https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-269
© 2019, Society for Imaging Science and Technology

Background Subtraction Using Multi-Channel Fused Lasso

Xin Liu and Guoying Zhao*
Center for Machine Vision and Signal Analysis, University of Oulu, Finland
xin.liu@oulu.fi, guoying.zhao@oulu.fi
*Corresponding author

Abstract

Background subtraction is a fundamental problem in computer vision. Despite having made significant progress over the past decade, accurate foreground extraction in complex scenarios is still challenging. Recently, sparse signal recovery has attracted a considerable attention due to the fact that moving objects in videos are sparse. Considering the coherent of the foreground in spatial and temporal domain, many works use the structured sparsity or fused sparsity to regularize the foreground signals. However, existing methods ignore the group prior of foreground signals on multi-channels (such as the RGB). In fact, a pixel should be considered as a multi-channel signal. If a pixel is equal to the adjacent ones that means all the three RGB coefficients should be equal. In this paper, we propose a Multi-Channel Fused Lasso regularizer to explore the smoothness of multi-channels signals. The proposed method is validated on various challenging video sequences. Experiments demonstrate that our approach effectively works on a wide range of complex scenarios, and achieves a state-of-the-art performance.

Introduction

Background subtraction can be defined as segmenting foreground in videos from static cameras. It plays a critical role in variety of computer vision applications, such as intelligent visual surveillance, content-based video coding, human-machine interface, and behavior understanding. Over the past decades, extensive work made remarkable efforts to background subtraction, while the background subtraction in realistic environments usually encounters many challenging situations, such as illumination changes, dynamic background motions, camouflage, camera shaking, low contrast, and high sensor noise. It remains an open problem to design a background subtraction model which can robustly handle a wide variety of scenes.

A popular framework for background subtraction is the sparse signal recovery, since the foreground signals in video are sparse. The basic idea is to factorize the given matrix of the accumulated frames into the low-rank background and sparse foreground as outliers, such as the famous Robus Principle Component Analysis (RPCA) [1], which uses the Principal Component Pursuit (PCP) to perform the low-rank and sparse matrix decomposition. A further prior for foreground is that the moving objects are spatially coherent clusters, namely, if a pixel is a foreground, its neighboring pixels would also belong to foreground, and vice versa. Therefore, variety of constraint has been utilized to enforce the spatial contiguity among the neighboring pixels of the foreground. In DECOLOR [2], Zhou *et al*. employed the Markov Random Fields (MRFs) to impose the smoothness on the foreground matrix. Also, the group lasso regularization was applied to model the foregrounds in GOSUS [3]. Liu *et al*. [4] proposed a low-rank and structured sparse decomposition where the (stacked frames) matrix is divided into overlapping groups of pixels to enforce structural sparsity constraints. Aim to enhance the continuity of foregrounds, the group clustering prior on nonzero coefficient was emphasized in the method [5]. In [6] [7], the local sparseness constraint was exploited by total variation (TV-RPCA) penalty and generalized fused Lasso (GFL) to better deal with corrupted data. However, existing methods had few considerations on the homogeneity of the channels. When dealing with color images, a typical option is to convert the RGB to gray frame [2] [3] [4], and another way is to apply sparse recovery independently to each of the three RGB channels [6] [7]. In fact, the pixel should be considered as a multi-channel feature, if a pixel is equal to the adjacent ones that means all the three RGB coefficients should be equal. So it is necessary to enforce homogeneity of the channels at group levels. This paper aims to explore the prior of multi-channel group-sparse for background subtraction.

