Foreground Segmentation for Moving Cameras under Low Illumination Conditions

Wei Wang, Weili Li, Xiaoqing Yin, Yu Liu and Maojun Zhang
College of Information System and Management, National University of Defense Technology, Changsha, Hunan, China

Keywords: Foreground Segmentation, Moving Cameras, Trajectory Classification, Marker-controlled Watershed Segmentation.

Abstract: A foreground segmentation method, including image enhancement, trajectory classification and object segmentation, is proposed for moving cameras under low illumination conditions. Gradient-field-based image enhancement is designed to enhance low-contrast images. On the basis of the dense point trajectories obtained in long frames sequences, a simple and effective clustering algorithm is designed to classify foreground and background trajectories. By combining trajectory points and a marker-controlled watershed algorithm, a new type of foreground labeling algorithm is proposed to effectively reduce computing costs and improve edge-preserving performance. Experimental results demonstrate the promising performance of the proposed approach compared with other competing methods.

1 INTRODUCTION

Foreground segmentation algorithms aim to identify moving objects in the scene for subsequent analysis. Effective methods for isolating these objects, such as the background modeling approach, have been achieved by using stationary cameras (Elqursh and Elgammal, 2012; Brox and Malik, 2010). However, the condition that a camera should be stationary limits the application of the traditional background algorithms in moving camera platforms such as mobile phones and robots (Jiang et al., 2012; Lezama et al., 2011; Liu et al., 2015a; Liu et al., 2015b). Furthermore, moving cameras are increasingly being used to capture a large amount of video content. Therefore, effective algorithms that can isolate moving objects in video sequences are urgently needed. Recently, foreground detection methods in dynamic scenes based on mixture of Gaussians modelling have been proposed(Varadaraja et al., 2015a; Varadaraja et al., 2015b).

Object detection has been studied rarely in low illumination conditions. This topic has gradually drawn the attention of researchers because of its wide applications in all-weather real-time monitoring. Results of foreground segmentation can be improved by enhancing the contrast of the images. Many effective contrast enhancement algorithms have been proposd to improve the visual quality and used as a preprocessing strategy for object detection. However, most of the methods are time and memory consuming in the application of the video enhancement research on simplification of the existing enhancement algorithms should be conducted.

Many effective methods have been proposed to handle the problem of trajectory classification, in which the foreground and background trajectories are generated according to the shape and length of feature point trajectories(Sheikh et al., 2009; Ochs and Brox, 2011; Ochs and Brox, 2012; Nonaka et al., 2013). Sheikh (Sheikh et al., 2009) used RANSAC to estimate the basis of 3D trajectory subspace by the inliers and outliers of trajectories corresponding to the background and foreground points, respectively. Ochs (Ochs and Brox, 2012) proposed a Spectral-clustering-based method, which uses information around each point to build a similarity matrix between pairs of points and implement segmentation by applying spectral clustering. Although the effectiveness of trajectory-based methods has been proven by experiments on various datasets, certain problems remain which affect video segmentation accuracy.

Various methods (Jeong et al., 2013; Zhou et al., 2012; Gauch, 1999) have been applied to video segmentation for moving cameras. Zhang (Zhang et al., 2012) proposed a video object segmentation method based on the watershed algorithm. However, the conventional watershed algorithm fails to explicitly pre-

Wang, W., Li, W., Yin, X., Liu, Y. and Zhang, M.
Foreground Segmentation for Moving Cameras under Low Illumination Conditions.
DOI: 10.5220/0005695100650071
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 65-71
ISBN: 978-989-758-173-1

(a1) Low illumination image (a2) Enhanced image (b1) Low illumination image (b2) Enhanced image

Figure 1: Results of image enhancement. (a1) (b1) are the original low illumination images. (a2) (b2) are the corresponding enhanced images.

