Moving object detection in dynamic background

Xiaoyun Ding

1 the School of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China

*Corresponding author’s e-mail: Dxy@sdust.edu.cn

Abstract. The application of RPCA model in moving object detection can accurately extract the moving foreground, but the effect of the model is not ideal under complex dynamic background conditions. Based on this, this paper proposes an improved RPCA model based on rank regularization and 3D-TV. The improved model uses regularization term to describe the low rank of video background, 3D-TV to constrain the spatiotemporal continuity of moving objects, and F-norm to eliminate the dynamic interference in video background. The experimental results show that the improved model proposed in this paper can effectively deal with the complex dynamic background and obtain a complete foreground object.

1. Introduction

Robust Principle Component Analysis (RPCA) model was proposed by Candes et al. [1] in 2009. It has been widely applied in moving target detection and has become one of the most important methods of moving target detection. The RPCA model decomposes the video scene matrix into low-rank background matrix and sparse moving foreground target matrix to realize moving target detection. In this model, the low-rank constraint on the video background is realized by using the minimum rank function, and the sparsity of the foreground target is described by using the $l_0$ norm. The specific formula is as follows:

$$\min_{L,S} \text{rank}(L) + \lambda \|S\|_0 \quad s.t. \quad X = L + S$$  \hspace{1cm} (1)

However, the rank function and $l_0$ norm are non-convex and non-smooth, so it is NP hard to solve the model. Considering that kernel norm and $l_1$ norm are convex hull of rank function and $l_0$ norm respectively, Wright et al.[2] proposed to use kernel norm and $l_1$ norm to replace rank function and $l_0$ norm approximately, and theoretically proved that RPCA can accurately recover low rank matrix as long as noise is sparse enough. The model is as follows:

$$\min_{L,S} \text{rank}(L) + \|S\|_1 \quad s.t. \quad X = L + S$$  \hspace{1cm} (2)

When the background in the video is static, RPCA can extract foreground target well and realize accurate foreground target detection.

However, the kernel norm only adds the singular values and treats each singular value equally, which leads to inaccurate description of the background. Therefore, scholars have made different improvements on the kernel norm for the background part, such as weighted kernel norm[3], sum of partial singular values[4] and so on. Secondly, for the foreground part of the video, the moving foreground has the characteristics of continuity in time and space, some scholars consider extending the
RPCA model by introducing spatial continuity constraints into the foreground\cite{5}–\cite{7}. In particular, three dimensional total variation (3D-TV) regularization term used in \cite{7} can make full use of the spatiotemporal continuity of the foreground when processing video sequences with dynamic background. Experiments show that the improved RPCA model with 3D-TV regularization term has better constraints on the foreground and can extract more complete foreground.

In this paper, the video matrix is divided into low rank background, sparse foreground and dense dynamic background noise by using the three-term decomposition technique of stable RPCA decomposition. The strong low rank of the rank – 1 regularization term is used to characterize the video background. Considering the temporal and spatial continuity of the foreground object, the 3D-TV regularization term is used to constrain the foreground object. Finally, the \( F \) norm is used to remove the interference of dynamic background noise. The experimental results show that the extraction effect of this model is better than that of similar detection methods, and it can improve the accuracy of moving object detection in complex dynamic background.

2. Related work

This section mainly introduces the rank – 1 regularization term and 3D-TV regularization term.

2.1. rank – 1 regularization

In the surveillance video, the static background of the video changes little, and the background in each frame hardly changes. The low rank of the video background matrix can be regarded as 1. Therefore, in the singular value of the matrix, the largest singular value contains most of the information in the background. The rank – 1 regularization term preserves the main information by minimizing the singular value of the noise except the target rank. The rank – 1 regularization term preserves the first singular value of the matrix and minimizes other singular values corresponding to the background noise. That is to say, the information of the video static background is preserved and the dynamic noise of the video background is minimized. The specific mathematical formula of rank – 1 regularization term is as follows:

\[
\|L\|_{r=1} = \sum_{i=2}^{r} \sigma_i(L)
\]

Where \( \sigma_i(L) \) is the ith largest singular value of a matrix.