In this work we take into account the prior of multi-channel group-sparse on foreground, and propose the Multi-Channel Fused Lasso (MCFL) regularizer, to enforce a multi-channel foreground to be piece-wise constant at group level, being adjacent groups equal in all the different channels at the same time. Inspired by the Fused Lasso penalty for preserving continuous structure on signals, we introduce a modification of Fused Lasso that uses the $\ell_{2,1}$ norm instead of the ℓ_1 norm to handle multi-channel structural smoothness. Furthermore, we propose a two-pass framework for background subtraction. Firstly, a low-rank and sparse matrix decomposition is utilized on video slices along X-T and Y-T planes, segmenting a video sequence into the low-rank background and the sparse foreground (a rough foreground candidate for next pass). Then, a sparse signal recovery with MCFL regularizer is used to refine the foreground. The main novelties and contributions are summarized as follows:

1. We propose a new formulation of sparse signal recovery via the Multi-Channel Fused Lasso (MCFL) regularizer. It explicitly reconstructs multi-channel foreground signals with a spatial structure that reflects smooth changes along the group features.

2. The experimental results on two benchmarks show that the proposed method works well on a wide range of complex environments, and achieves a state-of-the-art performance for background subtraction.

Multi-Channel Fused Lasso for Background Subtraction

From signal processing point of view, foreground detection can be regarded as separating a source signal from the mixture of sources, which can be expressed in general as:

$$Y = B + F + \varepsilon \tag{1}$$

where Y is the observed (a video frame) signal which are composed by individual sources, namely the background B, foreground F and noise ε. Given the assumption that foreground objects are usually sparse, then the signals of foreground and noise can be considered as the residual R between the frame and the background

$$R = Y - B \tag{2}$$

According to the framework of the sparse signal recovery, at time t, given a residual signal $R_t \in \mathbb{R}^s$ ($s = w \times h \times C$, where w, h and C are the width, height and number of channels of an input frame), its binarization for obtaining foreground mask can be modeled as a denoising process:

$$R_t = \Phi x + \varepsilon_t \tag{3}$$

where Φx accounts for the recovered foreground signal F_t, and ε_t is the noise. The $x \in \mathbb{R}^s$ is the coefficient vector, and x should be a k_x sparse vector and $k_x \ll s$. In other words, the computed nonzero part of x can be utilized to binarize the foreground mask. Here, we employ an identity matrix $\mathrm{I} \in \mathbb{R}^{s \times s}$ as the complete dictionary Φ for the foreground signals.

The moving foreground objects are spatially coherent clusters, namely, if a pixel is a foreground, its neighboring pixels would also belong to foreground, and vice versa. Therefore, variety of constraint has been utilized to enforce the spatial contiguity among the neighboring pixels of the foreground, such as the famous Fused Lasso [8]:

$$\min_{z} \frac{1}{2} \|R_t - \Phi x\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|Dx\|_1 \tag{4}$$

the term $\|R_t - \Phi x\|_2^2$ is the ordinary least squares minimization criterion for counting reconstruction error, where $\|\cdot\|_2$ denotes the ℓ_2-norm. The term $\|x\|_1$ is the sparsity constraint on coefficients, where $\|\cdot\|_1$ denote the ℓ_1-norm. The λ_1 and λ_2 are regularization parameters which control the relative contributions of the corresponding terms. The term $\|Dx\|_1$ is the Total Variation (TV) regularizer which penalizes the differences between consecutive coefficients, where $D \in \mathbb{R}^{(s-1) \times s}$ is the differencing matrix, that is, $D_{i,i} = -1$, $D_{i,i+1} = 1$ and $D_{i,j} = 0$ elsewhere

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

As discussed in the section Introduction, existing methods ignore the group prior of foreground signals on multi-channels (such as the RGB), commonly, a grey-scale operation is utilized to simplify the multi-channel task to a sole-channel one. However, the pixel should be considered as a multi-channel signal. If a pixel is equal to the adjacent ones that means all the channels' (three RGB's) coefficients should be equal. Considering F_t has N pixels ($N = w \times h$), and a pixel p_i of F_t has C channels, therefore

$$F_t = (\overbrace{p_{1,1}, p_{1,2}, \cdots p_{1,C}}^{p_1}, \overbrace{p_{2,1}, p_{2,2}, \cdots p_{2,C}}^{p_2}, \cdots \overbrace{p_{N,1}, p_{N,2}, \cdots p_{N,C}}^{p_N})^\top$$

