Moving Object Extraction in Infrared Video Sequences

Jinli Zhang[1,2](✉), Min Li[1], and Yujie He[1]

[1] Xi'an Research Institute of Hi-Tech,
Hongqin Town, Xi'an, Shaanxi Province, China
jlz_007@sina.com, {clwn,ksy5201314}@163.com
[2] Department of Information Engineering, Engineering University of CAPF,
Sanqiao Town, Xi'an, Shaanxi Province, China

Abstract. In order to extract the moving object in infrared video sequence, this paper presents a scheme based on sparse and low-rank decomposition. By transforming each frame of the infrared video sequence to a column and combine all columns into a new matrix, the problem of extracting moving objects in infrared video sequences is converted to a sparse and low-rank matrix decomposition problem. The resulted nuclear norm and L_1 norm related minimization problem can also be efficiently solved by some recently developed numerical methods. The effectiveness of our proposed scheme is illustrated on different infrared video sequences. The experiments show that, compared to ALM algorithm, our algorithm has distinct advantages in extracting moving object from infrared videos.

Keywords: Infrared video sequence · Moving object · Extraction · L_1 norm · Low-rank · Sparse

1 Introduction

With the prevalence of infrared cameras and infrared sensors, the infrared video plays an increasing important role on human life and production. For example, the infrared videos about the wild animal activity which were obtained from infrared video surveillance equipments have brought great convenience for the wild animal researchers. When the target and background brightness have not a distinct difference in the infrared video sequences or people's some need, it is very necessary to separate the moving object from the backgrounds. How to effectively extract the moving target in an infrared video is a problem worthy of studying.

This paper aims at developing an effective scheme to extract the moving objects in infrared video sequences. Motivated by the regularization models proposed in [1, 2] for other applications, we take a similar regularization approach for moving objects extraction from background.

Foundation Item: The National Natural Science Foundation of China (61102170).

Y.-J. Zhang (Ed.): ICIG 2015, Part II, LNCS 9218, pp. 491–499, 2015.
DOI: 10.1007/978-3-319-21963-9_45

2 Detailed Algorithm

In this section, we present the algorithm in details. First, each frame of the infrared video sequences is reformed to a column, then combining all columns to a new matrix $D \in \mathbb{R}^{m \times n}$. Because the video content is composed of the background and moving objects, the matrix D can be represented as the sum of background and moving objects. If we represent the background component as $A \in \mathbb{R}^{m \times n}$ and represent moving object component as $E \in \mathbb{R}^{m \times n}$, then the matrix D can be expressed as

$$D = A + E. \tag{1}$$

In video sequences, the adjacent frames have most of the same background information, especially in the video with high frame rate. Thus, matrix A has many same columns and it is a low-rank matrix. Because the size of the moving objects in each frame is far less than the size of frame, the number of nonzero is far less than the element number in matrix E. Thus, matrix E is a sparse matrix. Based on these observations, if we can accurately decompose the matrix D into the sum of a low-rank matrix and a sparse matrix, the moving objects can be extracted from background.

2.1 Notation

Before presenting the details of decomposing D into a low-rank matrix A and a sparse matrix E, we first define some notations for the simplicity of discussions. The L_1 norm and the Frobenius norm of a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ are defined by:

$$\|X\|_1 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |x_{i,j}| \text{ and } \|X\|_F = (\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |x_{i,j}|^2)^{1/2}, \tag{2}$$

respectively. Where $x_{i,j}$ is the (i,j)-*th* element of X. Assuming that r is the rank of X, the singular value decomposition of X is then defined by

$$X = U \sum V^T, \quad \sum = diag(\{\sigma_i\}_{1 \le i \le r}). \tag{3}$$

Where U and V are $n_1 \times r$ and $n_2 \times r$ matrices with orthonormal columns respectively. The nuclear norm of X is defined as the sum of singular values, i.e.

$$|X|_* = \sum_{i=1}^{r} |\sigma_i|. \tag{4}$$

The shrinkage operator $S_\tau : \mathbb{R} \to \mathbb{R}$ is defined by

$$S_\tau(x) = \text{sgn}(x)\max(|x| - \tau, 0). \tag{5}$$

Where $\tau \ge 0$. When S_τ is extended to matrices by applying it element-wise.