serve boundary fragments, thus leading to the deformation of the shape and contour of detected moving objects. Gauch (Gauch, 1999) proposed an efficient and unsupervised segmentation algorithm (Marker-controlled watershed segmentation) that can suppress the over-smooth problem and obtain accurate regions boundaries. However, the marker-extraction step of this algorithm is relatively complex because of the difficulty in accurately extracting markers. Given that foreground/background trajectory points can be considered markers, the combination of dense trajectories and watershed algorithm provides a new direction for foreground segmentation on moving cameras. In the trajectory-based segmentation algorithm proposed by Yin (Yin et al., 2015), combination of trajectory points and the original gradient minima of the images is used as markers to guide the segmentation. Foreground/background regions are segmented according to the trajectory points contained. However, regions without trajectory points should be processed by a time-consuming label inference strategy, which affect the real-time performance of the algorithm.

The paper is origanized as follows. In Section 2, a gradient field based image enhancement algorithm is proposed to pre-process the video sequence. Based on the point trajectory classification results obtained in Section 3, we combine marker-controlled segmentation with trajectory points to propose a new object segmentation algorithm in Section 4. Section 5 and Section 6 are experiments and conclusion.

2 GRADIENT FIELD BASED IMAGE ENHANCEMENT

Object segmentation in low illumination videos suffers from inaccurate boundaries because of low contrast and weak edges. Therefore, image enhancement is required as a pre-processing step of the low-contrast video sequences. A gradient-field-based image enhancement is designed to achieve high contrasts. The original image is first converted to the HSV format consisting of hue, saturation and value components. Gradient-field-based enhancement is then implemented on the value component. Based on the enhancement algorithm proposed by Zhu (Zhu et al., 2007), the real-time performance of the enhancement algorithm is further improved, while the visual quality of image enhancement is maintained.

The value component is assumed a $m\times n$ matrix V, and the gradient field is expressed in the Possion's equation form:

$$D = L_{mm}V + VL_{nn} \tag{1}$$

where L is denoted as the Laplacian operation matrix:

$$L = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \tag{2}$$

and L_{mm} and L_{nn} are the $m\times m$ and $n\times n$ Laplacian operation matrix, respectively; $D = div(G)$ is the divergence of value component, where div is the divergence function. Equation (1) is converted to the following form by matrix transformation:

$$P_1^{-1}A_1P_1V + VP_2^{-1}A_2P_2 = D \tag{3}$$

where A_1 and A_2 are the diagonal matrices with elements of L_{mm} and L_{nn}, with $[\lambda_1^{(1)},\lambda_2^{(1)},\ldots,\lambda_m^{(1)}]$ and $[\lambda_1^{(2)},\lambda_2^{(2)},\ldots,\lambda_n^{(2)}]$ being the diagonal elements, respectively. By multiplying P_1 and P_2^{-1} on both sides, Equation (3) is written as follows:

$$A_1P_1VP_2^{-1} + P_1VP_2^{-1}A_2 = P_1DP_2^{-1} \tag{4}$$

By using the Kronecker product, Equation (4) is converted to the following form:

$$
\begin{bmatrix}
\lambda_1^{(1)}+\lambda_2^{(1)} & & & & & & \\
 & \lambda_2^{(1)}+\lambda_2^{(1)} & & & & & \\
 & & \ddots & & & & \\
 & & & \lambda_m^{(1)}+\lambda_2^{(1)} & & & \\
 & & & & \ddots & & \\
 & & & & & \lambda_1^{(1)}+\lambda_2^{(n)} & \\
 & & & & & & \ddots \\
 & & & & & & & \lambda_m^{(1)}+\lambda_2^{(n)}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \vdots \\ \vdots \\ x_{mn-m+1} \\ \vdots \\ x_{mn} \end{bmatrix}
=
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ \vdots \\ \vdots \\ y_{mn-m+1} \\ \vdots \\ y_{mn} \end{bmatrix}
\quad (5)
$$

$$(I_{nn} \times A_1 + A_2 \times I_{mm}) v(P_1 V P_2^{-1}) = v(P_1 D P_2^{-1}) \quad (6)$$