From above, we can find that foreground signal F_t has a group structure, namely, F_t has NC components that come in N groups with C channels. As such, multichannel of a pixel should be considered as irrelevant or relevant as a whole, and not each component independently as in the traditional model. In other words, all the coefficients of a particular group (pixel) should be zero, or nonzero, at the same time, so the sparsity of x is achieved at the group level. Based on this observation, in this paper, we propose a Multi-Channel Fused Lasso model to explore the smoothness of multi-channels signals. The objective function is defined as:

$$\min_{x} \frac{1}{2} \|R_t - \Phi x\|_2^2 + \lambda_1 \|x\|_{2,1} + \lambda_2 \|Gx\|_{2,1} \tag{5}$$

where the $\|\cdot\|_{2,1}$ denotes the $\ell_{2,1}$-norm. Since the sparsity should be achieved at the group level, for a coefficient vector x, the term

$$\|x\|_{2,1} = \sum_{n=1}^{N} \|x_n\|_2 = \sum_{n=1}^{N} \sqrt{\sum_{c=1}^{C} x_{n,c}^2} \tag{6}$$

which is the group Lasso model, means the ℓ_1 norm of the ℓ_2 group norms. In contrast with the TV regularizer, the term $\|Gx\|_{2,1}$ enforces the similarity among the coefficients corresponding to nearby groups, namely the differences between consecutive groups should to be identically zero, as

$$\|Gx\|_{2,1} = \sum_{n=2}^{N} \sqrt{\sum_{c=1}^{C} (x_{n,c} - x_{n-1,c})^2}$$
$$\text{with } \mathrm{G} = \begin{bmatrix} -I & I & & & \\ & -I & I & & \\ & & \ddots & \ddots & \\ & & & -I & I \end{bmatrix} \tag{7}$$

where $G \in \mathbb{R}^{(N-1)C \times NC}$ is a group differencing matrix, and $I \in \mathbb{R}^{C \times C}$ denotes the identity matrix.

Optimization Method via Proximal Splitting

As we know the ℓ_1 regularizer is not differentiable, which rules out conventional smooth optimization techniques. In this paper, we introduce the proximal splitting method [9] for optimize Eq. (4), which can be formulated as convex optimization problem of the form

$$\min_{x \in \mathbb{R}^M} \theta_1(x) + \cdots + \theta_m(x) \tag{8}$$

The proximal splitting method is designed to split the objective into functions $\theta_1(x), \cdots, \theta_m(x)$ individually (minimizing them independently) so as to yield an easily implementable algorithm, and each non-smooth function in (8) is involved via its

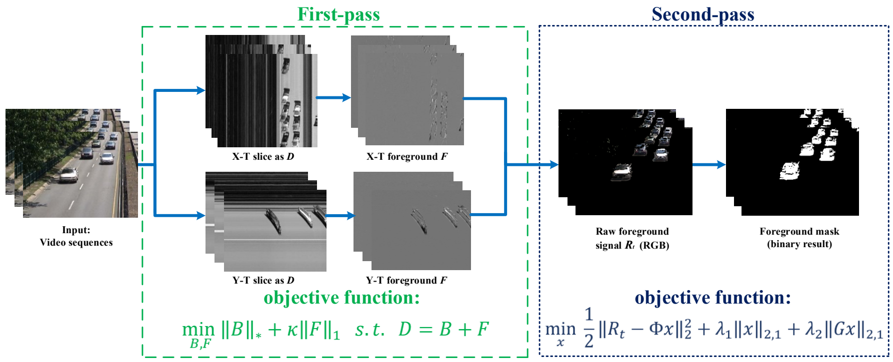


Figure 1. *Illustration of framework of the proposed method. The first-pass introduced a RPCA-PCP method to estimate the raw foreground signal (residual R_t), and a new sparse signal recovery with MCFL regularizer is proposed to obtain foreground masks in the second-pass.*

proximity operator [9]. Specifically, if θ_i is a convex, lower semi-continuous function, its proximity operator (denote as $\mathrm{prox}_{\gamma;\theta_i}$) at x with step $\gamma > 0$ is defined as

$$z_x = \mathrm{prox}_{\gamma;\theta_i}(x) = \arg\min_{z\in\mathbb{R}^M} \frac{1}{2}\|z-x\|_2^2 + \gamma\theta_i(z) \tag{9}$$