The singular shrinkage operator $D_\tau(x)$ is defined [3] by

$$D_\tau = US_\tau(\Sigma)V^T. \tag{6}$$

It is noted that $S_\tau(X)$ and $D_\tau(x)$ are the solutions of the following two minimization problems respectively

$$\min_Y \tau\|Y\|_1 + \frac{1}{2}\|Y - X\|_F^2, \ \min_Y \tau\|Y\|_* + \frac{1}{2}\|Y - X\|_F^2. \tag{7}$$

2.2 Sparse and Low-Rank Decomposing

In order to exactly extract the sparse matrix E and low-rank matrix A, we can solve the following minimization problem to estimate A and E:

$$\min_{A,E\in\mathbb{R}^{n_1\times n_2}} rank(A) + \lambda\|E\|_0 \text{ s.t. } D = A + E. \tag{8}$$

Where λ is a suitable regularization parameter. $rank(\cdot)$ denotes the rank for a matrix. $\|\cdot\|_0$ denotes the pseudo-norm that counts the number of non-zeros.

The minimization problem (8) is a non-convex problem. In general, it is very hard to solve. Referring to the approaches in [4, 5], we try to solve the follow minimization to estimate A and E.

$$\min_{A,E\in\mathbb{R}^{m\times n}} \|A\|_* + \lambda\|E\|_1 \text{ s.t. } D = A + E. \tag{9}$$

Where $\|\cdot\|_1$ is the element-wise sum of absolute values for a matrix.

The minimization model (9) above has been proposed in [1, 2] to extract low-dimensional structure from a data matrix. It could be viewed as a replacement of the Principal Component Analysis (PCA) method. The minimization approaches is termed as Principal Component Pursuit (PCP) for solving the problem of background subtraction in video surveillance. In their approach, the observed video matrix (array of image frames) is decomposed into the low-rank matrix structure (static background) and the sparse matrix structure (moving objects).

In our approach, we convert the minimization question (9) to an augmented Lagrange multiplier form:

$$\min_{A,E\in\mathbb{R}^{m\times n}} \|A\|_* + \lambda\|E\|_1 + \frac{1}{2\mu}\|D - A - E\|_F^2. \tag{10}$$

Here, the value of λ is set the same as [1] suggested:

$$\lambda = 1\Big/\sqrt{\max(m, n)}. \tag{11}$$

Where m, n are the number of rows and columns of the matrix D.

In recent years, there are some good methods on how to efficiently solve L_1 norm related minimization problem. One of them is the accelerated proximal gradient (APG) method, which shows a very good performance on solving L_1 norm and nuclear norm related minimization problems (e.g. [6–9]). Another promising approach is the ADMM (alternating directions method of multipliers) which also can efficiently solve such problems (e.g. [10–12]). In our approach, we used the APG method to solve the minimization problem (10).

The general APG method aims at solving the following minimization problem:

$$\min_{X} \quad g(X) + f(X) \tag{12}$$

Where g is a non-smooth function, f is a smooth function. Algorithm 1 describes the specific scheme of APG.

```
Algorithm 1.   APG method
1.  do
2.  $Y_k = X_k + \frac{t_{k-1} - 1}{t_k}(\mathrm{X}_k - \mathrm{X}_{k-1})$ ;
3.  $G_k = Y_k + \frac{1}{L_f}\nabla f(Y_k)$ ;
4.  $X_{k+1} = \arg \min_{X} g(\mathrm{X}) + \frac{L_f}{2}\|X - G_k\|_F^2$ ;
5.  $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{t_k}$, $k = k + 1$ ;
6.  until converged
```

Based on the APG method, the minimization problem (10) can be converted to (12) by setting

$$\begin{cases} X = (A, E) \\ g(X) = \mu\|A\|_* + \lambda\mu\|E\|_1 \\ f(X) = \frac{1}{2}\|D - A - E\|_F^2 \end{cases}. \tag{13}$$

When applying Algorithm 1 to solve the (10), the minimization problem in Step 4 of Algorithm 1 becomes (noticing $L_f = 2$ in our case)

$$\min_{A,E} \ \mu\|A\|_* + \lambda\mu\|E\|_1 + \left\|A - G_k^A\right\|_F^2 + \left\|E - G_k^E\right\|_F^2. \tag{14}$$

Since A and E are separable in the above minimization, their solutions can be obtained separately by applying singular value shrinkage operator on G_k^A and soft shrinkage operator on G_k^E, i.e. $A_{k+1} = D_{\mu/2}(G_k^A)$, $E_{k+1} = S_{\lambda\mu/2}(G_k^E)$.