2.2. 3D-TV regularization

3D-TV regular term has the function of smoothing signal light and has strong inhibition on discontinuous changes in signal. Therefore, 3D-TV is often used to constrain the spatiotemporal continuity of foreground. 3D-TV constraint is applied to 3D tensor \( X \). If each voxel in \( X \) is recorded as \( X(m,n,t) \), the difference operator of voxel in horizontal, vertical and temporal directions is:

\[
X_h = |X(m+1,n,t) - X(m,n,t)|
\]

\[
X_v = |X(m,n+1,t) - X(m,n,t)|
\]

\[
X_t = |X(m,n,t+1) - X(m,n,t)|
\]

For the convenience of calculation, the difference operator is vectorized along the horizontal, vertical and temporal directions

\[
D_h X = \text{vec}(X_h), D_v X = \text{vec}(X_v), D_t X = \text{vec}(X_t)
\]

the mathematical expression of 3D-TV is as follows:

\[
\|X\|_{D-TV} = \|DX\| = \|D_h X\| + \|D_v X\| + \|D_t X\|
\]
3. Model and algorithm

3.1. Proposed our improvement RPCA model
The proposed model decomposes the video sequence into low rank static background \( L \), foreground object \( F \) and dynamic background noise \( E \). rank – 1 regularization term is used to describe static background, 3D-TV regularization term is used to constrain video foreground, and \( F \) norm is used to remove dynamic background noise. To sum up, in this paper, we propose an improved RPCA model with rank – 1 regularization term and 3D-TV . To sum up, in this paper, we propose an improved RPCA model based on rank – 1 regularization term and 3D-TV regularization term:

\[
\min_{L,E,F} \|L\|_{r=1} + \lambda_1 \|E\|_F + \lambda_2 \|F\|_{3D-TV} \quad s.t. \quad D=L+E+F
\]  

(9)

Where \( \lambda_1, \lambda_2 \) is the trade-off parameters.

3.2. Algorithm
In this section, the augmented Lagrange multiplier method is used to solve the model (9). The augmented Lagrange function of the improved model is as follows:

\[
L(L,E,F,Y,\mu,\lambda_1,\lambda_2) = \|L\|_{r=1} + \lambda_1 \|E\|_F + \lambda_2 \|F\|_{3D-TV} + \langle Y, D - L - E - F \rangle + \frac{\mu}{2} \|D - L - E - F\|_F^2
\]

(10)

Where \( \mu \) is penalty parameter and \( Y \) is Lagrange multiplier. Assuming the current iteration times are \( k \), the basic iterative formula of the augmented Lagrangian multiplier method is:

1) \[
L^{k+1} = \arg \min_{L} \|L\|_{r=1} + \frac{\mu_k}{2} \|L - (D - E_k - F_k + \frac{Y_k}{\mu_k})\|_F^2
\]

(11)

The optimal solution of the subproblem can be obtained by the local singular value threshold (PSVT) operator:

\[
P_{1,\varepsilon}[Y'] = \sum_{i=1}^{\varepsilon} \sigma_i \text{diag}(\sigma_i,0,\cdots,0) D_{Y} V_{Y}^T = Y_1 + U_{Y_1} S_{Y_1} [D_{Y_1}] V_{Y_1}^T
\]

(12)

where \( \varepsilon = \mu_k, Y' = D - E_k - F_k + \frac{Y_k}{\mu_k} \), \( Y' \) can be regarded as the sum of two matrices

\[
Y' = Y_1 + Y_2 = U_{Y_1} D_{Y_1} V_{Y_1}^T + U_{Y_2} D_{Y_2} V_{Y_2}^T
\]

(13)

\( U_{Y_1} \) and \( V_{Y_1} \) are the singular vector matrices corresponding to the first singular value, \( U_{Y_2} \) and \( V_{Y_2} \) are the singular vector matrices corresponding to the second to the last singular value, \( D_{Y_1} = \text{diag}(\sigma_1,0,\cdots,0), D_{Y_2} = \text{diag}(\sigma_2,\cdots,0) \), \( \text{S[•]} \) stands for contraction operator.