Recall that in Eq. (4), $\theta_1(x) = \|F_t - \Phi x\|_2^2$ is differentiable, while $\theta_2(x) = \lambda_1\|x\|_{2,1}$ and $\theta_3(x) = \lambda_2\|Gx\|_{2,1}$ are convex but non-smooth functions. Here, we note $\theta_+(x) = \theta_2(x) + \theta_3(x)$. Since both functions $\theta_1(x)$ and $\theta_+(x)$ are convex, according to the proximal splitting method, the optimization of (4), namely the minimization the sum of $\theta_1(x)$ and $\theta_+(x)$, can be achieved through iterative minimization of $\theta_1(x)$ and $\theta_+(x)$ individually. Based on proximal gradient method [9] and Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [10], an optimal x^* can be obtained by optimizing the variable x and updating the variables z and t iteratively, which solves the following three sub-problems:

$$\begin{cases} x_k = \mathrm{prox}_{\gamma;\theta_+}(z_k - \gamma\nabla\theta_1(z_k)) \\ z_{k+1} = x_k + \frac{t_k-1}{t_{k+1}}(x_k - x_{k-1}) \\ t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2} \end{cases} \tag{10}$$

where ∇ denotes the differential operator, and $\gamma = 1/L$ where L is a Lipschitz constant for $\nabla\theta_1$. Here, take $z_1 = x_0$, $t_1 = 1$ for parameters initialization.

For the term $\mathrm{prox}_{\gamma;\theta_+}$, the proximity operators of the sum of $\theta_2(x)$ and $\theta_3(x)$ are needed. For that, we employ the Dykstra-like Proximal (DP) algorithm [9], which allows to compute the proximity operator of the sum of two (or more) functions combining their individual proximity operators in an iterative way. In our case, the problem is

$$z_x = \mathrm{prox}_{\theta_+}(x) = \arg\min_{z\in\mathbb{R}^M} \frac{1}{2}\|z-x\|_2^2 + \theta_2(z) + \theta_3(z) \tag{11}$$

Based on the DP algorithm, an optimal z^* can be obtained by alternating between optimizing the variables z, y and updating the variables α, β, which solves the following four sub-problems:

$$\begin{cases} y_k = \mathrm{prox}_{\gamma;\theta_2}(z_k + \alpha_k) \\ \alpha_{k+1} = z_k + \alpha_k - y_k \\ z_{k+1} = \mathrm{prox}_{\gamma;\theta_3}(y_k + \beta_k) \\ \beta_{k+1} = y_k + \beta_k - z_{k+1} \end{cases} \tag{12}$$

we set $z_1 = x$, $\alpha_1 = 0$ and $\beta_1 = 0$ for parameters initialization. In our case, for the term $\mathrm{prox}_{\gamma;\theta_2}$, namely the proximity operator of $\|x\|_{2,1}$ is the group soft-thresholding [11], defined as:

$$\mathrm{prox}_{\gamma;\|\cdot\|_{2,1}}(x_{n,c}) = \max\left(0, 1 - \frac{\gamma}{\|x_n\|_2}\right) x_{n,c} \tag{13}$$

which indicates that any group x_n with a ℓ_2-norm less than γ will be zeroed. For the term $\mathrm{prox}_{\gamma;\theta_3}$, we need to solve

$$\mathrm{prox}_{\gamma;\theta_3} = \arg\min_{z\in\mathbb{R}^M} \frac{1}{2}\|z-x\|_2^2 + \gamma\|Gz\|_{2,1} \tag{14}$$

