FOREGROUND ESTIMATION BASED ON ROBUST LINEAR REGRESSION MODEL

Gengjian Xue, Li Song, Jun Sun, Meng Wu

Institute of Image Communication and Information Processing,
Shanghai Jiao Tong University, Shanghai 200240, China
Shanghai Key Labs of Digital Media Processing and Communication, Shanghai 200240, China
{xgjsword, song_li, junsun, wmeng}@sjtu.edu.cn

ABSTRACT

Background subtraction is a basic task for many computer vision applications, yet in dynamic scenes it is still a challenging problem. In this paper, we propose a new method to deal with this difficulty. Our approach is based on robust linear regression model and casts background subtraction as a outlier signal estimation problem. In our linear regression model, we explicitly model the error term as a combination of two components: foreground outlier and background noise. The foreground outlier is sparse and can be arbitrarily large in most cases, while the background noise is relatively small and dispersed. In order to reliably estimate the coefficients under the constraint of sparse foreground outlier, we propose a new objective function. Then we transform the function to fit our problem by only estimating the foreground outlier and give the solution method. Experimental results demonstrate the effectiveness of our method.

***Index Terms*—** Background subtraction, robust linear regression, sparse outlier estimation

1. INTRODUCTION

Effective foreground detection is often the first step in video processing applications. Its output is usually as an input to a higher level process, such as object categorization, tracking or action recognition, making it as a critical part of the system. Although various techniques have been proposed for moving objects detection, it is still a challenging problem when the background scenes are dynamic, *e.g.* rippling water, swaying trees, flickering monitors.

Background subtraction is the most widely used technique for foreground separation. It consists of two steps: one is maintaining the background model, the other is subtracting the new frame from the background model and thresholding the difference value to determine the foreground. Based on this framework, many different approaches have been presented. Stauffer *et al.* [1] proposed to model each pixel in the background by a mixture of K Gaussians (GMM) and update the parameters using an on-line approximation. In [2], the extended work characterizing the adaptive component number for each pixel was presented. Another popular method is the nonparametric statistical approach. Elgammal *et al.* [3] proposed to adopt the kernel density estimation (KDE) technique to represent the background. This approach estimated the probability density of each pixel directly without any distribution assumptions. But these methods are not robust enough in dynamic scenes.

Recently, other approaches that did not rely on the tedious practice of trying to model the background have been paid much attention. In [4], background subtraction has been converted into a signal estimation problem under the sparse foreground assumption. The method viewed foreground objects as sparse corruption signals and estimated them by the sparse recovering method. Then, different base construction techniques have been compared and discussed in [5]. However, the sparse assumption on the total error in dynamic scenes may be inaccurate which degrades its detection performance.

In this paper, we propose a novel approach that detects foreground objects based on robust linear regression model. In this model, we regard foreground objects as outliers and consider that the observation error is composed of foreground outlier and background noise. In order to reliably estimate the coefficients, we should robustly remove the outlier and noise. Thus, the foreground detection task has been converted into a outlier estimation problem. Based on the observation that foreground outlier is sparse and background noise is dispersed in most cases, we propose a new objective function that simultaneously estimates the coefficients and sparse foreground outlier. We then transform the function to fit our problem by only estimating the foreground outlier and give the solution method. Experimental results demonstrate the effectiveness of our method.

The rest of the paper is organized as follows. In Section 2, we introduce the linear regression method, then we propose our linear regression model for foreground detection and give the solution method. In Section 3, foreground detection technique is presented. Experimental results are shown in Section

This work was supported in part by 973 Program(2010CB731401, 2010CB731406), NSFC(60702044,60902020,60902073), the 111 project and STCSM (10511500802).

4. Finally, we conclude the paper.

2. ROBUST LINEAR REGRESSION MODEL

2.1. Linear regression model

Linear regression attempts to model the relationship between two variables by fitting a linear equation to the observed data. A simple linear regression model is represented as follows:

$$\begin{aligned} y_i &= x_{i1}w_1 + \cdots + x_{iD}w_D + e_i \\ &= {x_i}^T \mathbf{w} + e_i \end{aligned} \tag{1}$$

where $i = 1, 2, \ldots, M$ means the number of observations, $x_i \in R^D$ stands for the explanatory variable, y_i is the dependent variable, $\mathbf{w} \in R^D$ is the regression coefficients, D means the regression model order with $M > D$ and e_i is the observation error. In the case of multiple variables, the linear regression model can be written as:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \tag{2}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_M)^T$, $\mathbf{X} = (x_1, x_2, \ldots, x_M)^T$ and $\mathbf{e} = (e_1, e_2, \ldots, e_M)^T$.