2) \[
E^{k+1} = \arg \min_{E} \|E\|_F + \frac{\mu_k}{2} \|E - (D - L_{k+1} - F_k + \frac{Y_k}{\mu_k})\|_F^2 = (I + \frac{2\lambda_1}{\mu_k})^{-1} (D - L_{k+1} - F_k + \frac{Y_k}{\mu_k})
\]

(14)

3) \[
F^{k+1} = \arg \min_{F} \|F\|_{3D-TV} + \frac{\mu_k}{2} \|F - (D - L_{k+1} - E_{k+1} + \frac{Y_k}{\mu_k})\|_F^2 = S_{\frac{\lambda_2}{2\mu_k}} [(D - L_{k+1} - E_{k+1} + \frac{Y_k}{\mu_k})]
\]

(15)

4) \[
Y^{k+1} = Y_k + \mu_k (D - L_{k+1} - E_{k+1} - F_{k+1})
\]

(16)

5) \[
\mu_{k+1} = \rho \mu_k
\]

(17)

where \( \rho \) is the step size. The purpose of \( \rho \) is to update the penalty parameterin \( \mu \) each iteration.
4. Experimental results

The video sequences selected in this paper are common complex scenes in daily life. The video sequences are "Skipping", "SnowFall", "Blizard", "Highway" and "PeopleInShade" in turn. In addition, the CD.net data set provides us with the ground truth corresponding to each image in each video sequence, that is, the ground truth, which provides a good standard and basis for our subsequent comparison and analysis of experimental results. The following is the result of visual experiment, as shown in Figure 1.

Comparing the ground situation of columns 3-6 and 2 in Figure 1, it can be seen that the algorithms of RPCA and WNNM-RPCA model will mistakenly detect the dynamic background (such as shaking leaves, falling snowflakes and light shadows) as the foreground, resulting in the extracted foreground containing too much noise; the algorithm of TV-RPCA model can greatly reduce the existence of noise, but the detection effect of the noise covering the detection object is not good, and there will be more holes; and this algorithm can detect the foreground target better, and greatly reduce the existence of noise, and can detect the complete foreground target better.

![Figure 1. Comparison of experimental results of different algorithms for different video sequences](image)

5. Conclusions

In this paper, an improved RPCA model based on rank –1 regular term and 3D-TV regular term is proposed, which can overcome the dynamic interference in the video background and recover the complete foreground target when dealing with the video under natural conditions. Note that as a general principle, for large tables font sizes can be reduced to make the table fit on a page or fit to the width of the text. In comparison with other methods, the proposed method also shows advantages. But in the face of small and fast moving objects, the effect of foreground extraction is not ideal, which is also the direction and focus of our next step.

References

[1] Candès E, Wakin M B, Boyd S P. Enhancing sparsity by reweighted L1 minimization[J]. Journal of Fourier Analysis and Applications, 2008, 14(5-6): 877-905.

[2] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, pp. 1–37, May 2011, doi: 10.1145/

[3] Gu S, Zhang L, Zou W, et al. Weighted Nuclear Norm Minimization with Application to Image Denoising[C]. IEEE Conference Computer Vision and Pattern Recognition, 2014.
[4] Oh T H, Tai Y W, Bazin J C, et al. Partial Sum Minimization of Singular Values in Robust PCA: Algorithm and Applications[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(4):744-758.

[5] Yang J, Sun X, Ye X, et al. Background extraction from video sequences via motion-assisted matrix completion[C]/2014 IEEE International Conference on Image Processing. IEEE, 2014: 2437-2441.

[6] Alexandre E B, Chowdhury A, A. X. Falcão, et al. IFT-SLIC: A General Framework for Superpixel Generation Based on Simple Linear Iterative Clustering and Image Foresting Transform[C]// Graphics, Patterns & Images. IEEE, 2015.

[7] Cao X, Yang L, Guo X. Total variation regularized RPCA for irregularly moving object detection under dynamic background[J]. IEEE Trans. Cybern, 2015, 46 (4): 1014-1027.