The detailed algorithm for solving the minimization problem (10) is described in Algorithm 2.

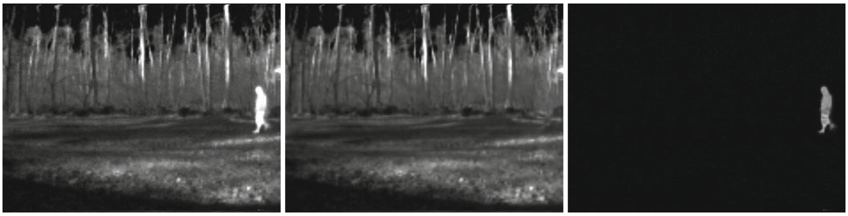
Algorithm 2. APG method for solving the minimization problem (10)

```
Initialization: $E_0 = A_0 = O^{m\times n}, Y_0^A = Y_0^E = \frac{D}{\sigma_1}, k = 0$
                $\mu_0 = \frac{5}{4}\sigma_1$, ($\sigma_1$ is the largest singular value of $D$);
                Set $\varepsilon > 0$, $K > 1$, $\rho > 1$, $\mu_{\max} \gg 1$;
do
    $Y_k^A = A_k + \frac{t_{k-1}-1}{t_k}(A_k - A_{k-1})$, $Y_k^E = E_k + \frac{t_{k-1}-1}{t_k}(E_k - E_{k-1})$ ;
    $G_k^A = Y_k^A - \frac{1}{2}(Y_k^A + Y_k^E - D)$, $G_k^E = Y_k^E - \frac{1}{2}(Y_k^A + Y_k^E - D)$ ;
    $(U, \Sigma, V) = svd(G_k^A)$, $A_{k+1} = US_{\mu k/2}(\Sigma)V^T$, $E_{k+1} = S_{\lambda k/2}(\mathrm{G}_k^E)$ ;
    $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{t_k}$, $\mu_{k+1} = \min(\rho\mu_k, \mu_{\max})$, $k = k+1$ ;
until $\frac{\|D - A_{k+1} - E_{k+1}\|_F}{\|D\|_F} < \varepsilon$ or $k \geq K$ .
```

After the low-rank matrix A and the sparse matrix E are obtained by Algorithm 2, the low-rank matrix A and the sparse matrix E will be reformed to the format of the original infrared video sequences.

3 Experimental Results and Analysis

In this section, we evaluate the performance of the proposed method on three infrared video sequences “irw1”, “irw2” and “plane”. In order to facilitate the evaluation, our algorithm is compared with the inexact augmented Lagrange multipliers (ALM) algorithm [11] for its high efficiency in solving minimization problems. For a fair comparison, in each algorithm, the error tolerance ε is set to 1.0×10^{-7} and the maximal iterations number K is set to 1000. 30 frames of each infrared video sequence were input to two algorithms in experiments. The sizes of each frame of the infrared video “irw1”, “irw2” and “plane” are 240 × 320, 240 × 320 and 200 × 256 respectively. All the experiments are performed on a desktop computer (CPU 2.30 GHz, RAM 3.25 GB) with the MATLAB R2012b software. Figures 1, 3 and 5 show the results of extracted

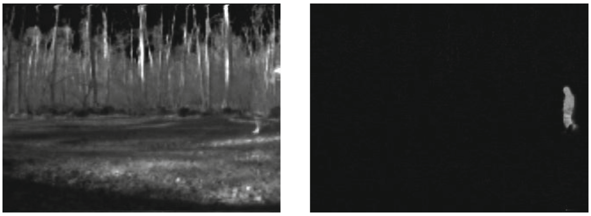


(a) one original frame (b) extracted background from (a) (c) extracted object from (a)