The essential purpose of linear regression method is to estimate the coefficients from a training dataset by minimizing a given objective function. The least squares criterion is the most commonly used objective and is represented as:

$$\hat{\mathbf{w}} = \underset{\hat{\mathbf{w}}}{\text{argmin}} ||\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}||_2 \tag{3}$$

But this criterion is sensitive to outliers which means the estimation coefficients are not reliable and the observation errors are inaccurately estimated, making it be unsuitable for our detection problem.

2.2. Robust linear regression model

In our real settings, foreground outlier and background noise have different properties. Foreground outlier is infrequently and its value can be arbitrarily large. Moreover, comparing to the whole background scene, foreground is sparse in most cases. By contrast, background noise is dispersed and independent, and the deviation of its value from the normal background is relatively small. Based on this consideration, the observation error is divided into two terms and the robust linear regression model (RLRM) is defined as:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{t} + \mathbf{n} \tag{4}$$

where $\mathbf{y}$, $\mathbf{X}$, $\mathbf{w}$ are same to (2), $\mathbf{t} = (t_1, t_2, \ldots, t_M)^T$ is a sparse vector corresponding to the foreground outlier, and $\mathbf{n} = (n_1, n_2, \ldots, n_M)^T$ represents the background noise. Since the entries of $\mathbf{n}$ are independent, we consider it as Gaussian noise $\mathbf{n_i} \sim \mathcal{N}(\mathbf{0}, \sigma^{\mathbf{2}})$. Thus, the objective function (3) under the sparse constraint is defined as:

$$(\hat{\mathbf{w}}, \hat{\mathbf{t}}) = \underset{\hat{\mathbf{w}}, \hat{\mathbf{t}}}{\text{argmin}} ||\mathbf{y} - \mathbf{X}\hat{\mathbf{w}} - \hat{\mathbf{t}}||_2 + \lambda ||\hat{\mathbf{t}}||_0 \tag{5}$$

where λ is a regularization parameter. This function is similar to [6], but the authors focused on the coefficients estimation while our aim is to detect the sparse foreground in dynamic scenes based on the robust linear regression model.

2.3. Outlier estimation based on RLRM

Simultaneously estimating two variables in function (5) is complicated and is not necessary in our problem. Since our goal is to detect the foreground outlier $\hat{\mathbf{t}}$, we transform function (5) into another form which contains only one variable of foreground outlier and can be easier to solve. Assuming the columns are linearly independent, the matrix $\mathbf{X}$ with size $M \times D$ can be regarded as a column space constructed by M-dimensional columns span a D- dimensional subspace. Then we compute the orthogonal matrix $\mathbf{F}$ to the column space of $\mathbf{X}$ such that $\mathbf{F}\mathbf{X} = \mathbf{0}$. $\mathbf{F}$ is of the size $(M - D) \times M$ with the rows forming an orthogonal basis for the left null space of $\mathbf{X}$. Multiplying (4) by F, we get

$$\mathbf{y}' \quad = \quad \mathbf{F}\mathbf{t} + \mathbf{n}' \tag{6}$$

where $\mathbf{y}' = \mathbf{F}\mathbf{y}$ and $\mathbf{n}' = \mathbf{F}\mathbf{n}$. For the orthogonality of $\mathbf{F}$, $\mathbf{n}'$ is still the Gaussian noise. This transform has no influence on the performance of $\hat{\mathbf{t}}$ in (5), but has removed the other variable $\hat{\mathbf{w}}$, which makes $\hat{\mathbf{t}}$ easier to be estimated. By replacing L_0 with L_1 norm, the corresponding objective function (5) has transformed into containing only one variable of sparse outlier and can be solved by the sparse recovery method.

$$(\hat{\mathbf{t}}) = \underset{\hat{\mathbf{t}}}{\text{argmin}} ||\mathbf{y}' - \mathbf{F}\hat{\mathbf{t}}||_2 + \lambda ||\hat{\mathbf{t}}||_1 \tag{7}$$

Here, we choose the parameter λ as $\sigma\sqrt{2 \log M}$ according to [7]. The parameter σ controls the background noise level.