which is a particular case of the more general problem $\inf_{z,y} \{\eta(z) + \gamma\delta(y)\}$ s.t. $y = Gz$ where $\eta(z) \equiv \frac{1}{2}\|z-x\|_2^2$ and $\delta(\cdot) \equiv \|\cdot\|_{2,1}$, as such $y \in \mathbb{R}^{(N-1)C}$. Then, we can get its Lagrangian as $\mathscr{L}(z,y;\mu) = \eta(z) + \gamma\delta(y) + \mu\cdot(Gz - y)$ with $\mu \in \mathbb{R}^{(N-1)C}$. Inspired by [12] [13], we can transform the equivalent saddle point problem $\inf_{z,y}\{\sup_u \mathscr{L}(z,y;\mu)\}$ into the dual problem, as

$$\inf_{\mu} \left\{\eta^*(-G^\top\mu) + \gamma\delta^*\left(\frac{1}{\gamma}\mu\right)\right\} \tag{15}$$

in terms of the Fenchel Conjugate [14] [13], the dual problem can be transform as

$$\min_{\mu} \left\{\frac{1}{2}\left\|G^\top\mu - x\right\|_2^2\right\} \tag{16}$$

which is quadratic with simple convex constraints [13], and it can be solved by projected gradient method. Thus, follows from the condition $0 = \nabla_z\mathscr{L} = z_x - x + G^\top\mu^*$, the proximity operator of (14) can be recovered from the dual solution μ^* through the equality [13]

$$z_x = \mathrm{prox}_{\gamma;\theta_3}(x) = x - G^\top\mu^* \tag{17}$$

Two-pass Framework for Background Subtraction

We propose a two-pass framework for background subtraction. The framework is illustrated in Fig. 1. In the first-pass, a low-rank and sparse matrix decomposition is introduced. In RPCA [1], Wright *et al*. considered background subtraction from

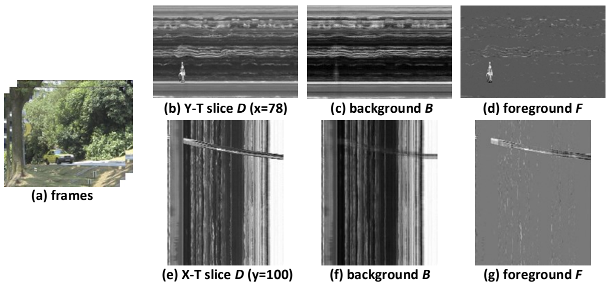


Figure 2. *Illustration of the matrix decomposition (first-pass) results on temporal slices Y-T and X-T.*

a viewpoint of matrix decomposition problem, which can be expressed as follows:

$$\min_{B,F} \|B\|_* + \kappa\|F\|_1 \quad s.t. \quad D = B + F \tag{18}$$

where $D \in \mathbb{R}^{s\times p}$ is the observed video matrix which stacked by p frames, and s is the size of a frame, κ is a regularizing parameter. B and F denote the background matrix and foreground matrix respectively. It is assumed that the background images are linearly correlated with each other, forming a low-rank matrix B ($\|\cdot\|_*$ is the nuclear norm). And the ℓ_1-norm is employed to constrain the foreground, since these regions should be a sparse matrix with a small fraction of non-zero entries. However, this method ignored the temporal continuity of foreground pixels.

Inspired by [15], we stack the temporal (T frames) slices along X-T ($D \in \mathbb{R}^{h\times T}$) and Y-T ($D \in \mathbb{R}^{w\times T}$) as the matrices D. Similar to Eq. (18), D can be decomposed into the low-rank part B represent the background and the sparse component F corresponds to the motion objects in the foreground. As illustrated in Fig. 2, since background motion is usually smaller and more regular than foreground object motion, the foreground object will form a distinct trajectory from the background in a temporal slice on the X-T and Y-T plane.