where $\times$ is the Kronecker product operation and $v(\cdot)$ is the vectorization operation. To reduce the memory cost of the algorithm, we convert Equation (5) to a linear system, as shown in Equation (6),where $[x_1, x_2, \ldots, x_{mn}]^T$ and $[y_1, y_2, \ldots, y_{mn}]^T$ in Equation (6) are the vector form of $P_1 V P_2^{-1}$ and $P_1 V P_2^{-1}$, respectively. Assuming X is the matrix form of $[x_1, x_2, \ldots, x_{mn}]^T$, the enhanced value component V' is obtained as follows:

$$V' = P_1^{-1} X P_2 \quad (7)$$

Combined with the enhanced value component with the former hue and saturation components, the enhanced HSV images are generated. The enhanced RGB images are then transformed from the HSV images and the results of enhancement are shown in Figure 1.

On the basis of the enhancement of contrast and edges, the following segmentation in Section 4 generates continuous and accurate segmentation results along the edges. The foreground/background boundaries are also precisely constructed.

3 POINT TRAJECTORY CLASSIFICATION

In trajectory classification, long-term trajectories are obtained on the basis of the analysis of dense points in long frame sequences and are used to accumulate motion information over frames. Thereafter, the foreground and background trajectories are distinguished by applying the trajectory classification approach.

Trajectory classification is implemented by using the cluster growing algorithm. The difference between trajectories T_j and T_k is measured by a shape similarity descriptor, including a motion displacement term and an Euclid difference term.

Figure 2: The results of trajectory classification. The points in red and blue represent the starting points of background and foreground trajectories, respectively.

If a trajectory is given with the start position and length, the motion displacement vector for the trajectory is expressed as follows:

$$\triangle T = [P_{s+l-1} - P_s] = [x_{s+l-1} - x_s, y_{s+l-1} - y_s] \quad (8)$$

$$T_j = [\triangle x_1^{(j)}, \triangle y_1^{(j)}, \ldots, \triangle x_{N_f}^{(j)}, \triangle y_{N_f}^{(j)}]^T \quad (9)$$

where j and N_f denote the index number of trajectories and number of frames in the video sequence, respectively. The variations of the coordinates of corresponding trajectory points are as follows.

$$\triangle x_i^{(j)} = x_{i+1}^{(j)} - x_i^{(j)}, \triangle y_i^{(j)} = y_{i+1}^{(j)} - y_i^{(j)} \quad (10)$$

$||T_i - T_k||$ is denoted as the Euclid distance between trajectories T_j and T_k. The shape similarity of trajectories T_j and T_k is measured by overall displacement and Euclid distance as follows:

$$S(T_j, T_k) = \alpha_1 ||T_j - T_k|| + \alpha_2 ||\triangle T_j - \triangle T_k|| \quad (11)$$

where α_1 and α_2 are the coefficients that determine the relative importance of each term. The similarity measurement is applied in the following clustering approach. Given the samples in the motionseg

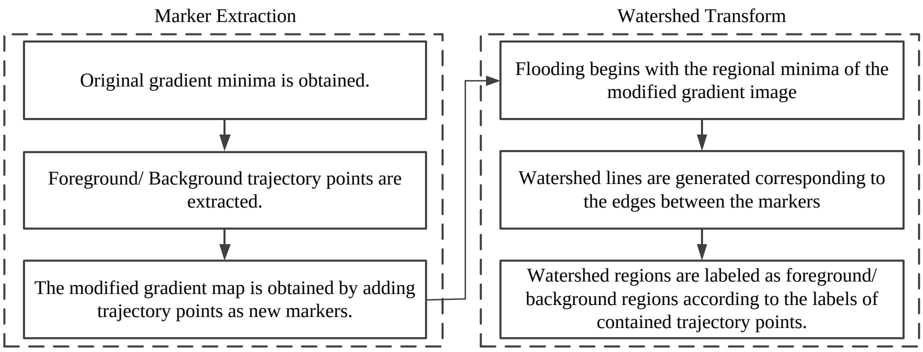


Figure 3: The procedure of the proposed segmentation algorithm.

database (Elqursh and Elgammal, 2012), the trajectories of the foreground and background are clustered. The representative trajectory classification results in the motionseg dataset (Elqursh and Elgammal, 2012) are illustrated in Figure 2. The points in red and blue represent the starting points of the background and foreground trajectories, respectively.