3. PROPOSED DETECTION METHOD

Our foreground detection method consists of two steps. In the regressor construction stage, we regard background frames as explanatory variables. The frame is translated into a 1D vector where each pixel corresponds to one variable. Next, the vectors of $\mathbf{D}$ background frames are stacked in column direction so that each variable is D-dimensional and the matrix $\mathbf{X}$ has been constructed. Then we compute the orthogonal matrix $\mathbf{F}$. Since the vectors corresponding to the frames in dynamic scenes are linearly independent with high probability, the condition for computing the orthogonal matrix is supposed to be always satisfied. In the foreground outlier estimation stage, we view each new coming frame as a dependent variable. We compute the new vector $\mathbf{y}'$ and estimate the sparse outlier according to (7). When no foreground is in

```
Set the parameters σ , model order D, number of subregions
L, Th value for foreground determination
➤ Regressor construction step
    For i = 1 to D
        For j = 1 to L
            get j_th subregion of i_th background frame and
               translate into 1D column vector x_ij
            set x_ij to i_th column of matrix X_j
        End
    End
    For j = 1 to L
        compute the orthogonal matrix F_j to matrix X_j
    End
➤ Foreground detection step
    For k = 1 to length(videos)
        For j = 1 to L
            get subregion and change into 1D vector y_kj
            compute the vector with F_j to get y_kj' by Eq.(6)
            estimate the outlier according to (7)
            change the estimation vector t_kj into subregion
        End
        Foreground = |t_k| > Th
    End
```

Fig. 1. Overview of the proposed framework

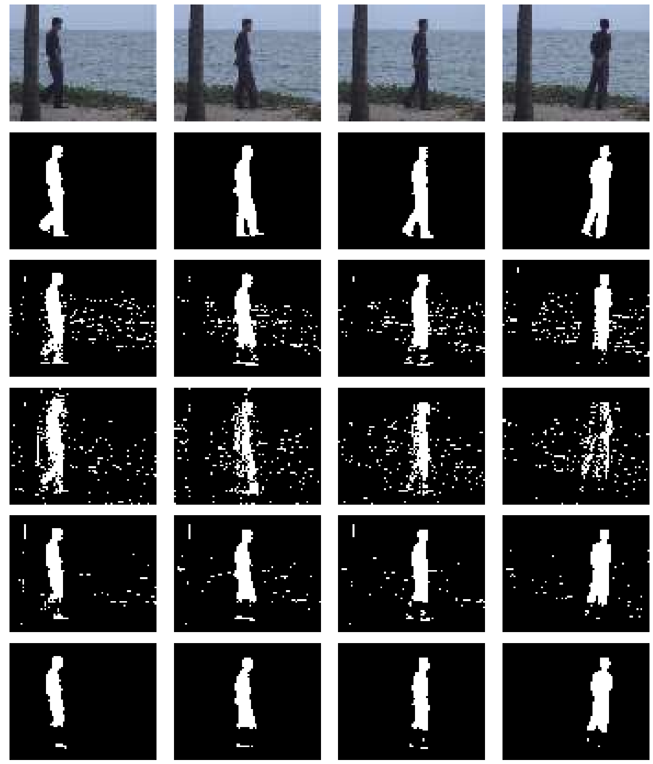

Fig. 2. Comparison results on the 1499^{th}, 1515^{th}, 1523^{th} and 1575^{th} frames of the rippling water sequence. The top and second rows are the original and groundtruth frames. The third, forth and fifth rows are results obtained by GMM, KDE and Sparse method respectively. The last row is the results obtained by our method.

the frame, all values of the estimated outlier are small and approximately equal. When foreground objects exist, the values of the estimated outlier corresponding to the foreground are much larger than the others. So we threshold the estimated outlier by value $\mathbf{Th}$ to get the final result. To effectively estimating and computing, we divide the frame into $\mathbf{L}$ subregions and implement our method on each region identically. The whole procedure of our algorithm is as Fig.1.

4. EXPERIMENTAL RESULTS

To evaluate the performance of our method, two challenging videos characterizing the dynamic background are adopted for testing. Both of them are from [8] and the groundtruth images are obtained by manually segmented. Our method is also compared with three widely used approaches including GMM [1], KDE [3] and Sparse [4] methods. Besides, the critical parameters ($\mathbf{L} = 4, \mathbf{D} = 20, \mathbf{Th} = 30$) are the same in our algorithm and Sparse method for fair comparison. No post processings are applied to all results. Both visual and numerical methods are used for comparison. The precision and recall rates are used for evaluation.

$$\text{Precision} = \frac{\text{Number of true positives detected}}{\text{Total number of positives detected}} \tag{8}$$

$$\text{Recall} = \frac{\text{Number of true positives detected}}{\text{Total number of true positives}} \tag{9}$$

The first sequence is about the rippling water background and the second experiment is conducted on the campus environment containing moving tree branches. As shown in the figures, we can see that GMM, KDE and Sparse methods all detect large number of dynamic background pixels as foreground and many foreground positives in the inner areas are not detected. Comparing with them, our method can effectively detect the foreground while suppress much of the background noise. Due to separating foreground outlier from background noise, our approach gets better results than Sparse method. As quantitative comparisons show in Table 1 and 2, although the recall rates detected by our method are slightly lower sometimes, the precision rates are all greatly increased. The quantitative comparison is in accordance with our visual evaluation.