The motion matrices obtained from the X-T and Y-T slices (planes) are integrated together as the residual R_t, namely the input of second-pass. Then, in second-pass, we utilize the proposed sparse signal recovery with MCFL regularizer to segment the foreground masks.

Experiments

The experiments are conducted on real video sequences from the I2R [16] and CDnet 2012 datasets [17]. To evaluate the effectiveness of the proposed method, we compare it with six state-of-the-art algorithms, including RPCA-PCP (PCP) [1], DECOLOR (DEC) [2], GOSUS (GOS) [3], TV-RPCA (TV) [6], GFL [7], and also including a deep learning model using the Convolutional Neural Network (CNN) [18]. For fair comparisons, all methods are using the same input frames (matrix), and without any post-processing (e.g., morphological operations). For parameters of other algorithms, we employ the default settings in their codes. Due to the pages limitation, in Fig. 3, we present several representative results of I2R and CDnet 2012 dataset for qualitative analysis. Here, we cannot provide the foreground detection results of CNN [18] since its implementation is not publicly available.

The first row of Fig. 3 (a) is the "Bootstrap" sequence from I2R dateset [16], which is a typical indoor surveillance environment where walking people are always occupied in the scene. In other words, there is no "clean" background frames in this sequence. The GOS lost a lot of foreground. DEC can detect the most foreground pixels, but it produces more false alarms due to the smoothness constraint of the MRFs. It is noted that GFL is a related work to ours since it based on the fused lasso. However, it still misses plenty of foreground pixels. The proposed MCFL, PCP and TV can achieve better foreground mask than others in "Bootstrap". The next three rows of Fig. 3 (a) include some typically dynamic backgrounds, such as trees shaken by wind in the "Campus", flickering water in the "Fountain", and the motions of "Escalator". Obviously, the PCP, GOS and TV yield a large number of false positives. DEC cannot obtain foreground objects completely in the "Campus". GFL with signal channel model of fused Lasso failed to detect the people in the "Campus" and "Fountain". In contrast with GFL, the proposed multi-channel fused Lasso model can promote the detection results obviously. The last row "Lobby" is a video with the light switch on/off, this typical illumination variation should be quickly updated into background model, and the same time, the system should not lose its sensitivity to detect real foregrounds. From Fig. 3 (a), we see that the proposed method can handle that background changes.

To provide a better evaluation, in Fig. 3 (b), we present the comparison results on another widely used CDnet 2012 dataset [17]. As shown in first two rows in Fig. 3 (b), the motion of vegetables seriously affects the foreground detection accuracy of PCP, DEC, GFL and TV. Although GOS can tolerance that background dynamics, it also filters out the real foreground objects. We would like to point out a weakness of the proposed method, as shown in "Fall", our MCFL cannot restrain the backgrounds movement entirely when they take up a large portion of frame. The third row of Fig. 3 (b) shows the results of "Parking" where a truck is stopping at the parking lot after a car just moved away. We can find that TV, GFL, and GOS lose the foreground completely. The last two (rows) sequences are used to test the ability of model to handle the shadows, it can be seen that the proposed method can depress the shadows without losing the sensitivity to segment the real foregrounds. Qualitatively, the results of MCFL are the closest to the ground-truth references.

Table. 1 Comparison Of F-Measure (%) on the I2R [16] and CDnet 2012 datasets [17] (best: bold, second best: underline).

Methods	F-Measure	
	I2R	CDnet
RPCA-PCP (PCP) [1]	58.74	72.86
DECOLOR (DEC) [2]	73.08	75.70
GOSUS (GOS) [3]	77.67	75.36
TV-RPCA (TV) [6]	74.89	73.42
GFL [7]	<u>85.15</u>	72.96
CNN [18]	-	**83.86**
Proposed	**86.39**	<u>83.81</u>

This is confirmed by the performance of the *F-Measure*[1] in

[1] *F-Measure* is defined as $2 \cdot precision \cdot recall/(precision + recall)$.