Fig. 1. Extracting result of background and object from “irw1” by Algorithm 2

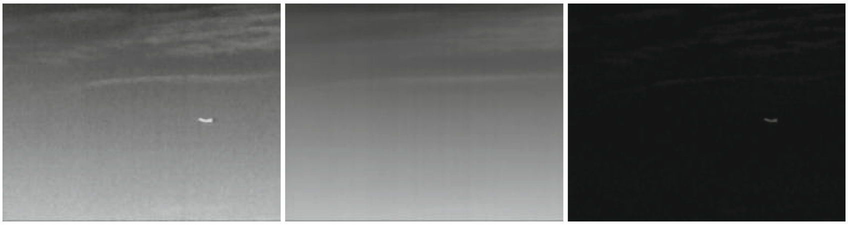
objects and background in three infrared videos by the Algorithm 2. Figures 2, 4 and 6 show the results of extracted objects and background by the ALM algorithm. The performance of two algorithms in terms of the runtime, iteration number and the rank of the extracted low-rank matrix *A* are listed in Table 1.

From above figures, it can be seen that, no matter big or small, quick or slow, the moving objects can be completely extracted by the Algorithm 2. In Fig. 2, one foot of the man had not been extracted to the moving object opponent by ALM algorithm. From Figs. 3 and 4, we can find that, for the small object plane, the extracted plane has clear edge by Algorithm 2 than that by ALM algorithm. From Figs. 5 and 6, it can be seen that, for the slow moving man, partial contour of the man was not extracted to the object opponent by ALM algorithm. As can be seen from Table 1, compared to ALM algorithm, Algorithm 2 has the following distinct advantages: the rank of the recovered



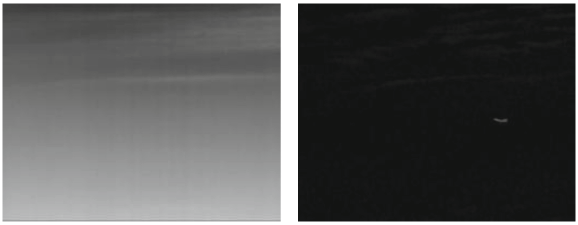
(a) extracted background from Fig.1 (a) (b) extracted object from Fig.1 (a)

Fig. 2. Extracting result of background and object from “irw1” by ALM algorithm



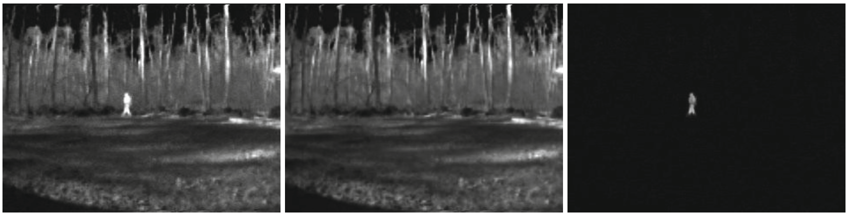
(a) one original frame (b) extracted background from (a) (c) extracted object from (a)

Fig. 3. Extracting result of background and object from “plane” by Algorithm 2



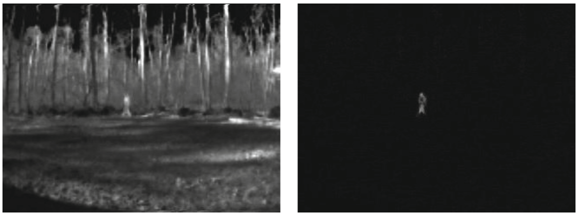
(a) extracted background from Fig.3 (a) (b) extracted object from Fig.3 (a)

Fig. 4. Extracting result of background and object from “plane” by ALM algorithm



(a) one original frame (b) extracted background from (a) (c) extracted object from (a)

Fig. 5. Extracting result of background and object from “irw2” by Algorithm 2

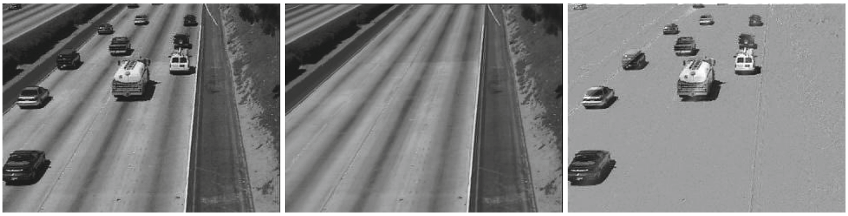


(a) extracted backgroundfrom Fig.5 (a) (b) extracted objectfrom Fig.5 (a)

