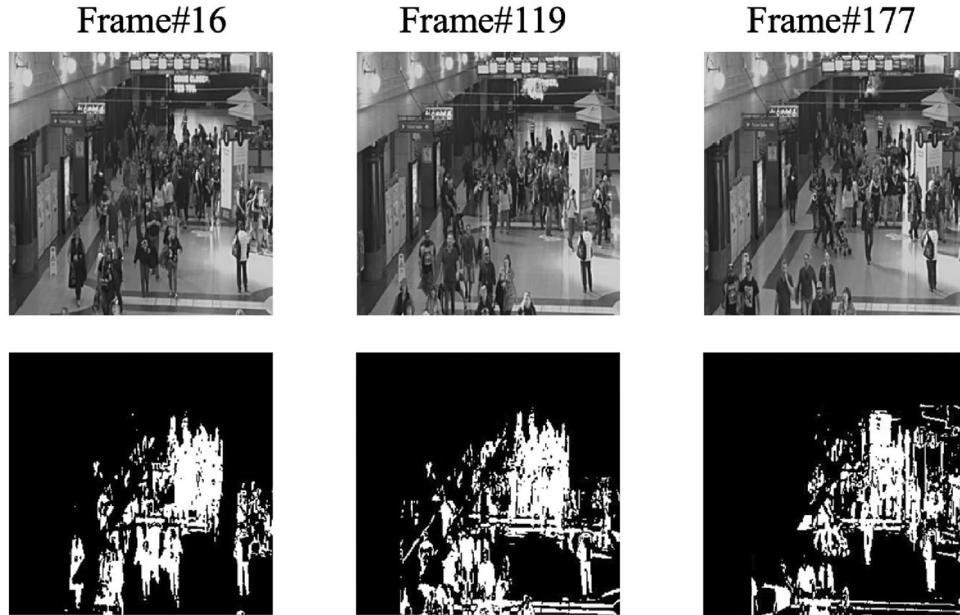
In the next experiment, we evaluated the performance of the proposed method against several deep learning methods using the parameter f . From Table 3, we can conclude that the proposed method performs well in ocean data which is simple background and contains slow-moving objects, and even surpasses Triple-CNN³⁶ and MsEDNET³⁷. However, when the foreground is relatively complex and the moving objects move with high speed, the detection accuracy decreases. The proposed method struggles to accurately detect fast-moving objects because it assumes that moving objects are sparse and smooth in the spatial dimension, where as in reality, they are not continuous in the spatial dimension.

In the final experiment, we evaluated the computational time of various methods on 4-dimensional video data from the MOTChallenge dataset. The proposed method and TQR-SVD method are sensitive to the rank parameter. During the experiment, we choose a small rank while ensuring relatively similar accuracy. As shown in the results presented in Table 4, the proposed method achieved the shortest computational time among all

| Algorithm | Ocean | Bridge | Pedestrian | Highway |
|--------------------------|-------|--------|------------|---------|
| Triple CNN ³⁶ | 0.89 | 0.84 | 0.79 | 0.86 |
| MsEDNET ³⁷ | 0.91 | 0.88 | 0.85 | 0.89 |
| MOD-CVS ³⁸ | 0.93 | 0.92 | 0.90 | 0.92 |
| TRLRTTV(Ours) | 0.91 | 0.87 | 0.83 | 0.87 |

Table 3. Performance comparison of different deep learning algorithms.

| Data Name | Size | TVPCRA | SCLR- $L_{1/2}$ | DNLRITV | DHLRTTV | TQR-SVD | TRLRTTV(Ours) |
|-----------|------------------|---------|-----------------|---------|---------|---------|---------------|
| MOT20-01 | 1920*1080*429*3 | 352.4s | 546.6s | 97.4s | 65.5s | 59.5s | 44.8s |
| MOT20-02 | 1920*1080*2782*3 | 2144.1s | 1251.4s | 193.5s | 109.4s | 98.4s | 86.4s |
| MOT20-03 | 1173*880*735*3 | 369.5s | 395.4s | 94.0s | 74.9s | 70.6s | 54.4s |
| MOT20-05 | 1654*1080*1211*3 | 425.6s | 341.1s | 92.5s | 68.4s | 64.9s | 55.1s |

Table 4. Execution time comparison of different methods.**Fig. 5.** The result of moving object detection under crowded road condition.

compared approaches. The majority of the computational cost in our method is attributed to the updates of the core tensor and the tensor total variation (TTV) term. In contrast, the main cost in DHLRTTV arises from the t-SVD and TTV computations, whereas in TQR-SVD it comes from QR decomposition and TTV updates. These findings indicate that the proposed method is more efficient and thus better suited for high-dimensional video data.