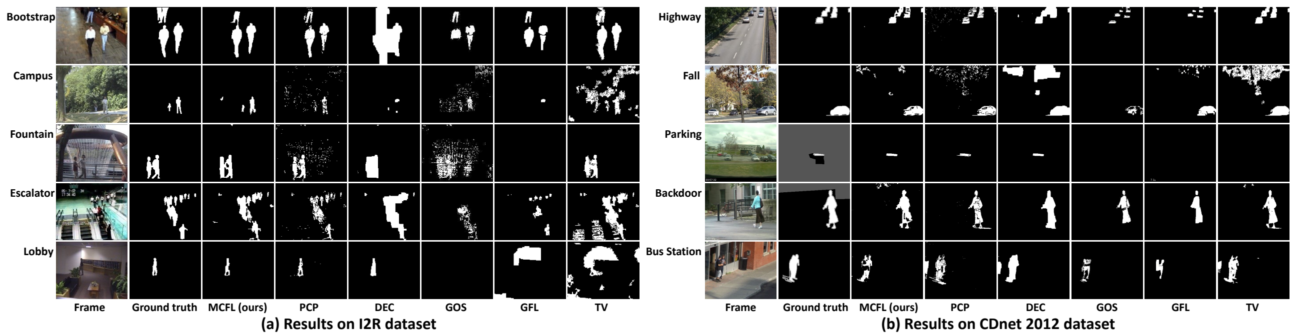


Figure 3. *Detected foreground results of videos from I2R [16] and CDnet 2012 [17] data-sets.*

Table. 1. In I2R dataset, the proposed method obtains the best average F-measure against all the other methods. In CDnet 2012 dataset, the performance of proposed MCFL is litter inferior to the deep learning model CNN [18], and better than others obviously.

Conclusion

In this study, we improved the accuracy of background subtraction in dynamic environments. This is achieved by enforcing the smoothness or continuity of foreground via the proposed multi-channel fused Lasso. Experimental results verified that our method effectively works on a wide range of complex scenarios, and achieves a state-of-the-art performance.

Acknowledgements

This work was supported by Academy of Finland, Tekes Fidipro Program, Tekes Project, Infotech, National Natural Science Foundation of China, Tekniikan Edistamissaatio, Nokia Foundation, Otto A. Malm Foundation, Tauno Tonning Foundation, and Riitta Ja Jorma J. Takanen Foundation.

References

[1] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In Proc. Adv. Neural Inf. Process. Syst., pages 2080–2088, 2009.

[2] X. Zhou, C. Yang, and W. Yu, Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 597–610, 2013.

[3] J. Xu, V. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh, "GOSUS: Grassmannian online subspace updates with structured-sparsity," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 3376–3383.

[4] X. Liu, G. Zhao, J. Yao, and C. Qi, Background subtraction based on low-rank and structured sparse decomposition. IEEE Trans. Image Process., vol. 24, no. 8, pp. 2502–2514, 2015.

[5] X. Liu, J. Yao, X. Hong, X. Huang, Z. Zhou, C. Qi, and G. Zhao. Background subtraction using spatio-temporal group sparsity recovery. IEEE Trans. Circuits Syst. Video Technol., 28(8):1737–1751, 2018.

[6] X. Guo, X. Wang, L. Yang, X. Cao, and Y. Ma, Robust foreground detection using smoothness and arbitrariness constraints. in Proc. Eur. Conf. Comput. Vis., 2014, pp. 535–550.

[7] B. Xin, Y. Tian, Y. Wang, and W. Gao, Background subtraction via generalized fused lasso foreground modeling. in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 4676–4684.

[8] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1):91–108, 2005.

[9] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In Fixed-point algorithms for inverse problems in science and engineering, pages 185–212. Springer, 2011.

[10] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2(1):183–202, 2009.

[11] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, et al. Convex optimization with sparsity-inducing norms. Optimization for Machine Learning, 5:19–53, 2011.

[12] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015.