On the basis of the point trajectories, which are divided into foreground and background trajectories, the movement of the objects is accurately represented. This movement is then used in foreground construction.

4 OBJECT SEGMENTATION USING MARKER-CONTROLLED SEGMENTATION

The watershed segmentation algorithm is an effective segmentation algorithm with relatively low computational complexity, generates watershed regions with boundaries closely related to edges of objects, and reveals structure information in images. However, the conventional watershed algorithm suffers from the imprecise location of region boundaries. The structural information of contours and details in the video sequences are affected by the inaccurate location of region boundaries.

To overcome the problems discussed above, we combine optical flow trajectory with a marker-controlled watershed algorithm (Gauch, 1999) to address the background subtraction problem for moving cameras. The procedure of the proposed segmentation algorithm is illustrated in Figure 3.

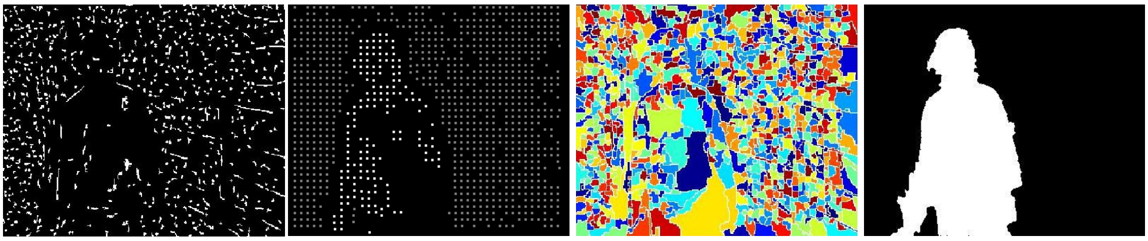
The original gradient minima of background/foreground parts, which are homogenous regions with similar gray values, are chosen as the initial markers. As useful prior knowledge for identifying the foreground and the background, trajectory points are considered components that mark the smooth regions of an image. The approximate shape and contour of the moving objects are represented by trajectory points. Therefore, the estimation of the segmentation should first be obtained. The sparsity of trajectory points also suppresses the over-segmentation problem. Trajectory points are selected as new markers to guide the generation of watershed regions, such as seeds, in the region growing process. Therefore, the problem of extracting markers is solved.

For an example frame in the video sequence in the motionseg dataset, the original gradient minima are first obtained (Figure 4a). Instead of using the combination of former gradient minima and the trajectory points as the gradient minima, which is proposed by Yin (Yin et al., 2015), trajectory points shown in Figure 4(b) are taken as the only markers that guide watershed segmentation and imposed as minima of the gradient function. This means that the markers are more sparse and the over-segmentation problem is effectively suppressed. The input marker image for watershed segmentation is a binary image that consists of marker points, where each marker corresponds to a specific watershed region. Morphological minimization operation (Soille, 1999) is applied to modify the initial gradient image, which takes the trajectory points as markers and adds them to the minima. The modified gradient image is obtained as follows:

$$G' = Mmin(G|P_T) \tag{12}$$

where G is the original gradient minima and $Mmin(\cdot)$ is the morphological minimization operation, with the trajectory points P_T being imposed as the gradient minima to guide the watershed segmentation. After the modified gradient image is obtained, watershed transform (Gauch, 1999) is applied to find the accurate contour S_w of the moving objects as follows:

$$S_w = W_{ts}Seg(G') \tag{13}$$



(a) Original gradient minima (b) Trajectory point image (c) Segmentation result (d) Binary labeling based on segmentation