Discussion

In this part, we tested the impact of the rank values of the proposed method. We assume that the tensor ring ranks are equal for every core tensor, i.e., $R_1 = R_2 = \dots = R_N$ and $R = (5, 10, 15, 20, 25, 30, 35)$ and the data is the first image of Fig. 3. As we set ranks from 5 to 30 we find that when $R = 5$ the parameter f is 0.52, and the best performance of rank value is 25, and then with the increase of rank, the performance does not increase, but it affects the efficiency of computing.

In another experiment, we tested the limitation of proposed method. MOTChallenge dataset is used to evaluate the performance of proposed method with multiple moving objections. The video is a hallway is in the mall, filled with pedestrians, and contains both sunshine and artificial lights, which further increases the difficulty of detection. From Fig. 5, we can find that most pedestrians can be detected, but the result is inaccurate, and the contours of pedestrians are not coherent. When it occurs to the multiple moving objections scenario and they occupy a portion of the video, the detection accuracy of our method will decrease cause the moving

subjects become less sparse. Our method has limitations on the number of moving targets and the proportion they occupy in the video.

Conclusions

The proposed algorithm combines the tensor ring low rank model and the total variation model to extract background information and dynamic foreground information from video data. The $l_{1/2}$ -norm regularization model is applied to separate the dynamic background, which is sparse and temporally non-continuous. Experimental results demonstrate that the proposed algorithm outperforms other methods, such as *TPCRA* and *DHLRTTV* in both background detection and moving target detection. Moreover, the proposed method demonstrates greater robustness against Gaussian and salt-and-pepper noise interference compared to other algorithms.

Data availability

The datasets analyzed during the current study are publicly available from the following sources: UCSD Background Subtraction Dataset: Available at http://www.svcl.ucsd.edu/projects/background_subtraction/. changedetection.net (CDnet) Dataset: Available at <https://doi.org/10.1109/CVPRW.2014.126>. MOTChallenge Dataset: Available at <https://motchallenge.net>. These datasets were utilized to evaluate the performance of the proposed algorithm for background subtraction and moving object detection. Additional experimental data generated during this study are available from the corresponding author upon reasonable request.