[13] C. M. Alaíz, A. Barbero, and J. R. Dorronsoro. Group fused lasso. In International Conference on Artificial Neural Networks, pages 66–73. Springer, 2013.

[14] H. H. Bauschke, P. L. Combettes, et al. Convex analysis and monotone operator theory in Hilbert spaces, volume 408. Springer, 2011.

[15] Y. Xue, X. Guo, and X. Cao. Motion saliency detection using low rank and sparse decomposition. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pages 1485–1488, 2012.

[16] L. Li, W. Huang, I. Gu, and Q. Tian, Statistical modeling of complex backgrounds for foreground object detection. IEEE Trans. Image Process., vol. 13, no. 11, pp. 1459–1472, 2004.

[17] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, Changedetection. net: A new change detection benchmark dataset. in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2012, pp. 1–8.

[18] M. Babaee, D. T. Dinh, and G. Rigoll. A deep convolutional neural network for video sequence background subtraction. Pattern Recognit., 76:635–649, 2018.

Author Biography

Xin Liu is currently a Ph.D. candidate with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. He received the B.S. and M.S. degrees in Computer Science, in 2003 and 2007, respectively. His research interests include human behavior analysis, 3D computer vision, image restoration, and object detection.

Guoying Zhao (IEEE Senior member) is currently a Professor with the Center for Machine Vision and Signal Analysis, University of Oulu. She has authored or co-authored more than 180 papers in journals and conferences. She was co-publicity chair for FG2018, and is associate editor for several journals. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, gait analysis, dynamic-texture recognition, human motion analysis, and person identification.

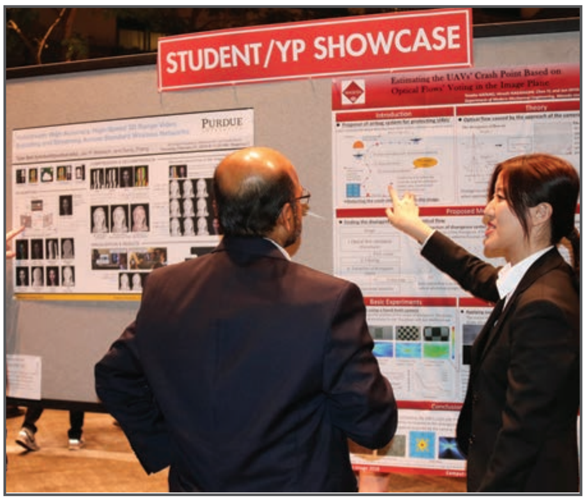
JOIN US AT THE NEXT EI!

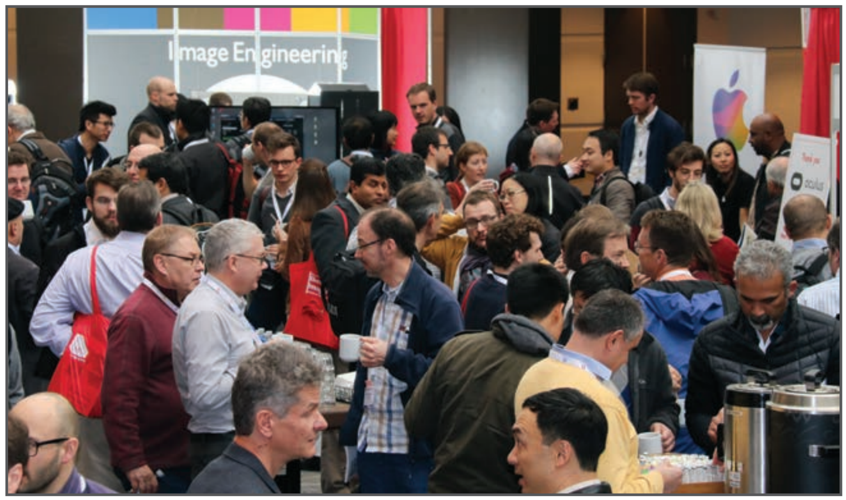
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org