Figure 4: Illustration of the proposed segmentation algorithm. (a) Original gradient minima image. White regions represent the regional minima. (b) Trajectory point image. White and grey points indicates the locations of foreground/background trajectory points, respectively. (c) Segmentation result. All the regions are painted as different colors. (d) Foreground/background labeling based on segmentation. The regions containing the foreground trajectory points are labeled as foreground regions (white parts). Background regions are represented as black parts.

where G' is the modified gradient minima, and $W_{ts}Seg$ is the watershed segmentation operation.The subsequent label inference strategy is unnecessary because the segmentation result is already satisfactory.

The segmentation result of watershed transform is shown in Figure 4(c). On the basis of the segmentation result, foreground/background labeling is performed for all regions. Watershed regions containing trajectory points are labelled as foreground/background regions according to the labels of the corresponding trajectory points. As shown in Figure 4(d), watershed regions with foreground and background trajectory points are indicated in white and black, respectively.

As shown in Figure 4(d), although most of the regions have been identified as foreground/background parts, the segmentation result for each frame still contains a few unlabeled regions that impair the completeness of moving objects. To improve the accuracy of background subtraction, a label inference procedure conducts binary labeling for each unlabeled pixel according to the probability belonging to the foreground/background (Section 5).

5 EXPERIMENTS

We validate the performance of our method on the motion segmentation dataset provided by Brox (Brox and Malik, 2010), which consists of 26 video sequences. The moving objects in this dataset are mainly people and cars. The PC for conducting the experiments has 2 GB of RAM and a 1.60 GHz CPU.

For evaluation purpose, we take precision and recall as metrics, which has been used by Nonaka (Nonaka et al., 2013). The numbers of true foreground pixels, false foreground pixels, and false background pixels are denoted as TP, FP, FN , respectively. The precision and recall metrics can then be obtained by Equation (14) as follows:

$$Prec = \frac{TP}{TP+FP}, Rec = \frac{TP}{TP+FN} \tag{14}$$

The comparison of average precision and recall metrics is shown in Table 1. Compared with the algorithms of Sheikh (Sheikh et al., 2009) and Nonaka (Nonaka et al., 2013), the proposed algorithm achieves the highest precision and significant recall, thus indicating less false foreground/background pixels and accurate segmentation result.

Table 1: Comparison on Average Precision and Recall.

	Average precision	Average recall
Proposed method	0.8239	0.8713
Sheikh (2009)	0.6957	0.8903
Nonaka (2013)	0.6135	0.8058
Zhang (2012)	0.8191	0.8270

Figure 5 shows some representative results of our method. The background parts are illustrated in red and the foreground parts remain the color in the enhanced images. As trajectory points are taken as the only markers to guides watershed segmentation, the sparsity of the markers suppresses the over-segmentation problem and improves the segmentation results. Thus the label inference strategy used in some papers (Sheikh et al., 2009; Nonaka et al., 2013; Yin et al., 2015) is unnecessary. Due to the errors of the optical flow algorithm, errors may exist in the locations of the trajectory points, which may affect the marker-controlled segmentation results and cause a few inaccurate contours of the moving objects. Although a few parts of the contours are slighted affected by the errors of optical flow, the proposed method generates satisfying segmentation results, as shown in Figure 5. In our experiment, the processing speed on the dataset is measured by the number of frames processed per second (fps). Compared with