Received: 13 March 2025; Accepted: 29 August 2025

Published online: 26 September 2025

References

- Yuan, J. et al. Independent moving object detection based on a vehicle mounted binocular camera. *IEEE Sens. J.* **21**, 11522–11531. <https://doi.org/10.1109/JSEN.2020.3025613> (2020).
- Amato, A., Huerta, I., Mozerov, M. G., Roca, F. X. & Gonzalez, J. Moving cast shadows detection methods for video surveillance applications. In *Wide Area Surveillance: Real-time Motion Detection Systems*, 23–47, <https://doi.org/10.1007/978-3-030-2122-3> (Springer, 2012).
- Yin, Q. et al. Detecting and tracking small and dense moving objects in satellite videos: A benchmark. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–18. <https://doi.org/10.1109/TGRS.2021.3130436> (2021).
- Chen, Y., Wang, J., Zhu, B., Tang, M. & Lu, H. Pixelwise deep sequence learning for moving object detection. *IEEE Trans. Circuits Syst. Video Technol.* **29**, 2567–2579. <https://doi.org/10.1109/TCST.2017.2770319> (2017).
- Yan, C. et al. Depth image denoising using nuclear norm and learning graph model. *ACM Trans. Multim. Comput. Commun. Appl. (TOMM)* **16**, 1–17. <https://doi.org/10.1145/3404374> (2020).
- Xie, X., Huang, W., Wang, H. H. & Liu, Z. Image de-noising algorithm based on gaussian mixture model and adaptive threshold modeling. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (ed. Xie, X.) 226–229 (IEEE, 2017). <https://doi.org/10.1109/ICICI.2017.8365343>.
- Yulita, I. N., Paulus, E., Sholahuddin, A. & Novita, D. AdaBoost support vector machine method for human activity recognition. In *2021 International Conference on Artificial Intelligence and Big Data Analytics* (ed. Yulita, I. N.) 1–4 (IEEE, 2021). <https://doi.org/10.1109/ICAIBDA53487.2021.9689769>.
- Dehariya, V. K., Shrivastava, S. K. & Jain, R. Clustering of image data set using k-means and fuzzy k-means algorithms. In *2010 International Conference on Computational Intelligence and Communication Networks* (ed. Dehariya, V. K.) 386–391 (IEEE, 2010). <https://doi.org/10.1109/CICN.2010.80>.
- Mane, S. & Mangale, S. Moving object detection and tracking using convolutional neural networks. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (ed. Mane, S.) 1809–1813 (IEEE, 2018). <https://doi.org/10.1109/ICICCS.2018.8662921>.
- Candès, E. J., Li, X., Ma, Y. & Wright, J. Robust principal component analysis? *J. ACM (JACM)* **58**, 1–37. <https://doi.org/10.1145/1970392.1970395> (2011).
- He, J., Balzano, L. & Szlam, A. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (ed. He, J.) 1568–1575 (IEEE, 2012). <https://doi.org/10.1109/CVPR.2012.6247848>.
- Javed, S., Narayananamurthy, P., Bouwmans, T. & Vaswani, N. Robust pca and robust subspace tracking: A comparative evaluation. In *2018 IEEE Statistical Signal Processing Workshop (SSP)* (ed. Javed, S.) 836–840 (IEEE, 2018). <https://doi.org/10.1109/SSP.2018.8450718>.
- Shijila, B., Tom, A. J. & George, S. N. Moving object detection by low rank approximation and l1-tv regularization on rpca framework. *J. Vis. Commun. Image Represent.* **56**, 188–200. <https://doi.org/10.1016/j.jvcir.2018.09.009> (2018).
- Zhu, L., Hao, Y. & Song, Y. $l_1 - \{1/2\}$ norm and spatial continuity regularized low-rank approximation for moving object detection in dynamic background. *IEEE Signal Process. Lett.* **25**, 15–19. <https://doi.org/10.1109/LSP.2017.2768582> (2017).
- Sabat, N., Raj, S., George, S. N. & Sunil Kumar, T. K. A computationally efficient moving object detection technique using tensor qr decomposition based trPCA framework. *J. Vis. Commun. Image Represent.* **92**, 103785. <https://doi.org/10.1016/j.jvcir.2023.103785> (2023).
- Xu, H., Fang, C., Wang, R., Chen, S. & Zheng, J. Dual-enhanced high-order self-learning tensor singular value decomposition for robust principal component analysis. *IEEE Trans. Artif. Intell.* <https://doi.org/10.1109/TAI.2024.3373388> (2024).
- Cai, H., Chao, Z., Huang, L. & Needell, D. Robust tensor cur decompositions: Rapid low-tucker-rank tensor recovery with sparse corruptions. *SIAM J. Imag. Sci.* **17**, 225–247. <https://doi.org/10.1137/23M1574282> (2024).
- Qiu, Y., Zhou, G., Wang, A., Huang, Z. & Zhao, Q. Towards multi-mode outlier robust tensor ring decomposition. *Proc. AAAI Conf. Artif. Intell.* **38**, 14713–14721. <https://doi.org/10.1609/aaai.v38i13.29389> (2024).
- Zhang, Z., Liu, S., Lin, Z., Xue, J. & Liu, L. A fast correction approach to tensor robust principal component analysis. *Appl. Math. Model.* **128**, 195–219. <https://doi.org/10.1016/j.apm.2024.01.020> (2024).
- Mandal, M. & Vipparthi, S. K. An empirical review of deep learning frameworks for change detection: Model design, experimental frameworks, challenges and research needs. *IEEE Trans. Intell. Transp. Syst.* **23**, 6101–6122. <https://doi.org/10.1109/TITS.2021.3077883> (2021).
- Ismail, A., Elpeltagy, M., S. Zaki, M. & Eldahshan, K. A new deep learning-based methodology for video deepfake detection using xgboost. *Sensors* **21**, 5413. <https://doi.org/10.3390/s21165413> (2021).