Fig. 6. Extracting result of background and object from “irw2” by ALM algorithm

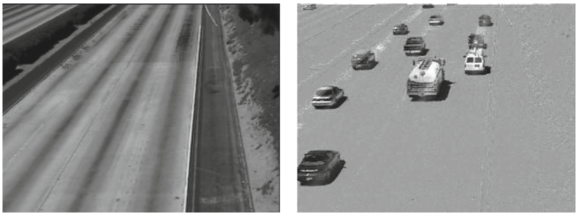
Table 1. Comparison of the results of extracting objects in different infrared video sequence by two algorithms.

Infrared video	Algorithm	Rank (A*)	Iteration number	Time (s)
“irw1”	Algorithm 2	10	20	12.7310
	ALM	15	34	21.6427
“irw2”	Algorithm 2	9	19	12.3445
	ALM	15	33	21.4404
“plane”	Algorithm 2	9	28	13.0447
	ALM	12	36	16.7718



(a) one original frame (b) extracted background from (a) (c) extracted object from (a)

Fig. 7. Extracting result of background and object from optical video “highway” by Algorithm 2



(a) extracted background from Fig.7 (a) (b) extracted object from Fig.7 (a)

Fig. 8. Extracting result of background and object from optical video “highway” by ALM algorithm

background more lower, running time more less and fewer iteration number to reach convergence. These advantages for rapid analysis and process large amounts of infrared video data is important.

In order to verify the validity of the proposed algorithm for optical videos, Figs. 7 and 8 show the results of extracted object and background in an optical video “highway” by Algorithm 2 and ALM algorithm respectively.

From Figs. 7 and 8, it can be seen that, the two algorithms are still able to extract the moving objects in an optical video. The extracted backgrounds by two algorithms have no obvious difference from the visual point of view, but there are more car tracks which belong to the background in Fig. 8(b) than that in Fig. 7(c).

4 Conclusions

In this paper, we presented a scheme to extract the moving objects from infrared video sequence. We convert the problem of extracting the moving object from videos to a sparse and low-rank matrix decomposition problem. The resulting L_1 norm related minimization problem can also be efficiently solved by many recently developed numerical methods. The effectiveness of our proposed algorithm is also validated to other types of video (e.g., optical videos). The experiments show that, compared to ALM algorithm, our algorithm has distinct advantages in extracting moving object from infrared videos and optical videos.

References

1. Candes, E.J., Li, X.D., Yi, M., Wright, J.: Robust principal component analysis? J. ACM **58**(3), 1–37 (2011)
2. Zhou, Z., Li, X., Wright, J., Candes, E.J., Ma, Y.: Stable principal component pursuit. In: IEEE International Symposium on Information Technology (ISIT) (2010)
3. Cai, J.F., Candes, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
4. Chandrasekaran, V., Sangavi, S., Parrilo, P.A., Willsky, A.S.: Sparse and low-rank matrix decompositions. In: IEEE 47th Annual Allerton Conference on Communication, Control, and Computing, pp. 962–967 (2009)
5. Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y.: Robust PCA: exact recovery of corrupted low-rank matrices via convex optimization. In: NIPS 2009, Whistler, BC, Canada (2009)
6. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
7. Shen, Z.W., Toh, K.C., Yun, S.: An accelerated proximal gradient algorithm for frame based image restoration via the balanced approach. SIAM J. Imag. Sci. **4**(2), 573–596 (2011). Technical report
8. Toh, K.C., Yun, S.: An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. Pac. J. Optim. **6**, 615–640 (2010)
9. Lin, Z., Ganesh, A., Wright, J., Wu, L., Chen, M., Ma, Y.: Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. UIUC Technical report UILU-ENG-09-2214 (2009)
10. Yuan, X.M., Yang, J.F.: Sparse and low-rank matrix decomposition via alternating direction methods. Pac. J. Optim. **9**(1), 167–180 (2013)
11. Lin, Z.C., Chen, M.M., Wu, L.Q., Ma, Y., The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. UTUC Technical report UILU-ENG-09-2215 (2010)
12. Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**(1), 57–81 (2011)