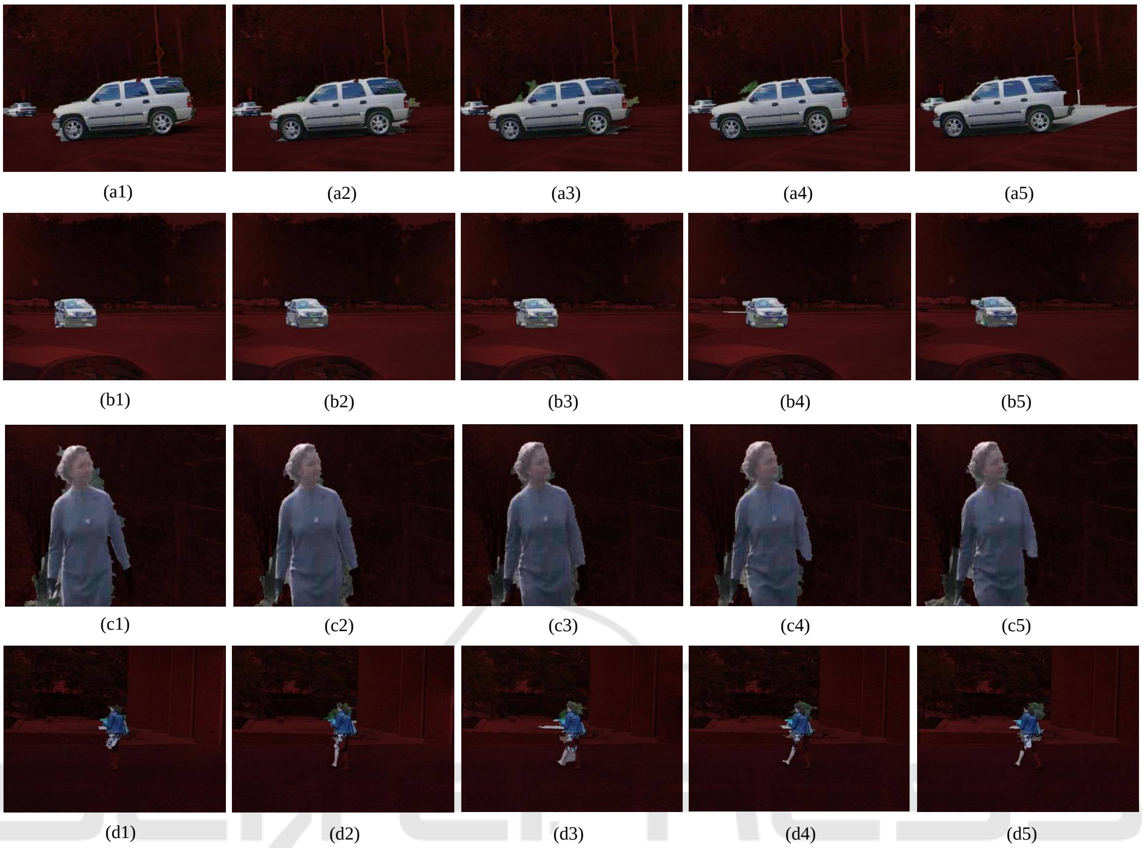


Figure 5: (a)(b)(c)(d) are 4 groups representative results on the sequences in motionseg dataset (Brox and Malik, 2010) using our method. The background parts are illustrated in red and the foreground parts remain the color in the enhanced images.

algorithms like label inference (Sheikh et al., 2009; Nonaka et al., 2013) and spectral clustering (Ochs and Brox, 2011; Ochs and Brox, 2012), the trajectory classfication and the watershed-based segmentation algorithm proposed in this paper show lower computing complexity and significant real-time performance, which effectively reduces the computing time of the proposed method. The comparison results are shown in Table 2. Although the proposed method performs slower than the algorithm of Nonaka (Nonaka et al., 2013), the precision and recall results of the proposed method outperforms their method. And the proposed method outperforms the algorithms proposed by Sheikh (Sheikh et al., 2009) and Zhang (Zhang et al., 2012) in terms of processing speed and segmentation results.

Table 2: Comparison on average processing speed (fps) on the motionseg dataset.

Algorithm	Average fps
Proposed method	0.538
Sheikh (2009)	0.096
Nonaka (2013)	1.042
Zhang (2012)	0.263

6 CONCLUSION

To cope with the problem of video segmentation on moving cameras under low illumination conditions, we present a new background subtraction method. Our work includes image enhancement, trajectory classification and object segmentation. The satisfactory performance of the proposed approach is shown in the comparison experiments. The segmentation consistency across the video sequences and algorithms efficiency for hardware implementation are considered the future research directions of this work.