22. Xie, T. et al. Moving object detection algorithm based on adaptive clustering. In *2022 3rd China International SAR Symposium (CISS)*, 1–5, <https://doi.org/10.1109/CISS57580.2022.9971223> (IEEE, 2022).
23. Benzer, R. & Yildiz, M. C. Yolo approach in digital object definition in military systems. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELF)* (ed. Benzer, R.) 35–37 (IEEE, 2018). <https://doi.org/10.1109/CISS57580.2022.9971223>.
24. Giraldo, J. H., Javed, S., Sultana, M., Jung, S. K. & Bouwmans, T. The emerging field of graph signal processing for moving object segmentation. In *International Workshop on Frontiers of Computer Vision* (ed. Giraldo, J. H.) 31–45 (Springer, 2021). https://doi.org/10.1007/978-3-030-81638-4_3.
25. Yuan, L., Li, C., Cao, J. & Zhao, Q. Rank minimization on tensor ring: An efficient approach for tensor decomposition and completion. *Mach. Learn.* **109**, 603–622, <https://doi.org/10.1007/s10994-019-05846-7> (2020).
26. Xuegang, L., Lv, J. & Wang, J. Hyperspectral image restoration via auto-weighted nonlocal tensor ring rank minimization. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5. <https://doi.org/10.1109/LGRS.2022.3199820> (2022).
27. Liu, Y., Chen, L. & Zhu, C. Improved robust tensor principal component analysis via low-rank core matrix. *IEEE J. Select. Top. Signal Process.* **12**, 1378–1389. <https://doi.org/10.1109/JSTSP.2018.2873142> (2018).
28. Xu, Z., Chang, X., Xu, F. & Zhang, H. $\ell_{\{1/2\}}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw. Learn. Syst.* **23**, 1013–1027. <https://doi.org/10.1109/TNNLS.2012.2197412> (2012).
29. Liu, Q. et al. A fast and accurate matrix completion method based on qr decomposition and $\ell_{\{2, 1\}}$ -norm minimization. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 803–817. <https://doi.org/10.1109/TNNLS.2018.2851957> (2018).
30. Chen, Q. & Cao, J. Low tensor-train rank with total variation for magnetic resonance imaging reconstruction. *SCIENCE CHINA Technol. Sci.* **64**, 1854–1862. <https://doi.org/10.1007/s11431-020-1851-5> (2021).
31. Cai, J.-F., Candès, E. J. & Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20**, 1956–1982. <https://doi.org/10.1137/080738970> (2010).
32. Giveki, D., Montazer, G. A. & SoltanShahri, M. A. Atanassov's intuitionistic fuzzy histon for robust moving object detection. *Int. J. Approximate Reasoning* **91**, 80–95. <https://doi.org/10.1016/j.ijar.2017.08.014> (2017).
33. Shijila, B., Tom, A. J. & George, S. N. Simultaneous denoising and moving object detection using low rank approximation. *Futur. Gener. Comput. Syst.* **90**, 198–210. <https://doi.org/10.1016/j.future.2018.07.065> (2019).
34. Tom, A. J. & George, S. N. A three-way optimization technique for noise robust moving object detection using tensor low-rank approximation, $\ell_{1/2}$, and ttv regularizations. *IEEE Trans. Cybern.* **51**, 1004–1014. <https://doi.org/10.1109/TCYB.2019.2921827> (2019).
35. Goyette, N., Jodoin, P.-M., Porikli, F., Konrad, J. & Ishwar, P. Changelogdetecion.net: A new change detection benchmark dataset. In *2012 IEEE computer society conference on computer vision and pattern recognition workshops*, 1–8, <https://doi.org/10.1109/CVPRW.2012.6238919> (IEEE, 2012).
36. Nguyen, T. P., Pham, C. C., Ha, S.V.-U. & Jeon, J. W. Change detection by training a triplet network for motion feature extraction. *IEEE Trans. Circuits Syst. Video Technol.* **29**, 433–446. <https://doi.org/10.1109/TCSVT.2018.2795657> (2018).
37. Patil, P. W., Murala, S., Dhall, A. & Chaudhary, S. Msednet: Multi-scale deep saliency learning for moving object detection. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (ed. Patil, P. W.) 1670–1675 (IEEE, 2018). <https://doi.org/10.1109/SMC.2018.00289>.
38. Sahoo, P. K. et al. An improved vgg-19 network induced enhanced feature pooling for precise moving object detection in complex video scenes. *Ieee Access* <https://doi.org/10.1109/ACCESS.2024.3381612> (2024).

Acknowledgements

Science and Technology Program of Zhejiang Province, China under Grant No.2022C01016.

Author contributions

Q.C. conceived the initial idea for the study. R.C. and X.L. designed the TRLRTTV model and its mathematical formulation. Q.C. and R.C. implemented the algorithm and optimized its computational efficiency. R.C. and J.C. designed and conducted the experiments on MOD benchmark datasets. R.C. and X.L. analyzed the experimental results and evaluated the performance against state-of-the-art methods. Q.C. and C.C. contributed to parameter tuning and stability testing. Z.S. and X.L. contributed to the discussion and interpretation of the results. R.C. and J.C. wrote the manuscript. All authors reviewed and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Q.C. or J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025