5. CONCLUSION

In this paper, we cast foreground detection in dynamic scenes as a sparse outlier estimation problem in the linear regression model. We divide the observation error into two different terms named foreground outlier and background noise, and

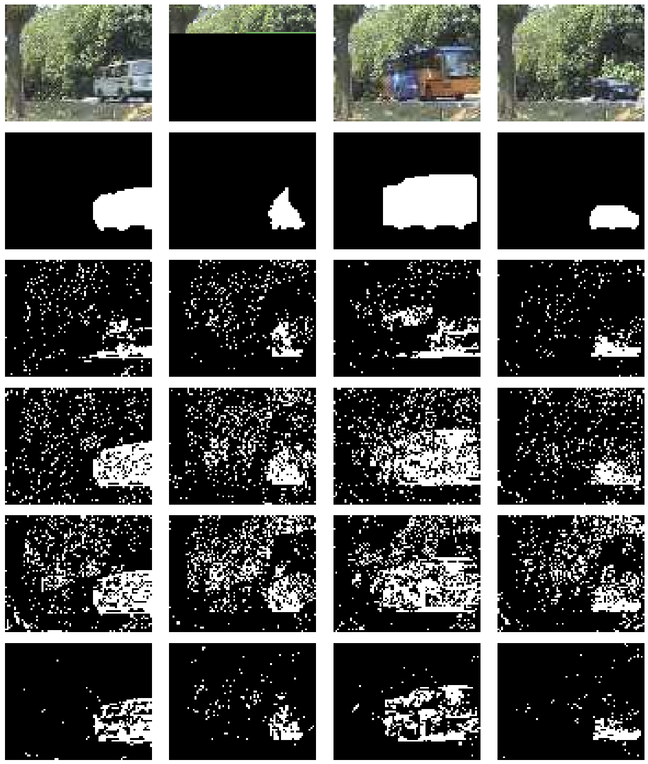

Fig. 3. Comparison results on the 1204^{th}, 1385^{th}, 1668^{th} and 1706^{th} frames of the campus sequence. The top and second rows are the original and groundtruth frames. The third, forth and fifth rows are results obtained by GMM, KDE and Sparse method respectively. The last row is the results obtained by our method.

then we propose a new objective function. By transforming the function to fit our problem, we get our foreground estimation result. Experimental results and comparisons show that our method can effectively detect the foreground under dynamic scenes. Our future work is to reduce the computation load of the algorithm.

Table 1. Quantitative comparison of Precision and Recall rates on rippling water sequence

Method		GMM	KDE	Sparse	Ours
Precision(%)	1499_{th}	64.22	59.82	83.72	96.24
	1515_{th}	67.22	56.31	81.55	99.32
	1523_{th}	63.38	59.13	82.51	94.56
	1575_{th}	61.52	45.36	89.22	97.48
Recall(%)	1499_{th}	89.25	87.63	77.42	68.82
	1515_{th}	80.00	69.38	75.31	72.59
	1523_{th}	82.03	74.22	78.65	72.40
	1575_{th}	65.43	37.42	79.65	76.15

Table 2. Quantitative comparison of Precision and Recall rates on campus sequence

Method		GMM	KDE	Sparse	Ours
Precision(%)	1204_{th}	43.11	52.94	42.66	85.71
	1385_{th}	31.43	21.07	17.22	55.30
	1668_{th}	63.61	62.99	60.86	86.68
	1706_{th}	47.81	28.68	20.96	66.11
Recall(%)	1204_{th}	33.96	80.66	72.64	59.43
	1385_{th}	56.30	78.74	68.90	57.48
	1668_{th}	27.95	66.69	62.29	43.43
	1706_{th}	59.93	65.41	68.84	54.11

6. REFERENCES

[1] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 1999, vol. 2, pp. 246–252.

[2] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[3] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, "Background and foreground modeling using non-parametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[4] Mert Dikmen and Thomas S. Huang, "Robust estimation of foreground in surveillance videos by sparse error estimation," in *19th International Conference on Pattern Recognition*, 2008, pp. 1–4.

[5] Mert Dikmen, Shen-Fu Tsai, and Thomas S. Huang, "Base selection in estimating sparse foreground in video," in *16th International Conference on Image Processing*. IEEE, 2009, pp. 3217–3220.

[6] YuzheJin and B.D. Rao, "Algorithms for robust linear regression by exploiting the connection to sparse signal recovery," in *International Conference on Acoustics Speech and Signal Processing*. IEEE, 2010, pp. 3830–3833.

[7] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[8] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu, and Qi Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004.