REFERENCES

Brox, T. and Malik, J. (2010). Object segmentation by long term analysis of point trajectories. In *Proceedings of the 11th European Conference on Computer Vision (ECCV)*. Springer Berlin Heidelberg.

Elqursh, A. and Elgammal, A. (2012). Online moving camera background subtraction. In *Proceedings of the 12th European Conference on Computer Vision (ECCV)*. Springer Berlin Heidelberg.

Gauch, J. (1999). Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transactions on Image Processing (TIP)*, pages 69–79.

Jeong, Y., Lim, C., Jeong, B., and Choi, H. (2013). Topic masks for image segmentation. *KSII Transactions on Internet and Information Systems (TIIS)*, pages 3274–3292.

Jiang, Y., Dai, Q., Xue, X., Liu, W., and Ngo, C. (2012). Trajectory-based modeling of human actions with motion reference point. In *Proceedings of the 12th IEEE European Conference on Computer Vision (ECCV)*. Firenze, Italy.

Lezama, J., Alahari, K., Sivic, J., and Lapte, I. (2011). Track to the future: Spatio-temporal video segmentation with long-range motion cues. In *Proceedings of the 24th IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Colorado Springs.

Liu, Y., Xiao, H., Wang, W., and Zhang, M. (2015a). A robust motion detection algorithm on noisy videos. In *IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane, Australia.

Liu, Y., Xiao, H., Xu, W., M.Zhang, and J.Zhang (2015b). Data separation of l1-minimization for real-time motion detection. In *IEEE British 26th International Conference on Machine Vision (BMVC)*. Swansea, UK.

Nonaka, Y., Shimada, A., Nagahara, H., and Taniguchi, R. (2013). Real-time foreground segmentation from moving camera based on casebased trajectory classication. In *IEEE International 2nd Asian Conference Pattern Recognition (ACPR)*. Okinawa, Japan.

Ochs, P. and Brox, T. (2011). Object segmentation in video: A hierarchical variational approach for turning point trajectories into dense regions. In *Proceedings of the 13th IEEE International Conference Computer Vision (ICCV)*. Barcelona, Spain.

Ochs, P. and Brox, T. (2012). Higher order motion models and spectral clustering. In *Proceedings of the 25th IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Rhode Island.

Sheikh, Y., Javed, O., and T.Kanade (2009). Background subtraction for freely moving cameras. In *IEEE 12th International Conference Computer Vision (ICCV)*. Kyoto, Japan.

Soille, P. (1999). *Morphological image analysis principles and applications*. Springer-verlag, Berlin, Germany.

Varadaraja, S., Hongbin, W., Miller, P., and Huiyu, Z. (2015a). Fast convergence of regularised region-based mixture of gaussians for dynamic background modelling. *Computer Vision and Image Understanding*, pages 45–58.

Varadaraja, S., Miller, P., and Huiyu, Z. (2015b). Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. *Pattern Recognition (PR)*, pages 3488–3503.

Yin, X., Wang, B., Li, W., Liu, Y., and Zhang, M. (2015). Background subtraction for moving cameras based on trajectory classification, image egmentation and label inference. *KSII Transactions on Internet and Information Systems (TIIS)*.

Zhang, G., Yuan, Z., Chen, D., Liu, Y., and Zheng, N. (2012). Video object segmentation by clustering region trajectories. In *Proceedings of the 25th IEEE International Conference on Pattern Recognition (CVPR)*. Rhode Island.

Zhou, J., Gao, S., and Jin, Z. (2012). A new connected coherence tree algorithm for image segmentation. *KSII Transactions on Internet and Information Systems (TIIS)*, pages 547–565.

Zhu, L., Wang, P., and Xia, D. (2007). Image contrast enhancement by gradient field equalization. *Journal of Computer-Aided Design and Computer Graphics*, page 1546.